Research | Penn State

VOLUME 22 NUMBER 3 SEPTEMBER 2001

A CABINET OF WONDERS

Steve Miller (4)

DNA is an elegant molecule. Purified, it strings like spider silk. Glassy clear, it hides reams of secrets: though we know its code — ATGC, the letters of its alphabet — we can't yet read it like a book. We may know what it *says*, but not what it *means*.

Yet a book is what the genome, the entire set of DNA in an organism, is often called: "life's instruction book," as the title of this year's Penn State Lectures on the Frontiers of Science put it.

Or it's a blueprint, the "blueprint for life." That analogy fails too: *If the architect you hired to design your home brought you a blueprint that solely consisted of a long list of parts that began "windowwabeborogovestaircasedoorjubjub,"* wrote a commentator in *Science*, referring, with his Jabberwocky talk, to the amount of "junk" DNA in a genome, DNA that isn't genes, *you might start to wonder if and when you will see your new house.*

Having sequenced the genomes of 599 viruses, 205 bacterial plasmids, 185 organelles, 31 eubacteria, seven archaea, one fungus (yeast), two simple animals (worm and fruitfly), one plant (mustard weed), and one mammal (human), with the mouse, rat, zebrafish, pufferfish, and rice sequences well on the way, we now appreciate how complicated genomes are. (How is a genome sequenced? See "Sort It Out.") Our comparisons, consequently, are little by little becoming more realistic.

According to a reporter for the Associated Press, the human genome is like the telephone directory for a major corporation: "You've got lots of names and locations. But what does each of these people do? How do they work together to get things done? If something goes wrong, what employees or teams are at fault?"

Even more aptly, it is, in the words of the editors of *Science*, who named the genetics revolution the "breakthrough of the year" for 2000, "a cabinet of wonders."

Explains Londa Schiebinger, "A 'cabinet of wonder' in early modern Europe was a splendidly crafted cabinet, itself made of rich materials, that the wealthy used to display natural and artificial oddities. The combination of divine and human craftsmanship displayed in the jumble of crystals, corals, shells, bits of bone, Amazon stones, or elaborately worked metals was to dazzle the observer."

As Schiebinger, a professor of the history of science at Penn State, writes in *Has Feminism Changed Science?*, feminists looking at the Human Genome Project in the early 1990s took exception to the idea that sequencing DNA was "the ultimate goal of biology." The emphasis on simplicity, on "reducing things to smaller and smaller units," was a holdover from World War II, Schiebinger argues. "Physicists, fresh from the Manhattan Project, imported to biology the attitude that mysteries can be solved. By reconfiguring life as the mechanism of genetic replication, they concluded that life itself was not complex, but alluringly simple."

It's not.

That, ironically, is the first result of the Human Genome Project. When *Science* and *Nature* published the complete human genome sequence last February, *Science* lead author Craig Venter wrote of "a major surprise: We have found far fewer genes (26,000 to 38,000) than the earlier molecular predictions (50,000 to over 140,000)." He concluded: "The modest number of human genes means that we must look elsewhere for the mechanisms that generate the complexities inherent in human development."

The central dogma has long been: DNA makes RNA makes protein.

"It has always struck me as a curious term," said Nina Fedoroff, "since *dogma* is not what science is about. Science is about theories: human constructs which fall and are replaced with other constructs."

Fedoroff, head of Penn State's Life Sciences Consortium, organized this year's Frontiers of Science lectures, a series sponsored by the pharmaceutical company Pfizer Inc., on which the articles in this special report were based.

It's true, she explained, that "the information flow is from DNA to RNA to protein. Think of DNA as a library — a rare books library. All you get to look at are microfiche or xerox copies. DNA is copied into messenger RNA, which is transported out of the nucleus and into the cytoplasm — which is the rest of the cell — where it

is translated or decoded, and the outcome is a protein." The work of the nucleus is to maintain DNA, to replicate it, and to transcribe it to RNA. Outside the nucleus, a ribosome catches hold of this messenger-RNA and reads groups of three letters at a time. Every three-letter "word" stands for one of 20 amino acids. Following the RNA code, the ribosome stitches the amino acids into a chain, which then coils and folds into a protein. "Proteins provide structure or do work," Fedoroff explained. "Each protein is different by virtue of its amino acid sequence."

