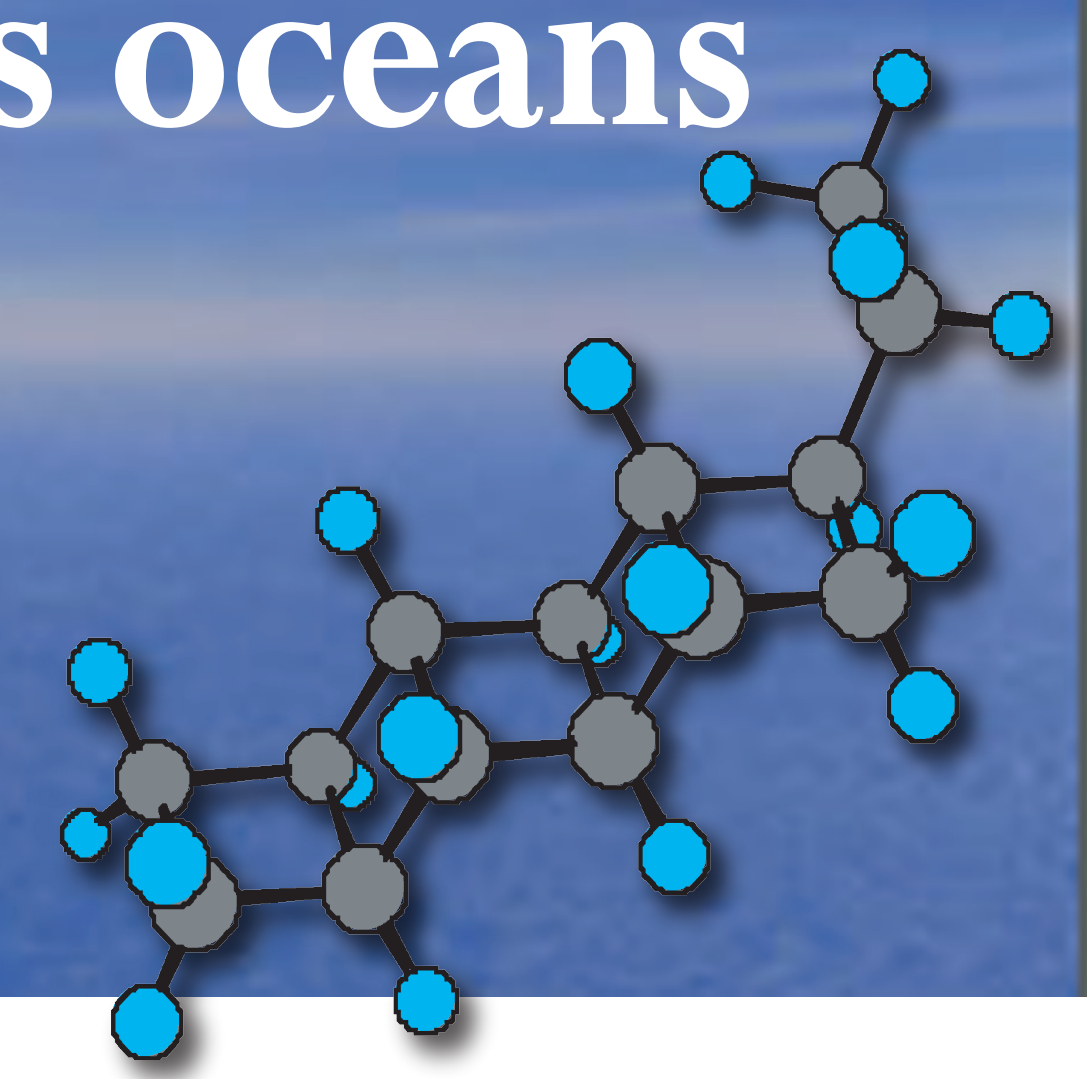


# The International Census of Marine Microbes (ICoMM) and a strategy for exploring microbial diversity throughout the world's oceans

Julie Huber, Hilary Morrison, David Mark Welch, Sue Huse, Phil Neal, and Mitchell Sogin  
Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole MA 02543  
sogin@mbl.edu



