

4 Historical Perspective — The Timeline Of The Human Genome Project

The beginning of the international human genome project (HGP) is a point of contention. There are numerous accounts.

4.1 Pre-genomic Sequencing Cottage Industry

Prior to the 1970s, scientists knew of no way of finding which part of the DNA molecule caused diseases such as Huntington's disease. The only detective techniques they had available were those associated with sex, or isolated cases with telltale signs. For example, one in 15 men is color-blind while almost all women are not. This fact indicates that this predominantly male disease must be caused by a defective gene somewhere on the single X chromosome. Women would have to have both X chromosomes defective to get the disease. Men, having one X and one Y chromosome, do not have the luxury of a spare X chromosome, and if the only X chromosome is defective, that is it. Other diseases, such as Down's syndrome, have telltale signs. Children with Down's syndrome have three copies of chromosome 21 (trisomy 21). This narrows down the location of the Down's syndrome gene to chromosome 21. This is a special case of aneuploidy, a defect that arises because of too many or too few chromosomes.

However, diseases with telltale signs are exceptions. Genetic diseases like Huntington's disease, a degenerative condition of the brain, leave no such clues for genetic detective Sherlock Holmes. In 1977, biologists made a breakthrough. They hit on a brilliant simple way of doing detective job to orient themselves in the DNA molecule using genetic markers. Any sort genetic differences will do, but by far the easiest kinds are tiny spelling differences in the DNA text. Much like milestones and road signs on freeways help highway patrols locate accident sites, these genetic markers help biologists navigate along the DNA and locate disease genes.

Despite the successes in diagnosing genetic diseases, biologists were novice at reading the genetic script as late as the 1980s. DNA sequencing — the job of deciphering short stretches of the four chemical letters of DNA — was a time-consuming, tedious, and error-prone task. Before the DNA could be sequenced, it had to be extracted from cells and grown up in bacteria colonies. Then it was chemically sliced into pieces of varying lengths. To sort the pieces into different sizes, they were dragged electrically through a gelatinous medium in a process called electrophoresis. Then after radioactively labeling the pieces, they were visualized on an X-ray plate. The resulting pattern of bars could be read off against a template, letter by letter, as the sequence. A tiny stretch of 10,000 letters could take a laboratory a year.

4.2 *Scaling up a la Big Science — A Biological Moonshot, A Biological Particle Smasher*

Biologists then were working in a cottage industry manner. They worked in small groups in independent laboratories. The human genome involves some 3 billion letters. At the rate of 10,000 per year per laboratory, it would take a laboratory working alone about 300,000 years to complete the task of sequencing the entire human genome! One way to tackle such a massive project was for biology to scale up. Physicists took on enormous international projects. CERN (European Organization for Nuclear Research) in Switzerland is one good example. SSC (Superconducting Super Collider) was a Department of Energy atom smasher. The construction cost was estimated at US\$12 billion and on top of that, it would have cost US\$500 million annually for operations. It went defunct when the U.S. House of Representatives decided in 1993 to halt the project after 14 miles of tunneling were completed and \$2 billion dollars had been spent.

Fifty years ago, finding a single gene was an interminable process. Genes are small sections of DNA that tell cells how to function. Each gene contains instructions on producing or controlling a specific protein. The proteins, in turn, carry out the cell's work in the body. In 1985, a biologist and chancellor at the University of California at Santa Cruz, Robert Sinsheimer had been searching for a large project to take biology to another level and to put Santa Cruz on the scientific map. He began to wonder if there were any large-scale biological projects that were being missed.⁴ Sinsheimer invited a dozen of the world's experts in molecular genetics to Santa Cruz to fathom their reactions to the idea of mapping and sequencing the entire human genome. At that point in time, many of those present thought it was a proposal way ahead of its time. On the face of it, Sinsheimer's idea seemed absurd.

Biologists in several quarters also began to campaign for a systematic gene discovery program. The proposal proved quite elusive because of divergent agendas and ideas. From a more cultural perspective, biology had always been a cottage industry science, populated by independent researchers and financed by modest grants. The culture of biology was hostile to the central planning implicit in a big project like unraveling the human genome.

