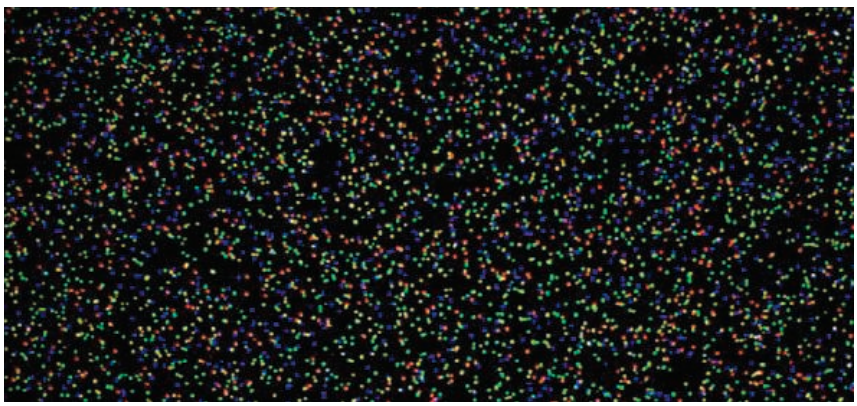


**Charting islands.** Glowing dots on a glass slide mark cloned DNA being sequenced.

GENE SEQUENCING

# The Race for the $1000 Genome

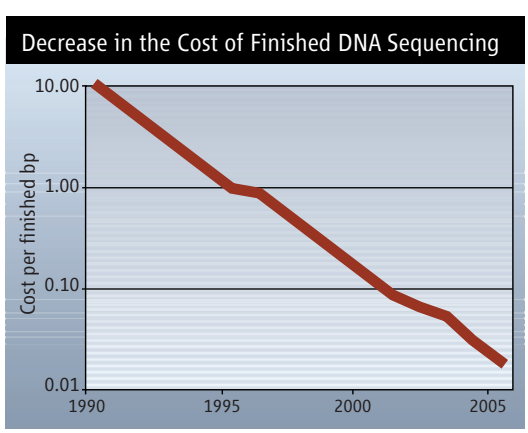
**Fast, cheap genetic analyses will soon become a reality, and the consequences—good and bad—will affect everybody**

**MARCO ISLAND, FLORIDA**—Computers aren't the only things getting better and cheaper every time you turn around. Genome-sequencing prices are in free fall, too. The initial draft of the first human genome sequence, finished just 5 years ago, cost an estimated $300 million. (The final draft and all the technology that made it possible came in near $3 billion.) Last month, genome scientists completed a draft of the genome sequence of the second nonhuman primate—the rhesus macaque—for $22 million. And by the end of the year, at least one company expects to turn out a full mammalian genome sequence for about $100,000, a 3000-fold cost reduction in just 6 years.

It's not likely to stop there. Researchers are closing in on a new generation of technology that they hope will slash the cost of a genome sequence to $1000. "Advances in this field are happening fast," says Kevin McKernan, co–chief scientist at Agencourt Bioscience in Beverly, Massachusetts. "And they are coming more quickly than I think anyone was anticipating." Jeffrey Schloss, who heads the sequencing-technologies grant program at the National Human Genome Research Institute (NHGRI) in Bethesda, Maryland, agrees. "People are roundly encouraged and nervous," Schloss says—encouraged because their own technologies are working, and nervous because their competitors' are too.

A host of these novel sequencing technologies were on display last month at a meeting here.* Although no one at the meeting claimed to have cracked the $1000 genome sequence yet, researchers are getting more confident that it's a real possibility. "From what I've listened to the last few days, there is no physical principle that says we shouldn't be able to do a $1000 genome," says Harvard University sequencing pioneer George Church.

