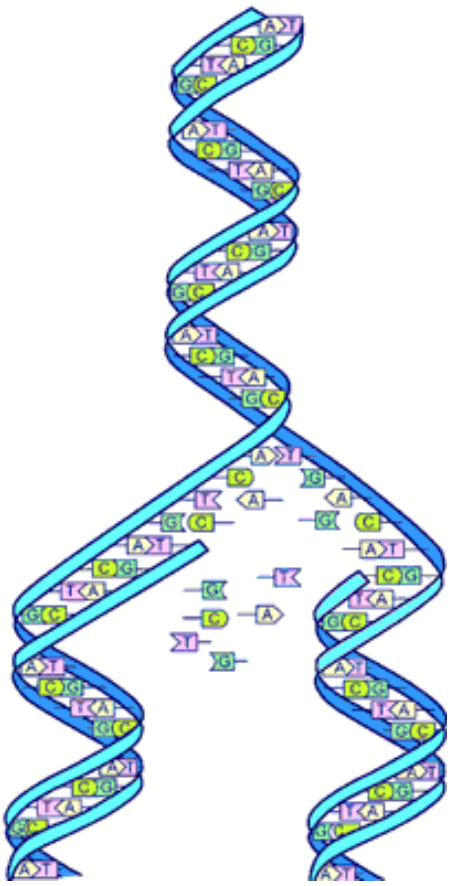
Human Genome Project

The **Human Genome Project** (**HGP**) is an international scientific research project with a primary goal of determining the sequence of chemical base pairs which make up DNA, and of identifying and mapping the approximately 20,000–25,000 genes of the human genome from both a physical and functional standpoint.[1]

The project began in October 1990[2] and was initially headed by Aristides Patrinos, head of the Office of Biological and Environmental Research in the U.S. Department of Energy's Office of Science. Francis Collins directed the US National Institutes of Health (NIH) National Human Genome Research Institute efforts. A working draft of the genome was announced in 2000 and a complete one in 2003, with further, more detailed analysis still being published. A parallel project was conducted outside of government by the Celera Corporation, or Celera Genomics, which was formally launched in 1998. Most of the government-sponsored sequencing was performed in universities and research centres from the United States, the United Kingdom, Japan, France, Germany and Spain. Researchers continue to identify protein-coding genes and their functions; the objective is to find disease-causing genes and possibly use the information to develop more specific treatments. It also may be possible to locate patterns in gene expression, which could help physicians glean insight into the body's emergent properties.



DNA replication

While the objective of the Human Genome Project is to understand the genetic makeup of the human species, the project has also focused on several other nonhuman organisms such as *E. coli*, the fruit fly, and the laboratory mouse. It remains one of the largest single investigative projects in modern science.

The Human Genome Project originally aimed to map the nucleotides contained in a human haploid reference genome (more than three billion). Several groups have announced efforts to extend this to diploid human genomes including the International HapMap Project, Applied Biosystems, Perlegen, Illumina, J. Craig Venter Institute, Personal Genome Project, and Roche-454.

The "genome" of any given individual (except for identical twins and cloned organisms) is unique; mapping "the human genome" involves sequencing multiple variations of each gene.[3] The project did not study the entire DNA found in human cells; some heterochromatic areas (about 8% of the total genome) remain un-sequenced.

Project

History

The project began with the culmination of several years of work supported by the US Department of Energy, in particular workshops in 1984[4] and 1986 and a subsequent initiative [5] of the US Department of Energy.[6] This 1987 report stated boldly, "The ultimate goal of this initiative is to understand the human genome" and "knowledge of the human is as necessary to the continuing progress of medicine and other health sciences as knowledge of human anatomy has been for the present state of medicine." Candidate technologies were already being considered for the proposed undertaking at least as early as 1985.[7]

James D. Watson was head of the National Center for Human Genome Research at the National Institutes of Health in the United States starting from 1988. Largely due to his disagreement with his boss, Bernadine Healy, over the issue of patenting genes, Watson was forced to resign in 1992. He was replaced by Francis Collins in April 1993, and the name of the Center was changed to the National Human Genome Research Institute (NHGRI) in 1997.

The $3-billion project was formally founded in 1990 by the US Department of Energy and the National Institutes of Health, and was expected to take 15 years.[8] In addition to the United States, the international consortium comprised geneticists in the United Kingdom, France, Australia, Japan and a myriad of other spontaneous relationships.[9]

Due to widespread international cooperation and advances in the field of genomics (especially in sequence analysis), as well as major advances in computing technology, a 'rough draft' of the genome was finished in 2000 (announced jointly by U.S. President Bill Clinton and the British Prime Minister Tony Blair on June 26, 2000).[10] This first available rough draft assembly of the genome was completed by the Genome Bioinformatics Group at the University of California, Santa Cruz, primarily led by then graduate student Jim Kent. Ongoing sequencing led to the announcement of the essentially complete genome in April 2003, 2 years earlier than planned.[11] In May 2006, another milestone was passed on the way to completion of the project, when the sequence of the last chromosome was published in the journal Nature.[12]