Walter Gilbert, a physicist-turned-biologist and a Nobel Laureate for inventing a method of sequencing DNA, argued that the project had to be done, even if it meant biologists have to change the way they work. At a heated meeting at Cold Spring Harbor, Gilbert wrote what he thought would be the cost of sequencing the entire human genome on a blackboard, \$3 billion. Many feared the size and cost would undermine ordinary cottage biology. At that time, the fear, especially among younger scientists, the ones who were really the key to make the genome project work, became palpable. They wonder what would happen to them because they

⁴ "Decoding the Book of Life", NOVA, WGBH Educational Foundation, October 31, 1989.

would find a situation in which not just the funding would be tight and the best ideas got funded, but also they would be required to join the sequencing effort or leave the field for good.

On a more technical level, many scientists thought it was a waste of time to study the entire genome because an estimated 95% of it is believed to be junk DNA, stretches of chemical letters with no apparent purpose. In the words of Sydney Brenner, then director of Molecular Genetics Institute at Cambridge, "More than 95% of the DNA is junk. But let me point out that it's not garbage because the difference between junk and garbage is exactly the same difference you make. Garbage you throw away and junk you keep because you think you might want to do something useful with it, and of course you never do. So, 95%, or more than 95% is junk, and I think that is a valid argument to say against the idea of sequencing the entire genome, because we'd spend a lot of time doing this. Against this, people said, well, you don't know until you've done it whether it is or isn't junk."⁵

In this regard, the human genome may be thought of as a software program evolved over 4 billion years. It is a language not too different from any human conversation. There is a lot of redundancy, a lot of hemming, hawing and stuttering. The minor difference is that the English alphabet has 26 letters, while the human genome alphabet has only 4 letters. Unlike most texts, the genome is a historical text in the sense that it has been passed on from generation to generation, much like in the Middle Ages, monks would make copies of a text one at a time and passed on to other monks. Despite their best efforts, over the course of the copying iterations, they introduced unintentional mistakes into the text. In an analogous way, the genome is a historical text recopied at each generation. Some of the variations that occurred during the reproduction process have turned out to be harmless (no visible effects), some have turned out to be advantageous (for example, adaptation), and some have manifested adverse effects (for example, deformities and diseases).

4.3 *Enters the Black Knight*

Improbable it may sound, the Human Genome Project (HGP) was saved from oblivion by the U.S. Department of Energy (DOE). DOE administered a dozen of national laboratories, including Los Alamos (LANL), Livermore (LLNL), and Lawrence Berkeley National Lab (LBNL). DOE was used to doing big science, had vast technical resources and unrivalled computing facilities. Its interest in biology stemmed from monitoring damage and inherited damage caused by radiation or other elements in the environment, especially after Hiroshima and Nagasaki. Despite all criticisms, DOE secured funding to start work on three chromosomes and set up three human genome centers at Los Alamos, Berkeley and Livermore. Soon, the initiative attracted the interest of the U.S. Congress for two possible reasons:

⁵ "Decoding the Book of Life", NOVA, WGBH Educational Foundation, October 31, 1989.

1. The intuitive feel that the initiative was going to make fundamental advances in human medicine, and
2. The belief that the initiative would serve as a stimulus to spin off a new industry for which the U.S. would have supremacy for decades.

At a 1988 National Research Council (NRC) committee meeting, headed by Bruce Alberts, settled the scientific debate. The committee recommended a 15-year, \$3-billion program to decipher the human genome. In the Summer of 1988, the National Institutes of Health (NIH), in a political and legerdemain move to regain its role in the center stage of biological science, appointed James Dewey Watson to head up the NIH new genome office. Watson is a Nobel laureate of 1962 for co-discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material with Francis Crick and Maurice Wilkins.

The U.S. Congress adopted the NRC strategy in 1990.

4.4 *Divide-and-Conquer Strategy*

As mentioned above, back in the 1980s, gene discovery was a labor-intensive and time-consuming exercise. Scientists used radioactive tags to reveal the sequence and it could take researchers a decade to decipher a single gene using the technique developed by the British scientist Frederick Sanger, who shared the 1980 Nobel Prize with Walter Gilbert for their contributions concerning the determination of base sequences in nucleic acids.

