

EBI (EMBL)



European Bioinformatics Institute of the  
European Molecular Biology Laboratory



## PRESS RELEASE

Embargoed until 2 pm US EST, Dec. 4, 2002

### The power of two

*The draft mouse genetic code provides a powerful  
tool for understanding our own genome*

**Contacts:**

**European Bioinformatics Institute, EMBL**

Cath Brooksbank, PhD  
Scientific Outreach Officer  
Hinxton Hall, Hinxton  
Cambridge CB10 1SD  
Great Britain  
Tel: +44 (0)1223 492525  
Fax: +44 (0)1223 494468  
email: [cath@ebi.ac.uk](mailto:cath@ebi.ac.uk)  
[www.ebi.ac.uk](http://www.ebi.ac.uk)

**EMBL Press Office:**

Office of Information and Public Affairs  
Meyerhofstr. 1  
D-69117 Heidelberg, Germany  
Tel: +49-6221-387252/452  
Fax: +49-6221-387525  
email: [info@embl.de](mailto:info@embl.de)  
[www.embl.de](http://www.embl.de)

**Policy regarding use:**

EMBL press releases may be freely reprinted and distributed via print and electronic media. Text, photographs and graphics are copyrighted by the EMBL. They may be freely reprinted and distributed in conjunction with this news story, provided that proper attribution to authors, photographers and designers is made. High-resolution copies of the images can be downloaded from the EMBL web site: [www.embl.de](http://www.embl.de)



## The power of two

### *The draft mouse genetic code provides a powerful tool for understanding our own genome*

London, Heidelberg, 4 December 2002

Today the Mouse Genome Sequencing Consortium, a collaboration between scientists across the world, announces the publication of a draft sequence of the mouse genome — the genetic blueprint of a mouse — along with an analysis that compares it with the human genome. This comparison has revealed 1200 new human genes and some intriguing clues about our evolutionary origins. The study is published in the December 5 issue of the journal *Nature*, and both the mouse and human sequences are freely accessible to the public at [www.ensembl.org](http://www.ensembl.org).

This is a major achievement because it marks the first time that scientists have been able to compare the human genome with that of another mammal. Despite obvious differences between ourselves and rodents, the close evolutionary relationship between the two species makes us very similar in terms of our body plan, physiology and reactions to disease. These similarities, coupled with a rich history of laboratory studies of mouse genetics, make the mouse an extremely useful ‘model’ organism to study human biology and disease. The mouse genome will greatly enhance this usefulness. For example, genes that have been linked to human diseases usually have close relatives in the mouse. Having both genomes will vastly reduce the amount of time needed to discover the mouse equivalents of such genes, giving researchers a head start on medically relevant research.

The draft sequence has allowed the consortium to get as much information about the mouse genome into the public domain as quickly as possible, so that scientists can begin to reap the benefits from the mouse sequence before it’s finished. And it’s already having a huge impact on how medical research is done. Professor Allan Bradley, Director of the Wellcome Trust Sanger Institute, says: “We download in a morning information that, only two years ago, would have taken months to determine in the lab. Experiments will remain the lynchpin of biology, but resources such as the mouse sequence transform the way we work.”



photo: EMBL Photolab



The sequence shows the order of the four DNA chemical bases — A, T, C, and G — making up the 20 chromosomes of a female mouse. The draft sequence was produced by a method known as the ‘whole genome shotgun’ approach, in which the mouse’s entire collection of DNA is broken into pieces, sequenced and then assembled by matching overlapping sequences. This was done mainly by scientists at the Wellcome Trust Sanger Institute in Hinxton (UK), the Whitehead Institute in Cambridge (USA) and Washington University in St. Louis (USA).

The next step was to find all the genes, which contain the recipes for protein molecules. This was mainly carried out by the *Ensembl* database team at the European Bioinformatics Institute (EBI, a branch of the European Molecular Biology Laboratory) and the Wellcome Trust Sanger Institute, both based in Hinxton, UK, and was primarily funded by the Wellcome Trust. Genes comprise only a small percentage of mouse and human DNA. The rest, whose function is largely unknown, is commonly called ‘junk DNA’. “Although a mouse cell knows which parts of the genome are genes and which parts are not, this is not obvious to us,” explains Ewan Birney, the EBI’s *Ensembl* Coordinator, “it’s rather like

**Mouse Genome Server**

**Ensembl Entry Points**

Search for  with

Display Chr  From  To

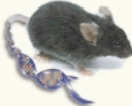
Retrieve a sequence

Advanced data retrieval tool

BLAST your sequence

For fast identity search try

**About MGSC (9.3.1 details)**

 **MGSC**  
The Mouse Genome Sequencing Consortium is a joint project between The Whitehead Institute/MIT Center for Genome Research, The Washington University Genome Sequencing Center, The Wellcome Trust Sanger Institute and EMBL - EBI to provide the Mouse genome sequence to the world. We work closely with other Mouse groups to provide an integrated resource [\[details\]](#).

