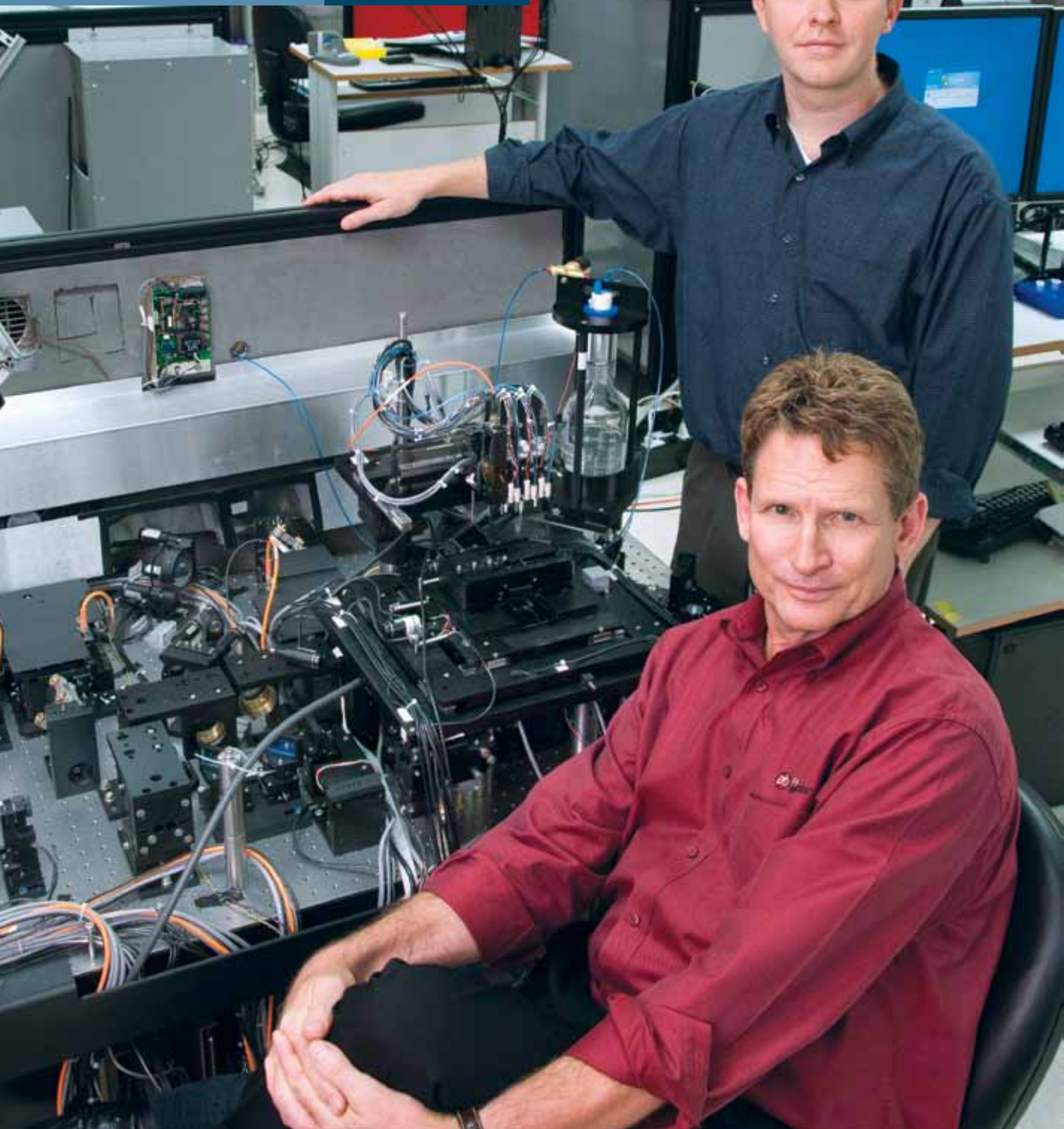


WHERE SCIENCE AND SOCIETY MEET

THE JOURNAL
OF
LIFE SCIENCES

SEQUENCING



Speed Freaks

The race is on to sequence whole genomes for \$1,000, but already companies are talking about shattering that mark and making scans no more expensive than routine procedures such as blood and urine tests.

