16. M. B. Bracken, *Soc. Psychiatry* **13**, 135 (1978).
17. J. D. Osofsky, H. J. Osofsky, R. Rajan, D. Spitz, *Mt. Sinai J. Med.* **42**, 456 (1975).
18. J. B. Rooks and W. Cates, Jr., *Fam. Plann. Perspect.* **9**, 276 (1977); N. B. Kaltreider, S. Goldsmith, A. Margolis, *Am. J. Obstet. Gynecol.* **135**, 235 (1979).
19. M. Bracken and S. Kasl, *ibid.* **121**, 1008 (1975).
20. M. B. Bracken, L. V. Klerman, M. Bracken, *ibid.* **130**, 251 (1978).
21. M. Eisen and G. L. Zellman, *J. Gen. Psychol.* **145**, 231 (1984).
22. S. Cohen and T. A. Wills, *Psychol. Bull.* **98**, 310 (1985); R. C. Kessler and J. D. McLeod, *Social Support and Health*, S. Cohen and S. L. Syme, Eds. (Academic Press, Orlando, FL, 1985), pp. 219–240.
23. M. B. Bracken, M. Hachamovitch, G. Grossman, *J. Nerv. Ment. Dis.* **158**, 154 (1974).
24. D. T. Moseley *et al.*, *J. Clin. Psychol.* **37**, 276 (1981).
25. J. M. Robbins, *Soc. Probl.* **31**, 334 (1984).
26. M. F. Scheier and C. S. Carver, *J. Pers.* **55**, 169 (1987); A. Bandura, *Psychol. Rev.* **84**, 191 (1977).
27. S. K. Henshaw, J. D. Forrest, J. Van Vort, *Fam. Plann. Perspect.* **19**, 63 (1987).
28. N. E. Adler, *J. Appl. Soc. Psychol.* **6**, 240 (1976); E. W. Freeman, *Am. J. Orthopsychiatry* **47**, 503 (1977).
29. E. C. Payne, A. R. Kravitz, M. T. Notman, J. V. Anderson, *Arch. Gen. Psychiatry* **33**, 725 (1976); E. M. Belsey, H. S. Greer, S. Lal, S. C. Lewis, R. W. Beard, *Soc. Sci. Med.* **11**, 71 (1977).
30. C. E. Koop, letter to R. W. Reagan, 9 January 1989.
31. A. C. Speckhard, *The Psycho-Social Aspects of Stress Following Abortion* (Sheed and Ward, Kansas City, MO, 1987).
32. C. B. Wortman and R. C. Silver, *J. Consult. Clin. Psychol.* **57**, 349 (1989).
33. One may also find more adverse consequences for the children born as a result of unwanted pregnancy [H. P. David, Z. Dytrych, Z. Matejcek, V. Schuller, *Born Unwanted: Developmental Effects of Denied Abortion* (Springer, New York, 1988)].
34. Committee on Government Operations, House of Representatives, *The Federal Role in Determining the Medical and Psychological Impact of Abortions on Women*, 101st Cong., 2d sess., 11 December 1989, House Report 101-392, p. 14.
35. This article is based on of a review conducted by a panel convened by the American Psychological Association. The authors were members of the panel. We thank J. Gentry and B. Wilcox for contributions to the manuscript and G. Markman and A. Schlagel for manuscript preparation.

The Human Genome Project: Past, Present, and Future

JAMES D. WATSON

This article presents a short discussion of the development of the human genome program in the United States, a summary of the current status of the organization and administration of the National Institutes of Health component of the program, and some prospects for the future directions of the program and the applications of genome information.

THE UNITED STATES HAS NOW SET AS A NATIONAL OBJECTIVE the mapping and sequencing of the human genome. Several other countries have joined in this exciting initiative, and we expect a number more to do so. Similar to the 1961 decision made by President John F. Kennedy to send a man to the moon, the United States has committed itself to a highly visible and important goal. Although the final monies required to determine the human DNA sequence of some 3 billion base pairs (bp) will be an order of magnitude smaller than the monies needed to let men explore the moon, the implications of the Human Genome Project for human life are likely to be far greater. A more important set of instruction books will never be found by human beings. When finally interpreted, the genetic messages encoded within our DNA molecules will provide the ultimate answers to the chemical underpinnings of human existence. They will not only help us understand how we function as healthy human beings, but will also explain, at the chemical level, the role of genetic factors in a multitude of diseases, such as cancer, Alzheimer's disease, and schizophrenia, that diminish the individual lives of so many millions of people.

The author is the director of the National Center for Human Genome Research, National Institutes of Health, Bethesda, MD 20892, and the director of the Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.

