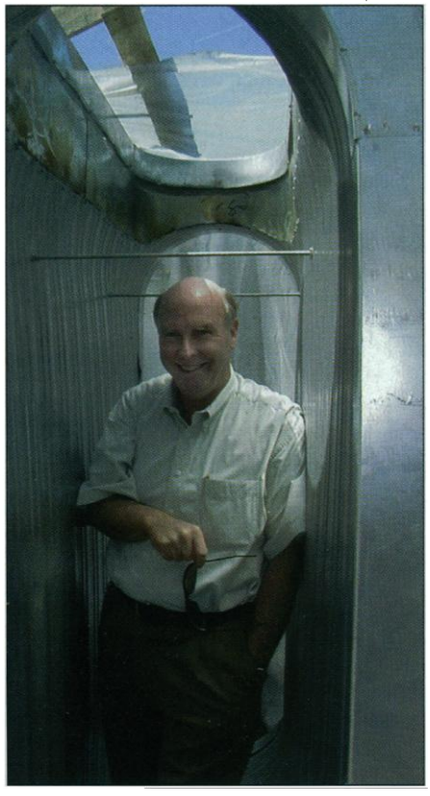
Craig Venter's commercial venture to sequence the human genome, and those of several other complex organisms, has shaken up the international Human Genome Project; but how will it make money?

A High-Stakes Gamble on Genome Sequencing

ROCKVILLE, MARYLAND—In nondescript office buildings here in the northern suburbs of Washington, D.C., where government contractors and other "beltway bandits" ply their trade, a remarkable experiment is taking shape. Celera Genomics, a new company that moved into its quarters just 10 months ago, is preparing to become the world's biggest producer of genome data. This year, it aims to start cranking out original biology on a truly industrial scale and marketing it to the world.

Before it can sell information, of course, Celera must obtain it. And to do so, it must finish installing a phalanx of robotic DNA sequencing machines that are still arriving from the factory, get them running smoothly, create a database, and devise software to mine the data. Somewhere along the line it must also figure out who its long-term clients are and how to make a profit. These are just a few of the challenges the company faces this year. Yet even as Celera tries to convert raw biology into a business, the president, J. Craig Venter, is promising that he will "give away" one of the first, and most important, fruits of this $300 million–plus venture—the DNA sequence of the human genome. The schedule calls for finishing the human genome—the Holy Grail of a separate, nonprofit international project, now in its 10th year—and those of at least three other complex organisms (see table) in just 18 months. How can Celera do this and make money?



Cool environment. Venter stands in cooling duct, part of a system to chill his 230 sequencing machines and state-of-the-art computers.

CREDIT: MARTY KATZ

Venter acknowledges that he hears this question a lot. Business people ask where Celera's profits will come from, while dubious academics ask whether the business agenda is compatible with collegial sharing of data. Venter—never one to mince words—responds that the questioners just "don't get it." Celera must succeed in two worlds, he said in a recent interview: "The scientific community thinks this is just a business project, and the business community thinks it's just a science project. The reality is, it's both." The "business model only works if [we do] absolutely world-class science," Venter explains, "and the science model only works if it's world-class business." In his view, he is implementing a "radical change" in biology, an approach that enjoys "the best of both worlds"—private funding and academic freedom. As a result, he says, he will be working more openly than most companies or academic labs, for both the science and the finances will be open to scrutiny: "Everything will be out in the open." This, Venter insists, "is the opposite of secret."

Many of Venter's peers in academic research don't buy it. They suspect that the science will prove more difficult than expected, or that Celera will have to scale back its promises to share data. And some are offended by Venter's style, especially his wisecracks about the public genome project—gibes he can't seem to resist—and his habit of predicting accomplishments in advance. As one genetics lab director says: "I tell my students they should never announce results before the experiment" is complete. It isn't the way science is done, he says.

But Venter seems to relish breaking the rules and battling skeptics. Over the next 18 months, his boasts will be put to the test in a glare of publicity. He stands to succeed or fail spectacularly.

