

SENSING THE LIVING OCEAN

By Ginger Armbrust

Executive Summary

A National Oceanographic Partnership Program (NOPP)-sponsored workshop entitled “Monitoring and Measurement of the Ocean Genome” was held in Washington, DC on March 7-8, 2005. About 50 participants attended the workshop, including representatives from academia, non-profit organizations, and government agencies. The overall goal of the workshop was to identify both near-term and future research areas and strategies to advance and apply the use of “genome-enabled” environmental biosensors in the ocean. These sensors are envisioned as being based either directly (DNA information to determine organism identity) or indirectly (RNA or protein information to determine organism function) on genome data. In addition, these sensors will likely be directly coupled to other biological sensors such as flow cytometers that detect and analyze single cells.

Three factors motivated the need for the workshop. First, the increasing availability of genome sequences for key microbial components of ocean ecosystems is transforming our understanding of this “hidden majority” of organisms. Second, ocean observatory facilities are a central feature of the future of ocean science, management and monitoring activities and will require associated instrumentation to serve as the necessary sensory organs. Third, it is generally recognized that biological sensors are the least well developed sensor type, and yet they are critically important because of the fundamental ways that biology integrates and records environmental information. All indications are that future genome-based technologies are heading towards “smaller, cheaper, faster,” which will make the genome-based sensors even more amenable to realization of the full potential of ocean observations. Collection of long-term biological data sets through use of this next generation of biological sensors, in combination with sensors that measure physical and chemical parameters, will provide new perspectives on the nature and effects of global climate change, pollution, and other environmental perturbations on the world’s oceans.

The general consensus of the workshop was that development of these sensors will require a bringing together of people with a broad range of expertise – for example, oceanographers, engineers, molecular and organismal biologists, computer scientists. Centers of Excellence are envisioned as a means of accomplishing this goal. A second critical recognition was the central role that databases/data streams/data storage will play in the actualization of these sensor types. The necessary cyber-infrastructure to accomplish these goals needs to be addressed up front simultaneously with sensor development. Workshop participants unanimously agreed that all data generated with these sensors would be made freely available in the public domain.

Introduction

The oceans constitute approximately 71% of the surface of the planet and are critical to almost all aspects of the global environment, including our own survival. The

oceans provide oxygen and food, corridors of commerce and are a key component of security. They modulate climate and drive major annual weather patterns and interannual events like El Nino and La Nina. In addition, the oceans have taken up about half the anthropogenic CO₂ generated from the burning of fossil fuels and cement manufacturing.

The biology of the ocean is a key component that determines ocean influence on the global climate system, and microorganisms are the major constituents of these biogeochemical processes. For example, phytoplankton carry out about half the photosynthesis that occurs on the planet despite the fact that these microscopic organisms represent less than one percent of total global photosynthetic biomass. The current state of knowledge of microbial processes in the ocean is derived from a variety of observations obtained from oceanographic vessels, near-shore platforms and coastal sampling. It is difficult and expensive to collect long-term time series observations at any given place and even more difficult to derive data sets with measurements of biogeochemical activity correlated with long-term ocean oscillations, so only a few data sets currently exist. Knowledge of the microbial oceans has been further hampered by the fact that most of the microbes cannot currently be cultivated and studied under laboratory conditions.

Recent advances in genomics, driven in large part by the Human Genome Project, have provided a rich new source of information and tools to understand ocean biology. Biological oceanographers and molecular biologists are using the existing large-scale genome sequencing capacity to determine the genome sequence and structure of a large number of marine organisms. Active programs are also in place to examine the ocean “meta-genome” and assess the wide genetic diversity of the ocean and demonstrate the wide ranging spatial distribution of this genetic diversity.

Genomics tools are empowering scientists to probe more deeply into key questions related to ocean biology in the face of global climate change and the potential effect of increased atmospheric CO₂. Genomics has opened a new window on understanding the capture of solar energy by ocean microbes and the flow of that energy through the ocean food web. Genomics-based approaches have provided much new knowledge even using the current sampling technologies. The rapid development of new sensor and array technologies has stimulated thinking about the potential of using sensors derived from the genomics information to make detailed measurements over longer time scales and under carefully monitored conditions. The potential impact of these tools is enormous. From sustained measurements of ecosystem changes to observing a variety of baseline parameters related to managing coastal and ocean areas, this new sensing paradigm could have an enormous impact. Genome-based sensors have a significant role to play in generating new knowledge about ocean biology and biogeochemistry, and in the operational observations of ocean conditions in marine managed areas, harbors and estuaries. The prospective prediction of toxic algal blooms and the potential to continuously monitor the influx of invasive species are additional operational benefits.