The possibility of knowing our complete set of genetic instructions seemed an undreamable scientific objective in 1953 when Francis Crick and I found the double helical structure of DNA (*1*). Then there existed no way to sequence even very short DNA molecules, much less any possibility of obtaining the totality of human DNA as a collection of discrete pieces for eventual chemical analysis. Only years later, with the 1973 birth of the recombinant DNA revolution, was it possible to think of routinely isolating individual genes (*2*). This breakthrough provided the incentive for Allan Maxam and Walter Gilbert (*3*) and Fred Sanger *et al.* (*4*) to develop their powerful sequencing techniques that now make it almost routine to establish in a single experiment 300 to 500 bp of DNA sequence.

The first complete DNA sequences to be established by these procedures were those of the smaller DNA viruses, such as the simian virus 40 (*5*) and the phage ϕX174 (*6*), each of which contains some 5000 bp. These sequences became known by 1977, and within the next 5 years the tenfold larger DNAs of the bacteriophages T7 (*7*) and lambda (*8*) were determined. Today, the more than 100,000 bp DNAs of several plant chloroplasts (*9*) and of the herpesvirus Epstein-Barr virus (*10*) have been established. The largest DNA now sequenced is that of cytomegalovirus (also a herpesvirus), which contains almost a quarter of a million base pairs (*11*).

Simultaneously, the sequences of a large number of individual genes have been worked out, with the total number of base pairs exceeding 37 million (*12*). The most completely known organism, in this regard, is the intensively studied bacterium *Escherichia coli*, with more than 800,000 bp of its 4.8×10^6 bp genome already established (*12, 13*). There are a number of academic laboratories in the United States and Japan geared up to complete the *E. coli* sequence, and there are good reasons for believing that success will come within the next decade. Today, DNA sequencing usually costs between $3 and $5 per base pair (*14*); so, at most, $25 million would be required—a large, but not unthinkable sum when spent

over an extended period. Most likely, as sequencing efficiencies improve, the *E. coli* sequence will finally cost a sum less than half this amount. Conceivably, we could know the complete sequences of several bacteria without the creation of a major research program especially aimed at working out complete genomes.

However, a completely different picture holds for the human genome, which is almost 1000 times larger than that of *E. coli* and is distributed over 24 different chromosomes (22 autosomes, X and Y). Here the cottage industry approach involving small groups of individuals, each working at a different site, seems unlikely to succeed. The time required would more than exceed the lifetimes of those who would like to work in this way. To be truly exciting for both our best scientific minds and the average citizen, we must aim to complete the job over a 15-year period. The project should be completed by those who start it, not their scientific descendants. To accomplish this, we must, from the beginning, design game plans where economies of scale are sought and found. Even so, most people assume that we shall, in the end, have to create DNA sequencing facilities that are far larger than any existing today and that more closely resemble industrial production lines than conventional university research laboratories.

The first serious proposal to start sequencing the human genome occurred at a meeting held in May 1985 (*15*). Robert Sinsheimer, then chancellor of the University of California at Santa Cruz, brought together a small group of scientists with the hope that the project might be centered in the Santa Cruz environment. Renato Dulbecco independently sensed the challenge and, later that year, he spoke glowingly of the prospects for cancer research offered by the knowledge of the sequence of our own DNA (*16*). By this time, the Department of Energy (DOE), particularly Charles DeLisi, director of the Office of Health and Environmental Research, was seriously thinking about taking on the project. After a meeting in Santa Fe, New Mexico, in March 1986 (*17*), DeLisi and his colleagues at DOE proposed that certain of the national laboratories should be the center of the U.S., if not the worldwide, human genome effort.

Over the next several months, rumors that DOE would soon commence a large-scale human genome program began to spread through the general biological community. The pros and cons of such a DOE-led project were first discussed before a more general audience at the Cold Spring Harbor Laboratory in 1986. During the course of a symposium on the "Molecular Biology of *Homo sapiens*," the question was taken up in a special afternoon session (*18*). Although several more senior scientists, including Walter Gilbert, Paul Berg, and me, voiced the opinion that it was then time to start the project, much less enthusiasm, if not downright hostility, was voiced by many younger scientists. They feared that a megabillion-dollar project would of necessity divert money away from single investigator-initiated research grants and slow down the pace at which our country does high-quality biological and medical research. Also troubling to many was the thought that DOE had never been a major supporter of recombinant DNA–based research and possessed few senior administrators familiar with the world of genetics. Concern was expressed about the way the Human Genome Project would be managed within a DOE whose leaders were invariably physical scientists and where biology, as a consequence, occupied a lower position on its list of priorities. It seemed to me that the safe course would be for the National Institutes of Health (NIH) also to participate in the Human Genome Project, provided that new monies would be appropriated by Congress to fund it.