A basic biology factory

What sets Venter's experiment apart is its scale and promised speed. When Celera's operation is running under full steam later this year, it will be the largest DNA sequencing center in the world. The project is built around an army of 230 freshly minted robots that determine the precise order of nucleotide bases in DNA—the genetic instructions. Getting DNA sequence has been a tedious and, until now, extremely labor-intensive process.

Celera has pinned its hopes on a new type of machine—the PRISM 3700, made by the PE [Perkin-Elmer] Biosystems Corp. of Norwalk, Connecticut—that greatly reduces the need for technical support. Instead of scanning DNA as it migrates through 96 lanes in a series of poured slab gels, it sends the DNA through 96 capillary tubes filled with polymer. In older machines, gels must be poured and reagents frequently reloaded, interrupting the sequencing. But the robot moves DNA and reagents through the tubes continuously, requiring attention only once a day. The system produces a steady flow of data—signals representing the DNA bases adenine, cytosine, guanine, and thymine. At Celera, these machines crowd the lab, yoked to a network of optical fiber, working in chilly silence.

The 3700 machine is central to the project in another way: Its manufacturer, Perkin-Elmer, came up with the idea and is bankrolling the venture. Venter says that early in 1998, Michael Hunkapiller, the PE president who developed the 3700, approached him. PE was ready to provide enough 3700s and "possibly the funding to sequence the human genome." Was Venter interested? Venter says he was intrigued. Venter thought he could use a shortcut method—a "whole-genome shotgun" strategy—that he had used in 1995 to complete the *Haemophilus influenzae* microbial genome with stunning

speed. Venter asked for a promise that, "once we had sequenced the genome, we weren't going to keep it secret." PE executives agreed, "but they turned it back to me," according to Venter, saying that "if you want to use a couple of hundred million dollars and sequence the human genome and give it away, come up with a business model that allows you to do that." Venter has been working on that conundrum ever since.

CONTRIBUTIONS OF NONREDUNDANT HUMAN GENOMIC DNA TO GENBANK

Center	Kilobases	Percent
Sanger Centre	90,908	26
Washington University	82,576	24
MIT/Whitehead Center	38,498	11
DOE Joint Genome Inst.	28,353	8
Baylor College of Medicine	24,660	7
Japan Science & Technology	14,960	4
TIGR	11,640	3
Inst. Molecular Biology, Jena	8,881	3
Univ. Texas Southwestern	7,716	2
Stanford Human Genome Ctr.	6,818	2
Univ. of Washington, Seattle	6,606	2
Other	26,281	8
TOTAL	347,897	100

After reaching an agreement, PE and Venter announced in May 1998 that they were creating a new company that would sequence the entire human genome by 2001. The target was several years earlier than the planned completion date set by the international Human Genome Project, funded mostly by the U.S. government and Britain's Wellcome Trust. It turned the sequencing world upside down (*Science*, 15 May 1998, p. 994).

The news broke on the eve of the annual meeting of academic genome scientists at the Cold Spring Harbor Laboratory (CSH) in New York. Some thought the timing was deliberate. They suspected that Venter, who had been snubbed by the genome research elite in the past, wanted to show them up. A year later, resentment still lingered: Attendees at the 1999 CSH meeting, held last month, warned that Celera was planning to use public resources and grab most of the credit. One center chief grumbled that Celera would leave others "the scut work" of filling gaps.

The sharpest on-the-record critique of Venter's plan came from biologist Maynard Olson, an intellectual leader of the genome project from its earliest days, now at the University of Washington, Seattle. Speaking at a House Science Committee hearing on 17 June 1998, Olson scorned Perkin-Elmer's "science by press release." He objected to a slapdash "biotech style," full of hustle and PR, taking over his field. He predicted there would be "over 100,000 serious gaps" in Venter's version of the human genome and expressed a concern that the publicly funded Human Genome Project would lower standards to keep pace. The latter is already happening: At last month's CSH meeting, leaders of the international project backed a plan to produce a rough "working draft" of the human genome by spring 2000—about a year before Celera's target delivery date—and to complete an accurate, finished version by 2003 (*Science*, 28 May, p. 1439).