State of completion

The Human Genome Project was declared complete in April 2003. An initial rough draft of the human genome was available in June 2000 and by February 2001 a working draft had been completed and published followed by the final sequencing mapping of the human genome on April 14, 2003. Although this was reported to be 99% of the human genome with 99.99% accuracy a major quality assessment of the human genome sequence was published in May 27, 2004 indicating over 92% of sampling exceeded 99.99% accuracy which is within the intended goal.[13] Further analyses and papers on the HGP continue to occur. [14]

Applications and proposed benefits

Potential benefits of sequencing the human genome expand to many fields from molecular medicine to a better understanding of human evolution. The Human Genome Project through its sequencing of the DNA can help us understand diseases including genotyping of specific viruses to direct appropriate treatment; identification of oncogenes and mutations linked to different forms of cancer; designing medications and predicting its response better; advancement in forensic applied sciences; biofuels and other energy applications; agriculture, livestock breeding, bioprocessing; risk assessment; bioarcheology, anthropology, evolution. Another proposed benefit is the commercial development of genomics research related to DNA based products, a multibillion dollar industry. [15]

The sequence of the DNA is stored in databases available to anyone on the Internet. The U.S. National Center for Biotechnology Information (and sister organizations in Europe and Japan) house the gene sequence in a database known as GenBank, along with sequences of known and hypothetical genes and proteins. Other organizations, such as the Genome Bioinformatics Group at the University of California, Santa Cruz,[16] and Ensembl[17] present additional data and annotation and powerful tools for visualizing and searching it. Computer programs have been developed to analyze the data, because the data itself is difficult to interpret without such programs.

The process of identifying the boundaries between genes and other features in a raw DNA sequence is called genome annotation and is the domain of bioinformatics. While expert biologists make the best annotators, their work proceeds slowly, and computer programs are increasingly used to meet the high-throughput demands of genome sequencing projects. The best current technologies for annotation make use of statistical models that take advantage of parallels between DNA sequences and human language, using concepts from computer science such as formal grammars.

All humans have unique gene sequences. Therefore the data published by the HGP does not represent the exact sequence of every individual's genome. It is the combined "reference genome" of a small number of anonymous donors. The HGP genome is a scaffold for future work in identifying differences among individuals. Most of the current effort in identifying differences among individuals involves single-nucleotide polymorphisms and the HapMap.

Findings

Key findings of the draft (2001) and complete (2004) genome sequences include http:// / www. genome. gov/ 12011238

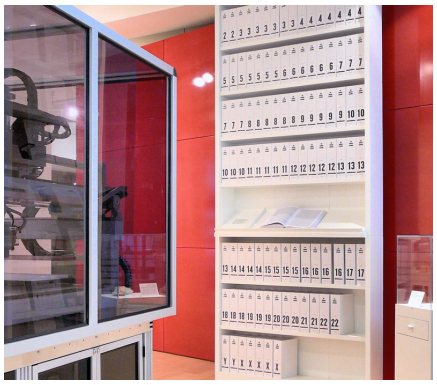
1. There are approximately 23,000 genes in human beings, the same range as in mice and roundworms. Understanding how these genes express themselves will provide clues to how diseases are caused.

2. The human genome has significantly more segmental duplications (nearly identical, repeated sections of DNA) than other mammalian genomes. These sections may underlie the creation of new primate-specific genes

3. At the time when the draft sequence was published fewer than 7% of protein families appeared to be vertebrate specific

How it was accomplished

The Human Genome Project was started in 1989 with the goal of sequencing and identifying all three billion chemical units in the human genetic instruction set, finding the genetic roots of disease and then developing treatments. With the sequence in hand, the next step was to identify the genetic variants that increase the risk for common diseases like cancer and diabetes.



The first printout of the human genome to be presented as a series of books, displayed at the Wellcome Collection, London

It was far too expensive at that time to think of sequencing patients' whole genomes. So the National Institutes of Health embraced the idea for a "shortcut", which was to look just at sites on the genome where many people have a variant DNA unit. The theory behind the shortcut was that since the major diseases are common, so too would be the genetic variants that caused them. Natural selection keeps the human genome free of variants that damage health before children are grown, the theory held, but fails against variants that strike later in life, allowing them to become quite common. (In 2002 the National Institutes of Health started a $138 million project called the HapMap to catalog the common variants in European, East Asian and African genomes.)

The genome was broken into smaller pieces; approximately 150,000 base pairs in length. These pieces were then ligated into a type of vector known as "bacterial artificial chromosomes", or BACs, which are derived from bacterial chromosomes which have been genetically engineered. The vectors containing the genes can be inserted into bacteria where they are copied by the bacterial DNA replication machinery. Each of these pieces was then sequenced separately as a small "shotgun" project and then assembled. The larger, 150,000 base pairs go together to create chromosomes. This is known as the "hierarchical shotgun" approach, because the genome is first broken into relatively large chunks, which are then mapped to chromosomes before being selected for sequencing.