Soon the controversy reached the attention of the Board of Basic Biology, Commission on Life Sciences of the National Academy of Sciences. After a meeting in Woods Hole, Massachusetts, in August 1986, a decision was made to appoint a special National Research Council (NRC) committee to prepare a report as to what our nation should do next (*19*). Chaired by Bruce Alberts, who had written articles expressing skepticism about the effectiveness of big laboratories in biology (*20*), the 15-member committee represented a diverse collection of viewpoints, including those who had voiced strong opposition to the project. Its 14-month-long deliberations led to a unanimous report (*21*), which urged the United States to begin the Human Genome Project and to work cooperatively with other nations who wished to jointly pursue the common goal.

Soon after the NRC committee began its deliberation, it became apparent that within the meeting room the project itself was not really controversial—who could be against obtaining the much higher resolution molecular genetic and physical maps of human DNA that would be needed before the sequencing itself would begin? Such maps themselves would be invaluable tools for finding human disease genes. It was proposed that such mapping efforts, plus the development of improved technology for DNA sequencing, should dominate the first 5 years of the project. What had generated much of the initial opposition was fear that the project would be divorced from the main currents of biological research. These initial opponents were concerned that exclusive concern would be focused on the human DNA sequence, most of which might prove uninterpretable in the absence of comparable information about the genomes of much simpler, but more easily studied, model organisms, such as *E. coli*, the yeasts, the roundworm *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster*, and the mouse. There were also strong reservations about any project in which the ultimate control of resources lay in the hands of administrators, as opposed to control by the scientific community itself.

In urging a major commitment to the Human Genome Project, the committee emphasized the need for technological improvements that would lead to a five- to tenfold increase in the efficiency of current gene mapping, sequencing, and data analysis capabilities. Only when the true cost of sequencing falls considerably should extensive sequencing begin (*22*); the committee expected that it would be at least 5 years before this could be envisioned. Federal funding, the committee urged, should rise quickly to $200 million a year, with the project planned to be completed in approximately 15 years. The sequencing of the model genomes should go hand-in-hand with, if not slightly ahead of, that of the human genome. Knowledge of the structures of the genes of organisms such as bacteria and budding yeasts would facilitate the task of distinguishing the amino acid coding regions of the human genome from the much more prevalent noncoding components.

Parallel with the deliberations of the NRC panel, the Office of Technology Assessment (OTA) of the U.S. Congress was commissioned by the House Committee on Energy and Commerce to prepare a report (*23*). The congressional interests on which the report focused centered on (i) assessing the scientific and medical reasons for genome projects, (ii) the potential funding—at what level and through what mechanisms, (iii) how to coordinate activities involving several federal and private agencies, and (iv) how to strike a proper balance between the virtues of international scientific collaboration and the need to promote the U.S. competitive position in biotechnology. Unlike the NRC report, the OTA document did not offer specific recommendations; it viewed its purpose as informing Congress on the options for future action. Despite this aim for neutrality, reading the OTA report left the unmistakable message that some form of human genome program was bound to proceed and that Congress had a role in seeing that it start off in the right direction.

In its 1988 budget request, DOE asked for $15 million, and later received $12 million (*24*), to continue its human genome effort that had started the previous year. There was no formal request from NIH for genome studies, but during the spring 1987 House

appropriations hearing, NIH Director James Wyngaarden, in response to a question, expressed the opinion that $50 million would be needed for a meaningful program (*24*). Later, in early May, David Baltimore and I visited key members of the House and Senate Appropriations Committees on behalf of the Delegation for Basic Bio-Medical Research, of which I was the official spokesman. We emphasized the need for a multihundred-million-dollar increase in research monies for acquired immunodeficiency syndrome (AIDS) and indicated that a $30-million appropriation would let NIH start a serious human genome effort. In the summer, when the respective committees marked up and then reconciled their NIH budgets, $18 million for genome studies was added to the appropriation for the National Institutes of General Medical Sciences (NIGMS) (*25*), which by then had issued announcements requesting grant applications on genome studies.

The Department of Health and Human Services (DHHS) appropriation, with its NIH component, was signed into law by President Reagan in early December, 2 months into the 1988 fiscal year. The $17.2 million finally earmarked for genome studies led Wyngaarden to convene an 18-member Ad Hoc Advisory Committee on Complex Genomes to propose priorities for the NIH genome program. This committee, which met in Reston, Virginia, in February 1988, was chaired by David Baltimore; it broadly backed the main features of the NRC report (*21*). In its final recommendations (*26*), it strongly supported Wyngaarden's proposal to establish within NIH an Office of Human Genome Research to be headed by a new associate director. Along with the office, there was to be a chartered advisory committee, as had been recommended by the NRC committee, to work with the head of the genome effort to keep the program on target. Emphasis was also given to the need during the early phases of the program for major training efforts that would develop scientists who were skilled in the use of the new technologies needed to generate, then assemble, and later interpret, the massive amounts of new information that would flow out of the genome program. At the Reston meeting, I strongly urged that the associate director position be filled by an active scientist, as opposed to an administrator. I argued that one person had to be visibly in charge and that only a prominent scientist would simultaneously reassure Congress, the general public, and the scientific community that scientific reasoning, not the pork barrel, would be the dominant theme in allocating the soon-to-be-large genome monies. I did not realize that I could be perceived as arguing for my own subsequent appointment. For many years, my most visible role had been that of an administrator dominated by the fund-raising activities needed to keep the Cold Spring Harbor Laboratory at the forefront of DNA-based science. Whether I was still a real scientist was not at all clear.

