

Unveiling the Ocean's hidden majority: a roadmap

December 17, 2003

Guidelines from a November, 2003 strategic planning workshop

Monterey Bay Aquarium Institute

Moss Landing, CA

Life flourishes within a thin veneer that amounts to less than 0.05% of Earth's diameter. The biota represents at most one ten-billionth of Earth's mass. But Life's effects, driven by metabolism and expressed as biogeochemical processes, impose an overwhelming force on planetary change and have shaped Earth's habitability. Microbes were the only form of life for the first 2-3 billion years of planetary and biological evolution. They originated in the oceans and have assembled there into complex consortia with enormous metabolic capability. Contemporary communities of diverse Bacteria, Archaea, and Protista account for more than 98 percent of oceanic biomass. These microscopic factories - aerobic and anaerobic - are the essential catalysts for all of the chemical reactions within the biogeochemical cycles. Macroscopic life and planetary habitability completely depend upon the transformations mediated by complex microbial communities.

We have reached an important juncture in studies of these processes. Molecular biology, genomics, and information technology have revealed unforeseen levels of microbial diversity and metabolic potential in the water column, sediments and deep subsurface. The new challenge is to investigate the natural dynamics of linkages between microbiological, chemical, and physical processes. We must explore the grand relationships between microbial physiology and ecology, marine chemistry, physical oceanography, and earth-system science. These interactions define the stage upon which evolutionary events occur, whether within isolated ecosystems or on a global scale.

A problem of this magnitude demands the coordinated application of many, often previously disparate disciplines. The focus is microbial oceanography, but the biogeochemical setting provides the context. The key details are those relating microbial genotypic diversity to functional dynamics. Understanding the responses of microbial assemblages to natural and anthropogenic perturbations will provide a framework for interpreting the past and for predicting future environmental change.

The most general statement of the problem is: *How do microbial diversity and its modes of change correlate with both stable and shifting biogeochemical processes?* Inherent in this paradigm is a need to know what kinds of microorganisms occur in a given community. For the systematist, a "kind" of organism is comparable to the concept of OTUs (Organism Taxonomic Units) for describing animal and plant species. For the microbial oceanographer, a "kind" of organism must harbor a definition for functional capacity or inducible phenotype. Molecular ecologists are able to identify the occurrence of a particular functional or structural gene as a marker of diversity within an isolate or for members of a naturally occurring microbial population. In a similar manner, post genomic technology can measure gene expression patterns as a means to differentiate between "kinds" of microorganisms.

Knowing what "kinds" of organisms exist within a microbial population and how the community structure changes in response to environmental shifts, tests even the most advanced genetic technology and evolutionary theory. Yet, this is only part of the challenge. If we are to make progress towards understanding the oceans as a biologically forced system, we must link increasingly sophisticated measurements of microbial and metabolic diversity with biogeochemical and physical processes. This enterprise demands interdisciplinary efforts to explore the dynamics of microbial population biology, genome diversity and the metabolic basis of biogeochemical processes. It will engage specialists in genome science, bioinformatics, microbiology, biogeochemistry and physical oceanography. The product will be a predictive modeling framework for understanding the interplay between members of complex microbial consortia and ocean biogeochemistry.

Marine Microbial Diversity

Recent studies of microbial diversity have produced spectacular discoveries of previously unknown microorganisms, many of which have major impacts on oceanic processes. Rich, chemosynthetic microbial communities thrive at deep-sea hydrothermal vents and cold seeps. Abundant Archaea populate oceanic midwaters. Very large populations of picoplankton are the primary catalysts in carbon fixation and in the cycling of nitrogen. Molecular techniques have identified SAR11 as a dominant clade in communities of ocean-surface bacterioplankton, and the sequencing of environmental genomes (metagenomics) provides evidence of hitherto unrecognized physiological categories among the planktonic microbes. Such discoveries reveal the richness of the problem.

In the same way that insufficient field observations have limited our understanding of microbial oceanography, under-sampling of genomic information has constrained our appreciation of phylogenetic diversity. Direct interrogation of microbial genomes has taken us far beyond traditional measurements of diversity such as morphological, physiological, and biochemical variation. Using molecular tools, we can quantitatively describe genetic variation and the distribution and the occurrence of different kinds of microbes – *phylotypes* – within marine populations. Results show that contemporary estimates of microbial diversity understate the number of microbial “kinds” by orders of magnitude. In contrast, deep-sea vents separated by thousands of miles often harbor anaerobic thermophiles that have nearly identical *phylotypes* even though these organisms have not been detected in open ocean waters. Mechanisms that might explain this biogeographical distribution await discovery.