Funding came from the US government through the National Institutes of Health in the United States, and a UK charity organization, the Wellcome Trust, as well as numerous other groups from around the world. The funding supported a number of large sequencing centers including those at Whitehead Institute, the Sanger Centre, Washington University in St. Louis, and Baylor College of Medicine.

The Human Genome Project is considered a Mega Project because the human genome has approximately 3.3 billion base-pairs.

If the sequence obtained was to be stored in book form, and if each page contained 1000 base-pairs recorded and each book contained 1000 pages, then 3300 such books would be needed in order to store the complete genome. However, if expressed in units of computer data storage, 3.3 billion base-pairs recorded at 2 bits per pair would equal 786 megabytes of raw data. This is comparable to a fully data loaded CD.

Public versus private approaches

In 1998, a similar, privately funded quest was launched by the American researcher Craig Venter, and his firm Celera Genomics. Venter was a scientist at the NIH during the early 1990s when the project was initiated. The $300,000,000 Celera effort was intended to proceed at a faster pace and at a fraction of the cost of the roughly $3 billion publicly funded project.

Celera used a technique called whole genome shotgun sequencing, employing pairwise end sequencing,[18] which had been used to sequence bacterial genomes of up to six million base pairs in length, but not for anything nearly as large as the three billion base pair human genome.

Celera initially announced that it would seek patent protection on "only 200–300" genes, but later amended this to seeking "intellectual property protection" on "fully-characterized important structures" amounting to 100–300 targets. The firm eventually filed preliminary ("place-holder") patent applications on 6,500 whole or partial genes [19]. Celera also promised to publish their findings in accordance with the terms of the 1996 "Bermuda Statement," by releasing new data annually (the HGP released its new data daily), although, unlike the publicly funded project, they would not permit free redistribution or scientific use of the data. The publicly funded competitor UC Santa Cruz was compelled to publish the first draft of the human genome before Celera for this reason. On July 7, 2000, the UCSC Genome Bioinformatics Group released a first working draft on the web. The scientific community downloaded one-half trillion bytes of information from the UCSC genome server in the first 24 hours of free and unrestricted access to the first ever assembled blueprint of our human species.[20]

In March 2000, President Clinton announced that the genome sequence could not be patented, and should be made freely available to all researchers. The statement sent Celera's stock plummeting and dragged down the biotechnology-heavy Nasdaq. The biotechnology sector lost about $50 billion in market capitalization in two days.

Although the working draft was announced in June 2000, it was not until February 2001 that Celera and the HGP scientists published details of their drafts. Special issues of *Nature* (which published the publicly funded project's scientific paper)[21] and *Science* (which published Celera's paper[22]) described the methods used to produce the draft sequence and offered analysis of the sequence. These drafts covered about 83% of the genome (90% of the euchromatic regions with 150,000 gaps and the order and orientation of many segments not yet established). In February 2001, at the time of the joint publications, press releases announced that the project had been completed by both groups. Improved drafts were announced in 2003 and 2005, filling in to ≈92% of the sequence currently.

The competition proved to be very good for the project, spurring the public groups to modify their strategy in order to accelerate progress. The rivals at UC Santa Cruz initially agreed to pool their data, but the agreement fell apart when Celera refused to deposit its data in the unrestricted public database GenBank. Celera had incorporated the public data into their genome, but forbade the public effort to use Celera data.

HGP is the most well known of many international genome projects aimed at sequencing the DNA of a specific organism. While the human DNA sequence offers the most tangible benefits, important developments in biology and medicine are predicted as a result of the sequencing of model organisms, including mice, fruit flies, zebrafish, yeast, nematodes, plants, and many microbial organisms and parasites.

In 2004, researchers from the International Human Genome Sequencing Consortium (IHGSC) of the HGP announced a new estimate of 20,000 to 25,000 genes in the human genome.[23] Previously 30,000 to 40,000 had been

predicted, while estimates at the start of the project reached up to as high as 2,000,000. The number continues to fluctuate and it is now expected that it will take many years to agree on a precise value for the number of genes in the human genome.

Genome donors

In the IHGSC international public-sector Human Genome Project (HGP), researchers collected blood (female) or sperm (male) samples from a large number of donors. Only a few of many collected samples were processed as DNA resources. Thus the donor identities were protected so neither donors nor scientists could know whose DNA was sequenced. DNA clones from many different libraries were used in the overall project, with most of those libraries being created by Dr. Pieter J. de Jong. Much of the sequence (>70%) of the reference genome produced by the public HGP came from a single anonymous male donor from Buffalo, New York (code name RP11).[24][25]

HGP scientists used white blood cells from the blood of two male and two female donors (randomly selected from 20 of each) -- each donor yielding a separate DNA library. One of these libraries (RP11) was used considerably more than others, due to quality considerations. One minor technical issue is that male samples contain just over half as much DNA from the sex chromosomes (one X chromosome and one Y chromosome) compared to female samples (which contain two X chromosomes). The other 22 chromosomes (the autosomes) are the same for both sexes.