I felt uneasy when I heard rumors that I was to be offered the position of associate director for Human Genome Research at NIH. My job at Cold Spring Harbor was already more than full time. If I ran the genome effort, I would hold two demanding positions simultaneously. Yet, if I turned down the job, it was not clear that any prominent scientist still active in the lab would take on the task. So, when in early May, Wyngaarden asked me to come to Bethesda to talk about working for NIH, I knew I would accept. By then I also realized that only once would I have the opportunity to let my scientific life encompass the path from double helix to the 3 billion steps of the human genome.

Officially, I started working for NIH in early October 1988 (*27*) and began a commuting life by which I would try to spend the beginning of each week at NIH. My major task was to help formulate a workable strategy for establishing the human genome sequence. Shortly thereafter, Secretary Otis Bowen of DHHS appointed the Program Advisory Committee on the Human Genome. Twelve members were chosen, with Norton Zinder as the chairman (*28*). The composition of the committee reflected a broad range of expertise, with strong representation from the world of pure science that had initially reacted so negatively to the Human Genome Project. This presence on the advisory committee was a strong message to the world of biology that NIH would not bring forth a narrowly construed effort. Furthermore, by having three members from industry, we believed we could ensure that our nation's competitive position in biotechnology would not be neglected.

Beginning with my opening press conference at NIH (*24*), and later through other meetings with the press, I made clear my concern for the ethical and social implications raised by an ever-increasing knowledge of human genes and of the genetic diseases that result from variations in our genetic messages. On the one hand, this knowledge undoubtedly will lead to a much deeper understanding of many of the worst diseases that plague human existence. Thus, there are strong ethical reasons to obtain this genetic knowledge as fast as possible and with all our might. On the other hand, the knowledge that some of us as individuals have inherited disease-causing genes is certain to bring unwanted grief unless appropriate therapies are developed. So it is imperative that we begin to educate our nation's people on the genetic options that they as individuals may have to choose among.

I believed we should put money behind these convictions and suggested that, at the start, at least 3% of the earmarked genome funds should go to support the ethical and social implications area. In doing so, we must be aware of the terrible misuses of the incomplete knowledge of human genetics that went under the name of eugenics during the first part of this century (*29*). There exists real fear among many individuals that genetic reasons will again be used to make the lives of the underprivileged even more disadvantaged (*30*). We must work to ensure that society learns to use the information only in beneficial ways and, if necessary, pass laws at both the federal and state levels to prevent invasions of privacy of an individual's genetic background by either employers, insurers, or government agencies and to prevent discrimination on genetic grounds. If we fail to act now, we might witness unwanted and unnecessary abuses that eventually will create a strong popular backlash against the human genetics community. We have only to look at how the Nazis used leading members of the German human genetics and psychiatry communities to justify their genocide programs, first against the mentally ill and then the Jews and the Gypsies (*31*). We need no more vivid reminders that science in the wrong hands can do incalculable harm.

At its inception, the Office of Human Genome Research had only advisory and staff functions. In fiscal years 1988 and 1989, the authority for the distribution of NIH genome funds belonged to NIGMS. Wyngaarden had told me when I first came to work with him that he wanted to change the Office of Human Genome Research into a "Center," (*32*) which would have the authority to make grants, as soon as we had sufficient funds to justify an independent program. The inclusion in the fiscal 1990 budget proposal of $100 million for genome studies (*25, 33*) gave Wyngaarden the go-ahead signal to request the DHHS secretary to effect that change. This request was approved by Secretary Louis Sullivan, and in October 1989 we became the National Center for Human Genome Research (NCHGR) (*34*).

For the 1990 fiscal year we will have a much smaller budget than initially proposed by the White House (*35*). The NCHGR program was appropriated just under $60 million, while the DOE program will be funded at a $28-million level. Because of commitments already made for future years, we will have a tight budget for this year, necessitating many difficult funding decisions over the coming months. We estimate that our award rate for 1990 will be only

slightly higher than the NIH average.

