

MARCH 24, 2000

## *Drosophila* Genome Sequence Completed

The common fruit fly, *Drosophila melanogaster*, has been the workhorse of biology and genetics laboratories for the past 90 years. Now the entire *Drosophila* genome has been sequenced through the collaborative effort of researchers from the *Drosophila* Genome Project Group, led by Howard Hughes Medical Institute (HHMI) vice president Gerald Rubin at the University of California Berkeley, and researchers led by J. Craig Venter at the Celera Genomics Corporation.

The *Drosophila* genome sequence was published in the March 24, 2000, issue of *Science*. The researchers report that they have sequenced 97 to 98 percent of the genome and perhaps 99 percent of the estimated 13,600 genes. The sequence data will be accessible to scientists worldwide through Genbank, the National Institutes of Health genetic sequence database.

---

"If you give people very efficient tools for figuring out the functions of genes, you can do it in a massively parallel way."

- Gerald M. Rubin

---

In an accompanying editorial in *Science*, Thomas Kornberg at the University of California, San Francisco, and HHMI investigator [Mark Krasnow](#) at Stanford University, report that the *Drosophila* sequence will be a "critical resource" for research in genetics, biology and medicine.

Over the years, *Drosophila* has been one of the most influential model systems for geneticists. "The conservation of biological processes from flies to mammals extends the influence of *Drosophila* to human health," write Kornberg and Krasnow. "When a *Drosophila* homology of an important but poorly understood mammalian gene is isolated, the arsenal of genetic

techniques in the *Drosophila* system can be applied to its characterization."

The *Drosophila* sequencing project was launched in 1991 when Rubin and HHMI investigator [Allan Spradling](#) at the Carnegie Institution decided, says Rubin, that the time was right to begin a fly genome project. In May 1998, the Berkeley *Drosophila* Genome Project was one year into a three-year NIH grant and had finished 20 percent of the sequencing, when Rubin was approached by Venter with what Rubin calls "an offer that was too good to turn down."

Venter proposed that his newly-formed company, Celera, would sequence the *Drosophila* genome free-of-charge using a controversial technique known as whole genome shotgunning. The technique requires shearing the *Drosophila* DNA into three million random clones with overlapping ends. These clones are then sequenced by automated DNA sequencing machines—at Celera, some 300 sequencers, each costing \$300,000—and then massive computing power is put to work to assemble the complete genome sequence in a process similar to reconstructing a jigsaw puzzle.

Venter formed Celera with backing from PE Corporation (formerly known as Perkin-Elmer Corporation), which makes the DNA sequencing machines, as a commercial venture to sequence the human genome by 2001, several years before the date projected for completion by the international Human Genome Project. While promising the data would be made available to researchers, Venter was also betting that Celera could make money by licensing early looks at the sequencing data to the pharmaceutical industry.

The *Drosophila* genome, says Mark Adams, Celera's vice president for genome programs, would be "a proof-of-principle" for the whole genome shotgun strategy. "It seemed like a good idea to do a medium-sized organism in which there was extensive scientific interest," he says, "and in which there was already a lot of good information available in terms of map and sequence data that we could use to validate the strategy."

While Rubin says he had some concern about working with Celera, he was delighted by the offer nonetheless. "Anyone who would help me get the *Drosophila* sequence done and out of the way was my friend," says Rubin. "They were offering to do all this work in a collaborative way and not expecting any money for it."

Celera started the sequencing last April and finished collecting the raw data in early September. "Since then," says Rubin, "we've been putting all the

pieces together, which is not trivial. It's the big challenge of the whole genome shotgun approach."

The finished genome already seems to be remarkably revealing. Of the 289 genetic flaws known to cause disease in humans, says Rubin, they have found *Drosophila* homologues for 60 percent and for 70 percent of the genes involved in human cancers. Among the genes that have already been identified are *Drosophila* homologues of genes involved in Parkinson's disease, and the long-sought *Drosophila* homologue of the *p53* tumor suppressor gene, which is implicated in a host of human cancers.

The biggest surprise to come out of the *Drosophila* sequencing project, says Rubin, is that flies have only twice as many genes as yeast. "Yeast is a simple, single-cell fungus," says Rubin, "and yet flies only need twice as many genes to make an animal that can fly around without crashing into walls, has tissues, nerves, muscles, memories and other kinds of complicated behaviors like circadian rhythms. The take-home message is that the higher complexity in animals like flies and humans comes without needing a lot of new parts. You can build them with the same parts list—with more of the same parts organized together—in much the same way a supercomputer can be built from a bunch of desktop PCs hooked together in parallel."

Rubin sees the genome drastically changing the pace of his research. With less than 15,000 genes in *Drosophila*, and some 5,000 researchers worldwide working on the organism, he says, "that's one human being for every three genes. If you give those people very efficient tools for figuring out the functions of genes, you can do it in a massively parallel way." Moreover, the full *Drosophila* sequence allows researchers to look at multiple genes simultaneously to understand the complex signal transduction pathways that regulate cellular processes. "That is where the genome project really comes into play," he says. "It enables us to know all the genes so we can look at all of them at once and see what they're doing."

At the Princess Margaret Hospital in Toronto, researcher Tak Mak says he has been working to understand the signal transduction pathways involved in cancer formation. "The easiest way to understand that would be some kind of a genetic screen." As a result he has recently dedicated one-third of his laboratory to *Drosophila* genetics in anticipation of the publication of the sequence. "It will make *Drosophila* genetics relatively easy," he says.

Whether the whole genome shotgun technique will work as impressively for the human genome is now the next question. Celera's Adams says the *Drosophila* work is obviously encouraging, and that Celera's human

sequencing work has already begun and should "start to look like a genome" toward the end of the year. Rubin says, "It worked better in *Drosophila* than most people expected it would. I think it will work for humans. But the problems are more complex for humans, so we'll have to wait and see."