

*Genomics provides a powerful tool for understanding life and the environment*



# DNA in the Oceans

BY CHUCK COLGAN

**T**ERRY GAASTERLAND, DIRECTOR OF SCRIPPS INSTITUTION OF OCEANOGRAPHY'S latest and most advanced computer facility for life sciences, beams like a proud mother. The Scripps Genome Center came online this year and is cranking out millions of lines of DNA sequences and other information critical to understanding marine organisms on the molecular level. Genomics is where the next revolution in biology is taking place, and Scripps is well positioned to lead the charge into the oceans with Gaasterland on board.

Sitting in a room full of computers, Gaasterland emphasizes that genomics is not just the next best thing—it is the crucial component when unraveling the basics of how life functions.

“In a sense, the genome is like the inventory of everything you could buy at Wal-Mart, if you could access all their stores anywhere in the world at the same time,” Gaasterland said. “So far, our reading of the genome is more a parts list of everything that could be made in a given cell. We have a long way to go before we are able to understand the directions on how to use those parts.”

Gaasterland is a pioneer in the emerging scientific field of bioinformatics, which combines mathematics, statistics, and database management to help answer biological research questions. She spent the past six years at The Rockefeller University in New York City collaborating primarily on cutting-edge biomedical studies. By coming to Scripps, she is setting her sights on investigations of the marine environment and life in the oceans by studying their genetic components. This move from unraveling DNA molecules for medical research to marine studies illustrates, according to Scripps Director Charles Kennel, “the unity of all biological systems.”

## **PARTNERSHIPS WITH FACULTY**

Establishing a marine genomics center is a natural extension of Scripps faculty's historic interests in microbial sea life. Microbiologist Claude E. ZoBell (1904–1989) was the first to confirm that there are unique bacteria in seawater, even at the deepest depths with the greatest effects of pressure. ZoBell is cited as

# Concept

## BIOSYNTHETIC PATHWAY

### BIRTH OF A METABOLITE



I.



II.