By Daniel S. Levine

On May 6, 1954, at a track meet in Oxford, U.K., Roger Bannister ran a mile in 3 minutes, 59.4 seconds. At the time, many had considered running a mile in under four minutes an impossible feat. Some viewed it as an insurmountable psychological barrier, if not one of physics itself. In fact, the breaking of the four-minute mile is still considered among the greatest sports accomplishments.

If the sequencing of a whole human genome—determining an individual's entire genetic code—for \$1,000 were the equivalent of running a four-minute mile 55 years ago, think of Hugh Martin as some kid in the early 1950s boasting that he would soon run a 40-second mile—and mean it.

The CEO of Menlo Park, California-based Pacific Biosciences has that jockey kind of swagger, typical of executives who have risen to the top in Silicon Valley. A veteran of the telecom and computer gaming industry, Martin was recruited by the venture capital firm Mohr Davidow Ventures to head Pacific Biosciences in 2005. The pick was a recognition that so-called third generation sequencing, new approaches to sequencing that promise to one day reduce the time and cost of sequencing to once unimaginable levels, was as much an information technology challenge as a biochemical one.

Built upon technology initially developed at Cornell University, Pacific Biosciences expects to go to market by the second half of 2010 with a device that can leave previous sequencing machines in the dust. The goal is to sequence 3 base pairs—the pairs of molecules that make up the rungs of the twisted DNA ladder—per second. With improvements to reagents, the chemicals used to make possible the reading of the DNA, the device will eventually be able

to complete 10 base pairs a second. That compares to second-generation machines today that sequence at the relatively plodding rate of just a base pair per hour, Martin says.

Eventually—about three years after the introduction of the first machine—Martin expects to produce a follow-up device that will be able to sequence a genome at a rate of 50 base pairs a second and at a cost of \$100. “When you go to the doctor and the doctor fills out the form and checks off blood, urine, test this, test that, there isn't anything on that form that cost more

When you go to the doctor and the doctor fills out the form and checks off blood, urine, test this, test that, there isn't anything on that form that cost more than 150 bucks. If we're going to have a box that says “sequence,” we are going to have to get it to the point where it's in that class of price.

—Hugh Martin, CEO, Pacific Biosciences

than 150 bucks,” says Martin. “If we're going to have a box that says ‘sequence,’ we are going to have to get it to the point where it's in that class of price.”

In the genome-sequencing field, it can be difficult to discern the true merit of a boast. Thank goodness there may be one way to gauge the technical and cost accomplishments of various efforts. It is the Archon X Prize, a \$10-million bounty being offered to the first team to successfully sequence 100 genomes (the full set of genetic material consisting of paired chromosomes, one from each parental set totaling

Pacific Biosciences CEO Hugh Martin (*seated*) with company founder and CTO Stephen Turner surround a prototype of their sequencer. The company expects its technology in the next several years to sequence an entire human genome for \$100.



It will turn out one of our major costs will be electricity for running our data center. The reagents cost is on its way to zero and the major cost will be electricity.

—Clifford Reid, CEO, Complete Genomics

6 billion base pairs) in 10 days, for less than \$10,000 a genome. So far, the prize has gone unclaimed. Although, at least seven teams have registered to vie for it. The prize sets a high bar. The winner must also demonstrate accuracy of no more than 1 error in 100,000 base pairs, with sequences accurately covering at least 98 percent of the genome. The completed results will need to include all insertions and deletions, all rearrangements, and other technical requirements aimed at measuring the accuracy and completeness of the sequencing. To date, none of the teams has successfully completed the task, let alone declared they were prepared to try.

The notion of the \$1,000-genome began to take shape before the completion of the mapping of the first human genome, according to Jeff Schloss, program director, technology development coordination at the National Human Genome Research Institute. As best as

Clifford Reid, CEO of Complete Genomics, is taking a different approach than his competitors. Rather than seeking to sell instruments and reagents, Complete Genomics is pursuing a service business model.