The collected new knowledge derived from the correlation of oceanographic measurements and biological and biogeochemical measurements from this new class of sensors will enable a new approach to ocean science -- predictive Microbial Oceanography. Changes in microbial populations are often early indicators of other physical changes in ecosystems and these can be used as surrogate markers for

impending changes. It is critical, therefore, that these genome-based sensors be developed, tested and deployed as components of integrated ocean observing platforms that will collect the variety of meta data that will enable the correlations that will lead to sound predictive science.

Current Ocean Observatory Efforts and Organizations

A number of recent reports, workshops, and National Research Council (NRC) studies have explored the general need and perspectives for ocean observatories from both research-oriented and ocean monitoring perspectives (see appendix for a list of references and web links). Two NRC Reports in particular discuss the science and management drivers, and operational and resource requirements in detail (*Illuminating the Hidden Planet: The Future of Seafloor Observatory Science*, National Research Council, 2000; *Enabling Ocean Research in the 21st Century: Implementation of a Network of Ocean Observatories*, National Research Council, 2003) Taken together, these reports indicate the general consensus that future ocean science, management and monitoring activities will require ocean observatory facilities, and the associated instrumentation that will serve as the “sensory organs” of these systems. The development of these sensory organs, especially those that leverage off the emerging molecular biological and genomic databases and technologies, are the main focus of this report.

Several federal agencies have been actively involved in planning efforts to define and scope the opportunities, challenges and resources required for the envisioned ocean observatory systems. The basic ocean observatory infrastructure includes undersea cables, buoys, deployment platforms, moorings and junction boxes, for power and two-way data communication to a wide variety of sensors at the sea surface, in the water column, and at or beneath the seafloor. Goals include the enablement of continuous observations at time scales from seconds to decades. Several interlinked initiatives from different agencies seek to complement and leverage technology, infrastructure and goals. In particular the National Science Foundation’s Ocean Observatory Initiative (OOI) represents research-driven interests that will be integrated with the National Ocean Partnership Program’s Integrated and Sustained Ocean Observing System (IOOS: www.ocean.us), an operationally focused national system for ocean monitoring. All these efforts are expected to contribute to the larger, international network of global ocean observatories (Global Ocean Observing System, GOOS; Global Earth Observing System of Systems; GEOSS: www.earthobservations.org).

Genome-enabled ocean science

Advances in microbial genome sequencing technologies are having a tremendous impact on the biological sciences, including biological oceanography. Over the past decade, the entire genetic blueprints of whole microbial genomes are available, and currently number in the hundreds. The list of published whole genome sequences now includes marine viruses, cyanobacteria, bacteria, archaea and protozoa from the photic zone, deep-water, benthic and hydrothermal vent habitats. These projects provide unprecedented insights into ways in which marine organisms are finely tuned to their environments. For example, the most abundant photosynthetic organism in the open ocean is the cyanobacterium, *Prochlorococcus*; comparative analyses of strains isolated from different depths reveals that adaptation to high light environments is correlated with a

reduction in gene content. Other studies have revealed that many microbes display unexpected combinations of metabolisms that characterize them as “opportunotrophs.” Less genomic information is currently available for marine eukaryotic organisms, primarily because of the increased complexity and size of the genomes. However, an increasing number of eukaryotic phytoplankton are making their way into sequencing queues (see for example, www.jgi.gov.org).

The most recent innovations in genomic approaches in the marine environment have been developed and applied to characterize whole microbial populations en masse, without requiring the cultivation of individual microbes. The environmental sequencing of the Sargasso Sea represents a spectacular example, resulting in millions of new genes discovered in the context of a single experiment. Applications include the genome analysis of uncharacterized taxa, expression of novel genes or pathways recovered directly from the environment, elucidation of community-specific metabolic properties, and cross comparisons of community gene content. These ongoing ocean genome survey efforts are contributing to rapidly growing genome databases that form a fundamental knowledge base for the development of genome-enabled biosensors.

