SCIENCEWATCH®
September/October 1997

TIGR's J. Craig Venter Takes Aim at the Big Questions

"I expect that between 50 and 100 genomes will have been completed worldwide by the end of the decade," says J. Craig Venter of The Institute for Genomic Research in Rockville, Maryland, ***"and all of those will transform their fields."***

SW Interviews

The story of how J. Craig Venter brought about a paradigm shift in genomic sequencing has now entered the mythology of science. In 1987, Venter, a former surf-bum and Vietnam medic, was researching adrenaline receptors at the National Institute of Neurological Disorders and Strokes (NINDS) when he read a paper by Caltech's Leroy Hood in *Nature* about an automatic gene sequencer. Venter's lab had just spent a year sequencing its first gene, so he got his hands on the first commercial automatic sequencer, and his lab may have been the first to make it work. He then pushed the idea of using cDNA libraries and Expressed Sequence Tags (EST) for sequencing, and when the National Institutes of Health (NIH) wouldn't fund his approach, he went out on his own.

Backed by venture capital dollars, Venter began a partnership with Human Genome Sciences (HGS). Venter would run the research institute known as The Institute for Genomic Research, or TIGR, while HGS would have first shot at the sequences.

For two years now, genomic information has been flooding out of TIGR. It started with the entire genomes for two bacteria in 1995: *Haemophilus influenzae* and *Mycoplasma genitalium*. The following year, it was *Methanococcus jannaschii*, a member of the archaea family, one of the most exotic forms of life on the planet. TIGR's latest complete genome was released in the summer of 1997: *Helicobacter pylori*, which causes stomach ulcers and stomach cancers, and may be the most widespread of all human pathogens. Now TIGR is at work on the genomes from a dozen different species, including the world's great killers—syphilis, tuberculosis, meningitis, malaria , and even cholera—not to mention plant genomes and, of course, the human genome, of which TIGR has so far deciphered over 40 million base pairs of genetic code from human genes and chromosomes.

Venter's impact, and that of TIGR, has been undeniable. The report on *H. influenzae*, for example, has been the hottest report in biology for the last year, with over 500 citations to date (see the table on page 4, paper #1). As Biology correspondent Jeremy Cherfas points out on page 8 of this issue, Venter and colleagues can claim two papers in the latest Biology Top Ten, with another lurking just beneath at #11. Additionally, since 1990 Venter has produced nine papers that have each tallied over 100 citations, including a 1994 paper, "Mutation of a *mutL* homolog in hereditary colon cancer," (see N. Papadopoulos *et al., Science,* 263[5153]:1625-9, 1994), which has garnered nearly 500 cites in just three years.

Venter, 51, received his B.A. in biochemistry from the University of California at San Diego in 1972. Three years later, he earned his Ph.D. in physiology and pharmacology studying with Nathan O. Kaplan. Venter spent the next 10 years as a professor of biochemistry at the State University of New York at Buffalo, and from 1982 to 1985 he was also an associate chief cancer research scientist at the Roswell Park Memorial Institute. In 1984, he joined NINDS, where he stayed until 1992. Since then he has been TIGR's President and Director/Chairman of the Board.

From his office at TIGR in Rockville, Maryland,

Venter spoke to Science Watch correspondent Gary Taubes.

What do you consider the key break throughs in genomic sequencing that made for the present extraordinary level of success?

Venter: Three components were necessary. Probably the most important was the idea to not just use cDNAs, but to do partial sequences of cDNAs. Having the automated sequencing technology and the computational abilities made the new idea reality. The three things had to go together. The ideas actually came from our initial attempts to try and sequence the chromosomes and then interpret that data. Basically there were no adequate software programs or computer approaches to interpret the sequences once we got them, because on our chromosomes, only 3% of the DNA actually codes for genes.

Even once we had the chromosome sequences, we had to go to cDNA libraries—libraries of expressed genes from tissues—to see if we could amplify something from the library. If we were able to, we would then sequence that amplified product. It was only then that we knew whether we in fact had a gene or not. So after two years of doing this—it took us two years to interpret about 200,000 base pairs in the genetic code, in contrast to 3 billion base pairs of the ultimate product—we found eight genes. It was very slow and cumbersome. It seemed crazy.