In 1986, Applied Biosystems (the new name is PE Biosystems) of Foster City developed a machine that improved upon Sanger's method by using fluorescent tags instead of radioactive ones. Fluorescent tags can be detected by lasers, which feed the data to computers, thus automating the sequencing process. Craig Venter, then at NIH, was quick to capitalize on the new machine. He used the machine to randomly hunt for gene fragments. At the time, the accepted approach was to look for specific genes whose functions were known, such as the insulin gene. Detractors scoffed that it was aimless to know the sequences of genes for which the functions were unknown. By the early 1990s, Venter had already found as many genes or gene fragments as all other scientists combined. In 1992, NIH caused a stir when it tried to patent the thousands of genes Venter had discovered.

Concurrently, a team of government scientists led by James Watson had already embarked on a different strategy to understand the genome. In contrast to Venter's approach to sequence genes at random, Watson's human genome project proposed to map the genome first. The original strategy for genome sequencing is a divide-and-conquer approach. In this approach, the genome is first dissected into smaller manageable pieces. During the dissection process, the order of the pieces is determined. An analogy will be taking apart a huge jigsaw puzzle in a sensible way so that the information of connectivity is not completely lost in the disassembly process. If the human genome is regarded as *Encyclopedia Humanica*, then the

dissection is equivalent to dividing the encyclopedia into sections of a volume (chromosomal fragments). Genetic mapping will be equivalent to dividing each book into chapters, a chapter for brown eye for example. Physical mapping will be like dividing the chapters into pages. Sequencing will then be reading text off each of the pages. In other words, mapping involved isolating genes, finding out something about them, and locating them on one of the 24 chromosomes.

The two polar strategies would soon lead Watson and Venter into irreconcilable disagreement. In April 1992 Watson left NIH, to be replaced by Acting Director Michael Gottesman, and Venter founded The Institute for Genomic Research (TIGR), where he continued pursuing his machine-driven, rapid-fire hunt for random genes.

4.5 Shotgun Strategy

Since its inception, the premise of the government project had been to find key genes and map their location on the 24 chromosomes. In other words, the more genes scientists could find and map, the smaller it would be the unknown stretches in between and the easier the eventual task of reading the entire genome.

Venter's vision was quite different. He envisioned the genome as a huge instruction book that was far too long for his machine to read. In his approach, the book is shredded into smaller readable fragments. By shredding several copies of the book and taking care to completely randomize the process, a solution is obtained. Each shredded book would result in a pile of fragments. The randomization process guarantees that some fragments in each puzzle will overlap. This is the so-called shotgun strategy.

The shotgun strategy involves the following steps. Genome center researchers first take a whole DNA from an anonymous donor. The DNA is chopped up into large strands of 150,000 bp using restriction enzymes or DNA text-cutters, which was discovered in 1968. Some strands constitute whole chromosome and others are parts of large chromosomes. If necessary, like in the international human genome effort which involves various institutions, these large strands are distributed to the five main and 11 contributing centers worldwide. These large unknown strands are cut into smaller pieces of roughly 2,000 bp. The smaller pieces are then cloned thousand of times. Technicians add in As, Ts, Gs, and Cs to match up with the strand under study. Done enough number of times, a fragment of every possible length will be created. The end of each fragment is tagged with a fluorescent dye. These fragments are put into a machine to sort them by size using a process called electrophoresis. In the electrophoretic process, smallest strand reaches a laser detector first, followed by one that is one bp longer, and so forth. The laser detects whether the fluorescent marker is an A, a C, a G, or a T. Eventually, it reads the whole fragment. The process is repeated 4–10 times for each 2,000 bp fragment. The result is a huge puzzle. Computers are used to detect patterns and assemble the

fragments in a way that reveals the sequence of the larger unknown strand. By computationally comparing overlapping segments, scientists can reassemble the complete book.

This shotgun process, first used in the late 1970s to read the genome of a virus with a DNA just a few thousand letters long, with the blessing of higher throughput of modern sequencing machines, proves to be more cost-effective and less time-consuming than the original divide-and-conquer strategy.

4.6 *The Traveling Salesman Is a Chinese Postman*

All sequencing projects involve breaking up a genome into manageable sizes and putting it back together again. For example, Celera sequenced the human genome by breaking it up at random and piecing together the resulting 27,271,853 sequences. Although the public Human Genome Project took a more structured approach, both groups faced similar problems when reassembling their sequences.