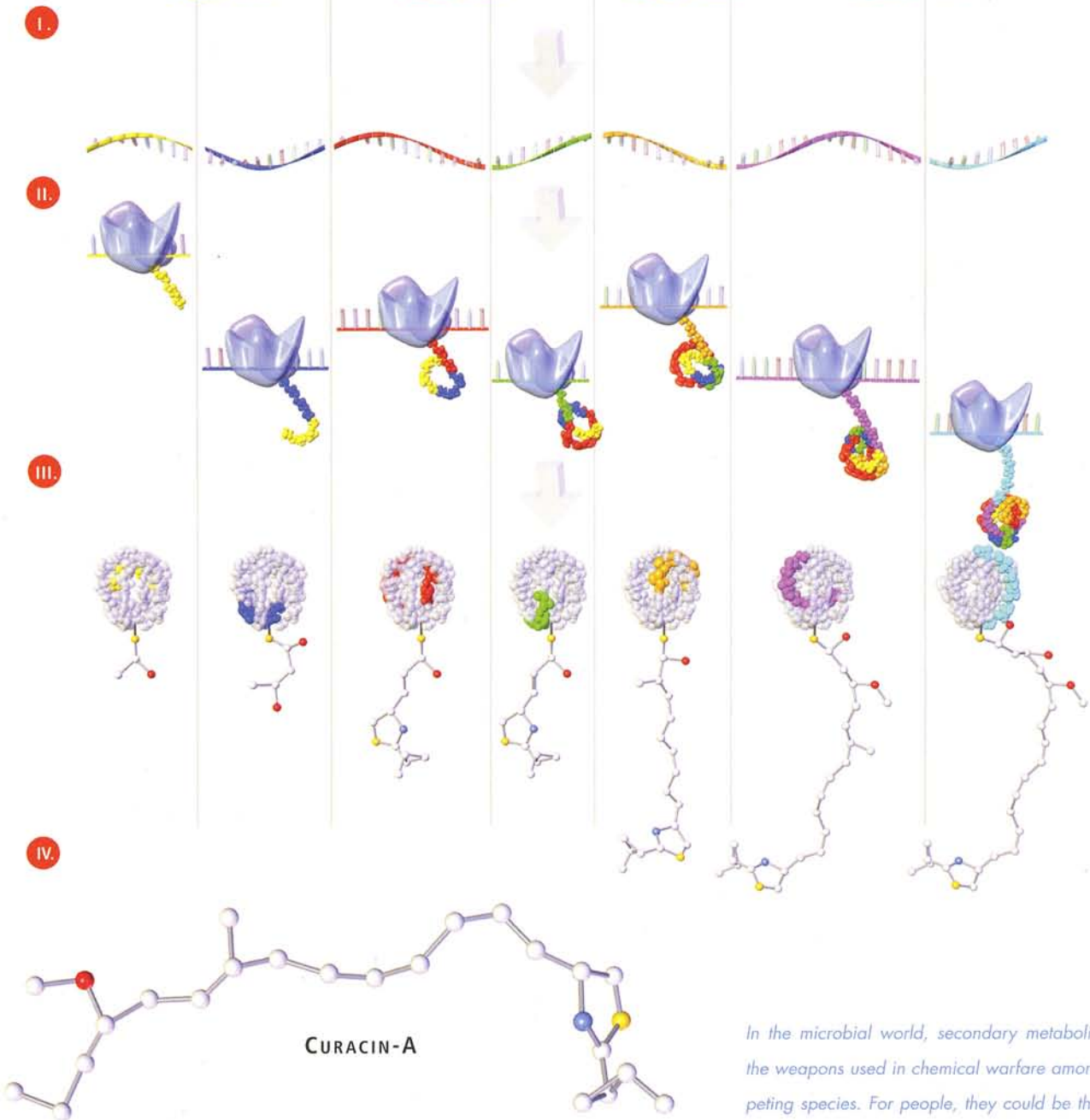


III.



IV.

EXPLORATIONS  
FALL 2005



*In the microbial world, secondary metabolites are the weapons used in chemical warfare among competing species. For people, they could be the basis of new medical treatments. William Gerwick is tracing the sequence by which cyanobacteria create*

*molecules like curacin-A, a potential anticancer compound. In strands of the organism's DNA, a cluster of genes encodes for a protein assembly line that sequentially catalyzes the construction of curacin-A, one step at a time. Through the process of transcription (I) to messenger RNA and then translation (II) to proteins using ribosomal machinery, the instructions from the DNA for curacin-A biosynthesis are "downloaded." The newly produced proteins aggregate to form an assembly line (III) where common biochemical "building blocks" such as acetate and amino acids are combined to make a single curacin-A molecule (IV).*

the founder of marine microbiology for his lifetime of research and his 1946 textbook *Marine Microbiology*, the reference cornerstone in the field. In the 1960s, marine physiologist A. Aristides Yayanos was the first to isolate and culture bacteria from the deep sea. He developed titanium high-pressure chambers for preserving samples that are still in use today and has kept pressurized microbial communities alive in his lab for decades—even bringing dried specimens back to life.