Even today, the declining cost of genome sequencing is triggering a flowering of basic research, looking at broad-ranging topics such as how the activation of genes is regulated and understanding genetic links to cancer. And as prices continue to drop, sequencing will revolutionize both the way biologists hunt for disease genes and the way medical professionals diagnose and treat diseases. In fact, some researchers say cheap sequencing technology could finally usher in personalized medicine in a major way. "The promise of cheap sequencing is in the understanding of disease and biology, such as cancer, where the genome changes over time," says Dennis Gilbert, chief scientist of Applied Biosystems, the leading gene-sequencing-technology company based in Foster City, California. "It will enable different kinds of science to be done." Of course, as with other forms of high technology, that promise brings new risks as well. Researchers expect cheap sequencing to raise concerns about the proliferation of bioterrorism agents as well as patient privacy.




**Free fall.** As with computer technology, the plunging cost of DNA sequencing has opened new applications in science and medicine.

## The race is on

The first group to produce a technology capable of sequencing a human genome sequence for $1000 will get instant gratification, as well as potential future profits: In September 2003, the J. Craig Venter Science Foundation promised $500,000 for the achievement. That challenge has since been picked up by the Santa Monica, California–based X Prize Foundation, which is expected to up the ante to between $5 million and $20 million. But the competition really began in earnest in 2004, when the National Institutes of Health launched a $70 million grant program to support researchers working to sequence a complete mammal-sized genome initially for $100,000 and ultimately for $1000. That program has had an "amazing" effect on the field, encouraging researchers to pursue a wide variety of new ideas, says Church. That boost in turn has led to a miniexplosion of start-up companies, each pursuing its own angle on the technology (see table, p. 1546).

All are racing to improve or replace a technology first developed by Fred Sanger of the U.K. Medical Research Council in the mid-1970s that is the basis of today's sequencing machines. The technique involves making multiple copies of the DNA to be sequenced, chopping it up into small pieces, and using those pieces as templates to synthesize short strands of DNA that will be exact complements of stretches of the original sequence. The synthesis essentially mimics the cell's processes for copying DNA.

The technology relies on the use of modified versions of the four bases that make up DNA, each of which is tagged with a different fluorescent marker. A short DNA snippet called a primer initiates the synthesis at a specific point on the template DNA, and the altered bases—which are vastly outnumbered by normal bases in the mix of reagents used to perform the synthesis—stop the process when one of them is tacked onto the end of the growing DNA strand. The result is a soup of newly synthesized DNA fragments, each of which started at the same point but ends at a different base along the chain.

Today's sequencers separate these fragments by passing the soup through tiny capillaries containing a gel; the shorter the fragment, the faster it moves through the gel. The process, known as capillary electrophoresis, is so effective that

CREDITS (TOP TO BOTTOM): OMEAD OSTADAN/SOLEXA; ADAPTED FROM GRAPH PROVIDED BY JEFFREY SCHLOSS/NHGRI

* Advances in Genome Biology and Technology Conference, Marco Island, Florida, 8–11 February 2006.

each fragment that emerges from the capillary is just one base longer than the one that preceded it. As each fragment emerges, it is hit by a laser, which causes the altered base at the fragment's tip to fluoresce. A computer records the identity of these bases and the sequence in which they appear. Eventually, the process generates billions of stretches of sequence that are fed into pattern-recognition software running on a supercomputer, which picks out overlaps and stitches the pieces together into a complete genome sequence.