Chief among the problems is that large genomes such as the human genome are very repetitive, like a jigsaw with many identically shaped pieces. Sequencing errors compound the problem, making it difficult to tell if one is looking at different stretches of DNA or not.

Sequence assembly is analogous to finding the shortest route through many cities that passes through each only once, like a traveling salesman. The path traversed is officially known as a Hamiltonian path. Mathematicians call problems like this NP-complete: the only way to solve them is to try every possible route. This requires massive computer power as the number of possibilities rises exponentially with the number of towns — or DNA pieces.

By breaking the chunks of DNA into smaller fragments of equal size, Pavel Pevzner of University of California, San Diego, and his colleagues have transformed the Hamiltonian path of genome assembly into a Eulerian path.⁶

In an Eulerian path, instead of visiting every city once only, the traveler must travel down every road once only — passing through each junction as often as needed. Finding the shortest route through this network is called the Chinese postman problem.

Chinese postman problems are mathematically much more tractable than traveling salesman problems. The traveling salesman problem is yet intractable except resorting to computers by forcing assemblers to make arbitrary decisions, which lead unavoidably to potential further errors. The Eulerian path is almost the same in formulation, but there is a dramatic difference in complexity.

In a play-off against other genome assemblers including PHRAP, used by the Human Genome Project, Pevzner's program, christened EULER, was the only one to make no errors piecing together fragments of the *Neisseria meningitidis* genome,

⁶ P.A. Pevzner, H. Tang, and M. Waterman, "An Eulerian path approach to DNA fragment assembly", *Proceedings of the National Academy of Sciences USA*, 98, 2001, pp. 9748–9753.

the bacterium that causes meningitis. Bacterial genomes are relatively nonrepetitive, so the researchers are in the process of giving EULER stiffer challenges using data from higher organisms.⁷

4.7 *The Race to the Finishing Line — Bigger is Better*

The turning point came in 1999 when Venter, now at Celera Genomics formed in 1988 by PE Biosystems, used shotgun sequencing to sequence *hemophilus influenza*, a microbe that caused ear infection and meningitis. The sequencing of *hemophilus* convinced the government camp to switch from mapping to shotgun sequencing. *Hemophilus* is very small in size compared to the human genome. In the book analogy, if *hemophilus* were a book page, then the human genome would be War and Peace. Scientists soon realized that size did not matter. To do shotgun sequencing of the human genome, what the community needs is stronger machinery.

When PE Biosystems came up with a new generation of faster sequencers, Francis Collins, who had assumed directorship of National Human Genome Research Institute (NHGRI) in 1993 from Michael Gottesman, decided it was time to halt mapping and go full force on sequencing. Collins managed to get \$250 million to buy hundreds of the new sequencing machines to turn the five main labs into production line factories running 24 hours a day, seven days a week, nonstop. These five main labs are:

1. Baylor College of Medicine
2. Washington University at St. Louis, Missouri
3. Whitehead Institute at the Massachusetts Institute of Technology.
4. The Sanger Institute, United Kingdom
5. The Joint Genome Institute (JGI) of the U.S. Department of Energy.

Francis Collins calls these main labs the G5 Alliance.

Together, the five labs were to produce 85% of the public program genome sequence. The remainder would be handled by international collaborators over time. Eventually, the genome project involves sixteen institutions from six countries that form the Human Genome Sequencing Consortium (HGSC):

1. Baylor College of Medicine, Houston, Texas, USA
2. Beijing Human Genome Center, Institute of Genetics, Chinese Academy of Sciences, Beijing, China
3. Gesellschaft für Biotechnologische Forschung mbH, Braunschweig, Germany
4. Genoscope, Evry, France
5. Genome Therapeutics Corporation, Waltham, MA, USA
6. Institute for Molecular Biotechnology, Jena, Germany

⁷ P.A. Pevzner, "A new approach to fragment assembly in DNA sequencing", Tenth Annual Bioinformatics and Genome Research, San Francisco, June 17–19, 2001, Chair: Hwa A. Lim.

