

PublisherInfo		
PublisherName	:	BioMed Central
PublisherLocation	:	London
PublisherImprintName	:	BioMed Central

The mouse genome

ArticleInfo		
ArticleID	:	4652
ArticleDOI	:	10.1186/gb-spotlight-20021205-02
ArticleCitationID	:	spotlight-20021205-02
ArticleSequenceNumber	:	318
ArticleCategory	:	Research news
ArticleFirstPage	:	1
ArticleLastPage	:	3
ArticleHistory	:	RegistrationDate : 2002-12-5 OnlineDate : 2002-12-5
ArticleCopyright	:	BioMed Central Ltd2002
ArticleGrants	:	
ArticleContext	:	130593311

Jonathan B Weitzman

Email: jonathanweitzman@hotmail.com

In the December 5 [Nature](#) the Mouse Genome Sequencing Consortium reports the draft sequence of the mouse genome and an initial analysis of its treasures (*Nature* 2002, **420**:520-562). It is widely hoped that comparative genomic analysis will enhance our understanding of the human genome and human disease.

The [genome sequence](#) was generated from the assembly of over 40 million sequence reads, representing seven-fold coverage, from the C57BL/6J strain (B6) of *Mus musculus*. The draft contains almost 225,000 contigs, about 96% of the euchromatic genome, amounting to 2.5 Gb (making it 14% smaller than the human genome). There are over half a million orthologous landmark sequences in the two mammalian genomes, allowing the definition of extensive regions of conserved synteny. The mouse genome has increased numbers of lineage-specific repeat sequences, but less ancestral repeats than the human genome. Analysis of repeat sequences indicates that mice have a two-fold higher nucleotide substitution rate.

The number of protein-coding genes is roughly equivalent in mice and men, at around 30,000, and less than 1% of these have no ortholog in the other species. The catalog of predicted mouse and human genes includes 1,200 new genes, several of which are associated with human diseases. Manual inspection suggests that many of the computer-predicted genes may be pseudogenes. Comparative analysis was also useful for improving the identification of authentic non-coding RNA genes and tRNA gene sets.

The mouse genome has several cases of expanded gene families; for example, the large olfactory receptor gene family in mice probably reflects the important role these genes play in reproductive behaviour and the pheromone response. A search for mouse-specific gene clusters identified over a hundred examples of local expansions, including clusters of genes involved in rodent reproduction and host immunity.

About 40% of the human genome could be aligned with mouse sequences at the nucleotide level. The authors speculate that both genomes have undergone considerable deletion during recent evolution. Analysis suggests that around 5% of the genome is under evolutionary selection, which is much more than the coding regions alone. These regions may include regulatory sequences and non-coding RNAs. The neutral substitution rate is estimated at about half a nucleotide substitution per site, with considerable variation between chromosomal regions.

Analysis of single nucleotide polymorphisms (SNPs) identified over 79,000 SNPs that vary between mouse strains (about 1 per 500-700 bp). The SNP collection will provide a key tool for genetic analysis studies in mice. The mouse genome sequence information is expected to contribute significantly to positional cloning projects, analysis of quantitative trait loci and the creation of knock-out, knock-in and transgenic strains. Comparative genomics is likely to provide key insights into the human genome and proteome, and mammalian biology in general. This resource will also contribute to improvements in the study of mouse models for human diseases.

A related article published by *Genome Biology* describes a comparison of the mouse genome sequence with that available from Celera genomics - see [Xuan et al.](#)

References

1. *Nature*, [<http://www.nature.com>]
2. Mouse genome server, [http://www.ensembl.org/Mus_musculus/]
3. Xuan Z, Wang J, Zhang MQ: Computational comparison of two mouse draft genomes and the human golden path. *Genome Biol* 2002, 4:R1., [<http://genomebiology.com/2002/4/1/R1>]