The insight that occurred then was one of those ideas that, once you hear it, you go, "Well, gee, I could've thought of that." I think that's why it upsets so many scientists. It was simply realizing that every cell in our body knows how to process this information. The heart cell knows how to go through the whole genetic code and find not just which are the genes, but which are the genes that are strictly specific for the heart. Our brain cells know how to do that for the brain. And maybe because I come from a physiology background, the idea seemed more obvious to me: why not use the cells as our supercomputer? So we take cells from the heart and isolate the messenger RNA and turn that into cDNA and sequence it. That gets us not just the 3% that are genes, but the fraction of that 3% that are specific to the heart. So we get even more information about it. And then the idea waso randomly pick 1,000 cDNA clones and to sequence those. Instead of getting part of one gene, it gave us almost 1,000 genes. We were stunned at how simple it was and how well it worked. In only a few months we doubled the number of human genes that were in the public database.

What do you see as your role at TIGR?

Venter: I try to set the overall agenda and the scientific goals. I ask the broad general questions that we're trying to answer. For instance, what is life? I don't think there are that many biologists trying to answer that one. They're so lost in the everyday detail of what they're doing that they're not asking those kinds of questions. I tried to set it up so that I have the freedom to do so.

Can you give us an idea of how you go about answering that?

Venter: Well, on one level, we're examining these whole genomes and trying to get as much insight as we can on life from seeing the whole picture. We're also working on a reductionist view of trying to take the smallest genome that we have, *Mycoplasma genitalium*, which is only 470 genes, and see if we can't understand how those 470 work together to create life.

Why M. genitalium?

Venter: As far as we know, it's the simplest life form that exists as a completely self-replicating entity. And when we sequenced it we learned that 25% of the genes in it were totally new to biology; they've never been seen before. So our view is, if we can't understand the 470 genes of *M. genitalium*, how are we going to understand the 60,000 in the human genome?

We now have a team that's been working to knock out each one of the *M. genitalium* genes to find out which ones are really essential for life. We found out that we can get rid of maybe 100 of those and still have something that's a living organism. To test that, we're trying to make an artificial chromosome with just those 370 genes in it to see if we can actually get life from it. And if we can do that at this minimalist level, then we can work forward and understand something like *Haemophilus influenzae*, which has about 2,000 genes.

SW You recently broke off your collaboration with Bill Haseltine and Human Genome Sciences. The press reports portrayed it as a not particularly amicable divorce. Could you tell us about your reasons?

Venter: It was all about a fundamental difference between trying to move knowledge forward versus the attitude that HGS took on how they were going to make money off of the knowledge we were generating. We felt it was very important to get our data to the broadest possible audience, and we wanted the broadest possible use and applications. That can only happen if the data's out there.

And that's what's happened: we put it out there and it's being used. Our website is used by thousands of scientists every day. But we generate so much information so quickly that there is no way that any one pharmaceutical company, let alone the entire pharmaceutical industry, can deal with all that information. And the mentality developed at HGS that they didn't want to let other people see the data because somebody else might make a discovery that they didn't make.

SW Didn't they get three months to look at the data, after which it had to be made public?

Venter: It was actually six months, but there was also a clause in our contract that said if there were one or two genes that were really going to have a profound impact on medicine, that they could hold things up for up to 18 months. It was meant to be for those one or two really special things that were going to make a difference. And what started happening was the lawyer mentality and the hoarding mentality. They tried to interpret those clauses and hold everything up for the 18 months, and for even longer. Instead of trying to protect one or two things that were really going to be developed and change medicine, they decided that they could sell what I call "gene futures": You know, "We'll let you see the data but you have to give us money for seeing it and you have to give us part of the money you make off discoveries." The hoarding became a substitute for inventiveness, and that became an increasing conflict with what we were doing. continued

Science Watch®, September/October 1997, Vol. 8, No. 5
Citing URL: http://www.sciencewatch.com/sept-oct97/sw_sep-oct97_page3.htm

Search | September/October 1997 Index | Archives | Contact | Home

Email this page

SCI-BYTES - What's NEW in Research
The Most-Cited Researchers in... | Analysis Of... | Site Map by Field | ! QUICK SCIENCE !
Alphabetized List of All *Essential Science Indicators* Editorial Features/Interviews