The historical events and underlying mechanisms that led to contemporary microbial diversity are uncharted. Genome-based studies suggest that large-scale genetic exchange corresponding to tens of thousands of base pairs from unknown genetic sources can occur over timescales required by microbes to adapt to shifts in environmental chemistry. Stunningly, we have only scratched the surface of marine environments but already learned that the correct conceptual framework for describing the dynamics of metagenome evolution and shifts in diversity might not yet be known.

Some of the fundamental questions that we must address include:

- 1) What accounts for large-scale genetic variation in microbial genomes that share a very recent common ancestry? Is there a cryptic source of genetic

information that selectively invades microbial genomes or are there undocumented mechanisms that can rapidly generate novel coding capacity within a bacterial chromosome?

- 2) What are the roles of the many hypothetical genes of unknown function that are either unique within or shared between nearly all sequenced genomes?
- 3) Do chemical environments select for lineages endowed with particular metabolic capabilities or does the unit of selection correspond to individual genes that can transfer particular metabolic functions between lineages?
- 4) How widespread is horizontal gene transfer? Do viruses mediate this process?
- 5) Why do complex microbial consortia retain functionally equivalent but genetically distinct lineages - for example, many different kinds of sulfate reducers or methanogens persist in anoxic settings – rather than selecting for a single “winner” with an optimal suite of metabolic activities?
- 6) When interactions occur between different kinds of microbes, is there intercellular communication at the genome level? What are the emergent functional properties and how do they transcend activities of the individual parts?
- 7) Does the diversity of a microbial guild relate to the stability of its functioning?
- 8) Is there a biogeography for distinct microbial lineages and, if so, what are the principal drivers or restrictors? What genomic changes, if any, are associated with relocation of dormant organisms over large distances?
- 9) How does genotypic diversity shape phenotypic diversity, and how does this diversity influence the functioning of ecosystems?

As a direct consequence of increased activity in marine metagenomics, bioinformatics can unveil evidence of microbes that have new traits, novel functions and unusual enzymes. In some cases, entirely new phyla are being discovered. These aid in understanding the evolution of life in this ancestral habitat and lead to sounder descriptions of new communities and species. Sequencing data will also be wedded to newly emerging molecular assays that incorporate automated sampling technologies and which will lead to finer temporal and spatial resolution of molecular diversity.

Genomic tools are most productive when they can be coupled with physiological investigations of important environmental isolates. New cultivation technologies, which must be extended wherever possible, are providing key organisms for experimental dissection of metabolic potential. These will serve as model systems and will provide genomic scaffolds for the assembly of shotgun environmental survey sequences into larger contigs. Ongoing studies of SAR11 and the establishment of the roles of *Synechococcus* and *Prochlorococcus* as primary producers at the base of the marine food web illustrate the success of this approach.

Where microorganisms are at present uncultivable, molecular methods can establish links between functions in nature and phylogenetic information about microbial

groups. The application of FISH combined with microautoradiography, quantitative measures of gene expression (*in situ* PCR), microarrays, and BAC libraries from large segments of DNA can link functional information to the phylogeny of novel organisms. Proteomics are needed to understand the adaptation of microbes at the levels of protein conformation and protein-protein interaction, *e.g.* light, temperature and pressure adaptation.

Information is also encoded in the metabolic and biosynthetic products of marine microorganisms. Isotopic analyses have pinpointed lipids produced by novel Archaea that oxidize methane anaerobically. Follow-up investigations at sites rich in these products have revealed abundant new phylotypes that are related to methanogens. The abundance of carbon-14 in lipids produced by planktonic Archaea proves that those organisms are assimilating large amounts of inorganic carbon from the ocean's midwaters and must be growing as autotrophs. Unprecedented lipid structures have been traced to previously unknown planctomycetes and the long-sought capability for anaerobic oxidation of ammonia.

Sampling and Analytical Strategies

Given the aim of integrating microbial diversity and evolution with biogeochemical processes, sampling strategies must satisfy both practical and theoretical requirements. It will be necessary to represent all biogeochemical provinces that bear on specific questions to be addressed. Sampling strategies must consider environmental heterogeneity, spatial and temporal scales to be sampled, size and definition of replicate samples required to achieve reproducibility, procedures for detecting and eliminating contaminants, potential artifacts associated with selective lysis of different kinds of microbial cells, and specific analytical methods to be employed. Because of the global nature of microbial oceanography, it will be necessary to develop higher-throughput techniques that provide enhanced resolution. New, technically-advanced sampling techniques and protocols appropriate to studies of microbial ecology must be developed. Interpretation of the results will require information about environmental conditions and biogeochemical processes and rates. Again, technological development of high-throughput methods that integrate measurements of concentrations and of rates or activities will be necessary.

