

Metagenomics Reveals Microbial Diversity

Author: BEARDSLEY, TIMOTHY M.

Source: BioScience, 56(3) : 192-196

Published By: American Institute of Biological Sciences

URL: [https://doi.org/10.1641/0006-3568\(2006\)056\[0192:MRMD\]2.0.CO;2](https://doi.org/10.1641/0006-3568(2006)056[0192:MRMD]2.0.CO;2)

BioOne Complete (complete.BioOne.org) is a full-text database of 200 subscribed and open-access titles in the biological, ecological, and environmental sciences published by nonprofit societies, associations, museums, institutions, and presses.

Your use of this PDF, the BioOne Complete website, and all posted and associated content indicates your acceptance of BioOne's Terms of Use, available at www.bioone.org/terms-of-use.

Usage of BioOne Complete content is strictly limited to personal, educational, and non - commercial use. Commercial inquiries or rights and permissions requests should be directed to the individual publisher as copyright holder.

BioOne sees sustainable scholarly publishing as an inherently collaborative enterprise connecting authors, nonprofit publishers, academic institutions, research libraries, and research funders in the common goal of maximizing access to critical research.

Metagenomics Reveals Microbial Diversity

TIMOTHY M. BEARDSLEY

Faster sequencing machines and computers are opening up new vistas for genomics. Better understanding of environmental processes could be one result.

A glorious Indian summer was having its last hurrah this past October as 400 research leaders gathered at a beachside resort in Hilton Head Island, South Carolina, for the annual conference organized by genomics innovator J. Craig Venter. Venter has directed the heady mix of frolic and big-budget biology for the

past 17 years, but last fall's meeting was to have a somewhat different emphasis. It had a new name—"Genomes, Medicine, and the Environment Conference"—that signaled Venter's developing interest in metagenomics: large-scale sequencing of genes sampled directly from the environment.

Venter is best known as the brash scientific rebel who, as head of Celera Genomics, had seemed poised to beat the government to a draft sequence of the human genome. He successfully adapted the technique of whole-genome shotgun sequencing for the task, a computationally demanding approach previously used only for much smaller genomes. But Venter has moved on since the politically engineered announcement at the White House in 2000 that the race had ended in a tie. Since his ouster as head of Celera two years later, he has increasingly turned his attention to applications of genomics that might yield solutions to global climate change and other such daunting challenges. To pursue that agenda, Venter established several institutes that have now merged into one, the J. Craig Venter Institute in Rockville, Maryland.

Kicking off the proceedings at Hilton Head was an overview of institute projects to shotgun sequence DNA from environmental microbes. Shotgun sequencing breaks up an organism's DNA into myriad short fragments that can each be quickly analyzed, then virtually reassembles the whole genome in a computer. The Venter Institute's effort includes a high-profile, round-the-world ocean sampling expedition in the 95-foot sloop *Sorcerer II*. Workers on the vessel have spent much of the past 18 months sampling and filtering seawater from a few feet down near coastlines and



J. Craig Venter's 95-foot sloop Sorcerer II recently completed a global circumnavigation sampling microorganisms in surface seawater. Shotgun sequencing of the DNA in the samples has revealed unprecedented levels of genomic diversity. Expedition leader Venter (left) and Anthony Knap, director of the Bermuda Biological Station for Research, are pictured on board.

Photograph courtesy of J. Craig Venter Institute.

at sites in the open ocean spaced 200 miles apart. The filters are frozen and sent to Rockville, where the DNA is shotgun sequenced by scores of automated analyzers and a nest of computers.

Because fewer than 1 percent of the microbes in sea water can now be cultured, DNA sequences are assembled and classified without, for the most part, the intrusion of traditional notions of species. Researchers group the sequences by their similarity to previously known genes, an exercise that yields pointers to new, possibly metabolically related, gene families and, not infrequently, apparent new species. A pilot project conducted in the Sargasso Sea, which had been thought relatively nondiverse because of its low nutrient levels, yielded 150 new species of bacteria and over 1.2 million new genes, including 782 new rhodopsin-like photoreceptors, according to a paper published in *Science* (2 April 2004).