Although the main sequencing phase of the HGP has been completed, studies of DNA variation continue in the International HapMap Project, whose goal is to identify patterns of single-nucleotide polymorphism (SNP) groups (called haplotypes, or "haps"). The DNA samples for the HapMap came from a total of 270 individuals: Yoruba people in Ibadan, Nigeria; Japanese people in Tokyo; Han Chinese in Beijing; and the French Centre d'Etude du Polymorphisms Humain (CEf) resource, which consisted of residents of the United States having ancestry from Western and Northern Europe.

In the Celera Genomics private-sector project, DNA from five different individuals were used for sequencing. The lead scientist of Celera Genomics at that time, Craig Venter, later acknowledged (in a public letter to the journal *Science*) that his DNA was one of 21 samples in the pool, five of which were selected for use.[26][27]

On September 4, 2007, a team led by Craig Venter published his complete DNA sequence,[28] unveiling the six-billion-nucleotide genome of a single individual for the first time.

Benefits

The work on interpretation of genome data is still in its initial stages. It is anticipated that detailed knowledge of the human genome will provide new avenues for advances in medicine and biotechnology. Clear practical results of the project emerged even before the work was finished. For example, a number of companies, such as Myriad Genetics started offering easy ways to administer genetic tests that can show predisposition to a variety of illnesses, including breast cancer, hemostasis disorders, cystic fibrosis, liver diseases and many others. Also, the etiologies for cancers, Alzheimer's disease and other areas of clinical interest are considered likely to benefit from genome information and possibly may lead in the long term to significant advances in their management.

There are also many tangible benefits for biological scientists. For example, a researcher investigating a certain form of cancer may have narrowed down his/her search to a particular gene. By visiting the human genome database on the World Wide Web, this researcher can examine what other scientists have written about this gene, including (potentially) the three-dimensional structure of its product, its function(s), its evolutionary relationships to other human genes, or to genes in mice or yeast or fruit flies, possible detrimental mutations, interactions with other genes, body tissues in which this gene is activated, and diseases associated with this gene or other datatypes.

Further, deeper understanding of the disease processes at the level of molecular biology may determine new therapeutic procedures. Given the established importance of DNA in molecular biology and its central role in determining the fundamental operation of cellular processes, it is likely that expanded knowledge in this area will

facilitate medical advances in numerous areas of clinical interest that may not have been possible without them.

The analysis of similarities between DNA sequences from different organisms is also opening new avenues in the study of evolution. In many cases, evolutionary questions can now be framed in terms of molecular biology; indeed, many major evolutionary milestones (the emergence of the ribosome and organelles, the development of embryos with body plans, the vertebrate immune system) can be related to the molecular level. Many questions about the similarities and differences between humans and our closest relatives (the primates, and indeed the other mammals) are expected to be illuminated by the data from this project.

The Human Genome Diversity Project (HGDP), spinoff research aimed at mapping the DNA that varies between human ethnic groups, which was rumored to have been halted, actually did continue and to date has yielded new conclusions. In the future, HGDP could possibly expose new data in disease surveillance, human development and anthropology. HGDP could unlock secrets behind and create new strategies for managing the vulnerability of ethnic groups to certain diseases (see race in biomedicine). It could also show how human populations have adapted to these vulnerabilities.

Advantages of Human Genome Project:

1. Knowledge of the effects of variation of DNA among individuals can revolutionize the ways to diagnose, treat and even prevent a number of diseases that affects the human beings.
2. It provides clues to the understanding of human biology.

Ethical, legal and social issues

The project's goals included not only identifying all of the approximately 20,000-25,000 [29] genes in the human genome, but also to address the ethical, legal, and social issues (ELSI) that might arise from the availability of genetic information. Five percent of the annual budget was allocated to address the ELSI arising from the project.

Debra Harry, Executive Director of the U.S group Indigenous Peoples Council on Biocolonialism (IPCB), says that despite a decade of ELSI funding, the burden of genetics education has fallen on the tribes themselves to understand the motives of Human genome project and its potential impacts on their lives. Meanwhile, the government has been busily funding projects studying indigenous groups without any meaningful consultation with the groups. (See Biopiracy.)[30]

The main criticism of ELSI is the failure to address the conditions raised by population-based research, especially with regard to unique processes for group decision-making and cultural worldviews. Genetic variation research such as HGP is group population research, but most ethical guidelines, according to Harry, focus on individual rights instead of group rights. She says the research represents a clash of culture: indigenous people's life revolves around collectivity and group decision making whereas the Western culture promotes individuality. Harry suggests that one of the challenges of ethical research is to include respect for collective review and decision making, while also upholding the Western model of individual rights.

References

[1] Robert Krulwich (2001-04-17). *Cracking the Code of Life* (http://www.pbs.org/wgbh/nova/genome/) (Television Show). PBS. ISBN 1-5375-16-9. .

[2] "About the Human Genome Project: What is the Human Genome Project" (http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml). The Human Genome Management Information System (HGMIS). 2011-07-18. . Retrieved 2011-09-02.

