

Scientists compare 12 fruit fly genomes

November 7 2007



Image by Broad Communications

In one of the first large-scale comparisons of multiple animal genomes, scientists at the Broad Institute of MIT and Harvard, the Computer Science and Artificial Intelligence Laboratory (CSAIL) at MIT, and many collaborating institutions, have analyzed the genomes of twelve species of the fruit fly *Drosophila* to reveal insights on the evolution of genes and genomes and to discern the functional elements encoded in animal DNA.

The work appears in the November 8 issue of *Nature* and in more than 40 accompanying papers in *Genome Research* and other journals.

The method of comparing the genomes of multiple related species, fly or

otherwise, not only reveals new insights into species evolution and identifies thousands of novel genes and other functional elements, but also provides a powerful tool for unraveling genome function that may help researchers unlock the secrets of our own genome.

In these papers, the international consortium reported the genomes of ten newly sequenced *Drosophila* species, some very closely related and others less so, and their comparison to two previously sequenced flies including *Drosophila melanogaster*, one of the most powerful model organisms for the study of animal biology and evolution. The availability of the many *Drosophila* genomes has enabled a great deal of new insights about genome function and aided the study of how genomes have changed across evolutionary time.

“Having the sequences of many closely related species allows us to study the evolutionary forces that have shaped the fruit fly’s family tree, and to discover the working parts of the fly genome in a systematic way,” said Manolis Kellis, associate member of the Broad Institute, assistant professor in MIT’s CSAIL, and one of the consortium’s project leaders.

On one hand, the researchers studied the differences across species to help elucidate how evolution has shaped fly biology over millions of years. Their analysis revealed that while many attributes of *Drosophila* genomes are in fact conserved across multiple species, each species has novel features not seen in any other. In fact, only 77 percent of the approximately 13,700 protein-coding genes in *D. melanogaster* are shared with all of the other 11 species. For example, the genes involved in interactions with the environment and in reproduction showed signs of adaptive evolution, meaning that they likely provided some survival advantage to the organism.

On the other hand, the researchers studied the similarities of the different species to help define the functional parts of the fly genome.

The parts of a genome that are unchanged (conserved) are those that have been kept by evolution, and are thus likely to play crucial roles. Thus, genome comparison can reveal which regions of the genome are functional, based on the degree to which evolution has conserved them.

“Focusing on the conserved part of the genome is a great way to discover what has been maintained by evolution,” said Kellis. “Moreover, by looking more closely at the subtle patterns of mutation within conserved regions, we can predict the functional roles they play.”

Indeed, at the level of DNA, several combinations of letters, or nucleotides, may encode the same function, in the way that a storyteller can use different combinations of words to tell the same tale. For example, four different nucleotide combinations – GTT, GTC, GTA, and GTG – all encode the same protein building block, or amino acid. Thus, a change in the third letter would leave the amino acid unchanged, one example of how DNA changes can be tolerated while still preserving the function of the corresponding protein.

Through these kinds of random mutations, evolution explores the space of possible nucleotide combinations that preserve function. This exploration produces unique patterns of genomic change, described by the researchers as “evolutionary signatures” that are specific to the function of that region of DNA. Protein-coding genes, for example, show frequent substitutions at every third nucleotide, due to the fact that one amino acid can be encoded by several nucleotide triplets. In contrast, some genes that don’t encode proteins — so-called RNA genes — show changes that preserve the overall structure of RNA while tolerating changes in the genes’ DNA sequence.

Like codebreakers turning their knowledge of biology into computational algorithms, Kellis and his colleagues identified evolutionary signatures associated with a variety of roles in the genome:

protein-coding genes, non-coding RNAs, microRNAs, and regulatory motifs. In each case, the researchers identified distinct evolutionary signatures associated with each function, based on the tolerated changes that still preserve that function.

The researchers then used these evolutionary signatures to systematically identify the functional elements encoded in the fly genome, leading to hundreds of novel functional elements and many new insights on animal biology.