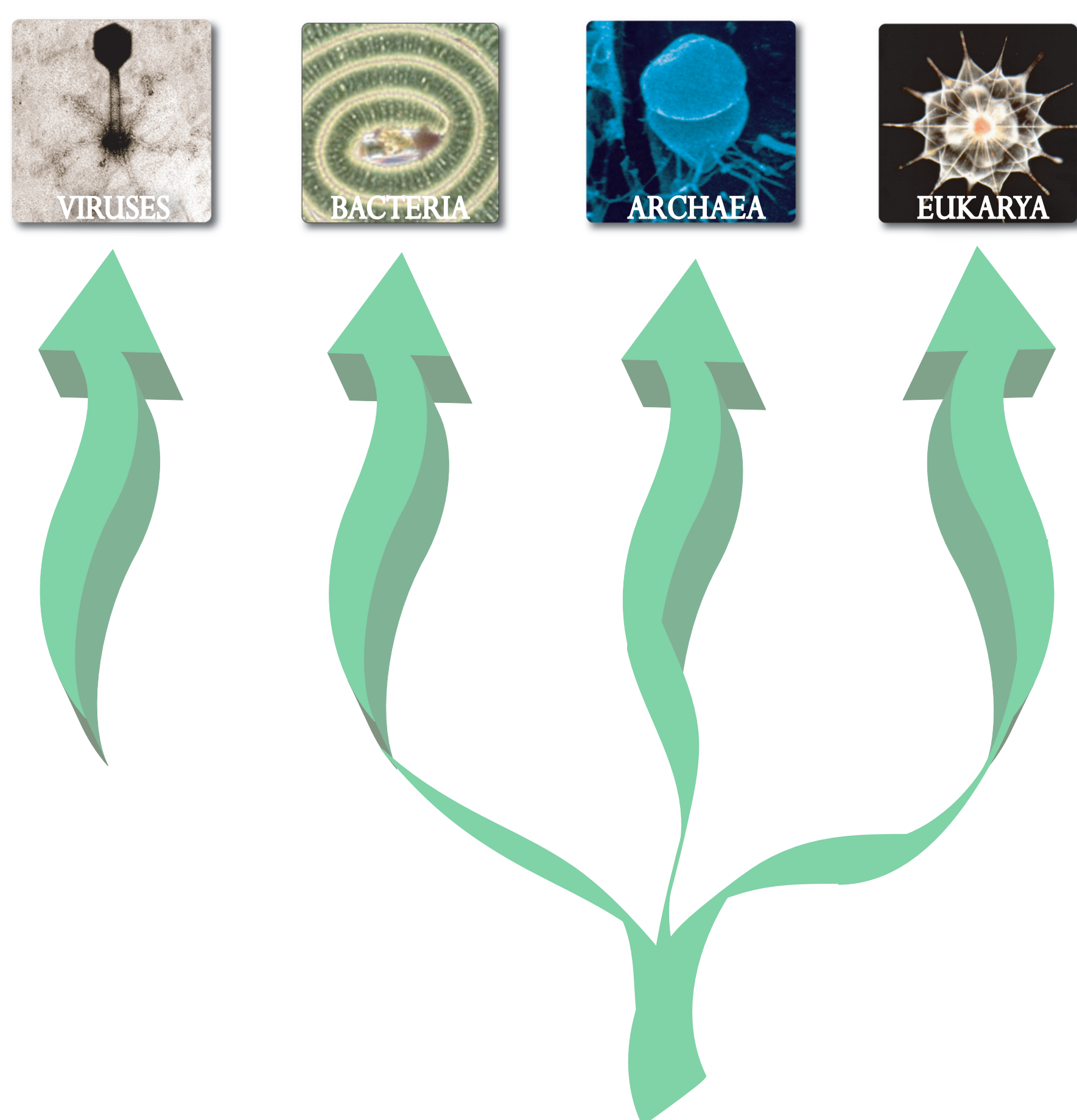
...Unveiling the Ocean's Hidden Majority