[3] Harmon, Katherine (2010-06-28). "Genome Sequencing for the Rest of Us" (http://www.scientificamerican.com/article.cfm?id=personal-genome-sequencing&print=true). Scientific American. . Retrieved 2010-08-13.

[4] Cook-Deegan R (1989). "The Alta Summit, December 1984" (http://www.ornl.gov/sci/techresources/Human_Genome/project/alta.shtml). *Genomics* **5** (3): 661–3. doi:10.1016/0888-7543(89)90042-6. PMID 2613249. .

[5] http://www.ornl.gov/sci/techresources/Human_Genome/project/herac2.shtml

[6] Barnhart, Benjamin J. (1989). "DOE Human Genome Program" (http://www.ornl.gov/sci/techresources/Human_Genome/publicat/hgn/v1n1/01doehgp.shtml). *Human Genome Quarterly* **1**: 1. . Retrieved 2011-05-02.

[7] DeLisi, Charles (2001). "Genomes: 15 Years Later A Perspective by Charles DeLisi, HGP Pioneer" (http://genome.gsc.riken.go.jp/hgmis/publicat/hgn/v11n3/05delisi.html). *Human Genome News* **11**: 3–4. . Retrieved 2005-02-03.

[8] http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml

[9] "A New Five-Year Plan for the United States: Human Genome Program" (http://www.genome.gov/10001476). 1993-10-01. . Retrieved 14 July 2012.

[10] "White House Press Release" (http://www.ornl.gov/sci/techresources/Human_Genome/project/clinton1.shtml). . Retrieved 2006-07-22.

[11] Noble, Ivan (2003-04-14). "Human genome finally complete" (http://news.bbc.co.uk/1/hi/sci/tech/2940601.stm). *BBC News*. . Retrieved 2006-07-22.

[12] "Guardian Unlimited |UK Latest | Human Genome Project finalised" (http://web.archive.org/web/20071012170819/http://guardian.co.uk/uklatest/story/0,,-5829253,00.html). *The Guardian* (London). Archived from the original (http://www.guardian.co.uk/uklatest/story/0,,-5829253,00.html) on October 12, 2007. . Retrieved 2006-07-22.

[13] Nature 429, 365-368

[14] http://www.ornl.gov/sci/techresources/Human_Genome/project/journals/journals.shtml

[15] genomics.energy.gov

[16] http://genome.ucsc.edu

[17] http://www.ensembl.org

[18] Roach; Boysen, C; Wang, K; Hood, L (1995). "Pairwise end sequencing: a unified approach to genomic mapping and sequencing" (http://linkinghub.elsevier.com/retrieve/pii/088875439580219C). *Genomics* **26** (2): 345–353. doi:10.1016/0888-7543(95)80219-C. PMID 7601461. .

[19] http://news.bbc.co.uk/1/hi/sci/tech/487773.stm

[20] "The Human Genome Project Race" (http://www.cbse.ucsc.edu/research/hgp_race). Center for Biomolecular Science and Engineering. . Retrieved 2011-05-02.

[21] International Human Genome Sequencing Consortium (2001). "Initial sequencing and analysis of the human genome" (http://www.nature.com/nature/journal/v409/n6822/pdf/409860a0.pdf) (PDF). *Nature* **409** (6822): 860–921. doi:10.1038/35057062. PMID 11237011. .

[22] Venter, JC et al. (2001). "The sequence of the human genome" (http://www.sciencemag.org/cgi/reprint/291/5507/1304.pdf) (PDF). *Science* **291** (5507): 1304–1351. Bibcode 2001Sci...291.1304V. doi:10.1126/science.1058040. PMID 11181995. .

[23] IHGSC (2004). "Finishing the euchromatic sequence of the human genome" (http://www.nature.com/nature/journal/v431/n7011/full/nature03001.html). *Nature* **431** (7011): 931–945. Bibcode 2004Natur.431..931H. doi:10.1038/nature03001. PMID 15496913. .

[24] Osoegawa, Kazutoyo; Mammoser, AG; Wu, C; Frengen, E; Zeng, C; Catanese, JJ; De Jong, PJ (2001). "A Bacterial Artificial Chromosome Library for Sequencing the Complete Human Genome" (http://www.genome.org/cgi/content/full/11/3/483). *Genome Research* **11** (3): 483–96. doi:10.1101/gr.169601. PMC 311044. PMID 11230172. .

[25] Tuzun, E et al. (2005). "Fine-scale structural variation of the human genome" (http://www.nature.com/ng/journal/v37/n7/full/ng1562.html). *Nature Genetics* **37** (7): 727–737. doi:10.1038/ng1562. PMID 15895083. .

[26] Kennedy D (2002). "Not wicked, perhaps, but tacky". *Science* **297** (5585): 1237. doi:10.1126/science.297.5585.1237. PMID 12193755.

[27] Venter D (2003). "A Part of the Human Genome Sequence". *Science* **299** (5610): 1183–4. doi:10.1126/science.299.5610.1183. PMID 12595674.