Currently, there are more than two dozen faculty affiliated with

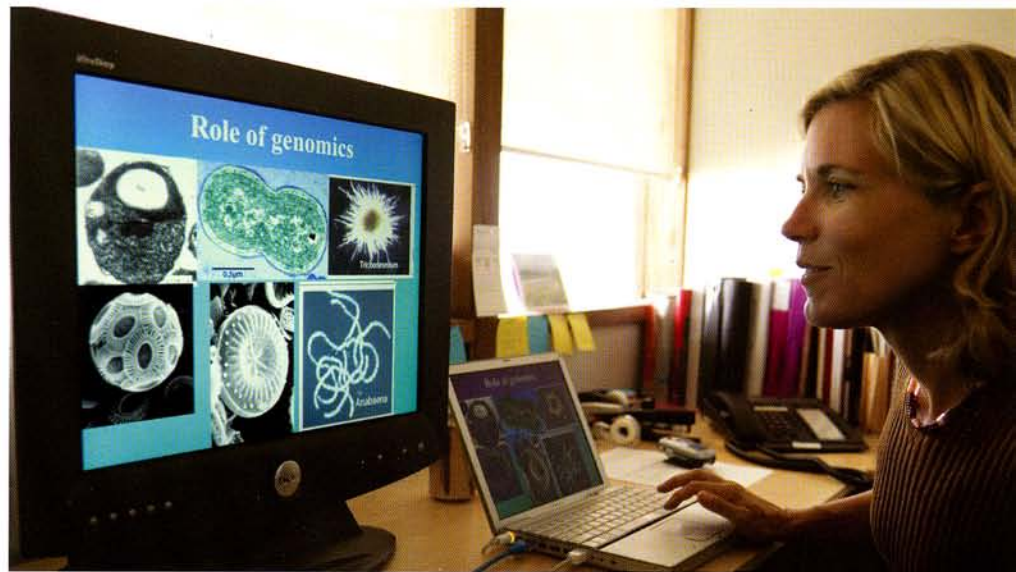
Scripps's Marine Biology Research Division specializing in cell and developmental biology, ecology and evolution, microbiology, and physiology, as well as 49 marine biology students, about one-quarter of the total Scripps graduate student enrollment. A growing number of



Claude E. ZoBell



A. Aristides Yayanos



Terry Gaasterland, director of the Scripps Genome Center, prepares for a lecture to the public at the Birch Aquarium at Scripps.

these traditional marine biologists and computer-savvy students are partnering with Gaasterland. It's a merging of conventional science with advanced information processing made possible with new computing power and expertise in a process she describes as "computational heads coming together with biology heads."

In the midst of aquariums and marine biological labs in Hubbs Hall, Gaasterland has built a half-million-dollar computer center specialized in handling and storing the enormous amounts of data necessary for analyzing DNA, gene expression, and genome sequencing. This was made possible by a million dollar donation from investment manager Louis Simpson of Rancho Santa Fe, California. Assisting with the scientific liaison, training, and computing at the center are research programmers Linda Maepa and Sheila Podell.

The lab has 64 Sun computer systems augmented with time-logic accelerator cards and about a terabyte (1,000 gigabytes or  $10^{12}$  bytes) of memory that simulate a cluster of a thousand computers. Apple computer systems serve as workstations for visualization and storage for 10 terabytes of data. One purpose of the center is to provide instant access to the international GenBank DNA database of sequenced genomes, which is updated every few days via the Internet from the National Institutes of Health's National Center for Biotechnology Information in Bethesda, Maryland. This is accomplished by linking with mass storage and high-speed connections through the UCSD-based California Institute for Telecommunications

## KEY TERMS

### BIOINFORMATICS:

Field of computer science that develops software and computational tools to organize large quantities of biological data.



and Information Technology.

Gaasterland is also recognized for developing computer software that races through genomes seeking repeated or specific patterns and in turn looks for comparisons among the patterns. One program sorts through billions of coded DNA instructions to identify sequences that have specific functions, such as producing proteins—the real workhorses of life. Another applies artificial intelligence to analyzing gene features and relationships, information that assists researchers in understanding what genetic materials are being used when and under what conditions and in preparing strategies for further analyses.

**This page, below,** Terry Gaasterland discusses sequencing projects with computer programmers Linda Maepa (right) and Sheila Podell (standing). **Opposite page, top,** Banks of automated DNA sequencers at the Sanger Center in Cambridge, England, were used for the Human Genome Project. **Bottom,** Human Genome Project Director J. Craig Venter.

