Harnessing the power of genomics for energy and the environment

Marvin Frazier and
J. Craig Venter

J. Craig Venter Institute
9704 Medical Center Drive
Rockville, MD 20850

The remarkable feat of sequencing the human genome, with technology and bioinformatics programs that were developed by scientific teams at Celera Genomics and elsewhere, resulted in a revolution in biology. At the time that the human genome sequence was published in *Science* in February 2001¹, the authors said that this was just the beginning of a long and exciting path to discovery. That was an understatement.

Since those exciting times, our colleagues at the Venter Institute have embarked on an entirely new venture in genome science—studying the environment and the microbes that populate the Earth's biosphere. This is an enormous challenge that will require imagination, perseverance, and, of course, financial resources. The Institute is proud to have many sponsors, in particular, the Genomics: GTL Program at the Department of Energy, The Gordon and Betty Moore Foundation, and the Alfred P. Sloan Foundation.

The GTL research program at the Venter Institute is multifaceted and spans the entire spectrum from discovery science, technology-enabling science, to technology development. The relevant focus areas are environmental genomics, synthetic biology, and bioenergy production.

Microbes in the sea: Biological diversity

The Venter Institute has mounted the largest, most comprehensive effort ever undertaken to identify, sequence, and understand microbes that live in the oceans, soil, and air. The Sorcerer II Global Ocean

Survey leads the Institute's research efforts in microbial ecology.

The first major achievement in ocean sampling occurred in the Sargasso Sea where, in collaboration with several groups, including colleagues at the Bermuda Biological Station for Research, the team discovered at least 1,800 species of microbes in nutrient-limited water. The data revealed ~150 new bacteria phylotypes and more than 1.2 million previously unknown genes². And, to our surprise, data analysis also demonstrated the presence of more than 650 proteorhodopsin-like genes that are of special interest because microbes may use light energy in new and interesting ways. One of the Institute's key partners throughout this marvelous journey has been the Department of Energy's Office of Biological and Environment Effects (OBER). OBER was the first Federal Agency to fund the Human Genome Project³ and whole shotgun microbial sequencing⁴, and it continues to promote cutting-edge science through the current GTL Program⁵.

At NIH in the late 1980s, the lab started with the production of ESTs⁶, and at TIGR, an extraordinary research team used shotgun sequencing to read the genome of *Haemophilus influenzae*, the first living organism ever sequenced⁷. Subsequently, our talented research groups perfected BAC end sequencing technology⁸, and then proved the value of our whole-genome shotgun technique with the completion of the genome sequences of *Drosophila*⁸, human¹, and several other species^{10,11}.

Genomics is having an incredible impact on science, but until recently only touched upon a minute fraction of life on earth. The enormous and largely untapped diversity of microbes and microbial communities on the earth and within its waters presents a scientific treasure of vast importance. Microbes make up about half of the earth's biomass¹², control earth's biogeochemical cycles, and affect the productivity of its soil and the quality of its water, as well as the global climate^{13,14}. Indeed, the ability of this planet to sustain life is completely dependent on microbial activity. Microbes produce half

of the world's oxygen supply and have the potential to produce unique, clean forms of energy, for fuel and other purposes.

Therefore understanding the microbial world is critical to our energy and climate futures, to developing sustainable modes of living and advanced industrial technologies, and to understanding life itself. Because science has not developed reliable methods for growing the vast majority of microbes in culture, our ability to study them in detail and harness their ability to produce energy has been limited. The Institute's massive sequencing program is the first step toward changing that.

The goals of the Venter Institute's ongoing environmental genomics research are to better understand microbial biodiversity; to discover new genes of ecological importance, including those involved in carbon cycling; to discover new genes that may be useful for biological energy production; and to establish a freely shared, global environmental genomics database that can be mined by scientists around the world. Using newly developed metagenomic methods, it is possible to examine whole communities of microorganisms and to identify the genes that enable them to capture energy from the sun, remove carbon dioxide from the air, take up organic carbon from other organisms, and cycle nitrogen in its various forms through the ecosystem. To date, we have discovered thousands of new species and millions of new genes, with no apparent slowing of the rate of discovery.

Synthetic Biology

The Venter Institute has launched a major program in synthetic biology that has already produced a significant result¹⁵. Venter Institute

scientists, funded by the DOE GTL Program, created with high accuracy a synthetic version of X174, a virus that is harmless to humans. Equally important, this achievement was accomplished in only 14 days, demonstrating the power of the techniques we developed.

Using chemically synthesized oligonucleotides, Institute scientists have developed ways to construct large DNA molecules, an approach that facilitates synthetic organisms of tremendous potential value to humankind. In theory, one can make entirely new cells by synthesizing their genomes, starting with the genes that are known to be essential for life. Other genes, with desirable properties, can then be added to create whole new classes of chemically synthetic organisms. As a result, scientists will gain not only a deeper understanding of living organisms, but also the ability to design organisms for specific commercial, industrial, and medical purposes.