[28] Levy S, Sutton G, Ng PC et al. (September 2007). "The Diploid Genome Sequence of an Individual Human". *PLoS Biol.* **5** (10): e254. doi:10.1371/journal.pbio.0050254. PMC 1964779. PMID 17803354.

[29] U.S. Department of Energy Genome Programs. (2011, July 25). In Human Genome Project Information. Retrieved March 20, 2012, from http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml

[30] Ms. Harry's comments in the International Forum on Globalization Teach-in held in New York City in February 2001 were based on her recent article, *Biopiracy and Globalization: Indigenous Peoples Face a New Wave of Colonialism*, published in the magazine Splice, January/April 2001 Volume 7 Issues 2 & 3 ((http://www.geneticsforum.org.uk))

Further reading

- Victor K. McElheny. *Drawing the Map of Life: Inside the Human Genome Project* (Basic Books; 2010) 361 pages. Examines the intellectual origins, history, and motivations of the project to map the human genome; draws on interviews with key figures.

External links

- Human Genome Project (http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml) official information page
- Delaware Valley Personalized Medicine Project (http://www.coriell.org/index.php/content/view/92/167/) Uses data from the Human Genome Project to help make medicine personal
- National Human Genome Research Institute (NHGRI) (http://www.genome.gov). NHGRI led the National Institutes of Health's contribution to the International Human Genome Project. This project, which had as its primary goal the sequencing of the three thousand million base pairs that make up human genome, was successfully completed in April 2003.
- Human Genome News (http://www.ornl.gov/sci/techresources/Human_Genome/publicat/hgn/hgn.shtml). Published from 1989 to 2002 by the US Department of Energy, this newsletter was a major communications method for coordination of the Human Genome Project. Complete online archives are available.
- Project Gutenberg hosts e-texts for Human Genome Project, titled *Human Genome Project, Chromosome Number #* (# denotes 01-22, X and Y). This information is raw sequence, released in November 2002; access to entry pages with download links is available through http://www.gutenberg.org/etext/3501 for Chromosome 1 sequentially to http://www.gutenberg.org/etext/3524 for the Y Chromosome. Note that this sequence might not be considered definitive due to ongoing revisions and refinements. In addition to the chromosome files, there is a supplementary information file (http://www.gutenberg.org/etext/11799) dated March 2004 which contains additional sequence information.
- The HGP information pages (http://www.doegenomes.org/) Department of Energy's portal to the international Human Genome Project, Microbial Genome Program, and Genomics:GTL systems biology for energy and environment
- yourgenome.org: The Sanger Institute public information pages (http://www.yourgenome.org/) has general and detailed primers on DNA, genes and genomes, the Human Genome Project and science spotlights.
- Ensembl project (http://www.ensembl.org/), an automated annotation system and browser for the human genome
- UCSC genome browser (http://genome.ucsc.edu), This site contains the reference sequence and working draft assemblies for a large collection of genomes. It also provides a portal to the ENCODE project.
- Nature magazine's human genome gateway (http://www.nature.com/genomics/human/), including the HGP's paper on the draft genome sequence
- Wellcome Trust Human Genome website (http://genome.wellcome.ac.uk/) A free resource allowing you to explore the human genome, your health and your future.
- Learning about the Human Genome. Part 1: Challenge to Science Educators. ERIC Digest. (http://www.ericdigests.org/2003-2/genome.html)
- Learning about the Human Genome. Part 2: Resources for Science Educators. ERIC Digest. (http://www.ericdigests.org/2003-2/genome2.html)
- *Patenting Life* by Merrill Goozner (http://www.prospect.org/cs/articles?article=patenting_life)
- Prepared Statement of Craig Venter of Celera (http://clinton4.nara.gov/WH/EOP/OSTP/html/00626_4.html) Venter discusses Celera's progress in deciphering the human genome sequence and its relationship to healthcare and to the federally funded Human Genome Project.

- Cracking the Code of Life (http://www.pbs.org/wgbh/nova/genome/) Companion website to 2-hour NOVA program documenting the race to decode the genome, including the entire program hosted in 16 parts in either QuickTime or RealPlayer format.
- Lone Dog L (1999). "Whose genes are they? The Human Genome Diversity Project" (http://www.haworthpress.com/store/ArticleAbstract.asp?sid=DS104W6XGPL58H9EKU06C2L1QSSP41VF&ID=5707). *J Health Soc Policy* **10** (4): 51–66. doi:10.1300/J045v10n04_04. PMID 10538186.