**Mouse Genome Assembly v 3**

This site presents the latest mouse draft sequence based principally on whole genome shotgun of around 7x coverage. This was frozen in Feb. 2002 and incorporates finished clone information where available. The sequence is estimated to cover 96% of mouse euchromatic DNA. The assembly consists of large supercontigs (sets of DNA contigs ordered and orientated on the basis of read pairs) aligned to chromosomes using the mouse genetic map. The supercontigs are extremely long, with the majority of the genome covered by under 100 supercontigs. The BAC map resource [\[details\]](#) has been aligned to this assembly.

We have predicted with high confidence 22,444 genes across the genome of which ~75% have a firm counterpart in the human genome. We provide an initial functional annotation of these genes and a DNA level comparison to the human genome.

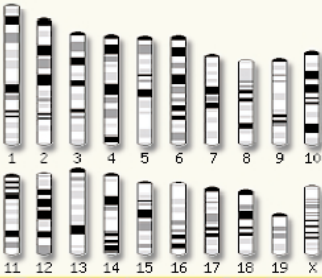
We will continue to improve the assembly and annotation over the coming year.

Credits for the mouse sequencing project can be found [here](#). Additional mouse resources can be found [here](#).

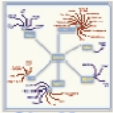
See the [press release](#) for more details.

**Other Species**

**Browse a Chromosome**



**Ensembl Links and Site Map**

 **Site Map**

**Help**

Click on any help icon to pop up a context-sensitive help window.

Questions or suggestions? Try our [HelpDesk](#).

Visitors to the Ensembl website have free and complete access to all the mouse and human genome data. [www.ensembl.org](http://www.ensembl.org)

looking at a magazine in a foreign language and trying to sort out what is news and what is advertisements.” *Ensembl* automatically estimates where the genes are, primarily by finding similarities with known genes — either in the mouse or in other organisms, such as humans.

Genes are especially difficult to find when you’ve only got one genome to work with, but when you’ve got two, it’s much easier — rather like piecing together an ancient manuscript from two partial copies. Using tools developed by *Ensembl* to search the human genome for sequences that look like mouse genes, the consortium uncovered 1200 new human genes. Similar methods, using sequences from many different organisms, were used to estimate the total number of genes in the mouse. By piecing together both the mouse and the human genomes in this way, the consortium has found that an amazing 99% of our genes are shared with the mouse.

What, then, underlies the difference between a mouse and a man? To get a handle on this question, the consortium used two other tools that are created and maintained by scientists at the EBI: InterPro, a database of functional units in proteins, and the Gene Ontology (GO), a dictionary that allows biologists to describe what proteins do in a consistent way. Searching the mouse genome with InterPro and GO is rather like searching an enormous pile of CVs to find out which candidates have the right qualifications to do a particular job. When they did this, the team found that mice have more genes concerned with smell, reproduction and immunity than humans.

As well as providing an invaluable resource for medical research, the mouse genome provides some tantalizing insights into our evolutionary origins. Nick Goldman, a researcher at the EBI, was involved in the comparative analysis of the mouse and human genomes.

One particularly exciting finding is that mice are evolving at about twice the rate of humans. In the past, some researchers had assumed that rodents were evolving at the same rate as humans, and this caused them to miscalculate the time at which the human and mouse lineages diverged. On the basis of molecular differences alone, they estimated this divergence to have taken place 100–130 million years ago. “For the first time, we have concrete proof that rodents are evolving faster than humans,” explains Goldman. “This supports research that has suggested that humans and mice are relatively closely related. A conservative estimate is that they actually diverged 75 million years ago, and some researchers would put it even later than that.”

Not only do species evolve at different rates, but different regions within an organism’s



Ewan Birney, coordinator of the Ensembl project at the EBI. Photo by EMBL Photolab



genome evolve at different speeds, as well. The study revealed that these variations are much larger than scientists expected. Every species' DNA steadily undergoes changes through mutations and other events. Usually these changes don't do any harm, because the vast majority of the genome consists of "junk". But some parts of the genome — such as genes — typically undergo fewer changes. Here, a single mutation could change the recipe by which a protein is built, and thus damage a molecule that plays a critical role in an organism's cells. Genes make up only about 2.5 percent of the mouse and human genomes. But very surprisingly, the current study revealed that about 5 percent of the mouse genome is evolving more slowly than can be explained by chance alone. This strongly suggests that parts of the DNA lying outside genes — another 2.5 percent of the genome — have important biological functions as well. Discovering their meaning will be a major goal for biomedical research in the coming years. These features might include protein-handling instructions, and information about how to pack and organize the DNA in the cell nucleus.

