

Geneticist from Baylor named as new head of UK's Sanger Centre

David Dickson, London

The Wellcome Trust announced this week that a mouse geneticist is to be the next director of the Sanger Centre. Based outside Cambridge, England, and founded jointly by the trust and the Medical Research Council, the centre is one of the world's leading sequencing facilities.

Allan Bradley has spent the past 13 years at the Baylor College of Medicine, Houston, and is currently Cullen Professor of Human and Molecular Genetics. He is also a Howard Hughes Medical Institute investigator. He will take over the leadership of the Sanger Centre from sequencing pioneer John Sulston, who has headed the centre since it was set up in 1992.

Bradley studied under geneticist Martin Evans at the University of Cambridge before taking up a post at Baylor in 1987. His main research interest is in elucidating the developmental genetics of the mouse, in particular by isolating stem-cell lines from pre-implantation embryos.

"Remarkable opportunities are arising from genome sequencing programmes to progress basic biology and its applications in medicine," says Michael Dexter, the head of the Wellcome Trust. "We are delighted that Allan Bradley is joining us and helping to drive forward the trust's mission."

The main task facing Bradley will be to map out the 'post genome' phase of Sanger's development, now that the sequencing of the human genome is nearing completion. His background suggests that both mouse genomics and stem-cell research will — as with the US National Institutes of Health — be high on the agenda. ■

Researchers take a gamble on the human genome

Paul Smaglik, Washington

Biologists are taking bets — literally — on the number of genes in the human genome. Enterprising attendees at the annual Cold Spring Harbor genome meeting last week opened a book, taking bets at \$1 a time. The results will be known in three years time, when sequencing is completed.

The spread of bets placed so far — from 27,462 genes at the low end to 153,478 at the high — represents two very different approaches to gene counting. Techniques that extrapolate from manually annotated portions of the parts of the genome that have already been fully sequenced are yielding estimates of around 35,000–40,000 genes. Those that use computer algorithms to scour random expressed sequence tags (ESTs) from the whole genome predict 100,000 or more.

Bets cost \$1 this year, \$5 at next year's meeting and \$20 in 2002. At stake — beyond a cash prize that will be awarded at the 2003 meeting — is an improved understanding of the complexity of the genome, and the relative importance of genes and regulatory regions. There is still confusion over how to count genes that don't code for protein, and uncertainty about what biological roles these pieces of DNA play.

Perhaps reflecting the absence of widely expected announcements on new sequencing milestones, the issue became the focus of heated discussion at Cold Spring Harbor last week. "Sequencing is like digging a gold mine," says John Quackenbush, a computational biologist with The Institute for Genomic Research (TIGR) in Rockville, Maryland. "How much gold is there to find?"

Quackenbush's own prediction of 120,000 genes, to be published in the June issue of *Nature Genetics*, is at the high end of the scale. DoubleTwist (formerly Pangea), a bioinformatics company in Oakland, California, last week announced that its own algorithms predicted a total of 100,000 genes.

Both use gene-hunting programs and DNA and protein homology searches to find candidate genes in GenBank. But each uses different software. Another genomics company, Incyte, predicts over 100,000 genes, but has used its proprietary EST database, rather than the public GenBank (see *Nature* 401, 311; 1999).

Tim Hubbard, who heads Ensembl, an automated annotation program similar to TIGR's and DoubleTwist's, has doubts. "Automatic annotation over-predicts the number of genes," says Hubbard, who is based at Britain's Sanger Centre. "This is due to false positives and cases where multiple genes are annotated when there is really only one."

The Ensembl total mirrors that of groups led by Philip Green, of the University of Washington, Seattle, and Jean Weissenbach, of the Centre National de Séquençage at Evry, France, both of which have separate papers appearing in June's *Nature Genetics*. Each used separate techniques that closely scrutinized a portion of the genome before scaling up their counts.

Green's group used either the curated, annotated genes from chromosome 22 or a filtered set of near full-length mRNA sequences from Genbank as a starting point. Weissenbach's team compared the bacterial artificial chromosome clone ends of a pufferfish genome, with a known, curated set of protein-encoding genes, then used algorithms to compare protein coding regions from the fish's genome with a human sequence.

Quackenbush notes that extrapolation has its weaknesses, too. For example, chromosomes 22 and 21 — whose sequence is published this week (see page 311) — are similar in size; but chromosome 21 has 225 genes, compared with 545 on chromosome 22. Using either total on its own to estimate the number of genes in the genome could be misleading.

Francis Collins, director of the US National Human Genome Research Institute, notes that totalling the genes in chromosomes 21 and 22, then scaling that figure up to account for the size of the whole genome, results in an estimate of 40,000 genes. That's close to the figure arrived at by the extrapolators and Ensembl. But he cautions not to put too much faith in this approach. ■

♦ <http://www.ensembl.org/genesweep.html>

Los Alamos labs are safe from fire

Rex Dalton, San Diego

The Los Alamos National Laboratory in New Mexico was reported this week as being out of fire danger, following a blaze that burned 43 square miles of nearby forest.

The homes of 216 laboratory employees — including many scientists — in nearby communities were destroyed in the fire. Officials say the blaze was started as a controlled burn by forestry agents, but that it then got out of control.

Laboratory employees began returning to work on Monday, after the facility had been closed and largely evacuated for nearly a week, leading to about \$3.5 million a day in operational losses. Although there was no



damage to any facilities housing nuclear material, two old, isolated structures from the Manhattan Project were destroyed. ■