Human Genome: Questions of Cost

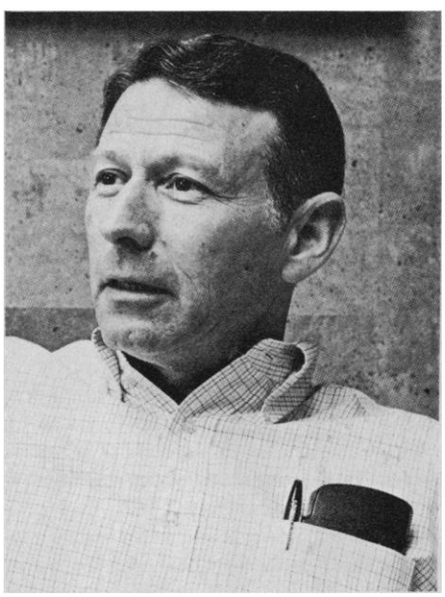
At a recent workshop scientists tried to tally the costs of the genome project; they came up with a hefty ballpark figure in the low billions

SINCE a massive effort to map and sequence the human genome was proposed, it has been the cost of the project—initially estimated at $3 billion—that has engendered the most concern, raising fears about "big science" and the effect a project of this magnitude might have on other areas of biological research. Since then, cost estimates have been steadily dropping. The latest, by a Department of Energy (DOE) subcommittee, put the total bill at $1 billion, which, though still big by biological standards, is greeted with more equanimity.

Last month, a couple of years after the project was first proposed, some of the key players met in Washington to take the latest tally. This workshop, which was organized by the Office of Technology Assessment (OTA) in response to increasing congressional interest, did not provide the definitive answer to what the project might ultimately cost. But it did offer glimpses into how thinking about the project has evolved, how it might be tackled, and what the scientific obstacles are. It also highlighted areas of disagreement, of which there are several.

The panelists compiled a menu of sorts for Congress on what components the project will likely include and how much each might cost: the genetic map, a guide to the loci of known genes; the physical map, a complete, ordered set of DNA fragments; and the sequence itself—the exact chemical order of the 3 billion nucleotide base pairs that make up the genome. And though estimates for the components varied widely, they can be combined for a ballpark figure of $1 billion to several billion dollars.

Work is well under way on the first component, the genetic map, which would consist of genetic markers—signposts to where genes are located—evenly spaced throughout the entire genome. A thousand or more genetic markers have been found over the past 5 years, and they have made possible the recent localization of genes for cystic fibrosis, Duchenne's muscular dystrophy, manic-depressive illness, and others. Blanketing the chromosome with markers, equally spaced, would allow an unknown gene to be located to within a relatively small stretch of DNA, providing a powerful tool in the search for disease-related genes. In addition, the genetic map is a necessary first step in the more difficult task of constructing a physical map of the genome. But because of its immense clinical value, work on the genetic map is proceeding independently of the larger genome effort.



Paul Berg. *"I think now everyone agrees this is a worthwhile project."*

How useful the genetic map ultimately is will depend on its resolution, or the distance between these evenly spaced markers, which is measured in centimorgans. A centimorgan is a indication of genetic distance (how often two markers are separated from each other during mating), but it roughly corresponds to a physical distance of about 1 million base pairs.

A coarse genetic map, with markers spaced about 10 or 15 centimorgans apart, is essentially in hand, according to Helen Donis-Keller of Collaborative Research. An intense effort is now under way to find more markers to create a map with greater resolution, perhaps a 5-centimorgan or a 1-centimorgan map. In the United States, most of this work is being done by Donis-Keller's group at Collaborative and Ray White's group at the University of Utah. Donis-Keller predicts that a 5-centimorgan map will be completed within 1 to 3 years for a cost of $11 million.

There was sharp disagreement, however, on whether a 5-centimorgan map offers enough resolution to be useful in completing the physical map of the genome or whether a 1-centimorgan map is needed. Donis-Keller maintains that a 5-centimorgan map is sufficient and that a 1-centimorgan map is simply not worth the cost and effort needed to obtain it. At least 3500 markers would be needed for a 1-centimorgan map, as opposed to 700 for a 5-centimorgan map, and it would be "orders of magnitude more expensive, tremendously expensive," she says.

James Watson of Cold Spring Harbor Laboratory disagrees. The National Academy of Sciences panel he sits on has concluded that a 1-centimorgan map is needed, he says, as has Maynard Olson of Washington University, who is developing new cloning techniques to use in physical mapping. The Howard Hughes Medical Institute (HHMI), which funds Ray White, has also set a 1-centimorgan map as its goal.

A 1-centimorgan map might in fact be "a lot less awesome" than Donis-Keller and others think, says Leroy Hood of Caltech, because it will be completed with automated technologies. "If you talk in terms of today's technology, it will give a false impression of cost that is way too high."

At this stage cost estimates for the 1-centimorgan map ranged from Donis-Keller's hundreds of millions of dollars to Harvard biologist Walter Gilbert's estimate of $50 million as an outside figure, and $25 million "if you stretch." OTA took the middle ground and settled on $50 to $100 million for a 1-centimorgan map, compared with $11 million for a 5-centimorgan map.

The keystone of the human genome project is the physical map, a set of overlapping DNA fragments, aligned in order as they would appear along the chromosome. Once the physical map is complete, researchers will be able to pinpoint a gene of interest to a specific fragment, perhaps 40 kilobases in

length, then pull out that fragment and sequence it, with vast implications for the treatment of genetically mediated diseases and for fundamental research. The ultimate physical map is the sequence itself.