Initially, there was much uncertainty over how the NIH and DOE programs would be coordinated. In the fall of 1988, the two agencies signed a Memorandum of Understanding (*36*). Among other provisions, this agreement created a joint NIH-DOE subcommittee with its members drawn from the advisory bodies of each agency. The most important initial task of the joint subcommittee was to draw up the National Genome Plan that Congress had requested of NIH by spring 1990, when it considers our fiscal 1991 appropriations. Our first planning meeting was held at Cold Spring Harbor in August 1989, and a smaller, follow-up meeting was held in October in San Diego. At these meetings, for the first time, the question before us was no longer whether to start a targeted genome program, but how best to do so. A draft of the report was approved by the parent advisory committees in January 1990. The report has now been submitted to Congress and is available to the public (*14*).

The report presents quantitative goals for the next 5-year period for the construction of genetic linkage maps, physical maps, and DNA sequence technology development. It also presents goals in the areas of informatics, ethics, training, technology transfer, as well as strategies for national and international cooperation. In the area of linkage mapping, the report calls for the expansion of the human linkage map to a resolution of 2 to 5 centimorgans (cM), meaning a map with an average spacing of 2 cM between markers and containing no gaps greater than 5 cM. As the NRC report had spoken of a 1-cM map, questions have been raised recently as to whether the linkage mapping goals of the project are already being retrenched (*37*). I believe not.

All along, the linkage map has been considered to have a dual role in the Human Genome Project. The first is as a tool for the identification and location of genes that can only be recognized by their phenotypic effects, such as those that are associated with certain diseases. For this purpose, a 3- to 5-cM map is generally considered to be of sufficient resolution. The second is as a tool to help assemble the physical map. Two years ago, at the time of the NRC report, the limits to the kind of physical mapping that involved construction of contigs (overlapping units of cloned DNA) seemed to be in the range of 1,000,000 bp [1 megabase (Mb)]. Thus, a 1-cM map was considered to be important to help order physically mapped units of 1 Mb. Recently, however, physical mapping techniques have been shown to be capable of generating continuous regions of at least 2 Mb. Therefore, a 2-cM linkage map may be adequate to play this role. Furthermore, it is not even obvious that the process laid out by the NRC committee just 2 years ago is how the high-resolution linkage maps will be achieved, as it may be more efficient to use a fine structure physical map to construct a detailed linkage map.

The "change" in policy was thus based on an assessment of recent technical improvements in mapping. In the same way, we intend to frequently reassess and update the goals of the program in the light of technological improvements in order to keep the cost of the effort as low as possible. This will not be the last time that a goal is reassessed in light of scientific advances. It would be a mistake to consider each redefinition of the goals as a failure of the program.

The 5-year goals for the physical mapping effort are to construct maps with STS markers (*38, 39*) spaced approximately 100,000 bp (100 kb) apart and to assemble 2-Mb contigs for large parts of the human genome. There are a variety of techniques and strategies that are currently being used for physical mapping. The so-called "bottom up" approach, in which long-range maps are assembled by identifying overlapping regions in randomly generated phage or cosmid clones, has worked rather successfully in generating partial maps for the genomes of organisms such as *E. coli* (*40*), *Saccharomyces cerevisiae* (*41*), and *C. elegans* (*42*). However, even in these cases, this approach has not been able to generate complete maps, and other means have been used to extend the contigs further (*43*). A number of groups, including those at the Los Alamos and Livermore National Laboratories and at the Imperial Cancer Research Fund in London, are exploring the usefulness of "bottom-up" strategies for building physical maps of human chromosomes.

Other groups, such as that of Maynard Olson and David Schlessinger at Washington University in St. Louis, are testing "top-down" approaches, in which a genome or chromosome is gradually subdivided into regions of smaller and smaller size, while determining the order of regions at each step along the way. Over the course of the next couple of years, I expect that the practical strengths and weaknesses of the various strategies, as well as the usefulness of particular techniques, will emerge and certain ones will be adopted by the community as the most efficacious.

In addition to working out the scientific strategy and methodology, in order to meet the goals set out in the national plan, we must, over the next 5 years, find groups of investigators large enough to oversee the detailed physical mapping and, over the subsequent decade, the sequencing of individual chromosomes. In my estimation, groups of about ten individuals are probably the appropriate size for a cost-effective analysis of bacterial and yeast chromosomes. For human chromosomes, which are anywhere from 10 to 50 times larger, it may be necessary to put together groups of up to 50 trained personnel. When it comes to sequencing, the output of finished sequences by such a group would need to be approximately a phage lambda equivalent of DNA (about 50 kb) per working day. In addition, the informational aspects of the genome program will, with time, become more important as an ever-increasing set of new DNA sequences begins to be compared with the sequences previously obtained. Although the key talents needed to start a successful genome program are likely to be those of the recombinant DNA chemist, our effectiveness at the end of the program may depend more on the computer skills that are applied.