Article Sources and Contributors

Human Genome Project *Source*: http://en.wikipedia.org/w/index.php?oldid=526041361 *Contributors*: (jarbarf), 1exec1, 24fan24, 63.206.105.xxx, 97198, AL SAM, APLAdam, Abdull, AbigailAbernathy, Adonovan0, Agni451, Agricola44, Ahoerstemeier, Aisaac, Aitias, Akriasas, Alansohn, Alasdair, AlexLevyOne, Alfie66, Ali, AlistairMcMillan, Allstarecho, Alpha7248, AlphaEta, Alphachimp, Amaury, Analbutt, Anaxial, Andre Engels, Andrew Delong, Andyjsmith, Animum, Anonymous3190, Antelopotamus, Antialias, Arigoldberg, Arjun01, Arthena, Astuishin, Atavism, Athouse88, Atif.t2, Auranor, BIsopp, Bambuway, BarretB, Bbpen, Beland, Bendy660, Betterusername, Big balls man 1, Bigwill121192, Biobunz, Blainster, Blessedmégots, Bluerasberry, Bobblewik, Bobo192, Bookwyrm404, Briaboru, Brightonr, Brteag00, Bsadowski1, Bucketsofg, Buckyball, Butterfly reflections, Bwithh, Byakuren, C.Fred, CSWarren, Calilasseia, Calmer Waters, Calor, Cammoore, Camw, Can't sleep, clown will eat me, CanadianLinuxUser, Capricorn42, Carstensen, Cashspencer, Cecillecarino, Ceyockey, Ched, Chewy3326, ChicXulub, Chokfung, ChrisGualtieri, Christian List, Cimorcus, Clemwang, ClockworkSoul, Colincbn, Color probe, Comatose51, Cometstyles, ConnorGriffin, Conversion script, Courcelles, Crc32, Crum375, Cspan64, CyberSkull, DARTH SIDIOUS 2, DJBarney24, DSachan, DVD R W, Dac04, Dac28, Dachannien, Dally Horton, Dan D. Ric, Dan100, Dan653, Daniel, Daniel,levine, Danny B-), David.Monniaux, Db099221, Dcirovic, Debresser, Delaney LaRosa, Delldot, Deltabeignet, Dennis Bratland, DerHexer, DesertCondor, Dgw, Dhollm, Diannaa, Discospinster, Dlohcierekim, DoSiDo, DocWatson42, Docboat, Dogcow, Dogface, DonSiano, DopefishJustin, DoubleAW, Draco Hunt, Dspradau, Dujo, Dullhunk, Duncan.france, Duncharris, Dwayne, Dylan Lake, E-at-windmills, Edcolins, Edgar181, Edokter, Elektrik Shoos, Elfguy, Eliz81, Elroch, Eric-Wester, EricLeMaster, Ewawer, Excirial, Faithlessthewonderboy, Fantastic fred, FastLizard4, Fastfission, Feureau, Feydey, Fireswordfight, Fireyair, Fitzwilliam, Fordmadoxfraud, Forluvoft, FrancoGG, Frankenpuppy, Freakofnurture, Fredrik, Fredwrigley, Froth, Furrykef, Fusionmix, Fvw, GLaDOS, GalaxyGirl99, Gamaliel, Genometer, George Church, Giftlite, Gilliam, Glacialfox, Glen, Glloq, Goodone121, Gorgan almighty, Grafen, Graham87, Greenshed, GreyCell, Gugilymugily, Gybem, HBow3, Hadal, Haham hanuka, Hair, Hateless, Haukurth, Hayfordoleary, Hbent, Hdt83, Hectorguinness, Heron, Honesty32, Human.v2.0, Husond, Hut 8.5, IMKatgrrl, Iain.mcclatchie, Iamthebob, Ibjt4ever, Idont Havaname, Ignoscient, Ii.libraaaa, Immunize, Indon, Infinium, Iramidi, Isaacvetter, Isayoolala, It Is Me Here, Ixfd64, Izalithium, J.delanoy, JHeinonen, JTN, JWSchmidt, Jackol, Jahub, Jakob Suckale, Jamw23, Jaredroach, Jazriel, Jazzeur, Jcbarret, Jeff G., Jeltz, Jeremyb, Jerry, Jersey emt, Jfdwolff, Jim1138, Jimkent, Jlorz, Jmason, Jmichaelrosenberg, JoeSmack, Joeyman365, John, John of Reading, JohnOwens, Johntex, Johnuniq, Jonny Grey, Jordyhoyt, Joshygoodwin, Jpeeling, Jtmac24, JuJube, Juliancolton, Junzhangyankees, Jusdafax, Ka Faraq Gatri, Kanyegirl, Karlpearson, Kblakes, Kbradnam, Keegscee, Keilana, Ketchie1, Kimiko, King2, Kingpomba, Kjhf, Klilidiplomus, KnowledgeOfSelf, Kristen Eriksen, Krol:k, Kungfuadam, Kurykh, Kyle Barbour, L33t 3lit3, L888Y5, LAX, La Pianista, Laurinaaaaa, LcawteHuggle, Leafyplant, Leevanjackson, Lexor, LilHelpa, Lina111, Llypde, Loom91, Lova Falk, Lowellian, Luna Santin, MONGO, Ma8thew, MacGyverMagic, Macy, Madhero88, Magnus Manske, Mahmutuludag, Malo, Mar4d, MarcoTolo, MarkSweep, Martin451, Matt.T, Matt1701D, Matthew