## Abstract

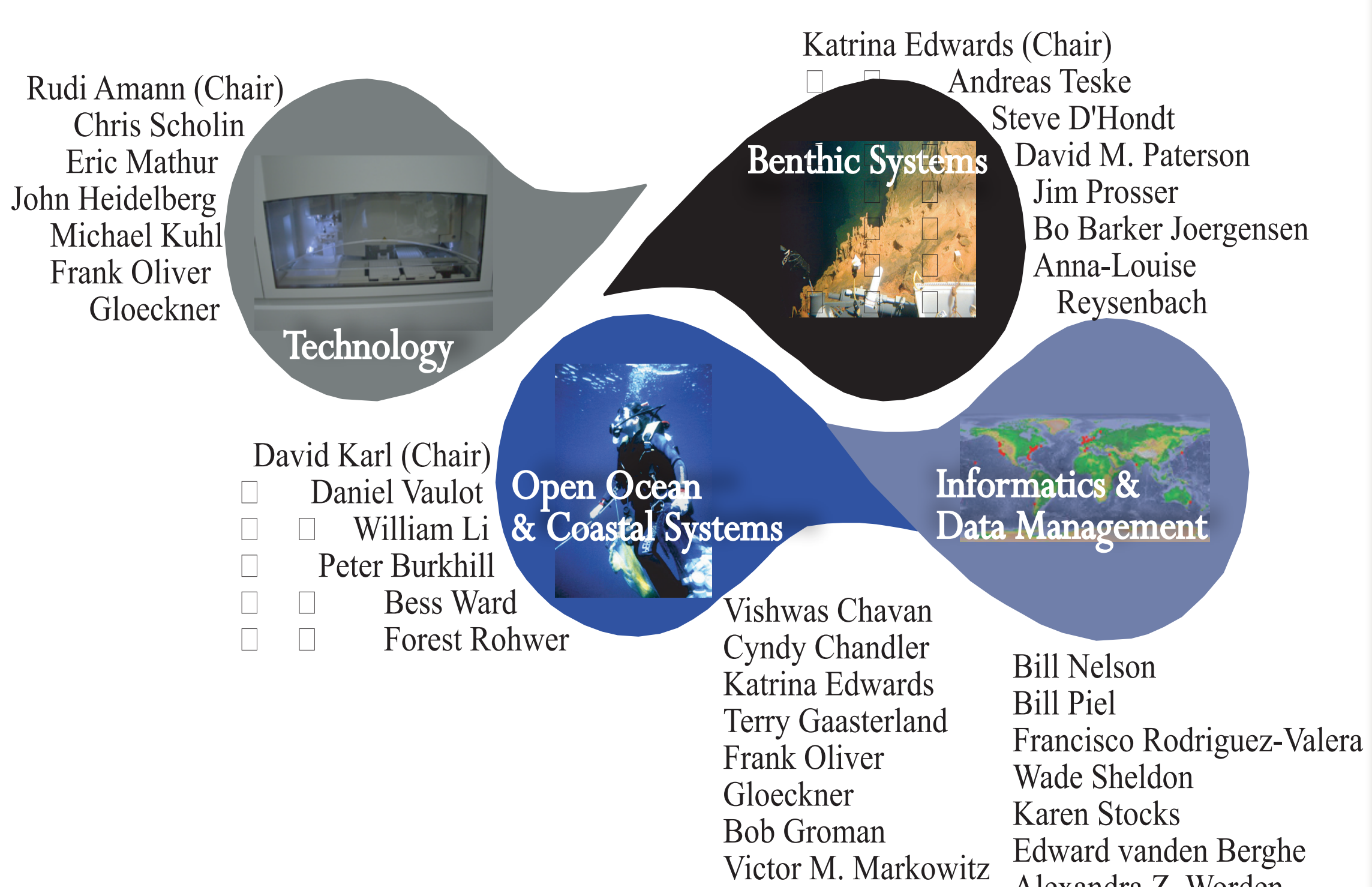
The world's oceans are teeming with microscopic life forms. For more than three billion years, microscopic factories - initially anaerobic and later aerobic - have served as the essential catalysts for all of the chemical reactions within the biogeochemical cycles. Although early cultivation studies described microbial population size to be only a few hundred cells per milliliter of seawater, the staining of cells with DNA-binding dyes (DAPI and acridine orange) coupled with epifluorescence microscopy demonstrated that nominal cell densities exceed 10<sup>5</sup>/ml of sea water. Extrapolations of these numbers predicts that the oceans harbor 3.6 x 10<sup>29</sup> microbial cells with cellular carbon of ~3 x 10<sup>17</sup> grams. Given the enormous populations of microbes with seemingly unlimited metabolic diversity, the accumulation of mutations during the past 3.5 billion years should have led to very high levels of genetic diversity and phenotypic variation. Yet, traditional microbiological methods have described only 30,000 protists and fewer than 5,000 kinds of microbes. The International Census of Marine Microbes (ICoMM) (<http://icomm.mbl.edu/>) is a field project of the Census of Marine Life (COML) (<http://www.coml.org/>) whose goal is to report what is known, what is unknown but knowable, and what may never be known about the biodiversity of marine micro-organisms. ICoMM supports four working groups: (i) Open ocean and coastal systems, (ii) Benthic systems, (iii) Databases, and (iv) Technology. To be successful, ICoMM must work through coordinated activities of marine microbiologists throughout the world and it will be important to develop a high-throughput, molecular-based paradigm for describing microbial populations. We have made a significant step forward through the use of tag sequencing strategies. In a single experiment we captured information about the phylogenetic diversity of more than 100,000 organisms. Rarefaction analysis of these data reveals that microbial diversity in the oceans is at least two orders of magnitude greater than previously reported.