The idea that the various human chromosomes will be divided among various laboratories is far from today's conventional wisdom. Arguing against this approach is the fact that most chromosomes are already being genetically mapped and several are being physically mapped in a number of high-quality laboratories. Deciding which one of these groups should be awarded the funds to complete the maps of their respective chromosomes at first appears an impossible political task. But closer inspection reveals that many of these gene mappers are primarily interested in locating specific disease genes. Once the gene of interest is mapped, they will want to go on to cloning and studying it and are likely to discontinue mapping per se. To my knowledge, the number of investigators wanting to make complete, high-resolution physical maps of a specific chromosome is less than ten. Extensive, multiple overall mapping efforts currently only exist for chromosome 21, the smallest human chromosome, which contains, among others, a gene that leads to increased susceptibility to Alzheimer's disease (*44*). Here we may witness a truly competitive race to clone the overlapping sets of DNA fragments that are needed to commence sequencing. To date, serious efforts have been started to make complete sets of overlapping DNA fragments for less than half the human chromosomes. Our real problem may be persuading capable teams to focus on those chromosomes that still have no champions, in addition to deciding among alternative proposals for total mapping and sequencing.

Will we have the money but not the talented scientists to bring the Human Genome Project home to completion within an acceptable time period? I think not, but in so arguing I want to focus on the chief motivation that attracts talented scientists to their goals. It is seldom fame or financial benefit. To be sure, when one makes a

great discovery, it is frequently rewarded by a major scientific prize and the prospect of a much better academic or industrial position. But the more important reward is satisfying one's curiosity about how nature operates, and for biologists this means a deeper understanding of the nature of living organisms.

The working out of a bacterial genome will let us know for the first time the total set of proteins needed for a single cell to grow and multiply. As soon as we have the *E. coli* DNA sequence, we will be able to determine the amino acid sequences of all those proteins that, for example, control its gene expression or function as channels through which ions and signal molecules move. A total understanding of *E. coli*, of course, will not fall out immediately from the possession of its instruction book, and hundreds of years are likely to pass before *E. coli* poses no further scientific challenges. But the mere statement that we will one day know completely how *E. coli* functions is an extraordinary scientific assertion.

Acquisition of the DNA sequences of multicellular organisms like *C. elegans* (100 Mb) and *D. melanogaster* (150 Mb) will be equally important scientific landmarks. Their much more complex genomes provide the instructions for the extraordinarily complex set of events that allow fertilized eggs to develop into functional adults. Until a decade ago, how multicellular organisms develop was virtually a black box at the molecular level. Then a number of molecular embryologists began to clone the regulatory genes that control the passage from one developmental stage to another. By now, a number of key steps in the development of genetically well-characterized organisms are understood at the molecular level. But if we are to integrate and understand all the events that lead, for example, to the differentiation of a nervous system, we have to work from the whole set of genetic instructions. So both the *C. elegans* and *Drosophila* worlds are starting to make plans for working out their respective DNA messages. The scientific group farthest along is that working with *C. elegans* (*42, 43*), where virtually all of the genome is available as cloned sets of overlapping DNA fragments.

I hope that the final sequencing effort in both of these cases will be shared between laboratories in Europe and the United States, with the final costs for these programs to be no more than $75 million each. There also are good reasons for believing that a joint U.S. and European effort will come together to work out the sequence of *Arabidopsis thaliana*, the mustard-type plant, with a genome of only 100 Mb, which increasingly is serving as the model organism for the plant molecular biology world (the National Science Foundation will probably play the lead role in organizing the U.S. efforts). Inherent scientific interest in the smaller model organisms could prove to be a major ingredient in attracting the appropriate high-level scientific talents to develop the production-line sequencing capabilities that we will need to tackle human chromosomes.

So far, the United States, the United Kingdom, the U.S.S.R., Italy, and Japan have announced definite human genome programs; France, the European Economic Community, Australia, and possibly Canada may also join. Whether the Federal Republic of Germany will mount an effort is problematic because of the negative connotations that human genetics research still has to many Germans. How to ensure that we, as nations, work together instead of indulging in costly competitive races for the same chromosomal objectives, is not yet settled. Although a number of prominent molecular biologists and human geneticists have formed the Human Genome Organization (HUGO) (*45*), it is not yet a free-standing organization capable of taking the steps that will make it a real, as opposed to paper, entity.

I hope that HUGO is successful, as it could greatly facilitate the free and open exchange of data that we would all like to have as an outcome of the human genome project. The alternative, of knowing the sequences of only half the human chromosomes, for example, without having access to the other half would be unbearably frustrating. Optimally, soon after new sequences are established, they will be added to a database that is accessible worldwide. Achieving this goal, however, will require great skill and imagination as there are problems of personal, financial, and national interests that have to be solved. For instance, on the one hand, laboratories generating large amounts of sequence will naturally, before passing them on to others, want to work out the genes located within them and find clues for where they function or how they are expressed. It would be naïve to expect that any extensive human sequence data will be released by a sequencing group until it has a reasonable time to explore its implications. However, making the sequences widely available as rapidly as practical is the only way to ensure that their full value will be realized and is the only acceptable way to handle information produced at public expense.