Venter, taking a broad swing at his critics, says they're worried about the old "academic funding order." As he told one interviewer last year, "If I were on the other side of this, I would feel upset and threatened, too." The U.S. National Human Genome Research Institute (NHGRI), which provides most of the U.S. funding for the Human Genome Project, may be defensive about having spent hundreds of millions of dollars to create genomic "sequence-ready maps" for a strategy that has now been abandoned, he says. Yet he points out that NHGRI cannot take credit for two key sequencing tools: bacterial artificial chromosomes or BACs, funded mainly by the U.S. Department of Energy, and capillary sequencers, developed by industry. He notes that "98% of the sequencing in the world is done on Perkin-Elmer machines," which even "the NIH [National Institutes of Health] labs are gearing up to buy."

CELERA'S FIRST FOUR TARGETS

Organism	Genome size (million base pairs)	Target for complete sequence*
Drosophila	160	Dec. 1999
Human	3500	Dec. 2001
Mouse	3500	2002
Rice	400	2001

* Celera plans to start releasing *Drosophila* data in July 1999 and human data in September 1999.

Largest Genome Sequenced to Date		
C. elegans	97	Washington U. and Sanger Centre, 1998

Largest Genome Sequenced by Whole-Genome Shotgun		
A. fulgidus	2	TIGR, 1997

In public, NHGRI director Francis Collins takes all this with a smile: He says he welcomes the private investment as "complementary" to the government's work. (Offstage, Collins isn't so chipper. For example, he warns that Celera's statements about releasing data to the public are "disturbingly ambiguous.") But some researchers have embraced Celera's arrival. Bioinformaticist Gabor Marth of Washington University in St. Louis says competition in science is "healthy" and that the field will benefit from Celera's quick pace. Richard Gibbs, director of the genome center at Baylor College of Medicine in Houston, says: "Things have been good since we got a shot in the arm from Craig's activities. The level of excitement is up. ... People will get their data more quickly. So, who should complain?"

World-class robotics

The jousting between Venter and the public genome project continues, but less noisily now that Venter's venture is up and running. Venter boasts, "I have $338 million in the bank" and an amazing scientific lab under construction. Celera got most of this money from PE after PE sold its Analytical Instruments division in January. In addition, Venter reports that the company has secured $100 million in income for 5 years from several clients, who will get an early look at the sequence data. And on 27 April, when Perkin-Elmer was reorganized under a new umbrella called the PE Corp., its stockholders got new shares—one share of PE Corp. and one-half of Celera—for each Perkin-Elmer share they had held. Celera's stock is traded publicly (at $17 a share, down from a peak of $29), but Venter says Celera isn't using stocks to raise money.

Celera's first objective is to finish converting the office buildings into a data factory—and to produce data. Huge cooling ducts, big enough to walk through, were still being finished in May. The sequencing and PCR machines generate lots of heat, and the 3700s require a cold environment. New roof supports were installed to hold the chillers. To prevent even a moment's interruption of electricity, Venter has installed a big generator and hired a fuel truck to deliver diesel. Plans call for 230 of the model 3700 machines to be installed by the end of June; at this writing, about 213 are on site.

A big challenge will be to get those machines to live up to their promise. The first 3700s were rushed to buyers in February with minimal testing or tinkering, and they have not yet achieved targeted efficiencies. At the CSH meeting last month, Washington University researcher Elaine Mardis reported that the performance of the 3700 was "a bit disappointing." They're hard to install, and some units are balky. PE has promised eight runs

SOURCE: (TOP TO BOTTOM) NCBI; CELERA

per day (each run is assumed to use all 96 capillaries per machine, and each capillary produces a "read length" of 500 or more bases). But in May, the machines at Washington University were averaging only five runs a day, with read lengths of 500 bases. Venter says Celera's are delivering six runs a day, with read lengths of 500 to 750 bases. He says he expects the machines will hit nine or 10 runs a day "in a few months." If so, Celera would generate more than 100 million base pairs of data a day. Already, in less than a year, its capacity has grown to 70 million base pairs a day—the world's largest.