## Who are the marine microbes?



The long-term goal of ICoMM is to coordinate an international census of marine microbes that will address grand questions in studies of marine microbial ecology, biogeochemical processes, and microbial evolution. ICoMM is an international community initiative and will establish the infrastructure for this work and obtain long-term funding for this effort.

ICoMM has established the following scientific working groups:



## What is the biogeography of marine microbes?

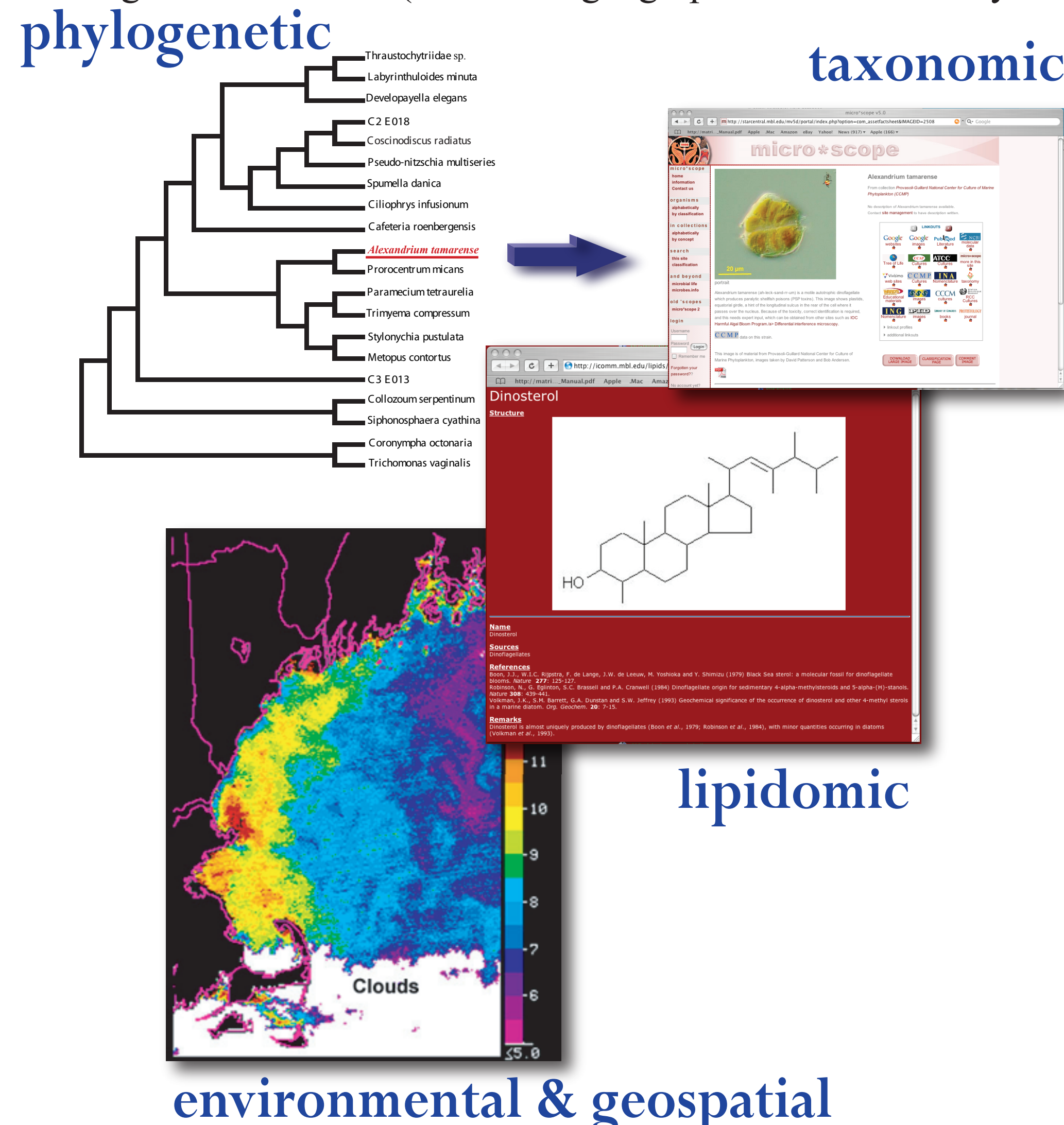
- What governs the evolution of marine microbial lineages within complex marine communities?
- Why do marine microbial consortia retain functionally equivalent but genetically distinct lineages?
- How does genotypic diversity shape phenotypic diversity, and how does this diversity influence the functioning of marine ecosystems?