Stannard, Maurice Carbonaro, Mav, Maxrider16, McSly, Mdd4696, Member, Menchi, Meno25, Mets501, Mgjdelgr, Michael Hardy, Mike Rosoft, Mikeo, Mild Bill Hiccup, Milnivlek, Minesweeper, Minesweeper.007, Minimac's Clone, MinorWiki, Mjd21, Mmm72392, Modest Genius, Modify, ModusOperandi, Monobi, Mr Rookles, Mr pand, Mr.Z-man, Mschel, Msctrojan, Mtd2006, Myanw, NTox, Nakon, NameIsRon, Naveenswiki, NawlinWiki, NeonMerlin, Neutrality, NewEnglandYankee, Nickshanks, Niels Olson, Nigborg, Nikofeelan, Nneonneo, Nofatlandshark, Nommonomanac, Nono64, Novangelis, Nsaa, Ntse, Nunh-huh, Nyvhek, Oblivious, Ocee, OldakQuill, Olivier, Olly1988, Optimale, Orlady, OverlordQ, Owen, Oxymoron83, PDH, Padgeman, Pandacomics, Paranoid, Pearle, Peeweeherman22, Per Olofsson, Peter.C, Pgan002, Phantomsteve, Pharaoh of the Wizards, Phil Boswell, Philip Trueman, Phippsyuk, Phoenix-forgotten, Pi, Pilotguy, Pouchkidium, Ppgardne, Prisonblues, Pro66, Prodego, Pvosta, Quebec99, Quinsareth, Qwfp, Qxz, R'n'B, RDBrown, RJB227, RJII, Radon210, Ragesoss, Ramuman, Rangoon11, Razor2988, Rdl381, Reach Out to the Truth, Reginmund, Regre7, Res2216firestar, Rex the first, RexNL, Rgoodermote, Rhobite, Rich Farmbrough, Richard001, Richi, Richwales, Rikuansem13, Rjd0060, Rjensen, Rjwilmsi, Rm1271, Robcarol, Robin klein, Rockpocket, Rory096, RoyBoy, Rtchicdpu, Russ London, Russoc4, Ruy Lopez, Ryan Postlethwaite, Ryan-McCulloch, SDC, SFairchild, SHP, SZielins, Sabedon, Saintrain, Saltwolf, Samosa Poderosa, Samsara, Sarindam7, Saseigel, Saul Greenberg, Savidan, Sayeth, Scarian, Sceptre, Seans Potato Business, Selket, Setanta747 (locked), Shalom Yechiel, Shimirel, Ships at a Distance, Shizhao, Shoeofdeath, Shoessss, Shultz IV, Sjjupadhyay, SkyWalker, Sljaxon, Smeggysmeg, Smileyborg, Sodium, Solitude, Some jerk on the Internet, Speedyboy, SpookyMulder, Spread da love lyk nutella, SpuriousQ, SriMesh, Sriram sh, Sterichinderance, Stolsvik, StuRat, Suidafrikaan, Sumankumardutta, Supahshadow, Superbeecat, Superborsuk, Superjack105, Superm401, Svick, Swid, Synchronism, TAnthony, TLG, Ta bu shi da yu, Tanzz, Tdtaylor, TedE, Tedder, Tellyaddict, Templarion, Template namespace initialisation script, TestPilot, TexasAndroid, The Nameless, The Thing That Should Not Be, The cuhjnts jot., The wub, TheIguana, Theda, Thedjatclubrock, Theguywhostheguy, Thingg, Thue, Tide rolls, TimKasoar, TimVickers, Titoxd, Toastthemost, Tollsjo, Tomer M, Tommy11111, Toussaint, Trainthh, TransControl, Treflipit23, TrevorRW83, Trevyn, Trimp, Turnstep, Tushar858, Tutec, Twilight, Twinsday, Typhoon94, Unclemikejb, UnitedStatesian, Urdaneta, Utcursch, Vegaswikian, Versus22, Viciouspiggy, Viddin5, Vigeesh, Vipinrajd, Waldir, Walkie, Warlordwolf, Wcfios, Weasel extraordinaire, Wet dog fur, WhatamIdoing, Wiki alf, Wiki-uk, Wiki13, WikiDao, Wild30002, William Avery, Wimt, Wonder al, Woppit, WriterHound, Wtmitchell, Xanthoptica, Xezbeth, Xooll, Yamamoto Ichiro, Ynhockey, Zackplanet42, ZayZayEM, Zeno Gantner, Ziyaddd, Zorahunter, 1564 anonymous edits

Image Sources, Licenses and Contributors

File:Dna-split.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Dna-split.png *License*: Public Domain *Contributors*: Conscious, Dietzel65, Juliancolton, LadyofHats, Madprime, Magnus Manske, Niki K, Pixeltoo, Roybb95, Samulili, Teetaweepo, Thomas81, 2 anonymous edits

Image:Wellcome genome bookcase.png *Source*: http://en.wikipedia.org/w/index.php?title=File:Wellcome_genome_bookcase.png *License*: Creative Commons Attribution-Sharealike 3.0 *Contributors*: Russ London (talk). Original uploader was Russ London at en.wikipedia

License