Motivated by a desire to understand life at its simplest, Venter Institute scientists are attempting to use this technology to build a synthetic minimal bacterial cell. The synthetic biology group is working toward synthesizing a minimal mycoplasma genome, with fewer than 400 genes, that has all of the machinery for independent life. A cell carrying this minimal genome might serve as a platform for designing and building more complex and useful cells by the addition of gene cassettes. To achieve this goal, Institute researchers are developing enabling technologies that will have many other applications, as well. Faster and more accurate methods for synthesizing chromosome-length stretches of DNA will find many applications in medical research and throughout the biotechnology industry.

Bioenergy production

The Institute's biological energy group is focusing on two promising alternatives to gasoline, more suited to a carbon-constrained world: photobiological hydrogen and cellulosic ethanol. One team is engineering photosynthetic cyanobacteria for sustainable, light-driven hydrogen (H_2) production. H_2 , when produced from renewable energy sources, is a clean alternative to gasoline and other fossil fuels, generating only water as a byproduct when it burns. Cyanobacteria and green algae have attracted considerable attention in recent years as a potential source of H_2 production, because they have the ability to split water photolytically through oxygenic photosynthesis.

However, nature did not design these microbes for H_2 production. The reducing energy and ATP generated from the photosynthesis process normally support CO_2 fixation via the dark Calvin cycle. The hydrogenase functions only as a valve to release excess reductant energy¹⁶, to prevent a slowing of electron transfer in the photosynthetic process. Thus, H_2 evolution in these microbes is generally very low. How to direct the reductant energy in these microbes mainly for H_2 evolution, instead of CO_2 fixation, remains a challenging problem that we at the Venter Institute are determined to solve.

The group working on cellulosic ethanol is developing organisms that can break down different types of complex biomass to sugars and also ferment those sugars to ethanol or other biofuels and bio-products. This technology represents a paradigm shift from the current approach to cellulosic ethanol production, which currently requires an expensive and complex enzyme cocktail. An integrated bioprocessing model using a single culture would eliminate the costs of setting up and maintaining an enzyme-production reactor, offering savings in both capital and operating expenses.

Both the H₂ production and cellulosic ethanol technologies may eventually make use of our minimal genome platform when the production stage is reached. These are obviously ambitious and long-range goals, but we believe such work is necessary in order to develop significant biological sources of clean and renewable energy.

Marvin Frazier is VP of research at, and J. Craig Venter is founder of, the J. Craig Venter Institute (Rockville, MD). Phone: (240) 268-2605. Fax: (240) 268-4000. Website: www.venterinstitute.org.

REFERENCES

1. Venter JC et al. The sequence of the human genome. *Science* 291, 1304-1351 (2001).
2. Venter JC et al. Environmental genome shotgun sequencing of the Sargasso Sea. *Science* 304, 66-74 (2004).
3. DeLisi C. The Human Genome Project. *American Scientist* 76, 488-493 (1988).
4. Fleischmann RD et al. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496-512 (1995).
5. Frazier ME et al. Realizing the potential of the genome revolution: the Genomes to Life program. *Science* 300, 290-293 (2003).
6. Adams MD et al. Complementary DNA sequencing: Expressed sequence tags and human genome project. *Science* 252, 1651-1656 (1991).
7. Fraser CM et al. The minimal gene complement of *Mycoplasma genitalium*. *Science* 270, 397-403 (1995).
8. Venter JC, Smith HO, Hood L. A new strategy for genome sequencing. *Nature* 381, 364-366 (1996).
9. Adams MD et al. The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185-2195 (2000).
10. Kirkness EF et al. The dog genome: survey sequencing and comparative analysis. *Science* 301, 1898-1903, (2003).
11. Gibbs RA et al. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428(6982), 493-521 (2004).
12. Whitman WB et al. Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* 95, 6578-6583 (1998).
13. Copley J. All at sea. *Nature* 415, 572-573 (2002).
14. Newman DK and Banfield JF. Geomicrobiology: Elucidation of the molecular-scale interactions that underpin biogeochemical systems. *Science* 296, 1071-1077 (2002).
15. Smith HO et al. Generating a synthetic genome by whole-genome assembly: phiX174 bacteriophage from synthetic oligonucleotides. *Proc Natl Acad Sci USA* 100, 15440-15445 (2003).
16. Appel J et al. The bidirectional hydrogenase of *Synechocystis* sp. PCC 6803 works as an electron valve during photosynthesis. *Arch Microbiol* 173, 333-338 (2000).