These are examples of some questions that drive the ICoMM initiative.

## The Known Unknown & Unknowable

Well-known microbial communities exist in the light-dominated surface waters of the ocean. However, much less is known about the mid- and deep-water column microbial communities. The TRANSAT cruises (indicated by yellow and red dots in the map to the left) have accumulated over 250 samples from just below the euphotic zone to the very bottom (> 4000m). These samples will allow us to examine how microbial populations vary at different depths and locations in the oceanic conveyor belt of the North Atlantic Deep Water (NADW). Deep-sea hydrothermal vents were only discovered in the late 1970's, and the emerging picture in the microbial ecology of hydrothermal vents is a phylogenetically and physiologically diverse population intimately linked with the geochemistry of the hydrothermal systems. However, little is known about the in-situ diversity or the spatial and temporal variability of these microbial populations, especially in the seafloor habitat and diffuse vents. Determining the diversity and variability of these communities is part of a current pilot project for ICoMM, focusing on diffuse vents from Axial Seamount on the Juan de Fuca Ridge. Shown at right is Marker 52, Sample FS396.

**microbis** is ICoMM's distributed database source that will bring together molecular environmental, geospatial, and taxonomic information to address key questions in microbial oceanography within a framework that integrates into OBIS (Ocean Biogeographic Information System).



## Emerging Technology

A new strategy for deeply sampling microbial communities is 454 Life Sciences technology implemented on a Roche Genome Systems 20 Sequencer (GS20). It combines pyro-sequencing of immobilized templates with a massively parallel high-throughput system for generating very accurate DNA sequences ~100-120 bp in length. The 454 Life Sciences protocols require few manipulations and each tag is two orders of magnitude less expensive to generate than any other sequence-based analysis. The system relies upon an emulsified amplification and a PicoTiterPlate™ device for sequencing individual PCR amplicons attached to beads. This allows the analysis of thousands of amplicons without the preparation of recombinant libraries or sequencing templates.

To evaluate the utility of 454 Life Sciences technology for tag sequencing, we generated amplicon libraries for six paired TRANSAT samples and two diffuse flow samples from Axial Seamount (Table 1). We employed bacterial-specific primers for the hypervariable V6 region of the 16S rRNA gene.

