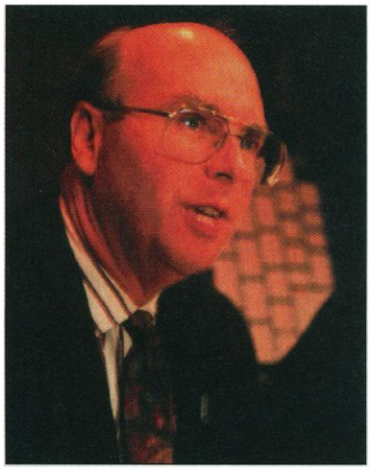
# Hubris and the Human Genome

A nonconformist sequencer has teamed up with a company in a crash project to sequence the human genome. Will the venture help or hurt the government's Human Genome Project?

J. Craig Venter, the DNA sequencing virtuoso who delights in rocking the establishment, shook the Human Genome Project to its foundations last week. In a surprise move, Venter and the Perkin-Elmer Corp. of Norwalk, Connecticut—the world's largest manufacturer of automated sequencing machines—announced on 9 May that they are forming a new company to "substantially complete the sequence" of the entire human genome in 3 years. If successful, this for-profit venture could preempt the work of the $3 billion Human Genome Project, a government-financed coalition of nearly a dozen academic and contract centers—and do it for a fraction of the cost, perhaps as little as $300 million.

Venter aims to pull off this feat by taking a brute-force approach, tackling the entire genome at once rather than in the painstaking, piecemeal fashion currently being followed by the Human Genome Project. The key to the venture, he says, will be a new generation of sequencing machines that Perkin-Elmer will donate to the new company (see sidebar). And he says he won't aim for perfection: He expects his sequence will contain "2000 to 5000 gaps of about 58 base pairs each," which would be expensive to close. The idea is to get the biologically important work done quickly and leave others to fill in the fine details.



R. KOZAK

**I'll do it my way.** J. Craig Venter plans to sequence the whole genome at once.

This bold plan has split the genome community. "It strikes me that this is a cream-skimming approach," says Robert Waterston, head of a large sequencing project at Washington University in St. Louis. "It's clearly an attempt to short-circuit the hard problems and defer them to the [research] community at a very substantial cost." Others welcome the venture as a quick way to get the most useful sequence data. "I think it's great," says David Cox of the Stanford Human Genome Center.

For now, federal officials are taking a polite but cautious attitude to the new venture. At a press conference on 11 May—3 days after he first learned of Venter's plan—National Institutes of Health (NIH) director Harold Varmus said "we are eager to get to the goal line," and any private venture that speeds up the process is welcome. And National Human Genome Research Institute (NHGRI) director Francis Collins praised this "significant new initiative." Like many researchers, however, they are skeptical about Venter's chances of meeting the promised cost and quality goals—especially in view of the serious problems federally funded centers have experienced in switching to new technologies and stepping up their rates of sequencing (*Science*, 8 May, p. 814). Moreover, they are concerned that the new company won't release sequence data as quickly.

In view of the uncertainties surrounding the new venture, Collins declined to take up a suggestion from Venter, who participated in the press conference, that the federal program concentrate on sequencing model organisms like the mouse genome while he tackles the human genome. Right now, "it would be vastly premature" for the Human Genome Project to change course, said Collins, adding, "we are 18 months away from assessing" the feasibility of Venter's project.

### Shotgun, or blunderbuss?

Some researchers caution that Perkin-Elmer's new machine has not yet been fully tested, but Venter says "I was blown away" by its capabilities when he saw it demonstrated earlier this year. "It took me about 15 minutes" to conclude that it would be feasible to sequence the entire genome in 3 years. He will resign as president of The Institute for Genomic Research (TIGR) in Rockville, Maryland—the nonprofit research center that he founded and directs—turning that job over to his wife, Claire Fraser. The new genome company he will head—so young it still has no name—will be located next door.