Genomics goes global

The *Sorcerer II* survey has now found astonishing levels of genetic diversity worldwide. Karin Remington recounted how the Venter Institute team was unable to put an upper limit on the species richness of all but the most extreme environments sampled, because even with the millions of sequences logged so far, additional sample sequences still increase the estimated number of species proportionately. Nevertheless, the clustering pattern of the sequence data is likely to yield continuing evolutionary insights, Remington said. One of her group's objectives is to use the DNA data to discern the shape of global "protein space": the full variety of proteins encoded in living organisms.

Another surprise relates to bacteriophages, the viruses that infect bacteria. Genomic data indicate that these predators on the lowliest life-forms (at least by size) play a far more important role in distributing genetic information among their prey than has been thought. Remington's group found 20-fold variations in the level of incorporation of a specific phage between closely related samples of the marine cyanobacterium *Prochlorococcus*, the most numerous photosynthesizing microorganism of tropical and



The marine cyanobacterium Prochlorococcus (typically 0.5 to 0.7 micrometers in diameter), the most numerous photosynthesizing microorganism of tropical and subtropical oceans, has been found to show high levels of diversity at the genomic level. Electron micrograph: Claire Ting.

subtropical oceans. That points to a dynamic predator-prey relationship.

The torrent of data from the project has stymied even the computers of Venter's institute, which perform at the level of trillions of floating point operations per second (known as "teraFLOPS"). That is one reason the Gordon and Betty Moore Foundation in California recently awarded the institute, along with the University of California-San Diego, a \$24.5 million grant to develop a cyberinfrastructure for analyzing marine microbes. The Moore Foundation has also bankrolled a project to sequence the complete genomes of over 150 of the organisms.

Not content with charting global marine biodiversity, Venter's institute has launched what it calls the air genome project. Researchers are shotgun sequencing DNA from pollen and microbes of different size classes trapped by air filters, including viruses, bacteria, and fungi. The biogeography of the sequences is likely to have important implications for health, because these motes of biological information may have profound effects on immune sys-

tems. As with the marine work, the distribution of phage genes is of special interest and may reveal much about hitherto uncharted natural processes. A pilot project is under way in New York City.

Sallie W. Chisholm of the Massachusetts Institute of Technology, a *Prochlorococcus* guru, took the conference podium to expand on that bacterium as the emerging "*E. coli*" of microbial ecology in the ocean. The estimated global total of 10^{25} *Prochlorococcus* cells, which have a very high level of diversity at the genomic level, may account for over half the chlorophyll in some ocean regions, Chisholm declared. Closely related coexisting lineages of *Prochlorococcus* differ in their degree of light-adaptedness, their ability to utilize phosphorus and nitrogen, their tolerance for trace metals, and, notably, their temperature sensitivity, she reported. Most surprising, however, has been the discovery that some phage genomes commonly found in *Prochlorococcus* encode photosynthetic pigments. Because photosynthesis is essential for the bacterium to generate maximal numbers of phages, it appears that, by infect-

ing *Prochlorococcus*, these phages are augmenting photosynthesis for the phages' own benefit.

E. Virginia Armbrust also noted the importance of captured genes in her talk on the whole-genome analysis of marine diatoms, unicellular algae with elaborate shells of silica. Armbrust estimates that diatoms are responsible for up to 40 percent of global marine photosynthesis, an amount comparable to all terrestrial rainforests. Moreover, diatoms play a particularly vital role in the global carbon cycle, because, on sinking, their remains take carbon out of circulation for centuries. Yet the genome sequence of the marine diatom *Thalassiosira pseudonana* reveals that its photosynthetic plastid was acquired from another eukaryotic cell through the process known as secondary endosymbiosis. (The sequence also reveals the unexpected presence of a urea cycle in the diatom, again suggestive of hijacked genes.) Armbrust is analyzing the activities of diatoms by using a flow cytometer at sea so that she can select specific diatom populations "right out of the water." She hopes to learn enough to use the organisms to monitor ocean health in real time with sensors connected by undersea cables.