ID	Site	Latitude, Longitude	Depth meters	Cells/ml	Volume Filtered
53R	Labrador Seawater	58.280°N, -29.133°W	1460	5.40 x 10 <sup>7</sup>	1000
55R	Oxygen Minimum	58.500°N, -29.133°W	500	1.78 x 10 <sup>7</sup>	1000
112	Lower Deep Water	50.400°N, -25.000°W	4121	3.92 x 10 <sup>7</sup>	1000
115	Oxygen Minimum	50.400°N, -25.000°W	550	1.49 x 10 <sup>7</sup>	1000
137R	Labrador Seawater	60.900°N, -38.616°W	1710	3.33 x 10 <sup>7</sup>	1000
138R	Labrador Seawater	60.900°N, -38.616°W	710	5.12 x 10 <sup>7</sup>	1000
FS312	Basin City	45.916°N, -129.983°W	1529	1.21 x 10 <sup>7</sup>	2000
FS396	Marker 52	45.943°N, -129.985°W	1537	1.57 x 10 <sup>7</sup>	1003

Roche generated sequences from both ends of the V6 hypervariable region and delivered ~118,000 tag sequences. The number of reads/sample ranged from 6,605 to nearly 23,000 sequences. For our entire collection of 454 reads, 25 percent are identical to at least one sequence in our V6 reference database, 40% are no more than 3% different, and ~75% are no more than 10% different than a sequence in our database. From this simple summary we conclude that it will be possible to identify the phylogenetic affinity of most of our tag sequences, as shown in the pie chart below for diffuse flow sample FS396. At the same time this analysis provides strong evidence of the considerable novel bacterial diversity we can detect. We also generated distance matrices that contain all pairwise comparisons of tag sequences from a sample. These served as input to DOTUR, a program that identifies OTUs according to similarity groupings. The species rarefaction curve shown below is from the DOTUR analysis of the diffuse flow sample FS396. Even when similarity groups represent relatively large genetic distances (5 or 10% similarity) the diversity will significantly increase with additional sampling.