So far the central dogma holds true. But, "each gene, we thought, makes one protein. That was the gospel," noted Paul Berg, the Stanford Nobel Prize winner (and Penn State alumnus) known as the "father of genetic engineering." In a lecture at Penn State last April, he spoke of the impact of genomics on science and society. When the Human Genome Project was expected to find 150,000 genes, for instance, the pharmaceutical and biotechnology industries were highly valued on Wall Street. "Every gene represented a potential target for a drug. Then, when only 30,000 genes were identified, the pharmaceutical industry was immediately downgraded!" The human genome suddenly represented only a fifth as many drugs.

"Yet those 30,000 genes probably make on the order of 200,000 proteins. One gene can turn out eight, ten, 15 proteins," Berg said. "There's an enormous amount of science that needs to be done to figure out who talks to whom and when," that is, to learn the reasons why a gene is translated into one protein and not another. These tasks define the developing fields of proteomics (the study of proteins) and bioinformatics (the combination of computer science and molecular biology).

"That's where we are now," said Berg. "We have the basic parts list. That puts us at the very beginning of the start of the problem of what makes a human being and why we're all different."

The Human Genome Project provided other humbling surprises. Of that string of three billion ATGCs, less than two percent is recognizable as protein-coding genes. These seem to cluster, leaving long stretches of "junk DNA" in between — at least, what used to be called junk. That some of this junk, the

SORT IT OUT

Sequencing DNA is not as overwhelming as it once was.

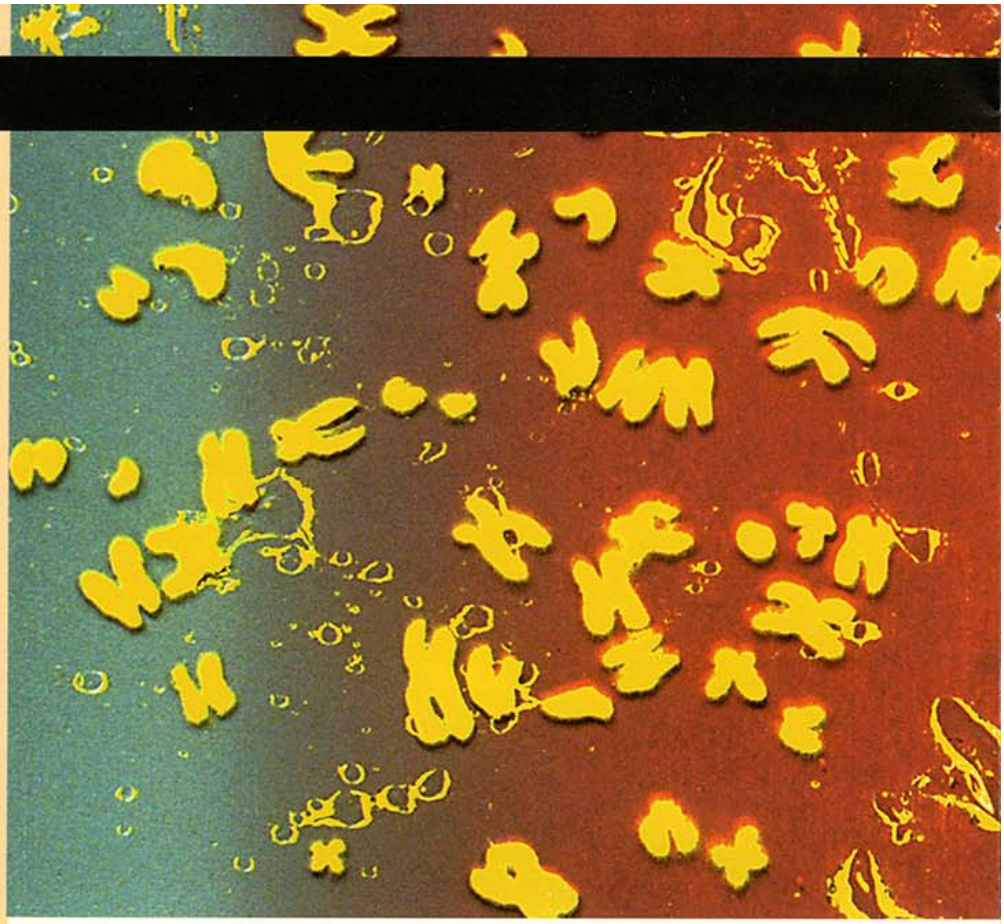
"The technology is really amazing," said Deb Grove of Penn State's Nucleic Acid Facility. "When I started here a few years ago the 'read length' for a DNA template was 300 to 400 nucleotides long. Today we can obtain read lengths of over 800 nucleotides, and can sequence templates up to 200k."