The work allowed the discovery of 1,193 new sequences that encode proteins, the flagging of 414 regions that were mistakenly labeled as protein-coding genes, and corrections to hundreds of previously annotated protein-coding genes. This allowed the researchers to revise the catalog of protein-coding genes for *Drosophila melanogaster*, with updates affecting 10% of all genes. The revision was confirmed through manual curation by scientists at the FlyBase consortium and through large-scale experimental validation led by the Berkeley *Drosophila* Genome Project.

In addition, the researchers identified hundreds of new RNA genes and structures, new microRNA genes, and new DNA sequences involved in the control of gene expression during embryo development and environmental changes. The twelve genomes also allowed the prediction of very small regulatory targets in the genome, which can help piece together the first regulatory network for an animal genome without having to perform intense and expensive experiments.

The work also led to many surprises. For example, the researchers found many protein-coding genes that defy the traditional rules of how the DNA code gets translated into protein. For example, 150 genes apparently bypass signals that would normally cause DNA to stop being translated, and other genes encode multiple proteins in a single RNA

transcript. Other findings include surprising evidence that a single microRNA gene locus can produce up to four functional microRNAs, each with distinct functions.

The team's analysis is the first time that such a diverse range of evolutionary signatures has been applied to identify the functional elements of a genome in a comprehensive way. "By comparing many closely related genomes, we were able to discover things we never thought were possible using one genome sequence alone," said Kellis. One intriguing possibility is that evolutionary signatures may even identify novel, yet unknown classes of functions. For example, although the fruit fly has been intensely studied for over a century, microRNAs were only discovered in the last decade, and are now known to play a central role in development. Many other classes of yet unknown functional elements may be hidden in the fly genome, and recognition of their common evolutionary properties may help lead to their discovery.

The study of the 12 flies has immediate implications for the discovery of functional elements in the human genome. "We are now using similar methods to analyze 32 mammalian genomes, in order to help understand the human genome," Kellis explained. "We should be able to apply the methodology of evolutionary signatures to any group of closely related species." Peering into the past and interpreting clues carved in the genome by evolution is yet one more way to make revelations about human biology. As the genome sequences of more organisms become available, the power to make discoveries about functions encoded in the genome will likely continue to increase.

On the whole, genome sequencing projects have given us a glimpse of the incredible variety of life, recording the genetic plans of organisms as wide-ranging as bacteria, algae, insects, and mammals and exposing common genes and functions conserved by evolution. The approach of sequencing many close relatives on the family tree of life provides a rare

view of the precise workings of evolution, giving scientists the tools to decipher the secrets hidden in our genome.

Papers cited:

Drosophila 12 Genomes Consortium. (2007) Evolution of genes and genomes in the Drosophila phylogeny. Nature
DOI:10.1038/nature06341.

Stark et al. (2007) Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. Nature
DOI:10.1038/nature06340.

Lin et al. (2007) Revisiting the protein-coding gene catalog of Drosophila melanogaster using twelve fly genomes. Genome Research
DOI:10.1101/gr6679507.

Stark et al. (2007) Systematic discovery and characterization of fly microRNAs using 12 Drosophila genomes. Genome Research
DOI:10.1101/gr6593807.

Stark et al. (2007) Reliable prediction of regulator targets using 12 Drosophila genomes. Genome Research DOI:10.1101/gr7090407.

Rasmussen, Kellis. (2007) Accurate gene-tree reconstruction by learning gene- and species-specific substitution rates across multiple complete genomes. Genome Research DOI:10.1101/gr7105007.

Ruby et al. (2007) Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs. Genome Research DOI:10.1101/gr6597907

Source: Broad Institute of MIT and Harvard, by Leah Eisenstadt

Citation: Scientists compare 12 fruit fly genomes (2007, November 7) retrieved 26 April 2024 from <https://phys.org/news/2007-11-scientists-fruit-genomes.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.