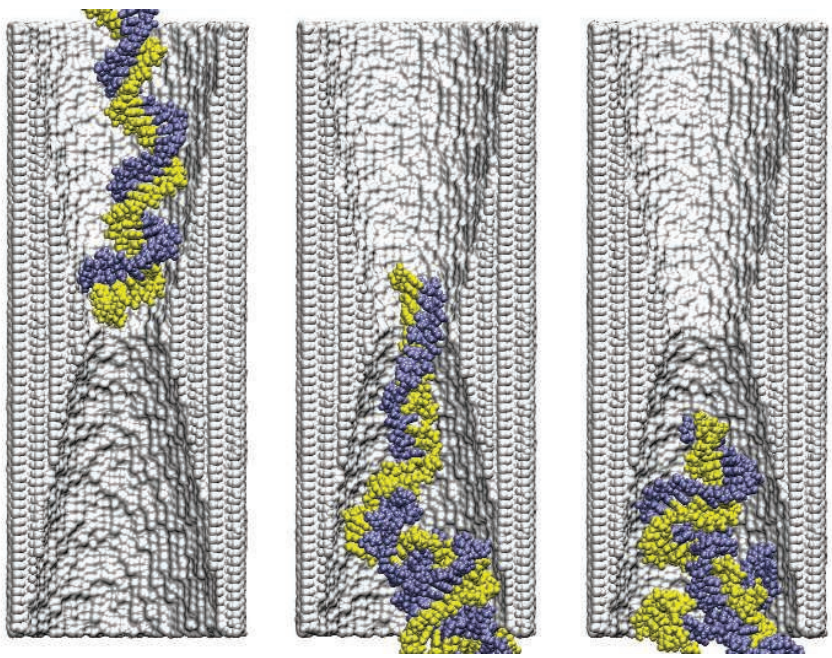
A long list of refinements in capillary electrophoresis systems, coupled with increased automation and software improvements, has driven down the costs of sequencing 13-fold since these machines were introduced in the 1990s. Most of the new technologies aim to miniaturize, multiplex, and automate the process even further. They fall into three main camps. The first, called sequencing by synthesis, tracks bases as they are added to a growing DNA strand. Second is a group of techniques that sequence single DNA molecules. Finally, nanopore-sequencing technologies coax DNA to wriggle through a tiny pore and read the bases either electronically or optically as they go by.

Sequencing-by-synthesis strategies have a head start. Indeed, one company, 454 Life Sciences Corp. in Branford, Connecticut, already has a commercial instrument; it sold 20 of them last year. The company's technique, called pyrosequencing, first chops a genome into stretches 300 to 500 base pairs long, unzips the double strands, discards one strand, and links the other to compounds tethered to a plastic bead—each bead gets just one strand. These snippets are then copied by the polymerase chain reaction (PCR) until the copies cover each bead. The beads are separated on a plate containing as many as 1.6 million wells and dosed with a series of sequencing reagents and nucleotides. Every time a nucleotide is tacked onto a growing DNA chain, the reaction triggers the release of a compound called pyrophosphate, which in turn prompts a firefly enzyme called luciferase in the well to give off a flash of light. By correlating the recorded flashes from each cell with the nucleotides present at the time, a computer tracks the systematic sequence growth of hundreds of thousands of DNA snippets simultaneously.

In August 2005, 454 Life Sciences researchers reported that they had sequenced the nearly 600,000-base genome of a bacterium known as *Mycoplasma genitalium* with an accuracy of 99.4%, as well as the larger 2.1-megabase genome of *Streptococcus pneumoniae* (*Science*, 5 August 2005, p. 862). At the Florida meeting, Michael Egholm, 454's vice president for molecular biology, reported that they had since sequenced four different microbial genomes, each with greater than 99.99% accuracy. "In a 6-month period, we have dramatically improved the data quality," Egholm says. Higher accuracy is critical because two genomes being compared, such as those of normal cells and cancer cells, could differ in only one part per million.

David Bentley, chief scientist for Solexa in Little Chesterford, U.K., also reported heady progress. Like 454's approach, Solexa's turns separate snippets into roughly 1000 exact copies. Instead of attaching individual DNA strands to a separate bead, Solexa researchers fix each strand to a different spot on a glass slide, much as they do in standard microarrays. They then duplicate those strands, creating myriad tiny DNA islands. Finally, in a step akin to Sanger sequencing, they use nucleotides with four different colors and standard microarray optics to simultaneously track the growth of strands complementary to those attached to the slide. Bentley reported that his team had sequenced a 162-kilobase stretch of human DNA and compared it to the standard reference sequence worked out by the Human Genome Project. Their sequencing turned out to be more than 99.99% accurate and spotted all 162 common mutation sites known as single-nucleotide polymorphisms known to exist in that stretch of DNA.