The first step in sequencing is to cut the DNA into those short "read" lengths and clone them using the PCR technique (see "Copy Shop"). Said Nina Fedoroff, director of Penn State's Life Sciences Consortium,"Cloning is the heart of genetic engineering. You can't sequence a single molecule — it's like figuring out how a molecule of sugar tastes. You need to get at least a quarter teaspoonful in your mouth to taste anything, and a quarter teaspoonful has millions and millions of molecules in it. The same problem confronts the scientist trying to find out the sequence of DNA. You have to have enough of it to actually carry out your analytical tools."

Once you have enough copies, you heat them to break the double helix into two strands: If DNA is like a spiral staircase, the nucleotides pair up to make the steps. The four nucleotides, or bases — adenine (A), thymine (T), guanine (G), and cytosine (C) — are complementary (C always bonds with G, and A always bonds with T) because C and G need three hydrogen bonds to hold them together, while A and T need two. These hydrogen bonds, Fedoroff said, are "sort of like Scotch tape. They hold things together but it's easy to pull them apart." (By comparison, the backbone of a DNA strand, made up of alternating groups of sugars and phosphates, "hangs together at high temperatures — it's more like crazy glue.")

Next you place the single strands into a solution along with four other ingredients: a primer, DNA polymerase, and two forms of free nucleotides, normal ones and so-called "stop" nucleotides (dideoxynucleotides). Each of the four stop nucleotides is labeled with a different fluorescent color.

The primer — a strand of DNA about 20 nucleotides long — gets things started. "DNA polymerase doesn't like to put the first two nucleotides together," Fedoroff explained, "but if you give it a place to hang onto, it'll go right to that place — which is extremely convenient for the molecular biologist." The primer, synthesized to match the end of the DNA strand you are sequencing, attaches, leaving an open hydroxl group for the next nucleotide to grab onto. Then the polymerase takes over. It reads along the original DNA strand, picking up a free nucleotide from the solution to pair with each



one it reads. It continues until it randomly picks up a "stop" nucleotide. These lack the hydroxyl group needed for the next to grab on.

Because choosing a stop nucleotide is random, the process stops at a different location on each copy. You separate the strands again and put the newly synthesized fragments into a sequencing gel. When you run electricity through the gel, the long pieces of DNA, which move slowly, remain near the top, while the short pieces run to the bottom. The difference in the length of each fragment from the bottom of the gel to the top is one nucleotide per fragment. Each fragment ends with a stop nucleotide tagged with one of four fluorescent colors; under a laser, the color molecule is excited and emits light. A sequencing machine, which can scan 96 gel lanes 12,000 times in 12 hours, detects the color of the light and translates it into a letter — A,T, G, or C.

When each of the read lengths is sequenced, they're compiled by a computer to match the original DNA.

—Teresa Rafacz

Deborah Grove, Ph.D., directs the Nucleic Acid Facility in the Life Sciences Consortium, 210 Wartik Lab, University Park, PA 16802; 814-865-3332; dsg4@psu.edu. Nina Fedoroff, Ph.D., directs Penn State's Life Sciences Consortium; nvf1@psu.edu.

On the cover of this special report, DNA in a "controlled drying" experiment undergoes a phase transition. Image from the Molecular Expressions website (microscopy.fsu.edu) of the Florida State University. Courtesy Michael W. Davidson. The image on the last page of this report also comes from that site.

On this and the preceding pages are details from "Genetic Portrait of Isabel Goldsmith," 1993, by Steve Miller; courtesy the artist and Universal Concepts Unlimited, New York, NY. As Miller wrote for the exhibition "Paradise Now," at Exit Art in New York, "Working on portraits at the electron microscope level was a logical extension of my previous investigations using new technologies in traditional art categories."

so-called Alu elements, also clusters in the gene-rich regions makes the designation suspect. As Berg noted in his lecture, DNA itself "was originally thought to be just baggage, just the structural component of a cell."