“AS WE FIGURE OUT WHAT THE FUNCTIONS OF THE PROTEINS ARE, WHAT CHEMICALS THEY MAKE, OR WHICH CHEMICALS THEY CHANGE—WE CAN TURN THAT INTO VERY PRACTICAL APPLICATIONS.” —TERRY GAASTERLAND

## INTO THE OCEANS

But why apply these bioinformatics techniques to studies at an oceanographic institution?

“As it turns out, marine organisms are excellent models for many human processes—fertilization, early development, cellular organization, and more,” Gaasterland explained. “That’s part of why it’s so exciting to initiate new questions about marine organisms on the DNA level.”

She also cites the potential for finding novel compounds with biomedical uses, such as drugs or cellular regeneration, and chemicals that could be useful in industrial processes or in cleaning up toxic environments. Information in marine genomes can also be used as a thermometer of environmental conditions.

“By going to the ocean and looking at the DNA that is there, we can find out much more about the diversity of life and the biochemical potentials,” Gaasterland said. “Being able to take all of these genomics technologies and apply them means that we’ve suddenly got the ability to look at the diversity of marine life at a level that we never could before.”





## KEY TERM

### DEOXYRIBONUCLEIC ACID

**(DNA):** An acid containing the genetic material of organisms. A major component of chromosomes, DNA is divided into genes that issue instructions for the development of life at the cellular level.

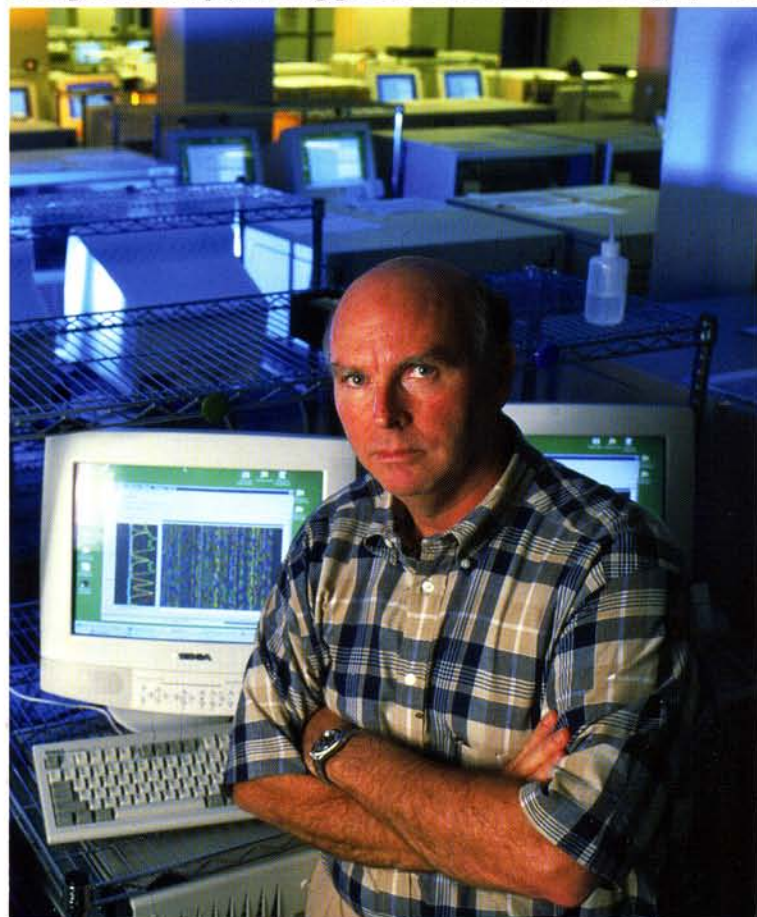
There are three general areas of research to be explored at the new center. Comparative genomics uses organisms or groups of organisms to understand the biochemical pathways they use to adapt and survive, especially protein production and diversity. Environmental genomics involves analyzing DNA from all organisms within a particular microenvironment, such as a liter (one quart) of seawater, and then reexamining the same environment at another time to look for changes and clues to what may have caused them. Developmental genomics builds upon a legacy of experiments conducted in marine embryology and gene regulation that will now be enhanced with improved investigative techniques.