Just as bold as the plans for generating raw sequence data are those for computer-based research. The space that holds Celera's main brain—where cables from hundreds of machines converge—is worthy of a sci-fi movie. To reach it, you pass a guard, flash a proximity card, punch a code on a key pad, hold your hand in a biometric scanner, then enter a glass cage watched continuously by TV cameras. Here the data funnel into a digital maze being created by a partnership of Celera and the Compaq Corp. of Houston.

Compaq is building Venter the world's most powerful civilian computer. Its retail value, according to Celera's computer chief, Marshall Peterson, is well over $80 million, although Celera is getting a big discount. The main system is a 64-bit machine powered by 1200 top-of-the-line Compaq alpha processors, connected in parallel and capable of crunching data at almost 1.3 teraflops, or 1.3 trillion floating point operations per second. The only rivals, says Compaq executive Ty Rabe, are those used for classified work by the U.S. government, notably a monster called "ASCI Red" built by Intel to model nuclear explosions.

Ultimately, Celera will employ its computer power to analyze the many organisms it plans to sequence. By aligning and comparing whole genomes, Celera hopes that archived genetic data on the mouse and fly will spotlight new human genes and reveal their functions. But to begin, the computers will be put to work piecing together DNA fragments into complete genomes. This is a lesser task, but hard enough that the consensus a few years ago was that this approach simply couldn't be used to complete the human genome. Some still think Celera may stumble at this hurdle.

Celera is taking a very different tack from the Human Genome Project. NHGRI has funded a multicenter effort that began with a massive investment in genome maps, sets of easily identifiable landmarks on the genome. The idea was that the maps could be used to coordinate the detailed work of sequencing, to come later. That strategy was scrapped, following Celera's announcement, for a faster coordinating method: using a common set of BAC clones into which human DNA has been inserted. Each BAC will have a unique DNA fingerprint and will be anchored by other identifiers to a location on the genome. Five big nonprofit labs have divvied up the genomic landscape and are conferring weekly on who's working on which areas. The challenge now is to get the BACs processed and machines running.

Celera, in contrast, is skipping the mapping and coordination for the whole-genome shotgun approach. (It will benefit, however, from the fingerprinted BACs, for they add structure to the data.) Celera is breaking the entire genome into random clones and sequencing each clone. The clones overlap, so each end sequence, like the shape of a jigsaw puzzle piece, should make a unique match with the end of another clone. With new pattern-recognition software, Celera plans to assemble the sequences of several hundred thousand clones into a complete genome.

Big brain. The nerve center of Celera.

Celera wasn't the first to suggest doing the human genome this way. In fact, James Weber, director for medical genetics at the Marshfield Medical Research Foundation in Marshfield, Wisconsin, proposed this strategy to NHGRI several years ago. With help from Eugene Myers, then a professor of bioinformatics at the University of Arizona, Weber proposed that NHGRI fund a pilot project to shotgun the human genome. But reviewers said it would be "expensive and risky," Weber recalls, and NHGRI declined. Weber and Myers eventually published their proposal in *Genome Research* in December 1997. The same issue carried a rebuttal by biocomputing expert Philip Green at the University of Washington, Seattle. He called the idea "extremely inefficient" and twice as costly as NHGRI's approach.

Even as NHGRI was rejecting the Weber-Myers approach for the human genome, Venter was busy trying it out on a small scale. The Institute for Genomic Research (TIGR)—a nonprofit outfit Venter founded, now run by his wife, Claire Fraser—used it to sequence microbe genomes that were 2 million bases long. That gave Venter confidence that the 3 billion base pairs of the human genome were within his sights. And, to help with the task, last year he hired Myers away from Arizona.