The business plan calls for full-scale genome sequencing to begin next April, after Perkin-Elmer has delivered 230 of its advanced machines. Venter expects to churn out a whopping 100 million bases of sequence data per day, every day. Within a year, he says, "we will have done 99% of the genome." The following 2 years will be spent fitting pieces together and filling key gaps.

The strategy employs a technique called whole-genome shotgun sequencing, an approach TIGR has used to sequence several bacterial genomes. In essence, Venter plans to mechanically blast the entire human genome—all 3 billion bases—into overlapping pieces each no more than about 5000 bases long, sequence the ends of these fragments, and assemble them into a complete genome using powerful computers to fit the sequence data together.

In contrast, the groups participating in the Human Genome Project work with much smaller stretches of the genome at a time. They first chop up the genome into pieces about 150,000 bases long, shotgun these into smaller fragments, and put them into bacteria to be copied many times over and sequenced. If all goes as planned, every base is contained in a half-dozen fragments, and computer programs piece together longer stretches by looking for overlaps between sequences of different fragments.

That final step of piecing together the fragments and filling in gaps accounts for most of the time and effort in sequencing. And it is here that Venter's critics have the biggest doubts about his approach. The task is daunting enough in traditional clone-by-clone sequencing, but "the problem gets bigger the bigger the stretch of DNA you are trying to put together," says Phil Green, a mathematician at the University of Washington genome center in Seattle who developed a widely used computer program for assembling short sequences into longer ones. Moreover, human DNA is worse than bacterial DNA; it contains so many stretches of repeating sequence that it is more difficult to sort out what goes where. "It's a million times harder to put the human genome together than a bacterial genome," says Green.

Venter responds that "we will have 10X coverage [10-fold overlapping clones] of the whole genome; that will be at least as good as or better than the best sequencing center now." Besides, he notes, many of the gaps left by his approach will be no different from the gaps created by the government-funded centers. But that still means, says Green, that "there will be a massive cleanup that somebody

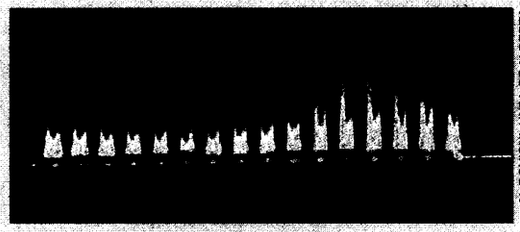
## Picking Up the Pace of Sequencing

Technology has been the key to many a scientific revolution. But the technology at the heart of a venture that may upend plans for the Human Genome Project (see main text) is more of a refinement than a revolution. Officials at Perkin-Elmer's Applied Biosystems Division say their new gene sequencers, which are still in the testing stage, simply apply robotics to a technology that has been improving gradually in recent years.

Known as capillary electrophoresis (CE), the technology is a direct offshoot of the conventional "slab gel" method for separating DNA fragments and reading out their sequence of nucleotide building blocks. For both technologies, researchers begin by chopping a section of DNA into small pieces, each just 1000 to 2000 base pairs long. Each piece is replicated many times, split into four batches, and mixed with the four nucleotide bases, ready to be copied again. But this time a small amount of each of the four bases—a different one for each batch—is modified so that when it is added to the growing nucleotide chain, the chain growth stops. The result is a set of partial copies ending at every position where the original sequence had that particular base—A, C, G, or T. The copies in each batch are labeled with a different fluorescent tag.

In a conventional DNA sequencer, researchers then manually load the fragments onto one end of a 30-centimeter-long polyacrylamide gel, sandwiched between glass slabs. When an electric field is applied, the negatively charged DNA migrates through a lane of gel. Because smaller fragments travel through the gel faster, fragments of different lengths gradually separate. This process is normally slow, taking about 4 hours to complete, as researchers must keep the electrical field low to prevent unwanted heat buildup. At the far end of the gel, a laser reads out the succession of colors, which indicates the sequence of nucleotides.