Schloss can tell running through documents at the institute, the first discussion of the \$1,000 genome took place in December 2001 during a planning workshop held by the National Human Genome Research Institute at the Arlie House in Warrenton, Virginia. There are staff notes that it had been proposed as a topic for discussion, but the record is unclear by whom. The institute, looking at the completion of human-genome sequencing within reach, had invited researchers, biopharmaceutical executives, and others with a stake in the emerging field of genomics. On the plate was identifying the problems the emerging science could solve as well as the tools that would be needed to do so. The institute at the time set goals for the next five to eight years.

Schloss admits that eight years ago, a \$1,000 genome was a “completely audacious” goal. He estimates that the cost of sequencing the first human genome—the genome itself excluding other costs related to the Human Genome Project—was about \$500 million. It took a year. “The idea that one could sequence genomes for much, much less money than it cost was an intriguing one,” he says. “I’m not sure anyone believed it could be done. To the extent that there was a discussion of the realities, people thought it would take between 10 and 25 years to get there.”

But the goal of the \$1,000-genome is not an arbitrary one. It reflects a reality that the mapping of a single human genome wasn’t by itself going to provide the insight into human disease contained in DNA. To do that requires sequencing hundreds of genomes from people with a specific disease and comparing their genomes to the genomes of healthy people. Sequencing so many genomes at a cost of \$500 million—or even \$100,000 as was the cost a few years ago—would be economically prohibitive.

That’s why the ability to sequence genomes for \$1,000 a piece, in simple terms, is a game changer. It promises to allow researchers to ask questions about biology and medicine in a completely different way. “Having one, you say ‘okay, this is what a human genome looks like,’ says Schloss. “What you really want to be able to do is look at the genomes of many individuals, particularly of people who have a particular disease, and compare that to people who don’t have that disease and see what’s different. That’s what you really want to see. And then you want to do that for all the major diseases.”

That’s precisely what Complete Genomics hopes to do. Already, the Mountain View, Cali-

fornia-based company says it is offering whole genome scans to customers seeking to scan eight or more genomes at a price of \$5,000 a piece. Unlike its competitors, Complete Genomics' business model is not to market instruments and reagents. Instead, it is pursuing a service model where customers send samples to its scanning facility and it returns the completed data. To date, the company has signed up a dozen customers, but it eventually envisions having scanning and data facilities throughout the world. "The key distinguishing characteristic between us and the other guys is scale," says Clifford Reid, the company's CEO. "We're built for scaling up. Everything we do is for scale."

The company's vision is to work with partners, such as countries, research organizations, or pharmaceutical companies to build sequencing centers over the next five years. Reid says he expects to open about 10 of these facilities around the world. Collectively, those 10 sequencing facilities over five years will have the capacity to sequence about 1 million genomes. "The way to think about 1 million genomes is that's 1,000 people in each of 1,000 disease studies," he says. "That's all of the important disease right there. In the next five years, we can understand the genetic basis of all of the important human diseases and that's going to change medicine."

Though the sequencing field is advancing very rapidly, it's difficult to pinpoint its exact

tion sequencing technology known as Sanger sequencing, despite room for improvement, would never be fast and cheap enough to make sequencing affordable for broad use. The method, named for its developer Fred Sanger, uses the molecular building blocks of DNA and the enzyme DNA polymerase to clone fragments of DNA. The cloning of the DNA fragments is repeatedly halted so that fragments of varying lengths are cloned. Fluorescent markers iden-



This is a young industry. We're just getting going. You are going to attract some smaller, newer players into it. This is a field that is moving very fast indeed. It is ideally suited to startups in some ways, although in the end you are going to need a big commercial infrastructure to cover the world.

—Mark Stevenson, president and COO, Life Technologies

progress in part because of the dizzying pace of change, and in part because people, quite simply, play fast and loose with the numbers. When it comes to sequencing costs, some focus on the cost of the reagents used while others include labor and amortization of equipment in the calculation. What is often left out of such calculations is the quality of the end result.