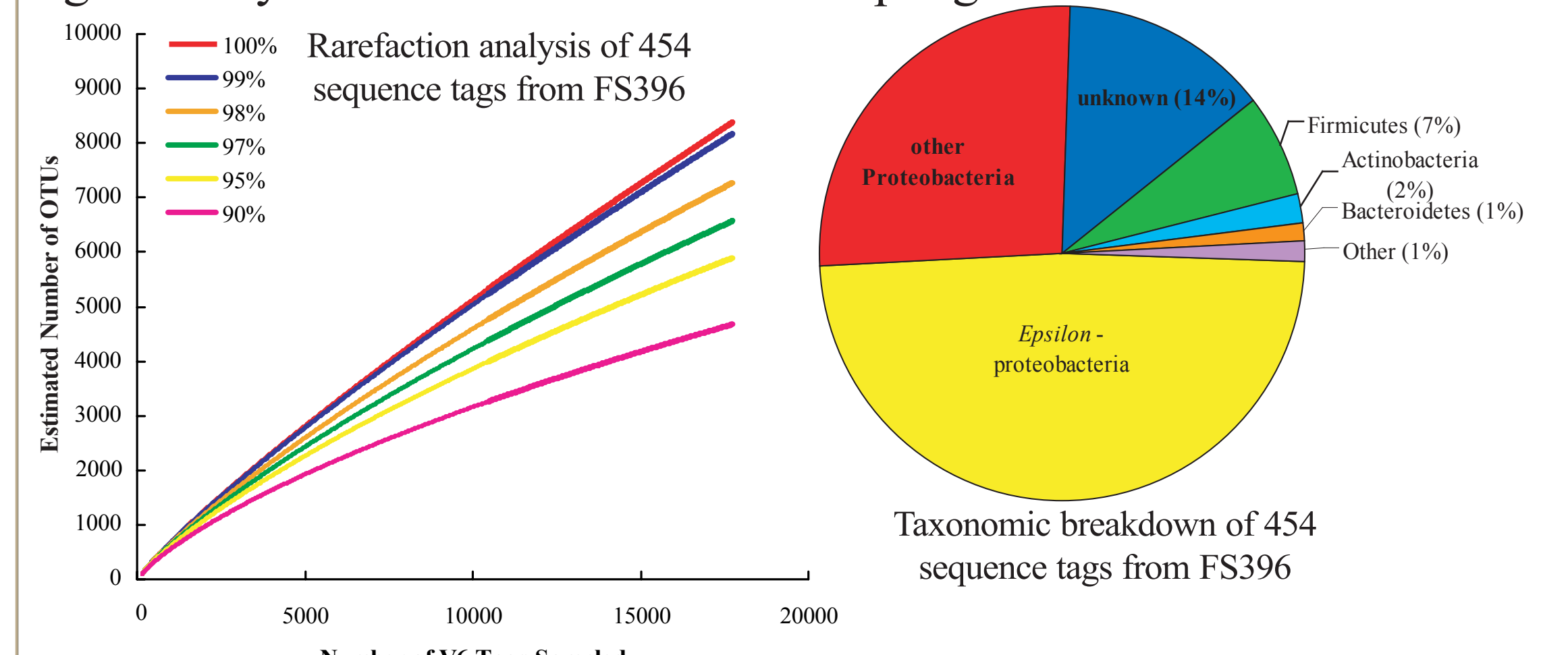


Table 2 reports the number of OTUs, ACE, and Chao1 estimators of species diversity for groups defined by unique sequences (no variation), 1 percent and 3 percent difference. The species diversity estimators ACE and Chao1 are orders of magnitude greater than previous DOTUR analyses of any microbial population, and the rarefaction curves predict even greater levels of diversity.

Dist	OTUs	ACE	Chao1	OTUs	ACE	Chao1	OTUs	ACE	Chao1
0	118	16,488	14,565	2,527	14,550	12,993	1,954	6,826	6,271
1%	4,995	2,595	16,488	14,565	2,527	14,550	12,993	1,954	6,826
3%	13,906	7,006	46,144	40,638	6,814	29,941	35,099	3,480	21,681
5%	21,270	9,459	56,213	51,948	9,271	41,276	47,988	4,238	26,620
10%	30,987	11,593	67,944	62,428	11,593	51,551	59,111	5,209	32,276
15%	38,898	13,591	81,477	76,003	14,401	60,822	71,557	6,782	37,747
20%	44,375	15,055	91,402	83,875	16,891	68,438	79,786	7,666	42,247
25%	48,845	16,061	99,201	91,202	18,611	74,611	85,910	8,310	46,276
30%	52,311	16,817	105,241	97,261	20,000	79,551	90,500	8,800	49,000

Using Roche's Genome Sequencer 20 System, we can now determine the range of genetic diversity and relative numbers of different microbial organisms at sampling sites throughout the world's oceans. In addition, we can now characterize marine microbial communities and document the presence of low abundance microbial taxa that traditional surveys fail to detect. With this technology, ICoMM will continue to engage the broader community of microbiologists with complementary interests in microbial diversity, evolution, biogeography, and their functional roles in marine systems.

## The ICoMM Membership

- Principal Investigators:**
- Mitchell L. Sogin, Marine Biological Laboratory (MBL), USA
  - Jan W. de Leeuw, The Royal Netherlands Institute for Sea Research (NIOZ)
- Secretariat:**
- Linda Amaral Zetler (MBL)
- Scientific Organizing Committee:**
- Gerhard Herndl (NIOZ)
  - David J. Paterson (MBL)
  - Stefan Schouten (NIOZ)
  - Lucal Stal (NIOZ)
- Scientific Advisory Council:**
- John A. Baross, University of Washington, USA
  - Robert Anderson, Bigelow Laboratory for Ocean Sciences, USA
  - Antje Boetius, Max Planck Institute (MPI), Germany
  - Victor Ariel Gallardo, University of Concepcion, Chile
  - Carlos Pedrés-Alto, Institut de Ciències del Mar (ICM), Spain
  - Francisco Rodriguez-Valera, Universidad Miguel Hernandez de Elche (UMH), Spain
  - Koki Horikoshi, JAMSTEC, Japan