Current state of the art for ocean-going biosensor instrumentation

Currently, there is a broad ‘spectrum of maturity’ with regard to available and routinely deployable ocean observatory sensors. At the mature end of the spectrum, instruments that measure meteorological parameters, salinity, temperature, pressure, current speed, light, and sound waves can now be fairly routinely deployed *in situ* in the ocean. Further along the ‘maturity spectrum,’ instruments for quantifying ocean chemistry are coming online, but others are still in developmental stages. (Continuous real-time sensors of CO₂, or bioactive compounds like nitrate for example, are now being deployed in the sea).

Arguably the least well developed instrumentation, and the most important from both scientific as well as societal perspectives are biological sensors. For ocean science, few if any robust bio-monitoring devices are currently available. Existing *in situ* biological instrumentation consists largely of bio-optical devices that measure light scattering as a proxy for biological particles, or chlorophyll *a* fluorescence as a proxy for phytoplankton. Two examples of in-development biological sensors are *in situ* flow cytometers that measure light scatter and fluorescent properties of individual particles (for a marketed version, see <http://www.cytobuoy.com/>) and an Environmental Sample Processor (EPS) that can detect the DNA signatures of different species. If ocean observatories are to reach their full potential, the sensitivity, specificity and reliability of ocean-going biological sensors will require intense and accelerated development. Development efforts for autonomous biological sampling and sensing instruments are nascent and much more effort will be required to produce readily available and robust biological sensors.

On land, the genomic revolution and spin-off technologies are now greatly enhancing detection and quantification of microbial species. Genome enabled sensing technologies include quantitative polymerase chain reaction (PCR) techniques that allow enumeration of small quantities of specific genes, DNA microarrays for simultaneously detecting tens of thousands of genes (and their expression), and “proteomic” approaches that can be used to characterize and quantify thousands of proteins in a single sample.

Moreover, current technologies are rapidly moving in the direction of single cell analyses. Many of these approaches are currently being applied to the study of marine organisms, but none are used to study organisms *in situ*.

Cyber-infrastructure issues

The ocean observatories as currently envisioned, even without the inclusion of genome based sensor systems, will generate enormous data streams that will require massive storage capacity and new algorithms for analysis and synthesis. When operational, the genomics enabled systems will be a component of that data stream. However the analysis of these data will need to go well beyond the data sets derived from routine operation. The identification of which microbes are present will require comparison to external (but linked) databases of genetic information, gene function and related factors. The databases required to support this effort are not a future development but need to be developed now as early sensor designs are contemplated. Some data sets already exist, but in formats that are difficult to use and others that are essentially not available as open resources. The database requirement can be broken into four areas, not all of which are specific to marine organisms and sensor technology, but sensor development and deployment will rely upon each of these.

1. Microbial genome sequences and their associated annotation. This includes the accurate identification of the source organism and represents as accurately as possible the complete identification of genes, gene families and metabolic pathways.
2. A database of gene annotation data that is not organism specific but is gene sequence specific. This database would include gene structure and function data associated with specific gene sequences from all well characterized systems. The goal is to enrich the gene prediction capabilities for annotating new sequences with unknown function. The goal of using genome-enabled sensor systems is very dependent on the accuracy of the annotation of the microbial targets and probes.
3. A database which fully captures the environmental meta-genomic data that is increasingly being collected by precursor projects to the envisioned sensor systems. At present, millions of gene fragments are being sequenced from locations throughout the world's oceans. These data constitute a large and valuable resource but are currently incomplete because of the loss of the associated meta-data from the collection and processing.
4. Once genomic enabled sensor systems are operational, a new database and data management system will be needed to monitor the continuous data streams and enable the adaptive sampling goal for these deployments. Current real-time analysis systems such as Road-net (<http://roadnet.ucsd.edu/>) could be adapted; however, the analysis algorithms will be dependent on the two marine microbe specific databases listed above. To answer the question of which organisms are present at a particular time and, even more challenging, to determine their current metabolic activity over time will require passing these data through the reference databases.

The development of internationally accepted database architectures interlinked from one location and enforced by scientific journals and funding agencies (analogous to

GenBank) will be a major factor that determines the success of genomics-enabled ocean sensors.