The sampling of microbial communities is often destructive. Accordingly, sampling strategies must provide reliable means of relating the collected material to the natural assemblage. Contamination must be absolutely minimized. Temporal variations and spatial inhomogeneities must be considered. Planktonic communities may be organized along spatial gradients, which can be sampled reproducibly. Biofilms and sediments are much more challenging. In them, spatial variations can be extreme and processes may be decoupled temporally. Key substrates, for example, the methane and acetate being consumed by subseafloor microbial communities, may have been produced thousands of years earlier.

Specific experimental objectives should be the primary determinants of sampling strategies. Some sampling modes will be simply observational: Who's there? How abundant are they? What are they doing? What is their influence on, and response to, surrounding biogeochemical process? Other approaches may rely quite strongly on comparisons. To be valid, these will require reproducible and strictly congruent sampling and analytical strategies. For example, comparisons of coastal and open-ocean

communities will require intercalibration of analytical methods, rigorous standardization of sampling, and full consideration of conceivably pertinent chemical factors. In other situations, experimental approaches may require environmental and/or biological manipulation. In these cases, the sampling strategies will need to distinguish carefully between natural and experimental variability and perturbations. In all cases, replicate sampling, confirmatory studies using independent methods and datasets, and sample archiving will be important. Particularly in cultivation-independent gene and/or genome surveys, archiving of replicate samples for confirmatory studies using alternative methods will be absolutely essential.

It is desirable to sample many stations at different times, latitudes and longitudes, and intervals throughout the water column and into the sub-seafloor. Using existing long-term, time-series stations, it will be possible to leverage historical data while providing new opportunities for detailed studies of coastal and open-ocean sites. Until we develop higher-throughput sampling and greater analytical capacities, it will be necessary to restrict investigations to selected areas of the oceans. Sites should be chosen according to the availability of correlative information including: temperature, salinity (any form of light) and pressure, marine chemistry, physical oceanography, and census data for marine life from the surface to the seafloor (benthic invertebrates, fishing grounds, reproduction grounds etc.). The integration of these data with microbial-centric measurements will accelerate our understanding of microbial diversity and biogeochemical processes in the oceans.

Analytical methodologies will also affect sampling strategies. In large-scale genomic studies, further considerations may apply: Should microbial cells be separated from solid matrices? If so, will this bias the recovery? Do the cells need to be concentrated in some way before extraction of genomic DNA? What size DNA fragments can be recovered, and does this dictate the particular cloning strategy to be used? What is the complexity, richness and homogeneity, and specific composition of the population/community in question? These ecological parameters will greatly influence the depth of coverage necessary for representative sampling. For example, it would take only a small proportion of eukaryotes in a sample, each having genomes 10 to 50 times the size of co-occurring prokaryotes, to dramatically affect the representation of prokaryotes in subsequent random shotgun genome libraries. In some cases, size fractionation or sample “normalization” may be desirable. This will be dictated by the ultimate goal of the sampling, analyses, and experimentation. For many types of analyses, the nature, complexity, and composition of both the environmental matrix and the microbial community will largely control experimental design and sampling strategy.

MICROBIS - A Center for Marine Microbial Bioinformatics?

An initiative in microbial oceanography that embraces genome-level information will produce unprecedented volumes of data and require the development and coordination of an effective informatics environment. Integrated microbial oceanography databases must provide analytical and interpretative services as well as resources for education and outreach.

Multiple databases exist and more will emerge. They will hold sequence, oceanographic, physico-chemical data, and be capable of integrating with information from research stretching back through 200+ years of ocean microbiology. A core component of this may be a new repository within GENBANK. MICROBIS is seen as a

coordinating structure capable of linking classical descriptive and biogeographical information to genomic data and relevant annotations. Such structures will enable different user groups to locate and extract new assemblages of data for different objectives. MICROBIS may be required to act as a resource center or to serve as a platform to provide technological support. A working group will be required to oversee the emergence of an interoperable informatics environment. The steering group will be international and work in close collaboration with the others such as the Ocean Biological Information System (OBIS) group representing the Census of Marine Life (COML) and the European MARS program.