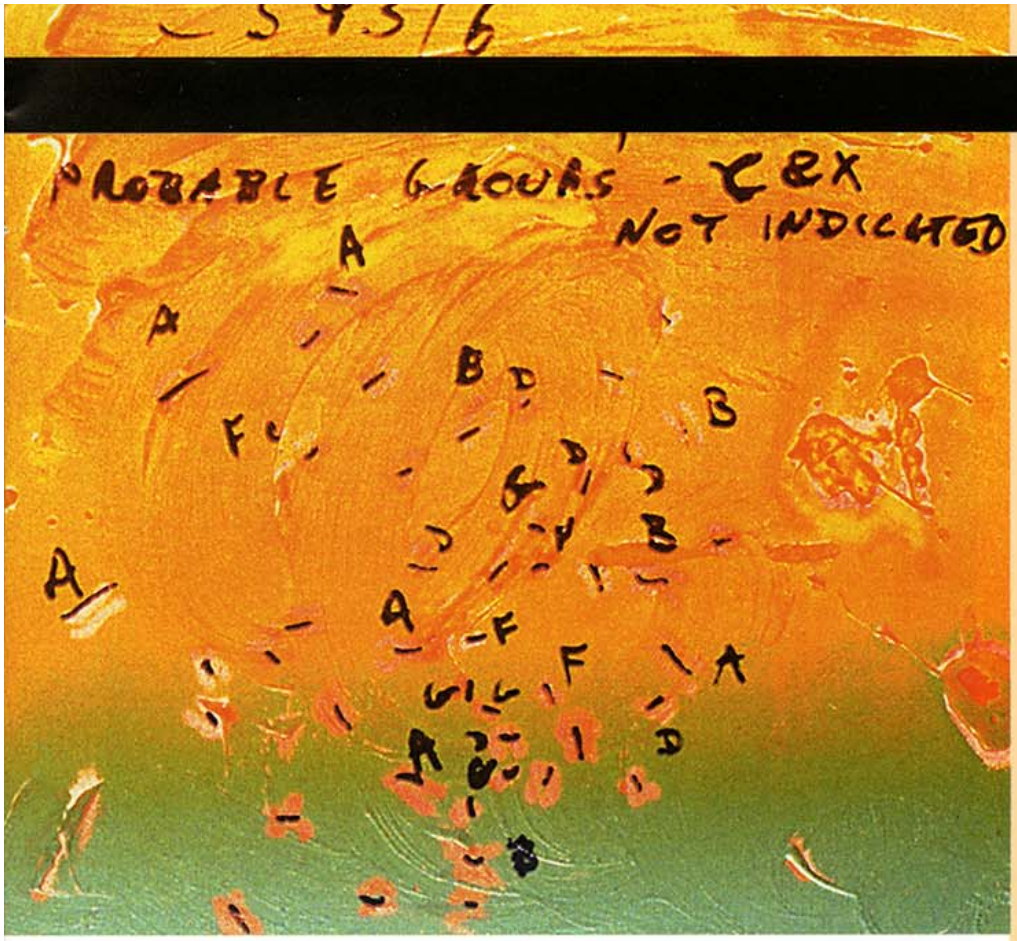
Another surprise is that some 200 genes in the human genome apparently come from bacteria. *Directly* from bacteria, not through millennia of evolution. It's startling enough, as Berg said, to know that "half the genes found in yeast are found in the human. And they're identical. You can take the gene from the yeast and put it into the human and it works perfectly well. What that tells us is that evolution has conserved information from the most primitive organisms to the most complex. What works in one place works everywhere."

David Baltimore of Cal Tech, writing in *Nature*, goes farther: "Apparently bacterial genomes can be direct donors of genes to vertebrates. So DNA chimaeras consisting of the genes from several organisms can arise naturally as well as artificially (opponents of 'GM foods' take note)."

Commenting on the news, one reporter quipped, "It is perhaps ironic that all humans, including those in the anti-GM lobby, are GM organisms." And it is perhaps coincidence that, the same week, the ban on genetically modified or "GM" foods was lifted by the European Union. "Gene-Altered Food Let In, But Europe Sets Strict Rules," the *New York Times* headline read. (The rules mainly involved labeling.)