Scientific Goals

There are a wide variety of applications (both research and long-term observational) that will benefit from the deployment of genomics-enabled sensors. However, in almost all cases the value of the sensor data is linked to the quality and completeness of the associated meta-data. In the scientific questions that follow it is assumed that a rich meta-data collection accompanies the sensor measurements. When marine scientists ask, for example, which microbes are present in a sample, it is implied that the full question includes “at this time, in this place, under these conditions, etc.” It is essential that the sensors discussed here be considered as components of a broader ocean observing environment. To be maximally useful, the genome-enabled sensor systems should be closely coupled with simultaneously collected physical and chemical oceanographic information. The data stream from the genomics-enabled sensor systems must be closely linked (cross referenced) to the base line data stream from a richly instrumented observatory platform. Likewise, the inclusion of these genome-enabled sensor systems in ocean observatory installations will expand the utility of the observatory and enable a far greater return on the infrastructure investment.

Scientific questions

1. Where are different organisms found and when?

Marine environments are dynamic and complex, with conditions that are rarely stable. Marine ecosystems are composed of diverse communities of interacting organisms. Identifying which organisms are where and when is critical for understanding how marine ecosystems function and therefore for predicting how ecosystems will respond to changes in environmental conditions from either shorter-term (e.g., eutrophication of coastal waters, over-fishing) or longer-term perturbations (e.g., global climate changes). The growth of marine microbes is directly linked to environmental conditions and thus monitoring changes in species composition over time provides the most basic “read-out” of the environment. The ability to collect this type of data over long time series will ultimately provide insights into the resiliency/stability of marine ecosystems. Furthermore, species distributions are key components of coupled ocean/ecosystem models to predict future carbon cycles.

There are multiple challenges to answering this simple question of “who is where and when?” The first critical step is to identify which organism is present in any sample of water. At this point, only unicellular cyanobacteria with their unique combination of size and pigments can be routinely identified in a high throughput manner using flow cytometry. None of the other prokaryotes can be routinely identified in real time and instead requires molecular probes based on DNA sequencing that logistically, can be used with only a subset of samples. The smaller size classes of eukaryotic algae are morphologically indistinguishable and the larger size classes of eukaryotic algae are frequently identifiable only with the use of electron microscopy. Identification of marine microbes is increasingly dependent upon the ability to determine the DNA sequence of targeted genes. The continue expansion of the molecular databases necessary for the identification of different microbes will facilitate using genome-enabled sensors to address the critical question of “who is where and when.”

2. What are they doing?

Marine microbes are the engines of the biosphere, driving the cycling of the elements and maintaining the balance of Earth's biogeochemical cycles. Their activities help regulate the concentration of greenhouse gases, oceanic productivity and global climate. However, our current understanding of which microbes are responsible for specific biogeochemical cycles, how fast they operate, how they interact with one another and how environmental variability influences their function is rudimentary, at best. Surveys of the genomic diversity in the ocean are now describing millions of new genes, and by inference many new microbial functions, whose significance is only now beginning to be understood. By using these genes as proxies for specific microbes, as well as microbial functions such as nitrogen fixation or sulfate reduction, new genome-enabled approaches can be used to track and quantify microbial activities in the environment.

Questions that might be addressed using genome-enabled sensors deployed in the ocean environment include: Are there specific keystone species in critical geochemical cycles that regulate material and energy flux in specific biogeochemical cycles? Can particular microbial species serve as and provide early warning systems, or effective environmental buffers of environmental perturbation, natural or anthropogenic? What are the short-term and long-term variabilities of these indigenous species, in relation to environmental fluctuations, both natural and anthropogenic? Is there functional redundancy in different ecosystems?

Despite the importance of these questions, the tools for sensitively and specifically monitoring and measuring microbial function and activity in the environment are still lacking.

