

Human Genome Sequence 2/3 Complete

March 29, 2000

Two Thirds of Human DNA Script Deciphered by Human Genome Project; Public Consortium To Complete "Working Draft" in June

The Human Genome Project international consortium announced today that two billion of the three billion "letters" that constitute the genetic instruction book of humans have been deciphered and deposited into [GenBank](#). GenBank, the public database of DNA sequence operated by the National Institutes of Health, is accessible freely and without restrictions to all scientists in industry and academia.

The two billionth "letter," or base pair, was deposited earlier this month by the Wellcome Trust's Sanger Centre in Great Britain. The "letter" was a "T," the abbreviation for thymine, one of the four chemicals or bases that make up DNA. The 2,178,076,000 unique base pairs now in GenBank have been mapped to their locations on the 24 human chromosomes.

The Human Genome Project is on track to complete the "working draft," which will include 90 percent of the human DNA sequence, with an accuracy of 99.9 percent, in June. The Human Genome Project worldwide will invest an estimated \$250 million in producing the working draft.

The finished, stand-the-test of time version of the human DNA sequence will be ready on or before 2003. Just four months ago, the Human Genome Project reached the one-billionth base pair milestone.

"It's good news that we're moving so fast but it's even better news that researchers throughout the world are using this data now to investigate the genetic underpinnings of health and diseases ranging from Alzheimer's to diabetes," said Dr. Francis Collins, director of NIH's National Human Genome Research Institute, in a speech today at the BIO 2000 annual international biotechnology conference in Boston.

"We are pleased to be contributing to the creation of the scientific infrastructure that will enable the next stage of the biotechnology revolution," said Dr. Ari Patrinos, director of the U.S. Department of Energy's Office of Biological and Environmental Research, which began the Human Genome Project in 1986 and which sponsors the Joint Genome Institute in Walnut Creek, CA.

Reaching the two billion base pair milestone is "a splendid achievement which will help doctors around the world in their quest to cure disease and advance knowledge," said Dr. Michael Morgan, chief executive of Wellcome Trust Genome Campus in Cambridgeshire, United Kingdom, which hosts the UK contribution to the Human Genome Project.

Sequencing, which is determining the exact order of DNA's four chemical bases, commonly abbreviated A, T, C and G, has been expedited in the Human Genome Project by technological advances in deciphering DNA and the coalition's collaborative nature, which has resulted in about 1,000 scientists worldwide working together effectively.

Today the Human Genome Project assembles 12,000 bases every minute. Twenty years ago, sequencing that many bases would have required one year or more. Three years ago, when pilot projects to evaluate

feasibility of large-scale sequencing were initiated by the Human Genome Project, deciphering 12,000 bases required 20 minutes.

Scientists throughout the world already are using the human sequence data in GenBank for basic research and disease related studies. Recently the genes responsible for hereditary deafness and cerebral cavernous malformations, an often-fatal vascular disease causing seizures and brain hemorrhages, were detected with data from GenBank.

Scientists are rapidly annotating the human DNA sequence in GenBank with information about the location of specific genes and the genetic variants (called Single Nucleotide Polymorphisms or SNPs) that can provide clues to various health disorders.

Almost 15 billion raw base pairs were sequenced to reach the two billion milestone. Human Genome Project scientists decipher each area of a chromosome at least four to five times to insure that the data deposited into GenBank is accurate. The "depth of coverage," as this repeat sequencing is called, also helps the scientists assemble the long stretches of the "A," "T," "C," and "G" bases. The finished version of the human DNA sequence that the Human Genome Project will complete in 2003 will have a greater depth of coverage, with at least eight to nine fold coverage for each chromosome region.

The international Human Genome Project consortium includes scientists at 16 institutions in France, Germany, Japan, China, Great Britain and the U.S. The five institutions that generate the most sequence are: the U.S. DOE Joint Genome Institute in CA; Baylor College of Medicine, Houston; Washington University School of Medicine, St. Louis; Whitehead Institute, Cambridge, Mass.; and the Sanger Centre in Great Britain. NHGRI funds the sequencing centers at Baylor, Washington University and Whitehead.

For more information contact:

- Department of Energy Human Genome Program, Jeff Sherwood, 202-586-5806
- NHGRI, Cathy Yarborough 301-594-0954