Washington
WASHINGTON·UNIVERSITY·IN·ST·LOUIS
School of Medicine

Office of Medical Public Affairs
Washington University School of Medicine
at Washington University Medical Center
Campus Box 8508, 4444 Forest Park Ave.
St. Louis, MO 63108-2259
(314) 286-0100 FAX: (314) 286-0199

News

Contact: Linda Sage
314-286-0100 or
Westin Hotel, 206-728-1000

A WORM'S EYE VIEW OF THE HUMAN GENOME PROJECT
(EMBARGOED UNTIL 2:30 PM PST, FEB. 15)

Seattle, Feb. 15, 1997 -- Collaborators in St. Louis, Mo., and Cambridge, England, have sequenced 63 million base pairs of roundworm DNA — by far the largest amount of sequence obtained from any species. They now are applying the techniques, tools and informatics developed with this and other model organisms to the human genome.

Robert H. Waterston, M.D., Ph.D., the James S. McDonnell Professor and head of genetics at Washington University School of Medicine in St. Louis, will describe the sequencing methods — and some surprising discoveries about the worm genome — on the afternoon of Saturday, Feb. 15, at the annual meeting of the American Association for the Advancement of Science in Seattle.

"The genes we are finding in this model organism are proving invaluable in understanding the corresponding human genes," says Waterston, who directs Washington University's Genome Sequencing Center.

The roundworm *Caenorhabditis elegans* has six chromosomes and an estimated 15,000 genes. With 100 million base pairs of DNA, its genome is one-thirtieth the size of the human genome and more than eight times larger than the biggest genome completed to date — the yeast sequence, to which Washington University also contributed.

In collaboration with the Sanger Centre in Cambridge, England, the St. Louis group began to sequence *C. elegans* in 1990, launching the first effort to complete the genome of a multicellular organism. The collaborators have finished 63 percent of the worm genome and posted data in progress for 80 percent on the World Wide Web. The accuracy rate of the finished data is 99.99 percent — just one error per 10,000 nucleotide bases.

A five-year $42 million grant from the National Center for Human Genome Research is supporting the project, which should be completed by December 1998.

The work is generating some unexpected findings. Some genes with related functions are clustered together on the *C. elegans* genome — such clusters previously were thought to be confined to bacteria. And many genes resemble each other in sequence, a discovery that should spur evolutionary studies. The large number of repetitive sequences in the *C. elegans* genome also was unexpected. The repeats can be categorized into families, some of which occur only on certain chromosomes.

About 40 percent of the *C. elegans* genes found so far have no known counterparts in other organisms. The remaining 60 percent are similar to known genes. "We predict that the *C. elegans*

-more-

genome will represent a subset of the genes in humans and other more advanced organisms," Waterston says.

The *C. elegans* DNA comes from cosmid libraries of genomic DNA collected by the Sanger Centre's Alan Coulson, Ph.D. Cosmids are circular pieces of bacterial DNA that can store samples of DNA from other species.

The researchers allow the bacterial colonies to multiply so they obtain a sufficient amount of each *C. elegans* segment. After physically cutting each sample into smaller lengths, they subclone the pieces into bacteria to obtain large quantities of sequence-ready snippets.

Sequencing begins by essentially tagging each of the four types of nucleotides in DNA with a fluorescent dye of a different color. The color-coded fragments run down a gel that is monitored by a laser. The machine that performs this operation, an Applied Biosystems DNA sequencer, therefore can determine the order of the nucleotide bases in the original sample. The information goes directly into a computer, which uses software developed by the collaborators to assemble the sequences of the small pieces of DNA into the sequence of the original 40,000-base-pair cosmid clone.

Automated sequencing is followed by finishing, a process that uses different sequencing techniques to clarify ambiguous parts of the sequence. The finished sequences are laid out on the physical map that marks the locations of the original cosmid clones, building a continuous sequence of all of the nucleotide bases in the *C. elegans* genome.

GeneFinder, another piece of software developed at Washington University, helps the researchers locate genes within the DNA sequence. It searches for code words that *C. elegans* tends to use and for sequences that direct the internal editing of genes. The researchers also compare the *C. elegans* sequence with sequences in GenBank, a repository of publicly available gene sequences maintained by the National Center for Biotechnology Information in Bethesda, Md.

The centers deposit their annotated *C. elegans* sequences in GenBank, which provides free access via the Internet to researchers around the world. They also post raw data on the World Wide Web at http://genome.wustl.edu/gsc.

Last April, the Washington University researchers received a three-year $24 million grant from the National Center for Human Genome Research for human genome sequencing. The team is sequencing chromosome 22, the smallest of the 24 human chromosomes, regions of chromosome 7, and parts of chromosome X, a sex chromosome that has been mapped at Washington University. They previously had sequenced a region of chromosome 13, enabling other researchers to pinpoint the location of the breast cancer gene BRCA2.

"Determining the sequences and precise locations of all the human genes will have a profound effect on the detection and treatment of disease," Waterston says. "It also will provide an invaluable framework for exploring normal biological functions."

###