Church has developed a slightly different sequencing approach, part of which Harvard has licensed to Agencourt. In this approach, called sequencing by ligation, researchers start with a short snippet of DNA bound to a bead or a surface. They then add a short stretch of DNA called an anchor primer that binds a known starter sequence on the DNA snippet. Additional nine-base primers, known as query primers, are then added to the mix. These primers come in each possible sequence combination, and each has a labeled A, G, T, or C at just one position. If a short primer with a correct complementary sequence binds to the DNA, an enzyme called ligase stitches it to the anchor primer to hold it to the surface, and the other primers, which bind less tightly, are washed off. The mix is then hit with a blast of laser light to reveal the color of fluorescence that gives away the identity of the newly bound base. Finally, the query and anchor primers are stripped away, and another anchor primer is added as the first step to identifying the next base in the template strand. Agencourt's McKernan said their version of the technology could currently sequence some 200 million bases a day and may reach 3 billion a day by August.



**Pore fit.** In this computer simulation, DNA wriggles through a 1.5-nanometer pore in silicon. Such devices hold the promise of sequencing DNA electronically.

CREDIT: IMAGE COURTESY OF ALEKSEI AKSIMENTIEV, KLAUS SCHULTEN, AND GREGORY TIMP/UNIVERSITY OF ILLINOIS, URBANA-CHAMPAIGN

### Slow start, strong finish?

Despite these advances, sequencing by synthesis has its drawbacks. One is that the techniques read relatively short DNA snippets—usually several hundred base pairs in length or less, compared with the 1000 or so in current capillary systems. That can make it hard to reassemble all the pieces into a continuous genome sequence. Another drawback is that they rely on PCR, which is expensive and can introduce copying errors. Greater experience with the new sequencing technologies may improve matters. 454's Egholm, for example, says his team has developed a prototype version of their technology that increases read lengths from 100 base pairs to 400. Several groups are developing ways to sequence a single copy of a long DNA strand, thereby achieving longer reads and avoiding PCR.

One approach being pursued by VisiGen Biotechnologies in Houston, Texas, anchors a polymerase—the enzyme that tacks new


Published by AAAS

nucleotides to a growing DNA chain—onto a surface and feeds it a template strand. As the polymerase then adds fluorescently labeled bases to a complementary strand, an advanced optical system detects the tiny flashes from the single molecule, allowing a continuous sequence to be read. A variation of this approach by LI-COR Biosciences in Lincoln, Nebraska, anchors single-stranded DNA and polymerase molecules to an electrode, and then uses an electric field to drive nucleotides linked to fluorescent nanoparticles in solution toward the polymerase. In the instant between the time when the polymerase latches onto the nucleotide and the time when it cuts it off the nanoparticle, the researchers reverse the electric field, driving away nucleotide-nanoparticle pairs not bound to the DNA. Then they snap a picture to see the color of the fluorescent particles still bound to the polymerase. Once the nucleotide is cut free, the nanoparticle drifts away, and the process is repeated to identify the next base. At the meeting LI-COR's John Williams predicted that this technique could produce read lengths of up to 20,000 bases.

But another technology altogether may hold the most revolutionary potential, Church says. Called nanopore sequencing, this family of techniques aims to sequence DNA strands as they thread their way through tiny synthetic or natural pores, each just 1.5 nanometers or so across. Numerous groups are pursuing nanopore synthesis techniques, but researchers acknowledge that they have far to go. "We're still learning about the science of nanopores," Schloss says.

No group has yet reported using such a setup to sequence DNA one base at a time. But in a series of papers beginning in 1996, researchers led by John Kasianowicz and Daniel Branton at the National Institute of Standards and Technology in Gaithersburg, Maryland, reported that they could use protein-based pores embedded in a lipid membrane first to detect snippets of DNA and then to differentiate snippets with all A's from those made up of C's. But because proteins and lipids are fragile, other groups have begun making their pores out of silicon and other electronic materials in hopes of developing a more robust technology that can also integrate transistors and other electronic devices. In most versions of nanopore technology, researchers use tiny transistors to control a current passing across the pore. As the four different DNA bases go through, they perturb that electric signal in different ways, causing the voltage to spike upward or drop in a way that identifies the nucleotide passing through.

At the meeting, for example, chemist Gregory Timp of the University of Illinois, Urbana-Champaign, reported that his team has generated electrical readings of DNA moving through nanopores. Unfortunately, the DNA wriggled back and forth so much that the researchers had trouble teasing out the sequence of bases in the chain. But Timp says he and his colleagues are finishing a second-generation device that uses electric fields to keep the movement of the DNA under control. If it works, the technology promises to read long stretches of DNA without the need for expensive optical detectors.