We envisage that MICROBIS will be a distributed network environment and that community participation through a secretariat will allow high rates of data assembly. A structure that allows for progressive editing and annotation of incoming information will promote consistency and high quality. The structure should embrace curated databases containing expression data and allow for re-annotation of certain microbial genomes, *etc.* The working group would identify other informatics initiatives with which MICROBIS must communicate. Interoperability has to be assured (*e.g.*, by protocols such as DiGiR, an IT protocol which allows interactions between different databases). MICROBIS should have capacity for journaling and rollback of additions or amendments.

MICROBIS *must* accommodate environmental documentation and traceable sample numbers that can be linked to all future studies. Data entries for all samples must include parameters of sampling, geographical location (GPS), depths, light conditions, temperature, oxygen, sample method and size, salinity, and pH. Pictures or short video sequences of the samples and of sampling will be preferable in some cases and necessary in others. MICROBIS must comply with Dublin core-metadata standards to promote harvesting by other informatics initiatives.

Original DNA samples and genetic libraries (when possible) should be deposited as reference samples or as templates to generate genes by PCR. A modified list of identifiers will distinguish samples from different environments (sediment, deep vent samples, microbial mats, open ocean, *etc.*). Where samples are derived from biological material, voucher specimens should be preserved to verify source material. MICROBIS should be able to evolve in response to future needs. Inclusion of taxonomic data will assist cross-linking between databases and ensure that the new results are linked to the present, classical understanding of microbial oceanography.

Coordination, education, outreach and infrastructures.

Large scale efforts to assess the nature and extent of marine microbial diversity necessarily span a broad range of disciplines, perspectives, taxa, environments and infrastructural requirements. One major challenge, but a prerequisite for transdisciplinary success, is to combine and coordinate the efforts, expertise and analyses of many different disciplines including: genomics, bioinformatics, microbiology, geochemistry, and oceanography. To some degree this could be realized *via* the formation of virtual “centers” linked by high-speed internet connections. Optimally, the essential conceptual and intellectual linkages would evolve from collaborations at “Centers of Excellence” which united the required specialists at a single site.

There are many oceanographic, biodiversity, biogeochemical programs and institutional initiatives that already offer collaborative and synergistic opportunities.

Many of these could benefit tremendously from, and contribute to, a coordinated, comprehensive focus on microbial diversity. These include observational oceanographic programs such as Global Ocean Observing System (GOOS), Gulf of Maine Ocean Observing System (GOMOOS), and the planned Oceanographic Observing Initiative (OOI). Also included are global environmental biodiversity programs like DIVERSITAS and European marine informatics initiatives such as MARS. Especially relevant to a marine microbiology initiative is the planned Integrated Marine Biogeochemistry and Ecosystem (IMBER), whose goal is to “bridge and merge the knowledge bases within the marine and biogeochemical disciplines”.

The matrix of collaborative opportunities that could link and coordinate an initiative in marine microbial diversity is extensive. Organizational efforts will be required to assess programs willing and able to tie into microbial diversity programs. The integration of outreach efforts, sampling, methodologies, databases and interoperability will need to be carefully coordinated, perhaps by a centralized Marine Microbial Biodiversity Scientific Advisory Committee (SAC).

The Marine Microbial Biodiversity SAC would identify activities that are best centralized and coordinated through initiatives such as MICROBIS. The matrix of coordination is multidisciplinary, and therefore requires a group of experts that could fully represent the marine microbiology community. Areas of concern include sampling-site priorities, levels of diversity to be sampled, coordination of across-taxon inventories, coordination with measurement of environmental contextual parameters, and databases. The group could focus on target areas of opportunity, coordinate the community, and help to assess and build infrastructure for supporting coordinated measurements, databases, communication, and education.

First and foremost, this would include representing and serving as a voice for the large community of marine microbiologists interested in the effort. The SAC would poll the community on a regular basis and distill a consensus that represents the community will. Other important tasks might include exploring potential synergistic programs and field efforts (see table below) and making appropriate preliminary contact to assess interests and establish connections. The group would explore and develop funding opportunities for these efforts including resources for research and support for both existing infrastructures and new facilities. Finally, this body would serve as a central conduit for coordination and education (*e.g.* webpages, books, outreach, *etc.*) that would communicate the importance and relevance of these activities to a broader audience. It would also maintain linkages to other microbial outreach programs such as those being initiated in the Microbial Observatory program at NSF and by the American Society for Microbiology, and ensure coordination with international initiatives.

