

MARCH 04, 2009

With Genomes, Bigger May Really Be Better

Biologists analyzing DNA in search of the molecular underpinnings of life have consistently favored species with small genomes, which are cheaper to sequence and lack the repetitive “junk” that clutters bigger genomes. But a new study by Howard Hughes Medical Institute scientists suggests that when it comes to figuring out how genes are controlled, bigger genomes are much more useful.

Animal genomes vary tremendously in size; worms have as few as 70 million “letters” of DNA, whereas salamanders have more than 100 billion. In a research article published in *Public Library of Science (PLOS) One* on March 4, 2009, Howard Hughes Medical Institute investigator Michael B. Eisen and colleagues report that large genomes can make it easier to find regions of DNA that control gene activity. “In small genomes, functional elements are packed tightly together. In bigger genomes functional elements are separated and therefore easier to find,” says Eisen, who collaborated on the study with scientists at the University of California, Berkeley, the University of Arizona, and the Pacific Basin Agricultural Research Center of the U.S. Department of Agriculture.

A genome is like a recipe for a meal that comes with two sets of instructions. One set shows how to make the ingredients for the meal -- the proteins that constitute living things. The second set shows how to measure, mix, and cook the ingredients -- that is, when and where proteins should be manufactured to carry out biological processes.

"While the idea that there is a fundamental difference in the complexity of vertebrate and invertebrate genomes fits with our anthropocentrism, it does not appear to be true."

- Michael B. Eisen

The first set of instructions is relatively easy to identify and read, but the second set has been more elusive. “We don’t understand how regulatory information is written in the genome,” says Eisen, “and in most cases we don’t even know where to look.”

Only a small fraction of the tens of thousands of regulatory sequences in the human genome have been identified. Most of these have emerged from studies comparing the human genome to those of mice, chickens, fish, and other vertebrates. Many of the small pieces of DNA shared by these distantly related species have proven to be involved in gene regulation.

To understanding the function of such regulatory sequences, Eisen and other geneticists have turned to model invertebrate species like the fruit fly *Drosophila melanogaster*. But the shortcut used to identify regulatory sequences in humans has never worked well in *Drosophila*. While comparisons among *Drosophila* genomes identify many shared sequences, the rapidly evolving DNA that separates these conserved sequences in vertebrates is largely absent in *Drosophila*, making it difficult to tell where one regulatory sequence ends and the next begins.

When Eisen and his Berkeley colleagues went hunting for regulatory sequences in the genomes of *Drosophila*’s distantly related fly cousins, they didn’t expect genome comparisons to be the key. But when graduate students Brant Peterson and Emily Hare compared pieces of the genomes of the medfly and the melon fly, two agricultural pests in the family Tephritidae, they noticed that these comparisons looked just like those seen in vertebrates.

The difference, Eisen says, is in the size of their genomes. *Drosophila* genomes are twenty times smaller than the human genome, and have been purged of non-functional DNA. But tephritid genomes are five times bigger than *Drosophila* genomes, and not nearly so streamlined.

“I’d love to say we chose the tephritids with this in mind, but it was totally serendipitous,” says Eisen. “The fact that the tephritids had big genomes was originally a nuisance because we had to do more sequencing and more screening. It was only after we got the data that we realized this might actually be an advantage.”

Based on earlier human work, Peterson hypothesized that the well separated blocks of conserved DNA in tephritids were regulatory sequences. Since there was no method available for testing these sequences in tephritids, Peterson inserted them into the laboratory mainstay *Drosophila melanogaster*. More than 150 million years of evolution separate tephritids from *Drosophila melanogaster*, but six of the nine pieces of conserved tephritid DNA functioned as regulatory sequences in the fruit fly. Furthermore, Peterson found matches for each of the tephritid sequences in the *Drosophila melanogaster* genome, and showed that the matched tephritid and *Drosophila* sequences drive the same patterns of gene expression.

Thus, it may be easier to identify regulatory sequences in the widely studied *Drosophila melanogaster* genome by sequencing and comparing tephritid genomes than sequencing more *Drosophila* genomes, Eisen says.

The findings have broader implications, too, Eisen says. Many biologists have been left with the impression that gene regulation is simpler in invertebrates than in vertebrates, since virtually all sequenced invertebrate genomes are small, with compact regulatory regions, and most sequenced vertebrate genomes are big. But Eisen points out that the sequenced invertebrate genomes are not representative. With limited funds available to study species not closely related to humans, and with the cost of genome sequencing scaling directly to genome size, the myriad invertebrate species with large genomes have been shunned.

“While the idea that there is a fundamental difference in the complexity of vertebrate and invertebrate genomes fits with our anthropocentrism,” says Eisen, “it does not appear to be true. It’s an illusion created by a bias towards sequencing small genomes whenever possible.”

Eisen is optimistic that observations from studies like this, together with the rapidly dropping cost of sequencing, will reverse this bias, allowing researchers to generate a clearer picture of the structure and evolution of animal genomes. To aid in that goal, he is working with scientists from the Department of Agriculture and Baylor College of Medicine to sequence complete tephritid genomes.