The final surprise of the Human Genome Project is that it changes how we think. Said Berg, "The implications and opportunities — as well as the concerns and problems — that come from it will convince you that it's at least as momentous as trying to put a man on the moon." From drug-design to issues of privacy, from family planning to food production, from evolution to environmentalism to medical ethics, the results of the Human Genome Project will color our arguments from now on.

Take the case of DeCode Genetics, a start-up biotechnology company in Iceland. In 1998, the American-trained CEO of deCode, Kari Stefansson, proposed linking the DNA, medical records, and genealogies of the entire population of Iceland, some 280,000 people. Similar databases are in the works in Britain, Denmark, Estonia,



Quebec, Singapore, Sweden, and Tonga, but none have the scale of the Icelandic. Genealogy has long been popular in Iceland, with reliable records back to 1650, even, for some families, to the ninth century, when the country was first settled. Medical records, too, are voluminous, as in many Scandinavian countries with a central, socialized healthcare system. Linking these records to each other and to DNA sequences would provide a remarkable resource for finding the genetic causes of health.

The pharmaceutical firm Hoffmann LaRoche immediately saw the possibilities; it signed on with a $200-million deal to study 12 common diseases.

Others saw, not medical opportunity, but "biopiracy." Icelandic anthropologist Gisli Palsson and Paul Rabinow of the University of California at Berkeley write, "For some people, commodification is inhumane and degrading, an offense against individuality and human dignity, but for others it represents a humanitarian effort with a potential to cure disease and reduce human misery." They continue, "The critical issue in post-genomic research, following the sequencing of the human genome, is the use and control of information that can be derived from bodily components rather than the bodily components themselves. There is a frequent confusion between the information and the component itself."

Only after months of public debate over questions of privacy and ownership was deCode's plan approved; ten percent of the Icelandic population, Palsson says, still object. The Icelandic state will receive fees and a share of the profits, recognizing "the common-pool nature of the information involved," says Palsson, while the records in the three databases "will only be combined in the context of specific research projects, monitored by ethics committees and public officials." Already DeCode has located genes related to eight diseases, and Hoffman LaRoche is developing a schizophrenia drug based on the information.

—Nancy Marie Brown

Londa Schiebinger, Ph.D., is the Edwin Erle Sparks Professor of History in the College of the Liberal Arts, 311 Weaver Bldg., University Park, PA 16802; 814-863-7303; lls10@psu.edu. Nina Fedoroff, Ph.D., is director of the Life Sciences Consortium and the Biotechnology Institute. She is the Willaman Professor of Life Sciences in the Eberly College of Science, 219 Wartik Lab; 865-5717; nvf1@psu.edu.

COPY SHOP

The technique called PCR, or Polymerase Chain Reaction, has been applied to every branch of DNA research, from sequencing to forensics.

To run PCR to find or copy a gene, you need to know the "flanking sequences," the 15 or 20 pairs of nucleotides that lie on either side of it. Flanking sequences are stored in DNA libraries in, for example, the Penn State Nucleic Acid Facility. Programming the sequence into a DNA synthesizing machine, you make millions of single-stranded primers, one for each strand. These you put into a solution along with your target DNA, millions of free nucleotides (A, T, G, and C), and the patented *taq* polymerase, a variety of DNA polymerase from a bacterium first found in a hotspring at Yellowstone National Park.

PCR has three basic steps: First the DNA is heated, causing the double helix to unzip into two single strands. As the mixture cools, the short primers quickly seek out and stick to their complements on the longer strands before the original DNA can zip back up. Now *taq* polymerase goes to work: If it reads an "A" on a target strand, it will insert a "T" on the incomplete strand. A copy of the target region is quickly made.

***Taq* polymerase can withstand high temperatures, so the solution can be reheated, causing the new copies to separate into individual strands. A chain reaction begins: from 4 copies to 8 copies to millions of copies in a few hours.**

PCR has been credited with making the Human Genome Project possible — but it may soon be obsolete. Penn State chemist Stephen Benkovic and post-doctoral scholar Frank Salinas filed a patent disclosure last April on a "PCR replacement" that will be both faster and cheaper to use.

—Daniel DeJoseph

Stephen J. Benkovic, Ph.D., is Evan Pugh Professor of Chemistry in the Eberly College of Science, 414 Wartik Lab, University Park, PA 16802; 814-865-2882; sjb1@psu.edu.