7. Joint Genome Institute, U.S. Department of Energy, Walnut Creek, CA, USA
8. Keio University, Tokyo, Japan
9. Max Planck Institute for Molecular Genetics, Berlin, Germany
10. RIKEN Genomic Sciences Center, Saitama, Japan
11. The Sanger Centre, Hinxton, U.K.
12. Stanford DNA Sequencing and Technology Development Center, Palo Alto, CA, USA
13. University of Washington Genome Center, Seattle, WA, USA
14. University of Washington Multimegabase Sequencing Center, Seattle, WA, USA
15. Whitehead Institute for Biomedical Research, MIT, Cambridge, MA, USA
16. Washington University Genome Sequencing Center, St. Louis, MO, USA

The previous mapping effort did not completely go to waste. The public's mapping program had already broken the 3-billion-letter genome into fragments corresponding to the 24 chromosomes. This enabled them to distribute the work among the sixteen public sequencing centers, and adopted a hierarchical shotgun strategy, a complementary strategy to the whole-genome shotgun of Celera Genomics.

4.8 *Complementary Strategies — Ally or Foe*

The public and the private sectors use similar automation and sequencing technology, but different strategies to sequence the human genome. The public project uses a “hierarchical shotgun” strategy in which individual large DNA fragments of known position are subjected to shotgun sequencing. Celera uses a “whole-genome shotgun” strategy in which the entire genome is shredded into small fragments that are sequenced and put back together on the basis of sequence overlaps.

Pros and cons of either approach have been a subject of debate.⁸ The hierarchical shotgun approach has the advantage that the global location of each individual sequence is known with certainty, but it requires constructing a map of large fragments covering the genome. The whole-genome shotgun approach does not require this step but has challenges in the assembly of fragments. Both approaches align the sequences along the human chromosomes by using markers contained in the physical map. Indeed, the two approaches are complementary and in the future, a hybrid of the two approaches may prove to be even more effective.

⁸ Hwa A. Lim, and T.V. (Venky) Venkatesh, "Bioinformatics in the pre- and post-genomic eras", *Trends in Biotechnology*, April 2000, Vol. 18 No. 4(195), pp. 133–135.

4.9 From Genome Marathon to Sprint to the Finishing Line

Officially, the genome project started in 1990 when the U.S. Congress decided to fund the initiative. The first ten years were a long marathon and progress was slow because the infrastructure and ancillary technologies were not in place. In 1999, partly because the development phase of the project had matured, and partly because of fierce competition from the private sector, the pace picked up. The healthy competition between the public and the private sectors fueled the sprint to the finishing line.

By November 1999, the public Human Genome Project had read one third of the genome. In March 2000, they reached the two-thirds milestone, on track to complete a first working draft in June 2000. On April 6, 2000, Celera announced that it had completed sequencing the human genome, but would need another three to six weeks to assemble the fragments into one book, thus completing a first working draft of the human genome. The working draft of the human genome will have thousands of blank spots or gaps. These genome gaps occur where the structure of DNA is hard for machines to read. It will take scientists another two years or so to go back and fill in the blanks.

By April 2000, the public effort had used \$1.9 billion. But one has to be careful in interpreting the expenses. Much of the money was spent on developing better tools, techniques and infrastructure in the years 1990–1999. Thus a fairer comparison will be the \$250 million Collins acquired to buy machines to pull even with Celera, which obtained \$300 million from PE Biosystems.

4.10 The Gutenberg Press of the Human Genome Project

Much as in the 15th century, Europe's awakened desire for learning created a huge demand for books, the genetic revolution has created a rapidly growing market for genomic research instrument and chemicals. According to a study by William Blair & Co., the market for genomic instrument and chemicals is about \$2.7 billion worldwide in 2000, and it is expected to grow at a rate of 22% per annum for the next 3–5 years.

The DNA sequencing machine, a gray box about the size of a small refrigerator, has positioned its manufacturer Perkin-Elmer Applied Biosystems of Foster City, California, in solid monopoly in the galvanic field of genomics. Central to the success is the premise that in most industries except biomedical instrumentation, standardized equipment is used to unify the process and churn out conforming data.