The new Scripps Genome Center creates more possibilities for research, according to marine biologist Douglas Bartlett, who studies microbes in deep-sea environments. Microbiologist Farooq Azam said the center provides powerful resources for understanding life and an environment that allows researchers and students to work in ways that integrate genomics into

existing biological research. The center, said marine biologist Brian Palenik, is an attempt to “make something that is more than the sum of individual research efforts.”

### HOW DOES IT WORK?

Genome analysis begins with extraction of an organism’s DNA followed by a lengthy process of adding chemicals, cloning fragments, letting cells multiply, adding genetic materials, and ending up with





The brightness of each dot on a microarray reveals how much of a specific DNA fragment is present in a sample.

billions of DNA strands with their four subunits, or bases, marked by fluorescent dyes. The bases of DNA are adenine, thymine, cytosine, and guanine—the familiar A, T, C, G letters associated with genome sequences that are always linked as pairs. The order of the base pairs is a code, part of which tells a cell the order in which to string together amino acids to make specific proteins. By encoding the type of proteins produced, DNA determines the cell's function—to make tissues or bones or antibodies—and what an organism looks like—the color of fur or the shape of a shell.

Until the late 1980s, the color-coded DNA was manually processed in the lab and painstakingly analyzed over weeks and months. Today, DNA samples are sent to facilities that operate automatic sequencing machines, such as Diversa Corporation in San Diego, where they can be read in just a few hours. The strands are placed in a gelatinlike substance in the

sequencers. An electric current is then applied, causing the DNA molecules to move through the gel. Smaller molecules pass through the gel more rapidly, separating the DNA into different bands according to their size. As the DNA pieces travel through the gel, the machine reads the order of A-Ts and C-Gs and records this information in a computer.

The DNA sequence in its raw form is all jumbled together, with gaps and ambiguities that are assembled by computer programs. Millions of trillions of comparisons are made by these programs to construct the genome—and even then it requires the talents of computer specialists

known as finishers to deal with any inaccuracies. Repeated sequencing and careful checking helps keep errors to a minimum.

The final output of genomic information can take two forms: a sequence or a map. A sequence lists the order of every DNA base pair in the genome, whereas a map shows a series of landmarks in the genome. Although the map is less detailed, it provides clues as to where the important parts of the genome sequence can be found and is used to help sequence the genome. It's extremely difficult to look at the whole sequence of millions, even billions, of bases and determine which parts are of interest and which parts are extraneous. Only about 25 percent of the DNA in a most genomes appears to do the majority of the work, leaving long stretches of “junk DNA,” so named because scientists don't yet know what, if anything, it does. The human genome contains three billion

## KEY TERM

**MICROARRAY:** A collection of genes or gene fragments affixed to a solid surface like glass in a specific pattern that enables researchers to identify relative levels of activity within a genome under various conditions.

base pairs—and other genomes are even larger. One species of toad has seven billion, and a lily has nearly 100 billion. Simpler life forms such as microbes, however, have much smaller genomes, in the range of tens of millions to hundreds of millions base pairs.

Once the maps and sequences are refined, scientists scour them, primarily looking for protein-making instructions and regulatory sequences that determine where, when, and in what amounts proteins are produced. For some studies, this might be all the information that is needed to determine how a microorganism functions or relates to another organism or its environment. Other scientists may take more proactive measures, such as mutating or knocking out parts of the genome to see what happens—will the microbe continue its normal biochemical functions or will it even survive? By doing this, they learn what parts of the genome perform which functions.

Part of what's bringing about this evolution in marine biology research is economics. It's estimated that to sequence a single base pair—two letters among millions—in 1990 was \$10 whereas that cost is about one-tenth of 1¢ today. "It took getting all of these tools down to the right price point before we could apply them to the oceans," Gaasterland said.