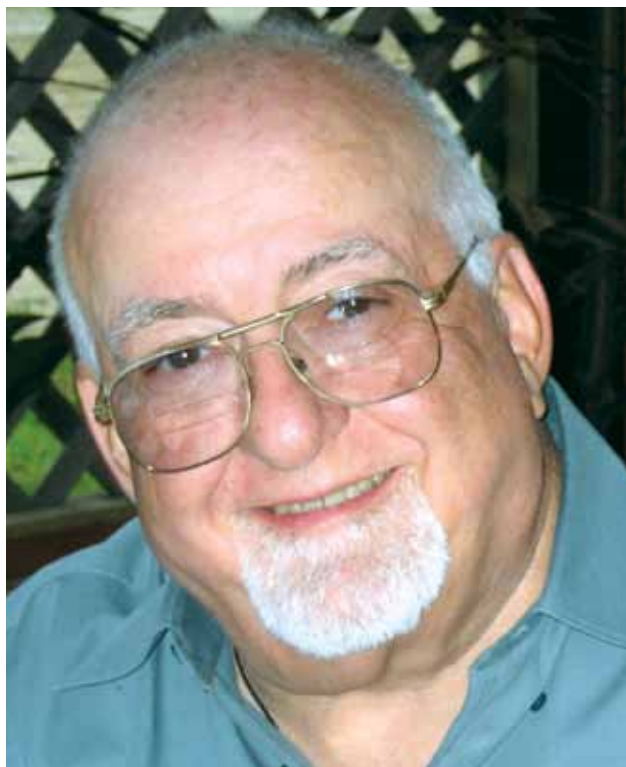
By the time the first human genome was mapped, improvements to the technology shrunk the cost of sequencing from to an estimated \$10 million from \$500 million. But there was wide belief that the first genera-

tify the location of the so-called nucleotides consisting of adenine, cytosine, guanine, and thymine. The fragments are sorted by length and then powerful computers assemble the data to stitch the information into a complete genome.

The second-generation technology involves the use of massively parallel sequencing carrying out thousands or millions of DNA fragments at once. It uses smaller fragments and has made sequencing more of a computing challenge. It has driven the cost down by most estimates to below \$100,000 a genome.

Mark Stevenson, president and COO of Life Technologies, is betting the company's size and reach will give it a big advantage over a new generation of startups when it comes to being commercially successful.

Costs are still falling—and fast. Life Technologies has said it expected researchers to be able to scan a genome for less than \$10,000 this year with its current system and thinks its second generation technology might eventually be able to reach the \$1,000 mark. But it is within the so-called third-generation sequencing technologies that the most radical cost breakthroughs are expected.



Larry Kedes, senior advisor to the Archon X Prize, doesn't think any current commercial technology will be able to grab the \$10 million bounty for sequencing 100 genomes in 10 days for \$10,000 or less per genome. In fact, he hasn't seen any technology in development that he thinks can yet meet the stringent requirements of the prize.

Third-generation sequencing is not a single technology. Rather, it's a range of approaches being pursued by various companies. What they share in common is that they read DNA a single molecule at a time. Some argue that Complete Genomics technology in this regard is not really third generation but instead a souped-up version of second generation technology. They argue the company is using a massively parallel approach where cost efficiencies have come about in part through the company's ability to densely pack DNA into arrays and minimize the amount of reagents needed. But such distinctions will not be as meaningful as results. Complete Genomics calls its technology third-generation because of the efficiencies it says it has been able to achieve. "It will turn out one of our major costs will be electricity for running

our data center," says Complete Genomics Reid. "The reagents cost is on its way to zero and the major cost will be electricity."

Among the technologies being looked at now is reading single DNA molecules in real-time as the enzyme polymerase is used to assemble a complimentary strand of DNA, mirroring the natural process of DNA replication that takes place within the cell. Fluorescent material bound to the different molecules that make up DNA, reveal the sequence. Pacific Biosciences, which calls its technology Single Molecule Real Time or SMRT, describes its approach as "eavesdropping" on a single DNA polymerase molecule as it assembles.

Unlike the truckloads of reagent needed for second-generation sequencing, this approach uses one molecule of reagent for each base pair, the same efficiency as inside the cell when DNA is replicated through natural processes. Life Technologies, through its acquisition of Houston-based VisiGen at the end of 2008, is pursuing similar technology.

Another approach is the so-called "nanopore" technology, which is being pursued by companies such as Oxford Nanopore Technologies, an Oxford, U.K.-based company backed in part with an \$18-million investment from San Diego-based Illumina made this January. That technology also reads DNA a base pair at a time, but by pushing DNA through a tiny hole formed in protein. Other companies using the nanopore approach use synthetic instead of protein nanopores. An enzyme is used to cleave DNA one base pair at a time and the cleaved base pair is read as it passes through the nanopore.

