



National Human Genome Research Institute
National Institutes of Health
Department of Health and Human Services
and
Office of Science
U.S. Department of Energy

Comparative Genomics

Sequencing the genomes of the human, the mouse and a wide variety of other organisms – from yeast to chimpanzees – is driving the development of an exciting new field of biological research called comparative genomics.

By comparing the finished reference sequence of the human genome with genomes of other organisms, researchers can identify regions of similarity and difference. This information can help scientists better understand the structure and function of human genes and thereby develop new strategies to combat human disease. Comparative genomics also provides a powerful tool for studying evolutionary changes among organisms, helping to identify genes that are conserved among species, as well as genes that give each organism its unique characteristics.

Using computer-based analysis to zero in on the genomic features that have been preserved in multiple organisms over millions of years, researchers will be able to pinpoint the signals that control gene function, which in turn should translate into innovative approaches for treating human disease and improving human health. In addition, the evolutionary perspective may prove extremely helpful in understanding disease susceptibility. For example, chimpanzees do not suffer from some of the diseases that strike humans, such as malaria and AIDS, even though chimps' DNA sequence is 98.8 percent identical to ours. A comparison of the sequence of genes involved in disease susceptibility may reveal the reasons for this species barrier, thereby suggesting new pathways for prevention of human disease.

In addition to its implications for human health and well-being, comparative genomics may benefit the animal world as well. As sequencing technology grows easier and less expensive, it will likely find wide applications in zoology, agriculture and biotechnology as a tool to tease apart the often-subtle differences among animal species. Such efforts might also possibly lead to the rearrangement of our understanding of some branches on the evolutionary tree, as well as point to new strategies for conserving rare and endangered species.

Although living creatures look and behave in many different ways, all of their genomes consist of DNA, the chemical chain that makes up the genes that code for thousands of different kinds of proteins. Precisely which protein is produced by a given gene is determined by the sequence in which four chemical building blocks – adenine (A), thymine (T), cytosine (C) and guanine (G) – are laid out along DNA's double-helix structure.

In order for researchers to use an organism's genome most efficiently in comparative studies, data about its DNA must be in large, contiguous segments, anchored to chromosomes and, ideally, fully sequenced. Furthermore, the data needs to be organized to allow easy access for researchers using sophisticated computer software to conduct high-speed analyses.

The successful sequencing of the human genome and the recent draft assemblies of the mouse and rat genomes have demonstrated that large-scale sequencing projects can generate high-quality data at a reasonable cost. As a result, the interest in sequencing the genomes of many other organisms has risen dramatically.

In addition to the human (*Homo sapiens*) and mouse (*Mus musculus*), organisms whose genomes have been sequenced include: rat (*Rattus norvegicus*); two fishes (*Takifugu rubripes* and *Tetraodon nigroviridis*), two fruit flies (*Drosophila melanogaster* and *D. pseudoobscura*); two roundworms (*Caenorhabditis elegans* and *C. briggsae*); yeast (*Saccharomyces cerevisiae*) and several other fungi; a malaria-carrying mosquito (*Anopheles gambiae*) along with a malaria-causing parasite (*Plasmodium falciparum*); two sea squirts (*Ciona savignyi* and *C. intestinalis*); a long list of microbes; and a couple of plants, including mustard weed (*Arabidopsis thaliana*) and rice (*Oryza sativa*). Sequencing work is well underway on the honeybee (*Apis mellifera*), and is just getting started or expected to begin soon on the chimpanzee (*Pan troglodytes*), the cow (*Bos taurus*), the dog (*Canis familiaris*) and the chicken (*Gallus gallus*).

The fledgling field of comparative genomics has already yielded dramatic results. For example, a March 2000 study comparing the fruit fly genome with the human genome discovered that about 60 percent of genes are conserved between fly and human. Or, to put it simply, the two organisms appear to share a core set of genes. Researchers have found that two-thirds of human genes known to be involved in cancer have counterparts in the fruit fly. Even more surprisingly, when scientists inserted a human gene associated with early-onset Parkinson's disease into fruit flies, they displayed symptoms similar to those seen in humans with the disorder, raising the possibility the tiny insects could serve as a new model for testing therapies aimed at Parkinson's.

NHGRI created a priority-setting process in 2001 to make rational decisions about the many requests being brought forward by various communities of scientists, each championing the animals used in its own research. The priority-setting process, which does not result in new grants for sequencing the organisms, is based on the medical, agricultural and biological opportunities expected to be created by sequencing a given organism. For more information on the status of model organisms in the priority-setting process, go to NHGRI's Web site at: www.genome.gov/10002154.

###