Forest Rohwer, a researcher at San Diego State University who sequences viruses from extreme environments and from human blood and stool, reported that marine sediments have much the largest number of viral genotypes in his samples. Rohwer, too, stressed the critical role of genes that wander between microbial hosts. Some 54 percent of the environmental sequences he detects are unknown, Rohwer said, and it seems that mobile DNA elements compose most of the uncharacterized diversity. To the totemic question, "Is everything everywhere?"—meaning, is there a rich stable of sequences in all environments?—Rohwer answered with an emphatic yes. The environment then selects those sequences that become numerically dominant. The population dynamics of phages and their microbial prey probably maintain the high diversity in the mobile gene pool, Rohwer suggested, and thus account for most of the differences between microbial species.



Nobel prize winner Hamilton O. Smith, vice president for synthetic biology and biological energy at the J. Craig Venter Institute, told the conference how his research group is designing and constructing a synthetic chromosome for the bacterial human pathogen Mycoplasma genitalium.

Photograph courtesy of J. Craig Venter Institute.



Marvin E. Frazier, vice president for research at the J. Craig Venter Institute, spoke movingly of how the anticancer drug imatinib mesylate (Gleevec) shrank his gastrointestinal stromal tumors when he had been thought to be only weeks from death. Researchers at the conference described ways in which genomic research could contribute to the design of other anticancer drugs. Photograph courtesy of J. Craig Venter Institute.

A similar notion was discussed by Jillian Banfield, a professor in the Department of Earth and Planetary Science at the University of California–Berkeley, who studies genomes of bacteria that produce biofilms in acid mine drainage. She has concluded that viruses are an important source of the novel genes the bacteria use to make a living in unwelcoming situations.

Metagenomics does not begin and end with DNA sequences. Jo Handelsman of the University of Wisconsin–Madison has taken the concept to the next level by cloning DNA from environmental samples into bacteria and testing the functional capabilities of the resulting transformants. Clones that share a biochemical activity are then analyzed by sequence. Handelsman maintains that her functional approach can flag activities of interest that would not be found by sequence analysis alone. She has used the technique to discover antibiotic resistance genes, for example, and novel quorum-sensing systems, which are communications protocols employed by bacteria to assess their population density.

Building up biology

The South Carolina event was not only about analysis. Several sessions focused on "synthetic biology"—attempts to build organisms with new capabilities through sophisticated genetic engineering informed by genomics. Environmental goals drive some of this. Venter himself, along with longtime colleague and Nobel laureate Hamilton O. Smith and others, is working on creating a stripped-down version of the already small genome of *Mycoplasma genitalium* G37 that would include only the components essential for life. The team is systematically deleting subsets of individually dispensable genes from a synthetic *M. genitalium* G37 chromosome to learn which groupings the organism can manage without. Smith said he expects there to be several answers that depend on conditions. The exercise should expose how the truly essential components of cells work, Venter maintained (see the interview on page 197).

Andrew D. Endy of the Massachusetts Institute of Technology described a dif-

ferent approach to creating novel life-forms. Endy is creating “biobricks,” minigenomes for functional chemical pathways that can be added to host organisms. The bricks are available to researchers by mail order, and the scheme is now being promoted by a nonprofit organization.

The ability to manipulate biology with that degree of sophistication may make plausible an idea Venter has entertained in recent years to engineer microorganisms to produce hydrogen, the cleanest possible fuel, from water and sunlight. The needed enzymes, or something like them, exist separately in nature, but to be useful they would need to be combined in a new type of microbe. Genomics may suggest other strategies for the great problems as well. Venter and some of his colleagues have published the genome sequences of methanotrophic bacteria, which metabolize the potent greenhouse gas methane to yield hydrogen. Venter said he has no breakthroughs to report, but allowed that he is driven by a desire to find new options to combat global warming and threats to the energy supply.