3. How do they respond?

The development of this new class of sensors, especially when embedded in a rich sensor system will enable a whole new sampling paradigm. Through real time detection of particular event-scale sampling, protocols can be modified "on the fly" in an adaptive sampling protocol. These changes can be event driven and linked to a variety of observations from the observing platform. One can envision changing the sampling protocol in response to the detection of a species of toxic algae to focus the sensor system on monitoring whether a full scale bloom is in progress or if this is simply a low level event. Likewise, unusual temperature patterns could trigger a new set of observations related to the change in microbial populations and their physiological activity. The power of such event-driven strategies is extraordinary and opens broad new areas of investigation and knowledge generation about marine ecosystems, their functions and the factors that influence their ability to provide the ecosystem services that influence the global climate. Inherent in this discussion is the ability to collect long-term biological data sets that will be provided by these new sensors.

Recommendations

The development of genome-enabled sensors will usher in a new era in ocean science. The convergence of scientific knowledge (e.g., genome-based information), technical capabilities (e.g., miniaturization, computing power, communication), and

societal needs (e.g., the increasingly apparent footprint of humans on ocean communities) have brought the potential development of these sensors to the forefront. Technologies developed for the marine-based sensors described in this report will also be transferable to the study of land-based ecosystems.

Enhanced innovation in biological sensor development requires interactions between engineers, oceanographers, molecular biologists, organismal biologists, computer scientists and mathematicians (to name just a few). Up-front considerations of the cyber-infrastructure necessary for data stream compatibility (both nationally and internationally) are essential to facilitate a rapid transition from data collection to data manipulation and modeling.

The following recommendations will help launch this new area of genome-enabled biological sensors.

- **Individual/small number of investigator projects**

Much of the innovation to drive genome-enabled sensor development will come from individual investigators or small groups of investigators. Projects of these types could include, for example, development of necessary probes for determining who is there (e.g., specific organisms or genes) and what are they doing (e.g., key biogeochemical cycles), new technologies for detecting biological functions, and new hypotheses about biological processes (e.g. ecosystem function, cellular processes).

- **Network of Centers of Excellence to enhance sensor development**

Development of genome-enabled oceanographic sensors will require the interface between investigators from different disciplines that do not typically interact: bio-instrumentation and biomedical engineers, marine engineers, oceanographers, biologists, molecular biologists and computer scientists, for example.

A primary recommendation is to support Centers of Excellence that bring together scientists from different disciplines, including investigators from the private sector. Not all investigators that participate in a Center need to be co-located. Investigators from different Centers will interact with one another to generate a Network that will set sensor standards/compatibility and therefore this Network will serve as a point of compliance for sensor use on observing systems. In addition to the tangible outcome of sensor development, the Network will also serve to train the next generation of researchers in cross-disciplinary research. This would serve as the starting point for this new program.

- Develop a **post-doctoral program** to train the next generation of cross-disciplinary scientists and engineers.
- **Inter-agency effort to build inter-dependent database** of annotated microbial sequences that are coordinated with environmental meta-genomic data and its associated environmental metadata. This database is critical as a reference resource for interpretation of sensor data. Importantly, these kinds of database needs expand well beyond the oceanographic community. This is a top priority that should begin immediately. This database is envisioned as a supplement to

the NCBI GenBank, which will be the ultimate permanent repository of genomic information.

- Promote the **transition from sensor development to sensor deployment** by facilitating the commercialization of sensors. A similar program known as “Topic 4B of the ONR BAA 04-022: The National Oceanic and Atmospheric Administration’s National Data Buoy Center (NDBC) partnerships with academic, industry/NGO and government entities” is already in place to enhance commercialization of sensors.

Scientific Milestones

- 1) Establishment of a repository for sequence data and metagenomic data that is generally available to the community.
- 2) Implementation of reliable platforms with diverse “sensor” types: for example, DNA, mRNA, metabolic indicators.
- 3) Implementation of sensors by broader community.
- 4) Incorporation of sensors into Ocean Observation Systems with appropriate data stream management.
- 5) Development of scientific questions as well as monitoring (addressing global change).