CE shrinks the lanes in the window-sized glass slabs down to a series of gel-filled glass tubes, each about the width of a human hair. Because of their small volume, the capillaries are better at dissipating heat, which allows them to be run at higher electric fields and thus achieve separations in as little as 2 hours. More important than that speedup, says Bill Efcavitch, who heads genetic analysis research for Applied Biosystems, is the ability of the new equipment to run continuously, 24 hours a day. The machines automatically load the samples, run the separations, detect the fluorescence, and clean out the capillaries between runs. "That considerably reduces your labor costs" and adds productivity, says Efcavitch.



N. DOVICHI, UNIV. OF ALBERTA

**Light work.** Laser light (bright line) triggers fluorescence from DNA samples as they emerge from 16 capillaries.

Other researchers who are working to automate CE say it's not easy. "Getting reproducible, robust runs over and over is a major hurdle," says Harold Swerdlow, at the University of Utah, Salt Lake City. Among the problems: The higher electric fields can disrupt the normally regimented migration of DNA through the polymer gel. Efcavitch says Applied Biosystems has largely solved such problems.

But Elaine Mardis, who heads the technology development group at Washington University's Genome Sequencing Center in St. Louis, doubts that all the bugs have been worked out. Mardis and her colleagues have been running initial tests on a 96-capillary sequencing machine made by Applied Biosystems' rival, Molecular Dynamics. The machine works well, she says, but to be used in full-scale sequencing effort, multicapillary gene sequencers must be 100% reliable, day in and day out. Says Mardis: "Capillaries are no way at that point yet."

–Robert F. Service

will have to do." Adds another genomics expert: "It's a rough draft of the genome. NIH scientists could easily do this at this price."

Indeed, the approach Venter is now pursuing was rejected by the genome community when it was proposed 2 years ago by James Weber of the Marshfield Medical Research Foundation in Wisconsin and Eugene Myers from the University of Arizona, Tucson. The idea was bandied about in meetings and discussed in *Genome Research* a year ago. "[It] was very carefully considered and rejected," says Waterston, "mostly because of the quality of the product [that would result]."

But Weber and several other genome sequencers think that quantity, not quality, is paramount at this time. "It will provide a lot of sequence data much sooner," says TIGR's Mark Adams, who has managed an NHGRI pilot human genome sequencing project for the past 2 years. And researchers have shown that even incomplete sequence information is useful for finding genes of interest and studying other genetics questions. "If we could get sequence information quicker, then in the long run, we'd be ahead," adds Weber.

**What's in it for Perkin-Elmer?**

When Venter and Perkin-Elmer made their dramatic announcement earlier this week, many researchers wondered why the company would plunk down several hundred million dollars for a sequencing venture. To Green, it's a worrisome development: "My guess is Perkin-Elmer wants to stake a claim on the genome." Michael Morgan, the official at Britain's Wellcome Trust who oversees funding of genome sequencing at the Sanger Centre near Cambridge, U.K., also fears that this project could erode an international agreement not to patent raw DNA data.

Venter and Michael Hunkapillar, president of Perkin-Elmer's Applied Biosystems Division, insist that's not the case. They say the new company will try to patent rare but pharmacologically interesting genes it discovers—perhaps "100 to 300"—but only those for which clear biomedical uses have been identified. It will also create a whole-genome database that it will market to academic researchers and companies on a subscription basis. And, because their approach will use DNA from many individuals, sequencing should reveal variations in DNA between individuals that could be valuable for use in clinical research and drug testing. The company will put together a proprietary set of about 100,000 of these single nucleotide polymorphisms, says Venter.

Partly to safeguard its proprietary claims, the company plans to release raw DNA data quarterly, rather than on a daily basis, as many federally funded genome centers are doing. Delays in releasing genome data are, however, a sensitive issue in the genome community, and the new company's plans are expected to spark heated debates at a genome sequencing meeting at Cold Spring Harbor Laboratory in New York this week.

In the meantime, although many scientists involved in the Human Genome Project remain skeptical that Venter can pull off this project, some old rivals say that it would be a mistake to underestimate him. Says William Haseltine, Venter's former business partner and president of Human Genome Sciences Inc. of Rockville: "I'm sure Craig can do it ... and it will be a great thing for science."

–Eliot Marshall and Elizabeth Pennisi