In the field long hindered by expensive and labor-intensive procedures, the machine has cut sequencing time by 60% and labor costs by 80% (year 2000 statistics). Its round the clock automation of once tedious tasks has embraced not just giant laboratories involved in sequencing genomes, but also average scientists who undertake more focused projects that were once off limit because of cost and time. Besides being the machine that scientists used to complete for the first time

the first draft of the human genome, it has also been used to identify the gene responsible for baldness and the gene responsible for aging. It also helped identify the subspecies of chimpanzee that is likely the source of AIDS epidemic. This machine is so ubiquitous it is hard at work in three-quarters of the international human genome laboratories. Analysts estimate that 40,000 laboratories in 100 countries rely on PE Biosystems instrumentation in 2000. In the USA, nine out of 10 university-based genetics laboratories use the instrument.

The DNA sequencing machine is usually paralleled with the 15th-century Gutenberg press.⁹ As far as necessity is the mother of invention is concerned, the analogy between the Gutenberg press and sequencing machine is good. However, we argue that the analogy is off in terms of reproduction. The subtle difference between the Gutenberg press and the sequencing machine is that in the former, the press mass prints books written by humans for other humans to read. In the latter, the machine reads the books of life of humans and other organisms for the benefit of humankind. In this sense, the press propagates, whereas the sequencer deciphers.

The real reproduction process in biology is the polymerase chain reaction (PCR), invented by Kary Mullis while at Chiron Corporation in the late 1980s. Mullis was awarded the 1993 Chemistry Nobel Prize for this invention. PCR revolutionizes biotechnology by allowing a genetic fragment to be reproduced indefinitely.

Table 1. Timeline of the genome project — from the beginning to the U.S. White House and U.K. Whitehall announcement.

1985:

The concept of the Human Genome Project (HGP) materializes. Robert Sinsheimer holds a meeting on human genome sequencing at University of California, Santa Cruz.

Charles DeLisi and David Smith commission the first Santa Fe conference to assess the feasibility of a Human Genome Initiative.

1986:

Following the Santa Fe conference, U.S. Department of Energy (DOE) announces the Human Genome Initiative with \$5.3M to develop critical resources and technologies at DOE national labs.

1987:

Senator Pete Domenici of New Mexico introduced legislation for about \$28 million to fund the Human Genome Project.*

Congressionally chartered DOE advisory committee recommends a 15-year, multidisciplinary, scientific and technological undertaking to map and sequence the human genome.

The National Institutes of Health (NIH) begins funding genome projects.

⁹ Lisa M. Krieger, "DNA sequencing machine puts PE Biosystems at forefront", *San Jose Mercury News*, June 10, 2000.

* Senator Pete Domenici calls the human genome project "the greatest wellness project".

Table 1. (Continued)

-
- Hwa A. Lim of Supercomputer Computations Research Institute (SCRI), a DOE supercomputer institute, coins bioinformatics.
- 1988:
National Research Council (NRC) recommends a concerted genome program.
DOE and NIH outline plans for cooperation.
Human Genome Organization (HUGO) founded by scientists to coordinate.
First Cold Spring Harbor meeting on genome mapping and sequencing starts.
- 1989:
DOE and NIH establish Joint ELSI Working Group. ELSI: Ethical, Legal, and Social Issues.
- 1980s:
Maynard Olson invents YAC. Leroy Hood invents four-color sequencing.
- 1990:
DOE and NIH present a joint 5-year plan to Congress.
The 15-year project formally begins in the U.S.
James Watson is appointed director of National Human Genome Project, NIH.
Hwa Lim convenes the very first “Bioinformatics & Genome Research” conference series, funded by DOE and NSF (National Science Foundation), Technology Research & Development Authority (TRDA) of Florida, and computer companies.
- 1991:
Genome Database (GDB) is established.
- 1992:
James Watson, Leroy Hood, Maynard Olson, Robert Waterston, Bill Gates, Paul Allen met in Seattle, USA to discuss HGP. Olson and Hood establish University of Washington Human Genome Center.
Low-resolution genetic linkage map of the entire human genome is published.
DOE and NIH establish guidelines for data release and resource sharing.
Watson and Craig Venter ran into disagreement on EST. Watson resigns from directorship. Venter leaves NIH to found the Institute for Genomic Research (TIGR).
The Sanger Centre is established in United Kingdom, funded by Wellcome Trust and the British Government. John Sulston and Mike Morgan are co-directors.
- 1993:
Genome centers in the U.S. are established: Washington University (WU), MIT, Baylor College, University of Washington (UW).
DOE and NIH revise 5-year goals of the genome project.
French Généthon provides mega-YACs to the genome community.
- 1994:
Genetic mapping 5-year goal achieved 1 year ahead of schedule.
Genetic Privacy Act is proposed to regulate collection, analysis, storage, and use of genetic information.
Sequencing by hybridization technology funded by DOE commercialized through Hyseq. Hwa Lim is appointed director of bioinformatics at Hyseq.