List of workshop participants, including agency reps

Last Name	First Name	Affiliation	Email
Steering Committee			
Armbrust	Virginia	University of Washington	armbrust@ocean.washington.edu
Bowler	Chris	Stazione Zoologica, Italy and CNRS, France	cbowler@biologie.ens.fr
DeLong	Edward	Massachusetts Institute of Technology	delong@mit.edu
Karl	David	University of Hawaii	dkarl@hawaii.edu
Kingsbury	David	Moore Foundation	david.kingsbury@moore.org
Vaulot	Daniel	CNRS, France	vaulot@sb-roscoff.fr
Meeting Participants			
Ammerman	Jim	Rutgers University	ammerman@imcs.marine.rutgers.edu
Amthor	Jeff	Department of Energy	jeff.amthor@science.doe.gov
Barber	Richard	Duke University	rbarber@duke.edu

Berges	John	University of Michigan	berges@uwm.edu
Boyen	Catherine	CNRS Roscoff, France	Boyen@sb-Roscoff.fr
Brezonic	Patrick	National Science Foundation	pbrezoni@nsf.gov
Bruyant	Flavienne	Dalhousie University	flavienne.bruyant@dal.ca
Caron	David	University of Southern California	dcaron@usc.edu
Cullen	John	Dalhousie University	john.cullen@dal.Ca
De Vargas	Colomban	Rutgers University	vargas@imcs.rutgers.edu
Delaney	John	University of Washington National Oceanic and Atmospheric Administration	jdelaney@u.washington.edu
Dickhoff	Walter	Administration	walton.w.dickhoff@noaa.gov
Drell	Daniel	Department of Energy	daniel.drell@science.doe.gov
Eccles	Catherine	European Commission	Catherine.Eccles@cec.eu.int
Eckman	James	Office of Naval Research	eckmanj@onr.navy.mil
Fell	Jack W.	University of Miami	jfell@rsmas.miami.edu
Garrison	Dave	National Science Foundation National Oceanic and Atmospheric Administration	dgarriso@nsf.gov
Goodwin	Kelly	Administration	Kelly.Goodwin@noaa.gov
Iglesias- Rodriguez	Debora	University of Southampton	dir@soc.soton.ac.uk
Isern	Alex	National Science Foundation	aisern@nsf.gov
Kane	Matt	National Science Foundation	mkane@nsf.gov
Lacher	Thomas	Conservation International	t.lacher@conservation.org
Massion	Gene	Monterey Bay Aquarium Research Institute	magene@mbari.org
Meldrum	Diedre	University of Washington	deedee@ee.washington.edu
Moran	Mary Ann	University of Georgia	mmoran@uga.edu
Olson	Rob	Woods Hole Oceanographic Institution	rolson@whoi.edu
Palenik	Brian	Scripps Institution of Oceanography	bpalenik@ucsd.edu
Palmissano	Anna	Department of Agriculture	apalmisano@csrees.usda.gov

Partensky	Frédéric	CNRS Roscoff, France	partensk@sb-roscoff.fr
Paul	John	University of South Florida	jpaul@marine.usf.edu
Perry	Mary Jane	University of Maine	perrymj@maine.edu
Preston	Christina	Monterey Bay Aquarium Research Institute	preston@mbari.org
Rappe	Mike	University of Hawaii	rappe@hawaii.edu
Raven	John	University of Dundee, Scotland	j.a.raven@dundee.ac.uk
Rocap	Gabrielle	University of Washington	rocap@ocean.washington.edu
Rohwer	Forest	San Diego State University	frohwer@sunstroke.sdsu.edu
Rokhsar	Dan	Joint Genome Institute	dsrokhsar@lbl.gov
Rusch	Doug	Venter Institute	drusch@venterininstitute.org
Saino	Toshi	University of Nagoya, Japan	tsaino@ihas.nagoya-u.ac.jp
Sayler	Gary	University of Tennessee	sayler@utk.edu
Schindel	David	Smithsonian Institute	schindelD@si.edu
Sosik	Heidi	Woods Hole Oceanographic Institution	hsosik@whoi.edu
Sussman	Michael	University of Wisconsin	msussman@wisc.edu
Taylor	Philip	National Science Foundation	prtaylor@nsf.gov
Tirpak	Liz	Department of State	TirpakEJ@state.gov
van den Engh	Ger	Institute for Systems Biology	engh@systemsbiology.org
Ward	Bob	CSIRO Marine Research	'Bob.Ward@csiro.au'
Zehr	Jon	University of California, Santa Cruz	zehrj@emerald.ucsc.edu

Program Office Staff

Beach	Reg	National Oceanographic Partnership Program	rbeach@coreocean.org
Feldman	Michael	National Oceanographic Partnership Program	mfeldman@coreocean.org