Clearly, it will be easier for a laboratory to release its own sequences if they can be exchanged for others of equal size. Early sharing of the human DNA database is much more likely to occur if large-scale mapping and sequencing efforts are undertaken by all those major industrial nations that will want to use this data. It is too early to ask what we should do if we identify one or more countries that have the economic clout to join in the effort, but that apparently do not intend to, hoping instead to take advantage of the information once it becomes publicly available. I do not like to even contemplate such a possibility, since Congress and the public are likely to respond by wanting to move us toward a more nationalistic approach to science. This alternative is counter to the traditions that have allowed me to admire and enjoy the scientific life. The nations of the world must see that the human genome belongs to the world's people, as opposed to its nations.

REFERENCES AND NOTES

1. J. D. Watson and F. H. C. Crick, *Nature* **171**, 737 (1953).
2. S. N. Cohen, *Sci. Am.* **233**, 24 (July 1975).
3. A. M. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
4. F. Sanger, S. Nicklen, A. R. Coulson, *ibid.*, p. 5463.
5. V. B. Reddy *et al.*, *Science* **200**, 494 (1978); W. Fiers *et al.*, *Nature* **273**, 113 (1978).
6. F. Sanger *et al.*, *J. Mol. Biol.* **125**, 225 (1978).
7. J. J. Dunn and F. W. Studier, *ibid.* **166**, 477 (1983).
8. F. Sanger, A. R. Coulson, G. F. Hong, D. F. Hill, G. B. Petersen, *ibid.* **162**, 729 (1982).
9. K. Ohyama *et al.*, *Nature* **322**, 572 (1986); K. Shinozaki *et al.*, *EMBO J.* **5**, 2043 (1986); J. Hiratsuka *et al.*, *Mol. Gen. Genet.* **217**, 185 (1989).
10. R. Baer *et al.*, *Nature* **310**, 207 (1984).
11. M. S. Chee *et al.*, *Curr. Top. Microbiol. Immunol.* **154**, 129 (1990); European Molecular Biology Data Library, accession number X17403.
12. GenBank, Release 62 (15 December 1989).
13. C. Burks *et al.*, *Methods Enzymol.* **183**, 1 (1990).
14. Department of Health and Human Services, Department of Energy, *A Five-Year Plan for the Human Genome Project* (DOE, Washington, DC, in press).
15. R. L. Sinsheimer, *Genomics* **5**, 954 (1989).
16. R. Dulbecco, *Science* **231**, 1055 (1986).
17. J. Palca, *Nature* **321**, 371 (1986); B. J. Barnhart, *Genomics* **5**, 657 (1989).
18. R. Lewin, *Science* **232**, 1598 (1986).
19. ———, *ibid.* **235**, 747 (1987).
20. B. Alberts, *Cell* **41**, 337 (1985).
21. National Research Council, *Mapping and Sequencing the Human Genome* (National Academy Press, Washington, DC, 1988).
22. Recently, the advisers to the NIH and the DOE genome programs agreed that large-scale sequencing of the human genome should not be considered until the true cost of sequencing falls to no more than 50 cents per base pair (*14*).
23. Office of Technology Assessment, U.S. Congress, *Mapping Our Genes—The Genome Projects: How Big, How Fast?* (OTA-BA-373, U.S. Government Printing Office, Washington, DC, 1988).
24. R. M. Cook-Deegan, "The human genome project: Formation of federal policies in the United States, 1986–1989," paper prepared for the Committee to Study Decisionmaking, Institute of Medicine, National Academy of Sciences, Washington, DC, February 1989.
25. U.S. Congress. House. Subcommittee of the Committee on Appropriations, *Departments of Labor, Health and Human Services, Education, and Related Agencies Appropriations for 1988*. 100th Cong., 1st sess., 1988.
26. National Institutes of Health, *Report to the Director, NIH, from the Ad Hoc Program Advisory Committee on Complex Genomes* (NIH, Washington, DC, 1988).
27. L. Roberts, *Science* **241**, 1752 (1988).