Table 1. (Continued)

1995:

Los Alamos and Livermore National Labs announce high-resolution physical maps of chromosomes 16 and 19, respectively.

Moderate-resolution maps of chromosomes 3, 11, 12, and 22 are published.

Sequence of the smallest bacterium *Mycoplasma genitalium* completed and provides a model of the minimum number of genes needed for independent existence.

First nonviral whole genome is sequenced, bacterium *Haemophilus influenzae*.

1996:

Saccharomyces cerevisiae (yeast) genome sequence is completed by international consortium.

Methanococcus jannaschii genome is sequenced and it confirms existence of third major branch of life on Earth.

Healthcare Portability and Accountability Act prohibits use of genetic information in healthcare insurance eligibility decisions.

PE 377 sequencing machines (48 lanes and 64 lanes) first came on the market.

Venter used a shotgun strategy to sequence bacteria.

1997:

Escherichia coli genome is sequenced.

High-resolution physical maps of chromosomes X and 7 are complete.

DOE forms Joint Genome Institute (JGI) consisting of DOE genome centers.

National Center for Human Genome Research (NCHGR) becomes National Human Genome Research Institute (NHGRI).

UNESCO adopts Universal Declaration on the Human Genome and Human Rights.

1998:

Caenorhabditis elegans genome sequence is complete.

Mycobacterium tuberculosis sequence is complete.

DOE and NIH revise 5-year plan through 2003.

Human Genome Project passes mid point.

Hospital for Sick Children, Toronto, assumes GDB data collection and curation.

Incyte Pharmaceuticals announces to sequence human genome in two years.

1998, May:

PE invests \$300M and 300 automated 96-lanes PE 377 to form Celera with Venter. The plan is to use shotgun sequencing to finish the human genome in 3 years, most of which will be made public, others will be patented.

1998, Aug:

Pharmacia 96-lane MegaBACE capillary sequencing machine debuts on market.

1998, Oct:

The U.S. and U.K. agree to speed up HGP, planned to have a first draft complete by Spring of 2000, and a detail sequence by 2003.

1999, Jun:

With 300 PE 377 and high performance computers, Celera announces completion of rice genome in 6 weeks, and completing of the human genome in 2–3 years.

Table 1. (Continued)

Dec 1, 1999	Chromosome 22 is the first human chromosome to be completely sequenced. Biotechnology stocks skyrocketed.
Jan 10, 2000:	A day before the International Human Genome Conference, Celera announces using shotgun sequencing to cover 1.8x80% of the human genome, more cost effective and less time-consuming than HGP can do the work.
Jan 11, 2000	International Human Genome Meeting in San Francisco. Participants talk a lot about Celera. NIH and Celera enter into confidential negotiations to share data.
Mar 6, 2000:	NIH-Celera negotiation breaks down.
Mar 12, 2000:	U.S. President Bill Clinton and U.K. Prime Minister Tony Blair jointly announce that genetic data should be freely available. News media misconstrue their message. Biotechnology stocks crash.
Mar 13, 2000:	The U.S. Patent Office announces that Bill Clinton speech is misconstrued. It does not forbid patent applications.
April 6, 2000:	Venter and Waterston appear at a U.S. Congress Hearing. Celera announces completion of 90% of first human genome draft.
May 8, 2000:	To avoid the dilemma of January Celera announcement, HGP announce entering the second phase of sequencing a day before the Cold Spring Harbor Meeting.
June 26, 2000:	The U.S. and U.K. jointly announce completion of the first draft of the human genome.
Feb 12, 2001	Science publishes online the data from the private sector, Nature ¹⁰ publishes online data from the public sector.

5 A Small Step For Genome, A Giant Step For Humankind

Deciphering the human genome has been an epic task. Scientists have used the most powerful computers and developed a new generation of automated machines to read the 3 billion chemical letters. The task has been likened to putting a man on the moon. But is it?

¹⁰ <http://www.nature.com/genomics>