Says Nathan Goodman, a bioinformaticist at Compaq: "When I heard he had hired Gene Myers, my confidence that you could do it shot up." It seemed no longer the "vague mutterings of a genius" but "a plan to be taken seriously."

Green still has doubts. He thinks even Celera's crack informatics team will have trouble figuring out where certain bits of human DNA belong. The human genome contains many "repeat" sequences that may lack unique identifiers. Celera could have a "very big problem" locating the repeats, he says.

Myers, however, is confident that Celera's number-crunching power, combined with new software, will prove the critics wrong. "Truthfully," says Myers, "we have code in place and running that will do human [genome assembly] in less than 3 months," although he admits, "we're still working on getting those repeats resolved."

A crucial test of Celera's strategy will come in the next few months. As a kind of muscle-flexing exercise, the company is assembling the genome of the fruit fly, *Drosophila melanogaster*, in collaboration with an academic team led by Gerald Rubin of the University of California, Berkeley. It is supposed to meet some exacting deadlines. According to a memo issued in February, Celera should begin releasing raw data to the public—and simultaneously to Rubin's team at Berkeley—starting "about July 1, 1999." The first data are now expected in late July. The job is to be finished by January 2000.

Venter says Celera will simultaneously begin sequencing the human genome, the rice genome, and then—possibly as the human work is winding down—the mouse genome. He wants to get started on the mouse "as soon as possible"—perhaps in March—he says, because it will be "essential for interpreting human data." Celera plans to align the genomes, one atop another, for detailed analysis. Boasts Venter: "We are going to discover the circuitry of biology here."

What's for sale?

If much of this information is to be publicly available, how can Celera make a profit sell-

CREDIT: MARTY KATZ

ing it? Venter answers this question in several ways. First, he points to the support of Celera's "early access" clients. He says the $5 million annual fees they pay show that, even though genomic data have been available for years, important customers will pay a premium for Celera's work. Second, a lot isn't being given away: Celera intends to patent several hundred human genes and a large set of human single nucleotide polymorphisms for use in individually tailored medicine. Venter hasn't set the terms for data release of three of the four initial genome projects, or for others on a list of possible targets in agriculture, such as the cow, corn, wheat, soya, and apples.

Finally, Venter says he's not interested in exclusivity because he doesn't want to peddle intellectual property. He wants to create "a big information company," not just for sequence data but for the analyses pouring out of his computers—something like Bloomberg News, which distributes financial data over a closed network. His future customers, Venter likes to say, are not just companies and universities, but "anybody with a genome." He wants to reach the masses.

Celera's first customers, however, are elite: They include the pharmaceutical companies Amgen Inc. of Thousand Oaks, California; Novartis Pharma of Basel, Switzerland; and Pharmacia & Upjohn (P&U) of Bridgewater, New Jersey. William Boyle, who heads Amgen's cell biology lab, says it's easy to explain his interest in Celera: "It's not so much what the information is," says Boyle, but "the timing and the pace with which it will become available. ... It's getting a first look" at the entire human genome. Over time, Boyle says, Amgen expects to collaborate with Celera in comparative genome projects to elucidate the hierarchy, organization, and function of genes.

Les Hudson, head of global research at P&U, says his company is not interested in raw data but in building tools to analyze diseases. Like the other two early clients, P&U has its own, fenced-off computer server located at Celera, which it can access remotely. Hudson expects that the company's bioinformatics group, located at the Karolinska Institute in Stockholm, will use the system to search genomes for "druggable targets." And Novartis's chief of research, Paul Herrling, says that Celera's "key aspect is speed." Novartis, he adds, will use the collaboration to develop "new high-powered computer tools to analyze and annotate" genomic data.

Venter sees his company as an elaboration of a model "validated" by Incyte Pharmaceuticals of Palo Alto, California. This is the concept that you can make a profit marketing nonexclusive access to genomic data, selling information services rather than information ownership. He says he plans to offer data at "a reasonable price" to everyone—including university scientists and citizens who want to learn about their health. But Celera hasn't disclosed the terms.

