Another Human Genome Project

A private company's plan shocks the genetics community

By JOHN TRAVIS

Consider it modern biology's equivalent of the fabled race between the tortoise and the hare—except that the prizes at this finish line are the priceless secrets of the human body and the tortoise may not repeat its legendary victory.

In the role of tortoise is the international Human Genome Project, a $3 billion worldwide effort—funded by U.S. agencies such as the National Institutes of Health and the Department of Energy—to decipher the complete human genetic sequence by 2005. Playing the hare are J. Craig Venter, a maverick scientist with a knack for startling the genetics community, and Perkin-Elmer Corp. of Norwalk, Conn., the leading maker of automated DNA sequencing machines. With a pistol shot signaling the start of a scientific race that few people anticipated, Venter and Perkin-Elmer announced May 9 their intention of creating a new company whose goal is to unravel the human genome in just 3 years—and for a measly $200 million to $300 million.

"This is a private company paying to sequence the human genome and give it to the public," says Venter.

Knowing little about the new effort, many scientists, particularly officials overseeing the worldwide genome project, are still contemplating how to react. Some express skepticism about the strategy, known as whole-genome shotgun sequencing, that the as-yet-unnamed company has embraced. They are also concerned about what the company plans to do with the massive amount of genetics data it will generate and what access other scientists will have to it.

"What we're doing is to produce the complete human genome sequence to a very high accuracy and with nothing missing," says John Sulston, director of the Sanger Centre in Cambridge, England, which plans to sequence one-third of the human genome as part of the now 8-year-old worldwide effort. "I'm sure that the product they're going to produce will be of lower quality, and we consider it inadequate."

Supporters of the shotgun strategy respond that speed is of the essence. "The whole-genome shotgun approach would leave gaps and regions where the sequence is less certain, but I still think it's the best approach," says James L. Weber of the Marshfield (Wis.) Medical Research Foundation, who last year coauthored a paper arguing that the strategy should be applied to human DNA. "The greatest cost of sequencing the genome is the cost of not having the sequence. It's the cost of missed opportunity. Pharmacologists and biochemists need new targets for diabetes, obesity, epilepsy, asthma, etcetera in order to develop better treatments, and it can take 5 to 15 years to develop a new drug. These factors overwhelm concerns about incomplete information and imperfections."

"This is going to be an incredibly complete, incredibly accurate genome sequence," Venter says.

Venter also dismisses criticisms of the shotgun strategy, arguing that the data produced by his new company will equal, if not surpass, that generated by the worldwide effort. "This is going to be an incredibly complete, incredibly accurate genome sequence," he says.

Venter has a record of backing up such forceful claims. While an investigator at NIH, he hit upon a technique to identify quickly most of the 50,000 to 80,000 genes in the human genome. The method used bits of single-stranded DNA, so-called expressed sequence tags (ESTs), which represent part of a gene's sequence, as lures to identify a gene's complete protein-coding sequence. NIH skepticism about the method prompted Venter to leave and establish the Institute for Genomic Research (TIGR) in Rockville, Md., where he pursued the project privately. He and other scientists have since used ESTs to identify an estimated 80 percent of all human genes.

The EST method ignores much of the genome, however, including the regulatory DNA sequences that control the protein-coding portion of a gene. Nor can ESTs identify every gene.

Consequently, ESTs can't provide all the useful information contained in the complete human genome—and sequencing full genomes happens to be another of Venter's talents. In 1995, TIGR and a research group led by Hamilton O. Smith of Johns Hopkins Medical Institutions in Baltimore published the complete genetic sequence of the bacterium *Haemophilus influenzae*, the first living creature to have its full genome unveiled (SN: 6/10/95, p. 367). Venter and his colleagues have since finished off genomes of many more microorganisms.

Their remarkable productivity stems in large part from whole-genome shotgun sequencing. In essence, the strategy involves shattering an organism's DNA into many thousands of fragments, each containing about 2,000 base pairs, the chemical subunits of DNA. Sequencing machines then read the identity of about 500 base pairs at the ends of every fragment. If this is done with enough copies of a genome, the resulting data enable sophisticated computer software to match overlapping fragments and ultimately reassemble a complete, correctly ordered sequence of DNA.