#### BACK TO THE OCEANS

Among those attracted to microbiology in the oceans is genomics visionary Dr. J. Craig Venter, leader of the Human Genome Project that successfully sequenced human DNA in 2000. Venter, a UCSD graduate, searches for interesting DNA around the world, often aboard his 95-foot (29-meter)

yacht that he converted for use as a research vessel. In a 2004 *Science* magazine article, Venter reported on the collections that he and his research team made in the Sargasso Sea off Bermuda where they found an estimated 1,800 species of microbes, based on their DNA structures, including 150 new bacteria.

A grant from the U.S. Department of Energy and the Gordon and Betty Moore Foundation is providing funds for sequencing the DNA of these microbes and creating an online database as part of a large-scale marine microbiology research initiative. To understand gene expression and protein production in the Sargasso Sea data, Gaasterland is working with biochemist Susan Taylor at the UCSD School of Medicine and systems biologist Adam Godzik at the Burnham Institute in San Diego. Other researchers at Scripps are using the data to make comparisons with the genomes of microbes collected elsewhere.

The most important aspect of the Scripps Genome Center, according to Gaasterland, is that it provides a place where there is sufficient computing power and knowledge about genomics to take advantage of the new kinds of information and technologies resulting from the genomics revolution in biology. Whether it's investigations of the environment, animals, or organism diversity, she sees her job as "helping marine biologists ask questions at a scale unavailable before." 🌐

[www.sgc.ucsd.edu](http://www.sgc.ucsd.edu)



"BEING ABLE TO TAKE ALL OF THESE GENOMICS TECHNOLOGIES AND APPLY THEM MEANS THAT WE'VE SUDDENLY GOT THE ABILITY TO LOOK AT THE DIVERSITY OF MARINE LIFE AT A LEVEL THAT WE NEVER COULD BEFORE." —TERRY GAASTERLAND



# ANNA NEILL

FIRST BIOINFORMATICS PH.D. AT SCRIPPS



**F**OR ANNA NEILL, THE EMERGENCE of genomic sciences in marine biology could not have come at a better time in her education. It allowed her to not only become the first bioinformatics Ph.D. graduate of Scripps, but also to accomplish much more research.

During studies on sea urchin fertilization processes in Victor Vacquier's biochemistry lab, Neill worked with one protein for four years. But at the end of her fifth year, she worked on 400 proteins, and in greater detail.

Among the proteins she studied was one nearly identical to a protein associated with human polycystic kidney disease. According to Neill, that is the serendipity you find when looking into the genomes of marine organisms.

Her doctoral dissertation on the structure and function of proteins in sea urchin sperm was the first at Scripps in which genomics was used as the basis of research. Neill had switched from traditional embryonic development studies to genomics research in 2003, and soon after made a connection with newly appointed Scripps Genomic Center Director Terry Gaasterland. Gaasterland and Vacquier were Neill's co-advisors, even though having two advisors is unusual, especially when they are from two very diverse areas of science—one being a biologist and the other a computer scientist. Her course of study, according to Vacquier, shows how flexible the graduate program is at Scripps. "We don't stamp them out of a mold," he said. "Students have a lot of flexibility—the program isn't set in stone."

Neill discovered her interest in marine invertebrates while taking undergraduate courses through Duke University's marine program in Bermuda. In 1999, she graduated from the University of Richmond with two bachelor degrees, in biology and in Spanish. Neill then received a full scholarship at Scripps as a Howard Hughes Medical Institutes pre-doctoral fellow. It's not unusual for Vacquier's students to be headed toward medical pursuits. Two of his students went on to medical school, but Neill has gone further afield. She is currently enrolled in law school at Stanford University, where she plans to apply her scientific background to legal issues related to intellectual property and products in molecular biology. 🍷



"SEA URCHINS PROVIDE AN EXCELLENT MODEL FOR THE STUDY OF FERTILIZATION."

—ANNA NEILL