As well as these slowly evolving regions, some parts of the human and mouse genomes are evolving faster than expected. The vast amount of new data provided by the human and mouse genomes is making scientists reevaluate their previous understanding of the evolutionary process. "Our current mathematical models are inadequate to explain the kinds of differences that we see at the level of base-pairs between the mouse and human genomes," Goldman says. "New models of evolution are likely to result from this information."

The final goal of the Mouse Sequencing Consortium will be to produce a finished sequence that is richly annotated. This means that scientists who 'read' the genome will find a wealth of information about genes and other features that have been discovered in experiments. In the meantime, scientists are already making the most of the freely available draft sequence. The *Ensembl* team has worked hard to make the draft as easy to use as possible, including many features that allow quick cross-references between the human and mouse genomes. A 'beginner's guide' to surfing the *Ensembl* site can be found at [www.embl.de/ExternalInfo/oipa/t041202.pdf](http://www.embl.de/ExternalInfo/oipa/t041202.pdf).

"The availability of the mouse genome sequence is a real leap forward in making everyday lab work simpler, quicker and more effective," says Professor Azim Surani from the Wellcome Trust/Cancer Research UK Institute in Cambridge, UK. "Significant information can now be obtained by computer analysis. The sequencing of the whole genome has turned out to be of major significance for biomedical research."

### Websites:

*Ensembl* mouse: [www.ensembl.org/Mus\\_musculus/](http://www.ensembl.org/Mus_musculus/)

InterPro: [www.ebi.ac.uk/interpro](http://www.ebi.ac.uk/interpro)

GO: [www.geneontology.org](http://www.geneontology.org)

US National Center for Biotechnology Information

[www.ncbi.nlm.nih.gov/genome/guide/mouse](http://www.ncbi.nlm.nih.gov/genome/guide/mouse)

A comparison between the mouse sequence and the human sequence, which is expected to be completely finished in about a year, can be found at <http://genome.ucsc.edu> at the University of California, Santa Cruz, USA.

**About EMBL**

The European Molecular Biology Laboratory is a basic research institute funded by public research monies from 16 member states, including most of the EU, Switzerland and Israel. Research at EMBL is conducted by approximately 80 independent groups covering the spectrum of molecular biology. The Laboratory has five units: the main Laboratory in Heidelberg, Outstations in Hinxton (the European Bioinformatics Institute), Grenoble, Hamburg, and an external research programme in Mouse Biology in Monterotondo near Rome. The cornerstones of EMBL's mission are: to perform basic research in molecular biology, to train scientists, students and visitors at all levels, to offer vital services to scientists in the member states, and to develop new instruments and methods in the life sciences. The Laboratory also sponsors an active Science and Society programme. Visitors from the press and public are welcome. For more information see the EMBL website at:

<http://www.embl.de>

or contact:

the Office of Information and Public Affairs  
European Molecular Biology Laboratory (EMBL)  
Tel: +49 (0)6221 387252  
Fax: +49 (0)6221 387525  
email: [info@embl.de](mailto:info@embl.de)

The EMBL member states are:

Austria, Belgium, Denmark, Finland, France, Germany, Greece, Israel, Italy, the Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the United Kingdom

**About the EBI**

The European Bioinformatics Institute (EBI) is part of the European Molecular Biology Laboratory (EMBL), and is located on the Wellcome Trust Human Genome Campus in Hinxton near Cambridge (UK). The EBI grew out of EMBL's pioneering work in providing public biological databases to the research community. It hosts some of the world's most important collections of biological data, including DNA sequences (EMBL-Bank), protein sequences (SWISS-PROT and TrEMBL), animal genomes (Ensembl), three-dimensional structures (the Macromolecular Structure Database) and a new resource to hold data from microarray experiments (ArrayExpress). The EBI hosts several research groups and its scientists continually develop new tools for the biocomputing community.

contact:

Cath Brooksbank PhD  
Scientific Outreach Officer  
European Bioinformatics Institute, EMBL  
Hinxton Hall, Hinxton  
Cambridge CB10 1SD  
Great Britain  
Tel: +44 (0)1223 492525  
Fax: +44 (0)1223 494468  
email: [cath@ebi.ac.uk](mailto:cath@ebi.ac.uk)

[www.ebi.ac.uk/](http://www.ebi.ac.uk/)