HOW TO PROCEED

A new field of science is forming at the intersection of microbial genomics, physiology and ecology, biogeochemistry, and physical oceanography. Its connections to all of those fields will lead to a far-more-thorough and potent understanding of the processes that regulate the outputs from Earth’s carbon cycle. That in turn will elucidate interactions between the carbon cycle and the cycles of redox partners like oxygen and sulfur and of nutrients like nitrogen, phosphorus, and iron.

This synthesis of paradigms amounts to a major restructuring in the science of the global environment. It will succeed only when marine microbiologists, biogeochemists, and oceanographers embrace the ongoing revolution in genomics. It is essential to identify the most important questions and to reach consensus about an international strategy. Unanswered challenges include but are not restricted to:

- Determining the role and importance of recently discovered microbial photosynthetic-like pathways based on rhodopsin rather than chlorophyll.
- Understanding the energy sources used and processes catalyzed by the abundant archaea recently discovered in oceanic mid waters.
- Identifying relationships between basin-scale processes like the *El Niño* Southern Oscillation (ENSO) and the populations and activities of microorganisms.
- Determining the net significance of the large number of chemoautotrophic activities now recognized in diverse habitats.
- Establishing relationships between biogeochemical gradients (concentrations of trace micronutrients and of redox substrates) and microbial populations.
- Explaining interactions between prokaryotes and eukaryotes (symbioses, gut microorganisms) and among prokaryotes (quorum sensing, syntrophy, biofilms).

In each case, new genomic information, in combination with chemical and physical data, will yield important advances comparable to the discovery of the widespread distribution of bacterial rhodopsins in marine microbial genomes.

A global sampling of “the marine microbial genome” can provide fundamental information about selection and evolution in the microbial world. With more extensive exploration, it may be possible to complete a fully inclusive census of marine microorganisms and thus a complete roster of microbial processes. Ultimately, a continuum of systems at scales ranging from cellular to oceanic may be defined, with all of the included processes being treated quantitatively.

The field is flourishing intellectually but faces expensive challenges. Until now, the genomic techniques applied have not included high-throughput sequencing of total DNA. But the Institute for Biological Energy Alternatives (IBEA) has just completed a sampling and sequencing of total DNA from prokaryote-size particles filtered from Sargasso Sea surface water (southeast of Bermuda). More than one million genes have been recognized. The assembly to date indicates the presence of at least 1800 *phylotypes*. To extend this work, the IBEA is planning similar studies at depths to 100 m at more than 70 stations on a cruise track from the North Atlantic to the Caribbean Sea, through the Panama Canal, across the South Pacific to Easter Island, then back around Cape Horn and northward through the Atlantic.

The effort will provide a random survey of genes related to photosynthetic processes but will entirely omit both the deeper waters that comprise the bulk of the oceans and the whole range of seafloor environments, including hot and cold vents. Each sample will represent a single point in time, contrasting sharply with the time-series sampling that has proven so crucial to progress in the understanding of marine ecosystems. In future surveys, a much greater scientific return per dollar spent can result

from carefully organized sampling strategies developed in consultation with a broad community of stakeholders..

The number of workers in the field is relatively small. Few institutions have more than two or three research groups focused on marine microbiology. Funding is required to develop centers capable of addressing problems of the magnitude and range identified here. Such centers will include a large community of specialists so that graduate students are made fully aware of the scope and scale of microbial oceanography. They will provide multidisciplinary training and a choice among diverse research options. Attainment of those goals will require daily interactions between microbiologists and colleagues in physical oceanography and marine chemistry who can serve as consultants and collaborators.

Ongoing governmental programs favor extension of established techniques and lines of inquiry. For microbial oceanography, however, a *transformation* is desirable. The integrative studies envisioned here could be energized, and the transformation catalyzed, by private funding. Such funds can nourish investigations based on fruitful, smaller-scale techniques such as bacterial artificial chromosomes *and* plan and manage much more ambitious projects. Private funds could support a planning group that would develop and articulate the transformative vision. That group would inform itself through a series of meetings or conference sessions designed both to enhance communication among existing research groups and to recruit workers from related fields such as proteomics, bioinformatics, and conventional microbiology. It would establish and maintain a prototype MICROBIS that worked with and extended microbiological and marine-chemical data (an example – and starting point – is provided by the databases established at Woods Hole by the Joint Global Ocean Flux Study). Additional high priorities for this group would be the expansion of culturing facilities capable of handling marine microorganisms and the development of sea-going instrumentation for the collection and study of microbial samples.

These contributions would yield the foundation for a new era of microbial oceanography. In response, the scientific community would be drawn forward. The vision outlined here would be elaborated and amplified. Our understanding of the marine microbiota would indeed be transformed.