High-speed chemistry

Companies touting new, faster DNA sequencing technologies were much in evidence at Venter’s beachside jamboree. Almost all large-scale genomic sequencing to date has been done with machines made by Applied Biosystems, now part of Applera Corporation. The company’s technology employs what is known as Sanger chemistry (named for its inventor, double Nobel laureate Frederick Sanger) and fine capillaries to separate tagged DNA fragments. But several upstart entrants to the field employ quite different schemes that, although not yet as reliable as Applied Biosystems’ workhorse machines, promise quantum jumps in sequencing speed. These could make feasible medical applications that require answers to specific questions about a patient’s genome in hours or days. Notable among the new players, to judge by gossip in the exhibition hall, were machines introduced by 454 Life Sciences and by Solexa, which promises its machine can sequence a billion bases of DNA in two



Aristides Patrinos (left), now president of Synthetic Genomics, Inc., confers at the Hilton Head conference with Andrew D. Endy, of the Massachusetts Institute of Technology. Endy, a pioneer in the field of synthetic genomics, makes segments of DNA encoding chemical functions available to researchers as “biobricks.”

Photograph courtesy of J. Craig Venter Institute.

days. Both sequence DNA by synthesis rather than by separation, the principle underlying traditional approaches.

Medical applications of new genomics technologies were not ignored. Several sessions focused on innovative ways in

which genomic data are being employed to gain insights into the complex process of tumor formation—and ideas for possible therapies. Imatinib mesylate (Gleevec), a powerful drug approved for the treatment of chronic myelogenous leukemia, is one product of rational drug design whose effectiveness may be extended as a result of genomic studies. Mutations in tumor cells that affect the cellular sites where the drug binds can make a tumor resistant to the drug. The ability to quickly detect such mutations would allow physicians to select imatinib variants that would still be effective.

Officialdom observes

The emerging themes were witnessed by officials with some clout, who seemed enthusiastic about the gains reported. Maryanna Henkart of the National Science Foundation said her agency sought to develop tools that would make it possible to simulate changes in microbial populations and the resulting impacts on other organisms and the environment. Aristides Patrinos, at the time associate director of the office of biological and environmental research at the US Department of Energy (and broker of the tie agreement on the human genome that took Venter to the White House),



David T. Kingsbury, deputy chief program officer of the Gordon and Betty Moore Foundation, chaired a session on environmental metagenomics at the conference. Kingsbury’s foundation has supported several of the Venter Institute’s major programs, including one that aims to sequence over 150 genomes of marine microbes. Photograph courtesy of J. Craig Venter Institute.

observed that metagenomic data are arriving so rapidly that “our ignorance is growing as fast as our knowledge.” Patrinos said he thought that biology had hitherto been unjustly neglected in the US Climate Change Science Program and announced that he was now “in the repentance stage.” But he told the assembled metagenomicists they were “sloppy” and “fall very short” in organizing to develop the needed teraFLOPS of scientific computing resources. Even so, he was convinced that “we will be able to use this new science to get a better understanding of Earth systems.”

With concerns about climate change growing, some participants were in-

evitably drawn to speculate about how metagenomics may help in the search for practical solutions to environmental problems. One researcher asked, for example, about the availability of grants to study how oceanic diatoms might be engineered to synthesize long carbon polymers and thus sequester more carbon from the global cycle (he was encouraged to enquire at the National Science Foundation). But Patrinos was cautious about large-scale biological engineering. There is, he pointed out, a widespread reluctance to sanction such research, and, especially, “we probably shouldn’t diddle with the oceans.”

Patrinos was backed up by David T. Kingsbury of the Gordon and Betty Moore Foundation, who declared that despite the plethora of new studies, there is still much scope for learning from metagenomic surveys of specific ecosystems. There are, Kingsbury said, “many environments we haven’t really touched yet.” By the time metagenomicists meet this fall at the 2006 Venter Institute conference in Hilton Head, it’s a fair bet that there will be fewer.

Timothy M. Beardsley (e-mail: tbeardsley@aibs.org) is editor in chief of BioScience.