While Venter proved that the shotgun method works for bacterial genomes, many scientists have been skeptical that the strategy would work to decipher the much larger human genome, with its estimated 3 billion base pairs. Indeed, such projects have been proposed and rejected several times. Among the concerns is that the long stretches of repetitive DNA in the human genome, which don't code for proteins and have no known function, would make it more difficult to connect the fragments.

"People have felt that you might gather all that data and not be able to assemble it [into an ordered genome]. It takes . . . faith," says Frederick R. Blattner of the University of Wisconsin-Madison, who argued for applying a shotgun effort to the human genome more than a decade ago.

In contrast, the international human genome project takes a much more deliberate approach. Its strategy, known as the clone-by-clone method, first divides the genome into small chunks, each of which is copied, or cloned, many times by incorporating it into a replicating microorganism. Before sequencing these clones, investigators determine where on the chromosomes they belong. This mapping has consumed most of the project's

time so far; only about 3 percent of the human genome has been sequenced.

As scientists in the international project begin to concentrate on the sequencing of their clones, they will use the mapping data to order them. Moreover, they can go back to individual clones and resequence any uncertain or missed areas. This finishing phase of the project is actually the most expensive and time-consuming, say researchers, but it's necessary to ensure a trustworthy final sequence. They've set a goal of 99.99 percent accuracy.

The shotgun strategy skips the mapping stage, relying mainly on a computer to line up the many thousand DNA sequences. While the method facilitates quick spotting of new genes, Sulston and other scientists worry that Venter's company plans to skip the finishing phase and will therefore offer a final human genome sequence of inferior quality. "By their own admission, they're going to produce an incomplete product. They're deferring costs and passing them on. It's simply not a sensible way to proceed. In the end, it will consume more resources than going our way," says Sulston.

Venter acknowledges that his company's strategy will leave some 3,000 to 4,000 gaps of about 50 base pairs in length, but he suggests that most will be in regions of repetitive DNA that hold little interest. He also flatly denies that his company's final genome sequence will be less accurate, pointing out that he and his colleagues plan to sequence 10 copies of the human genome to provide the data needed to correctly put together one final sequence.

Furthermore, Venter notes that not all of those copies will be from the same person. By using copies of several different genomes, the shotgun method should reveal common genetic variations among people. Such information is valuable in many ways, from identifying disease genes to determining the appropriate drug for a patient. In the step-by-step approach, each clone comes from a single genome and so will not offer any data about genetic variation, says Venter.

While the shotgun strategy is crucial to the new company's plans, Venter points out that the automated DNA sequencing machines just developed by Perkin-Elmer are equally important. They "provide for a substantial decrease in the cost of large-scale sequencing," says Perkin-Elmer Vice-President Michael W. Hunkapiller. By operating 24 hours at a time without human intervention, the machines dramatically lower labor costs and make possible the sequencing of 100 million base pairs per day, he explains.

The new company will house 230 of the Perkin-Elmer machines in a Maryland facility that Venter calls a super genome sequencing factory. He estimates the cost of electricity alone at $4,000 to $5,000 a day.

Venter and officials at Perkin-Elmer say they will probably patent 100 to 300 of the new genes they expect to discover, but they contend that the majority of the profits will come from establishing sequence databases that researchers will pay to consult and from applying their data in novel ways. Venter stresses that he and Perkin-Elmer plan to release all the data quickly. "There was never any disagreement that if we were going to sequence the human genome that it would be morally wrong to hold that data hostage and keep it secret," he says.

While some scientists remain concerned, others give Venter the benefit of the doubt, noting that he recently gave up nearly $40 million in order to sever a partnership with a biotech firm and make public the genomes of several microorganisms (SN: 7/12/97, p. 29). "He walked away from a lot of money," notes Arthur L. Caplan, a bioethicist at the University of Pennsylvania in Philadelphia. "He understands the public welfare side of this genomic information."

As for the doubts about the shotgun strategy, Venter points to his track record. "The human genome is a much bigger project, but just 3 years ago many of these same scientists didn't think we could do *Haemophilus influenzae*," he says. "Scientists are supposed to be skeptical, but skepticism can also inhibit breakthroughs."

Let the race begin. □