Physical mapping of selected chromosomes is already under way in several laboratories, but obtaining a complete map of the human genome will be difficult. The most likely approach involves breaking the chromosomes into fragments, cloning those fragments, and then using overlapping sequences to determine where each belongs along the chromosome. Problems arise, however, because some stretches of DNA appear to be "unclonable." In addition, ordering the hundreds of thousands of clones that will be developed will be a herculean task, even with the help of the genetic map.

According to Anthony V. Carrano of Lawrence Livermore National Laboratory, the actual production and ordering of the clones will cost about $50 million and take about 150 person-years: $10 million to produce them and $30 million to $50 million to order them. The first, and simplest, part of the task—producing the clones—should be complete within 3 years.

Participants were clearly astonished to hear that the most expensive part of the endeavor might be maintaining the 600,000 or so clones in a repository. Robert E. Stevenson, who directs the American Type Culture Collection and is thus in a position to know, estimated the storage cost, for 30 years, at $250 million, which would raise the entire cost of the physical map to $300 million. At that rate, it would be cheaper to generate the clones again than to store them, Gilbert pointed out. Without disputing Stevenson's figures, the panelists generally agreed that as a rule of thumb, the costs of storing the map should not exceed the cost of generating it and thus settled on $60 million for storage, for a total of $110 million for the physical map.

The cost of the physical map would drop dramatically, to perhaps a few million dollars, if Olson's new technique for cloning large DNA fragments can be applied. Using conventional approaches, the largest fragment that can be cloned is about 40 kilobases in length, so perhaps 600,000 clones will be needed to span the entire genome. But with this new approach, which involves cloning in yeast artificial chromosomes (YACs), Olson and his colleagues have successfully cloned fragments 400 kilobases and larger. And the larger the fragments, the fewer there will be to order. Opinion was divided, however, on whether YAC clones can replace the conventional cosmid clones for mapping and sequencing.

By far the most expensive component of the project is the actual sequencing. Although everyone at the meeting agreed that the current sequencing cost of $1 to $2 per base will decline as technology advances, just how much is an open question. As Hood reported, Japanese researchers have automated major parts of the process, dropping the cost to about 17 cents per base. The automated sequencer Hood and his colleagues developed, and which is being designed for commercial use by Applied Biosystems, Inc., now does the job for 6 to 8 cents per base. Hood predicts the cost with this machine will drop to 1 or 2 cents per base within the next 6 months. The Japanese researchers also predict a cost of 2 cents per base with their equipment.

Given that there are 3 billion base pairs, a few cents make a tremendous difference in the overall cost. For the figures Hood mentioned, the cost for sequencing a single strand of DNA (3 billion bases) ranges from a low of $60 million (for 2 cents per base) to $500 million (for 17 cents per base) to $3 billion ($1 per base). What is often ignored in these estimates, however, is the need for multiple sequencing runs to check accuracy. It has been commonly assumed that sequencing both strands of DNA, thus doubling the price, would suffice for an accuracy check. But Paul Berg of Stanford University pointed out that with today's technology (and relatively high error rate), the sequence might have to be redone ten times to ensure accuracy.

"But should we be worried about $600 million, over probably 15 years?"

Hood thinks they will be able to get by with three runs, provided that good software is available, the error rate is reduced to 0.5% or 0.1% , and a good physical map is complete. If so, the final cost might range anywhere from $180 million (2 cents per base) to $1.5 billion (17 cents per base), to perhaps more, depending on what the cost per base ultimately is.

Gilbert, who has launched a company to sequence the genome, thinks that the entire sequence can be done and checked for $300 million, using a different way of preparing the DNA for sequencing, known as nested deletions, and existing technology. Others are clearly skeptical. "Even with nested deletions, I think we would be irresponsible to claim we will ever get it down to less than 10 cents a base pair," said Watson, who believes a more realistic figure for the finished product is $600 million. "But should we be worried about $600 million, over probably 15 years? That's $45 million a year, 1% of the NIH [National Institutes of Health] budget. It's not a colossal relative amount of money."

But that figure is just for working out the sequence—it does not include the costs of analyzing it or developing the technologies necessary to obtain it. Costs for data handling and analysis alone were estimated at an additional 15% of the total budget. "Fifteen percent of what?" asked Berg, who, like many of the others, was having difficulty keeping track of whether the costs were for a 5-centimorgan or a 1-centimorgan map, for sequencing one strand of DNA or two, with or without additional error checking.

After some wrangling, the panelists also agreed to add a surcharge of 25% for developing automated technologies for mapping, sequencing, and data analysis. Some panelists think that a separate technology development is not necessary at all. Carrano and Hood, on the other hand, thought 50 to 60% could be profitably used. And then there are the costs for quality control, administration of the project, and training new scientists.

When all these costs are added in—along with such unglamorous but necessary factors as capital costs, overhead, and salaries—it looks as if the original estimate of $3 billion, which so polarized the biological community, might not be that far off. What is different, however, is how biologists view the project, as Berg described. There has been an enormous change in thinking about the project," he said, recalling his "abortive attempt" a year ago at Cold Spring Harbor to lead a discussion on the topic. "We could hardly get to the science because of the ominous views people had about the project. I think now everyone agrees this is a worthwhile project, and we can get on to talking about how one might go about it in the most cost-effective and scientifically effective way."

The other major question yet to be resolved, participants agreed, is who will oversee the project to ensure that these funds are well spent—that it is run as "an A effort, not a B effort," as Watson described. Momentum seems to be gathering behind a scientific advisory board to oversee the project, whether it is run by DOE or NIH. Watson thinks the project may require the equivalent of a NASA administrator. "There are going to be bad years, the program is going to be under economic attack, and there's got to be someone there who sees all the components and fights for it," he said. "I couldn't think of a job I'd like less." ■ **LESLIE ROBERTS**