Because Celera's plans for releasing human genome data remain cloudy, some researchers suspect Venter is having trouble finding a solution that satisfies the company's business plan. One genome center leader predicts Venter will have to curb his academic ambitions to protect the company's investment in data. Celera may grant access, this researcher predicts, but only to those who sign a contract promising not to share information or use it commercially. Indeed, although Celera originally talked about putting human data in GenBank, the public repository at NIH, NIH officials report that discussions are stalled. NIH's stance on the need to share research tools without such strings, issued last month (*Science*, 28 May, p. 1445), may make it harder to work out an agreement.

The never-ending questions about public data release are irksome to Venter. It is "inappropriate for us to be discussing what might or might not be happening with human data vis-à-vis GenBank right now," Venter says. "We're going to make the data available to the scientific community on our Web site, like we've always promised." At the moment, he says, "our goal is to get *Drosophila* done. We're going to let our accomplishments speak for themselves." He adds: "That's the beauty of genomics: Sooner or later you have to come up with the data. If you do, you win; if you don't, you lose."

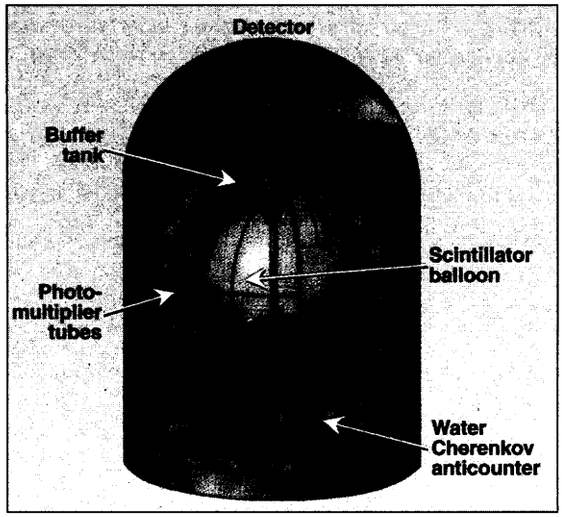
In just 1 month, Celera is scheduled to begin releasing sequence data from the *Drosophila* genome; in 3 months, it plans to start putting human genome data on its Web site. Soon, everyone will be able to judge for themselves who won. —ELIOT MARSHALL

PARTICLE PHYSICS

Experiment Uses Nuclear Plants To Understand Neutrinos

Physicists hope a novel facility being built in a Japanese mine will shed light on the elusive neutrino—and Earth's radioactive heat source

Neutrino research and nuclear reactors go back a long way. The first neutrinos ever detected, in a 1956 Nobel Prize–winning experiment by physicists Clyde Cowan and Frederick Reines, emanated from a nuclear plant. But since then the relationship has cooled. In recent years, physicists trying to understand these elusive particles have targeted the high-energy neutrinos coming from space or from accelerators at high-energy physics labs because of the logistical problems of siting detectors at the right distance from enough reactors. Now the old flame is reviving in a Japan–United States collaboration that is building a massive underground snare for neutrinos emitted by Japan's nuclear power plants, which may hold the key to neutrino puzzles that are hard to unlock with other approaches.



CREDIT: KAMLAND COLLABORATION

Mining knowledge. This detector in a Japanese mine will use a special chemical soup to capture evidence of neutrinos.

Called KamLAND (Kamioka Liquid scintillator Anti-Neutrino Detector) and located beneath the mountains of central Japan, the detector will catch antineutrinos—the antimatter counterparts of neutrinos—from the country's 51 nuclear power reactors, as well as neutrinos directly from the sun. By studying how the neutrinos behave on their way to the detector, the project members hope to add to recent evidence that neutrinos—assumed until recently to be massless—do have mass. And because nuclear reactors produce neutrinos in similar energy ranges to those produced in the sun, KamLAND may help physicists explain the so-called solar neutrino deficit: the shortfall—by up to one-half—