

AUGUST 30, 1999

Genome Tools Pass Big Test in Fruit Flies

With the effort to sequence the entire human genome speeding toward completion, some researchers are now focusing their energy on developing the next generation of tools that can be used to extract valuable scientific information from the unabridged human genetic sequences.

Howard Hughes Medical Institute investigators, Allan Spradling at the Carnegie Institution of Washington, and Gerald Rubin at the University of California, Berkeley, and more than two dozen colleagues have developed and used several types of tools to analyze the genome of the fruit fly *Drosophila melanogaster*, which has provided a treasure trove of information about genes and their function.

"Our goal was not to identify a function for every gene, but to show that it was feasible to generate large numbers of P element mutations randomly throughout the *Drosophila* genome and to use them to find new genes. We did that."

— Gerald M. Rubin

"When the Human Genome Project started eight years ago, the organizers had the foresight to sequence a number of other organisms to serve as models and as interpretive guides to the much larger human genome," said Rubin.

"The biggest problem in the post-genomic era will be determining what each of the thousands of human genes does," said Spradling. "Fortunately, it has become increasingly clear that many gene functions have been conserved during evolution, and that crucial insights into gene function can often be gained by studying the genomes of experimentally favorable, non-human organisms."

In 1982, Spradling and Rubin discovered how to use a transposable element, a piece of DNA that can jump from place to place in the genome, to engineer changes in the fly's genome. By 1988, Spradling's group found a way to turn

this piece of DNA the P element into a powerful tool for determining gene function. The idea, says Spradling, was to create large numbers of mutant fruit flies, each containing one P element inserted within a different gene. The observable characteristics of each such mutant strain would help indicate the function of the particular gene that contained the single inserted P element.

Many laboratories have been using this technique to generate mutant flies, which Rubin and Spradling have collected under the aegis of the Berkeley *Drosophila* Genome Project (BDGP), a program backed by HHMI. Integrating the mutant fly strains within the genome project has greatly enhanced their value, for example, by determining which gene is disrupted in each strain and by eliminating redundant or damaged strains. BDGP makes the fly strains available to researchers around the world, and scientists have used the flies to characterize more than 250 *Drosophila* genes.

In a research article published in the September, 1999 issue of the journal *Genetics*, Rubin, Spradling and their colleagues present the results of generating and characterizing mutant fly strains that contain P element disruptions in 1,052 different genes, more than 25 percent of *Drosophila's* 3,600 vital genes. The researchers examined each fly strain for obvious changes from wild-type flies, including phenotypes such as lethality, near-lethality and other readily apparent physical abnormalities. Thus, says Spradling, the current collection focuses on genes that have a major effect on the survival and appearance of the adult organism, rather than subtle influences on its behavior and metabolism.

"Our goal was not to identify a function for every gene, but to show that it was feasible to generate large numbers of P element mutations randomly throughout the *Drosophila* genome and to use them to find new genes," said Rubin. "We did that." More sophisticated examinations for behavioral deficits or biochemical changes should turn up functions for even more genes since mutations in only about a third of *Drosophila* genes show phenotypes we would have recognized in the current study, he added. Since the current study showed that P elements can target a wide range of *Drosophila* genes, the researchers now believe that their method can be used to determine the function of essentially all of the 12,000 genes that are thought to reside in the *Drosophila* genome. Rubin and Spradling are now planning to greatly expand the P element approach.

Of course, the P element disruption technique is not the only tool that researchers are developing to analyze genomes. In a second research article that will appear in the same issue of *Genetics*, Rubin, Spradling and more than two dozen colleagues at several institutions in the United States and England, used a wide variety of tools to probe a 2.9 million base pair region of the *Drosophila* genome.

"We chose a region of the *Drosophila* genome that had been characterized in substantial detail already from a genetic perspective, partly through the use of P elements, and for which we now have a sequence. The idea, then, was to apply all the available tools to this wealth of genetic and molecular

information in an attempt to understand this piece of DNA what genes are there, what they do, how are they organized as completely as possible," said Rubin.

"One very interesting finding for evolutionary biology came out of this analysis," explained Rubin, "Genes with mutant phenotypes are far more likely to have counterparts in other organisms, including humans, than are genes with no known mutant phenotype."

Their analysis also provided a test for the tools being developed by computational biologists. For example, researchers have obtained gene sequence information from dozens of organisms and have drawn some general conclusions about what those genes "look" like when buried within the millions of consecutive As, Ts, Gs, and Cs, the four bases that make up an organism's chromosomes.

Computational biologists have used this information to create software programs that scan large stretches of raw sequence data the exact order of the four bases in DNA and predict where functional genes might lie within the large stretches of DNA that have not been well studied. Members of the BDGP's informatics team organized a workshop at a recent international meeting of computational biologists where they compared the experimentally generated data with the predictions generated by several different software programs.

"This kind of competition showed that some programs worked better than others, but the important outcome is that people will be able to take these results and improve their software," said Rubin. "We will have a better set of computational tools as a result."