But there are still other approaches being pursued. Some of the teams that have registered to compete for the Archon X Prize seem to have technologies that are unique. North Reading, Massachusetts-based ZS Genetics is using heavy elements to label DNA molecules and make them visible to modified transmission electron microscopes. Reveo, an Elmsford, New York-based company, is creating an electro-optic sequencer using what the company calls "nano-knife edge" probes to measure the frequency at which each base of DNA vibrates when excited by an electrical charge. It's a potentially rapid and inexpensive means of sequencing. Industry observers say it's likely that multiple technologies will emerge with strengths and weakness that make them well suited for one application but not another.

But it is not just a new generation of startups that are chasing low-cost sequencing. The big

players in the field today are also in the game. Companies such as Life Technologies, Illumina, Roche's 454 Life Sciences, and Helicos have been driving down the cost of their current technology and investing in next-generation technology.

The \$1,000-genome will likely be available sometime in 2011 or 2012, predicts Mark Stevenson, president and COO of Carlsbad, California-based Life Technologies, the company formed in 2008 through the merger of Invitrogen and Applied Biosystems. The first to reach the goal will surely gain some recognition for hitting a milestone, he says. But he argues it will be more important to have a product that can be integrated into the entire biomedical ecosystem from research centers to electronic health records used by doctors. Some of the startups in the field may be successful at developing their technology, he adds. But it will take much more than that to be commercially successful.

"This is a young industry," Stevenson says. "We're just getting going. You are going to attract some smaller, newer players into it. This is a field that is moving very fast indeed. It is ideally suited to startups in some ways, although in the end you are going to need a big commercial infrastructure to cover the world. This is partly why Invitrogen and AB came together."

Larry Kedes, a senior advisor to the Archon X Prize for Genomics, the group offering a \$10 million-bounty to the first group that can sequence 100 genomes in 10 days for \$10,000 or less per genome is not surprised that no one has yet claimed the prize. In fact, he doesn't think any of the current commercial technologies are capable of winning. The professor emeritus of biochemistry and molecular biology at the University of Southern California also says he doesn't think the majority of the technologies in development that he's aware of will be able to meet the prize requirements for cost, speed, completeness, and accuracy.

The good news is, the marketplace initially may not require third-generation technologies to meet the X Prize goals. Some third-generation technologists and companies are going to be satisfied delivering less than the capability of the prize because they feel the market can tolerate that depending on the application of the sequencing and the improvements in cost and efficiency new technologies may offer initially. But ultimately, Kedes says what the X Prize is asking for is essentially what's needed for a medical payoff from sequencing. That, he says, is because there are lots of diseases



and disease variants that are just not going to get picked up in less sophisticated scans unless you have that more granular, so to speak, information. The whole genome must not only be just sequenced—but done so with a high level of accuracy and completeness—to provide an understanding of genetic variation and what it means, he argues.

"There are a lot of very smart people out there with extraordinarily exciting technologies that have a theoretical shot at being able to do this," says Kedes. "I just don't think anyone has published anything yet or revealed anything yet that says they are even close to accomplishing the goals of the X Prize."

It is worth noting that when the runner Bannister did break the four-minute mile and set a new world record, it stood for a mere six weeks. On June 21, 1954, Australian John Landy bested Bannister's time by setting a new record of 3:57.9 at an international competition in Finland. Since then, it has become common for top runners to break the mark.

Though we have yet to see a commercial whole genome sequenced for \$1,000, breaking that mark may well become routine in a few years. And while such a goal may have seemed beyond reach when scientists sequenced the first human genome, people are now talking about sequencing genomes for less than \$100. It now seems likely that within a few years, whole genome sequencing will be affordable to most researchers who would want access to the technology. The question will be how quickly resulting discoveries gets translated into new understandings of disease and new treatments for patients. [tjols](http://www.tjols.com)

When Roger Bannister broke the four-minute mile, his astounding record stood for just six weeks and the mark is now routinely broken by world class runners.