**Searching for Cheaper Genome Sequencers**

| Company | Format | Read Length (bases) | Expected Throughput MB (million bases)/day |
|---|---|---|---|
| 454 Life Sciences | Parallel bead array | 100 | 96 |
| Agencourt Bioscience | Sequencing by ligation | 50 | 200 |
| Applied Biosystems | Capillary electrophoresis | 1000 | 3–4 |
| Microchip Biotechnologies | Parallel bead array | 850-1000 | 7 |
| NimbleGen Systems | Map and survey microarray | 30 | 100 |
| Solexa | Parallel microchip | 35 | 500 |
| LI-COR | Electronic microchip | 20,000 | 14,000 |
| Network Biosystems | Biochip | 800+ | 5 |
| VisiGen Biotechnologies | Single molecule array | NA | 1000 |

**Generation next.** Companies racing for the $1000 genome sequence strive simultaneously for low cost, high accuracy, the ability to read long stretches of DNA, and high throughput.

### "We have to worry now"

No matter which technology or technologies make it to market, the scientific consequences of lower sequencing costs are bound to be enormous. "I think it's going to have a profound impact on biology," says Yale University molecular biologist Michael Snyder.

Some early progress is already on display. At the Florida meeting, for example, 454's Egholm reported that he and his colleagues used their technology to identify as many as four genetic variants of HIV in single blood samples, in contrast to today's technology, which identifies just the dominant strain. The technique, Egholm says, could eventually help doctors see the rise of drug-resistant HIV strains in patients at the earliest stages. In another study, they quickly analyzed the sequence of non–small cell lung cancer cells and identified the specific mutations that give rise to drug resistance.

In similar studies, Thomas Albert and colleagues at NimbleGen Systems, a biotechnology firm in Madison, Wisconsin, used their version of sequencing-by-synthesis technology to identify the mutations in *Helicobacter pylori*—the microbe responsible for ulcers—that cause resistance to a drug known as metronidazole, as well as the mutations in the tuberculosis-causing bacterium that trigger resistance to a new TB drug. The power of such studies is "unbelievable," Snyder says, because they hold out the hope of enabling doctors to tailor medicines to battle diseases most effectively. Some personalized-treatment strategies are already in use: Herceptin, for example, is targeted to patients with a specific genetic form of breast cancer. But cheap sequencing should make them far more widespread, Church says.

Basic researchers are looking at the early benefits of cheap sequencing as well. At the meeting, for example, Snyder talked about his team's use of gene chips to map the sites where transcription factors—proteins that control when genes are turned on—bind to the genome. The technology is effective, but gene chips are expensive. So Snyder is turning to cheap sequencing technology to rapidly sequence the millions of DNA fragments needed to identify transcription factors. Church says he is also using cheap sequencing techniques to propel his group's synthetic-biology efforts to create an extensive tool kit of microbial "parts" that can be mixed and matched to enable microbes to perform new functions.

Like most new technologies, ultracheap sequencing casts shadows as well. For starters, Church says, it's hard to imagine what privacy will mean once anyone with a sample of your DNA can determine who you are, your heritage, and what diseases you're likely to inherit. Celebrities and politicians may soon face a world hungry to scrutinize their genes. Among ordinary people, many analysts worry that insurers and employers will use genetic information to screen out those at high risk for disease. Finally, the same sequencing technology that could potentially help create beneficial new microbes, such as ones tailored to turn out large amounts of hydrogen gas to power fuel cells, could also make it easier to create new bioterrorist pathogens.

"We have to worry about these issues now, because we will be sequencing with very high throughput in 10 years," Timp says. Schloss notes that NHGRI has long supported research on ethical, legal, and social concerns. However, he adds, "it's very hard to do it in the abstract." With technology advancing at a rapid clip, neither the benefits nor the concerns of ultracheap sequencing are likely to remain abstract for long.

**–ROBERT F. SERVICE**


Published by AAAS