28. J. D. Watson and E. Jordan, *Genomics* **5**, 654 (1989).
29. D. J. Kevles, *In the Name of Eugenics* (Univ. of California Press, Berkeley and Los Angeles, 1985); R. Proctor, *Racial Hygiene* (Harvard Univ. Press, Cambridge, MA, 1988).
30. Committee for Responsible Genetics, *Position Paper on Human Genome Initiative* (Committee for Responsible Genetics, Boston, MA, 1990).
31. B. Muller-Hill, *Murderous Science* (Oxford Univ. Press, New York, 1988).
32. J. Wyngaarden, personal communication.
33. J. Palca, *Nature* **337**, 200 (1989).
34. ———, *Science* **245**, 131 (1989).
35. C. McGourty, *Nature* **340**, 496 (1989).
36. L. Roberts, *Science* **241**, 1596 (1988).
37. ———, *ibid.* **247**, 281 (1990).
38. M. Olson, L. Hood, C. Cantor, D. Botstein, *ibid.* **245**, 1434 (1989).
39. STS is an acronym for sequence-tagged site. An STS is basically a short DNA sequence that has been shown to be unique within a genome.
40. Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987).
41. M. V. Olson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7826 (1986).
42. A. Coulson, J. Sulston, S. Brenner, J. Karn, *ibid.*, p. 7821.
43. A. Coulson, R. Waterston, J. Kiff, J. Sulston, Y. Kohara, *Nature* **335**, 184 (1988).
44. P. H. St George-Hyslop *et al.*, *Science* **235**, 885 (1987).
45. V. A. McKusick, *Genomics* **5**, 385 (1989).

Orchestrating the Human Genome Project

CHARLES R. CANTOR

The Human Genome Project is under way. The Department of Energy and the National Institutes of Health are cooperating effectively to develop organizational structures and scientific priorities that should keep the project on schedule and within its budget.

THE HUMAN GENOME PROJECT IS BIOLOGY'S FIRST LARGE science project with a definite end point. Although it is small compared to most other Big Science efforts, many biologists are still somewhat fearful of the impact this project will have on biology research traditions and funding priorities. Here I outline how the project has evolved from its earliest conceptions to the present, rather different structure. My intention is to convince the reader that a productive, sensible, scheme is in hand to manage this effort and to achieve the goals of the project within a reasonable budget and time period. The short-term cost to traditional biology should be small, but the long-term benefits should be almost unmeasurable.

The Human Genome Project appears to have had several independent origins. One started in a meeting in Alta, Utah, in 1984, when a number of scientists began thinking about the prospect of sequencing all the DNA in the human genome (*1*). The meeting was not called for this purpose. Under the auspices of the U.S. Department of Energy (DOE), Ray White and Mortimer Mendelsohn had convened a small group of experts, mostly molecular biologists, to try to solve a problem. The DOE has a congressional mandate to monitor inherited damage caused by low-level exposure to radiation and other environmental hazards. Existing methods simply were not capable of detecting mutation rates in exposed human populations. Tools were needed that could detect a single altered nucleic acid base in, say, 10^8. However, doing that would be almost as much work as sequencing the human genome.

Other significant origins of the Human Genome Project include a meeting organized by Robert Sinsheimer (*2*) at Santa Cruz in 1985 and an article by Renato Dulbecco (*3*) in 1986. All these roots seem to have coalesced for the first time at a meeting in Cold Spring Harbor in 1986 when the current model of the project as a multicenter, multinational cooperative effort reached full bloom.

More than 5 years after the first conceptualizations, we remain a long way from sequencing any complex genome, and even complete bacterial sequences have still been elusive. However, many initial skeptics have become convinced that mapping and ultimately sequencing the human genome and other complex genomes is a practical and worthwhile task. Our perspective of how to organize it has changed considerably, partly in response to concerns about the costs involved, concerns from the biological community, and changes in technology and strategies.

In the years immediately after the Alta meeting, a major stumbling block was finding people who would want to do such a seemingly boring and tedious task as sequencing the genome. Indeed, Sydney Brenner has jokingly suggested establishing a penal colony where sentences consisting of large-scale sequencing projects would be carried out (*4*). A popular model was a large center, highly integrated and organized along industrial lines. Walter Gilbert made a strong case that there was no reason for delay as the technology was in hand to do the project at a cost that would be dwarfed by the ultimate benefit (*5*). However, a majority of the early enthusiasts for the project felt that initial, major investments in improvements in technology would soon result in much more efficient gene mapping and sequencing methods. This would greatly increase the power of individual investigators and obviate the need for a massive central structure. This model, with evolving technologies playing a major role, fit in much better with the spirit of contemporary biological research, and it ultimately became the accepted framework. It carries the explicit assumption that the cost of DNA sequencing must be reduced by at least an order of magnitude before the major sequence production aspects of the final project can commence.

Several research developments helped stimulate broader interest in generating complete human genomic maps on a reasonably short time scale. The completion of physical maps of *Escherichia coli* showed the feasibility of such projects (*6*), and the immediate usefulness of these maps in a variety of biological experiments ranging from finding genes to characterizing DNA rearrangements made the project seem less onerous. Excitement was generated when several important human disease genes were located by a combination of genetic mapping and molecular biological analysis. Howev-

The author is the director of the Human Genome Center, Lawrence Berkeley Laboratory, Berkeley, CA 94720.