

## Water Flea Boasts Whopper Gene Count

Packed into a body no bigger than the letters on this page is a whale of a genome. The body belongs to *Daphnia pulex*, a crustacean common in lakes and ponds around the world. Since 2004, the Department of Energy Joint Genome Institute in Walnut Creek, California, and a consortium that now numbers 350 investigators from 17 countries have been sequencing and analyzing the 200-million-base genome from a *Daphnia* that lived in a pond along the Pacific coast of Oregon. It is one of just two noninsect arthropods to be deciphered to date.

At first glance, the genome seemed to have about 25,000 genes—a lot, but no record-breaker. Eventually, however, gene-finding programs found 31,000, John Colbourne of Indiana University, Bloomington, reported at the meeting. And a variety of experiments have revealed as many as 8000 more genes that gene-finding annotation programs missed, he said. That tops the gene count of the newly sequenced genome of another tiny creature: the pea aphid, which sports 34,600 genes (see p. 1253).

“It’s a big surprise that critters that you think wouldn’t have a high gene count do,” says Eric Green of the National Human Genome Research Institute in Bethesda, Maryland. These findings are further evidence that biological complexity does not directly correlate with gene number. But we are also “probably naïve in defining what is biological complexity,” Green adds.

Part of the appeal for sequencing the genome of *Daphnia* is its ability to adapt—it



**All-purpose genome.** An extraordinary number of genes may help *Daphnia* cope with diverse environments and predators.

usually clones itself but reproduces sexually under certain conditions. Eggs can hatch right away or lay dormant for more than a century. *Daphnia* thrive on algae and in turn are fodder for many fish and other predators, making them a key link in aquatic food webs. But, depending on the predator, they can sprout helmets, tail spines, or ridges called neck teeth. The genome is already helping researchers get to the genetic basis of this plasticity.

Colbourne and his colleagues first made a microarray of all the organism’s genome

sequence rather than just pieces of the genes themselves. “The array is blind relative to the annotation,” Colbourne explains, and thus can pick up expressed DNA that gene-finding programs might miss. They have started using the array to study how gene expression changes under different conditions. In experiments that looked at *Daphnia* exposed to different predators, for example, they found “a set of genes that were hidden,” Colbourne reported. “And we expect there are still more” to be unearthed as they evaluate different environmental conditions. The new genes they’re finding seem to code for proteins but not for any that look familiar.

Colbourne wonders whether this gene-packed genome arose in part because of the complexity of the aquatic environment—this species can live where it’s salty, acidic, hot, and so shallow that exposure to the sun is hazardous. “It’s almost as if they have more than one genome” to be able to cope with this diversity, he said. Another reason for the high gene count could be a consequence of having both sexual and asexual reproduction.

Whatever the cause, the high gene count comes from having many duplicated genes. This stands in contrast to the human genome, which generates a variety of different proteins by splicing genes into different configurations. Why the *Daphnia* genome evolved this way is a mystery, says John Werren of the University of Rochester, New York. “Some feature of the genome of *Daphnia* has pushed it to make more copies as opposed to [evolving] splicing.”

## Some RNA May Play Key Role in Repressing Genes, Slowing Cancer

Protein-coding genes have long been the stars of the Human Genome Project, but now RNA is moving into the limelight. Over the past 3 years, researchers have come to realize that protein-coding genes account for barely a quarter of the DNA that gets transcribed. The rest leads to RNA strands of various lengths—but toward what end has been a mystery, because that RNA doesn’t seem to lead to any proteins. Some experts have even argued that this RNA is little more than “transcriptional noise.” Yet, just as junk DNA proved to be more than junk,

at least some of this “noise” translates into meaningful molecules that may play key roles in turning genes on and off.

“For the past 5 to 10 years, researchers have been cataloging the presence of the noncoding RNAs,” says Thomas Gingeras, a molecular biologist at Cold Spring Harbor Laboratory in New York. “Now people want to understand what they do.” Chris Ponting of the University of Oxford in the United Kingdom and his colleagues took some of the first steps in that direction in 2007 by showing that 3000 long

noncoding RNAs were conserved in evolution, with sequences that were quite similar among mice, rats, and humans—an indication that they serve some vital function.

At the meeting, another team described progress in quantifying and assessing the function of a particular group of long RNA molecules. John Rinn of Harvard Medical School in Boston and his colleagues presented further evidence that at least some of these molecules seem to be important to a cell’s survival, and they reported that by studying the molecules these RNAs associate with, they are beginning to glean how some of them may actually work. One, for example, seems critical to helping the tumor suppressor gene *p53* keep cancer in

## The Bug and the Bacterium: Interdependent Genomes

Any successful relationship demands sacrifices. The partnership between the pea aphid and a tiny bacterium called *Buchnera aphidicola* is no exception. The newly sequenced DNA of this tiny insect, a common pest of legume crops, reflects a long history of give-and-take between the genomes of the bug and the bacterium. "The bargaining chips are genes, and the inventory reflects concessions during the course of negotiations," says John Colbourne, an evolutionary biologist at Indiana University, Bloomington.

Like other aphids, *Acyrtosiphon pisum* live off plant sap, a sugary mix low in protein. To make up for this nutritional shortfall, the insects depend on their microbial guests to supply essential amino acids. In return, the pea aphid has given up some of the genes that normally help fend off infections by Gram-negative bacteria such as *Buchnera*, Stephen Richards of Baylor College of Medicine in Houston, Texas, reported at the meeting. This loss "might account for the evolutionary success of aphids to obtain beneficial symbionts," reported aphid consortium collaborator Shuji Shigenobu of Princeton University.

*Buchnera* bacteria have tiny genomes, and genes number about 640. But they include key ones for providing the aphid with about nine amino acids that are missing from the sap that aphids feed on. A big surprise however,



**Sappy diets.** The aphid genome revealed that making essential amino acids was a joint venture with this plant pest's microbial partners.

found about 11. At least two are important to *Buchnera* for making the microbe's cell wall, and these are active in the nuclei of aphid cells specialized to house the microbes. Surprisingly, those genes didn't come from *Buchnera*, Richards reported: They appear to have come from a different type of microbe altogether, an alpha protobacteria.

The aphid was rife with duplicated genes, with an estimated 34,604 protein-coding genes in all, double the number in *Drosophila*. It also has several new genes, not known in other species, that code for saliva proteins that likely help keep plant juices flowing once the aphid has broken into the plant. "If you ask how we are going to control aphids on plants, this is the interaction that you have to stop," Richards says. "And now you have a molecular entry into that." **—E.P.**

check. "There are a lot more of these [RNAs], and they are probably more important than we thought," says Richard Myers of the Hudson-Alpha Institute for Biotechnology in Huntsville, Alabama.

Rinn and his colleagues study what he calls "large intervening noncoding RNAs" (lincRNAs), 2300 to 17,200 bases long, that are coded for in DNA between genes. Until recently, researchers knew of only about a dozen lincRNAs, notably XIST, an RNA that turns off the extra X chromosome in females, and HOTAIR, an RNA that directs the specialization of skin cells.

Rinn's team has searched systematically for lincRNAs by looking outside gene boundaries for chemical signatures that they know mark the coding regions of active genes. They initially looked in four types of mouse cells, assuming that such marks signaled additional transcription. The survey initially turned up about 1500 candidates, Rinn graduate student Mitchell Guttman and colleagues reported online 1 February in *Nature*.

At the meeting, Guttman reported that the team has expanded the search to 10 human tissues and has come up with 4000 definite lincRNAs. He estimates

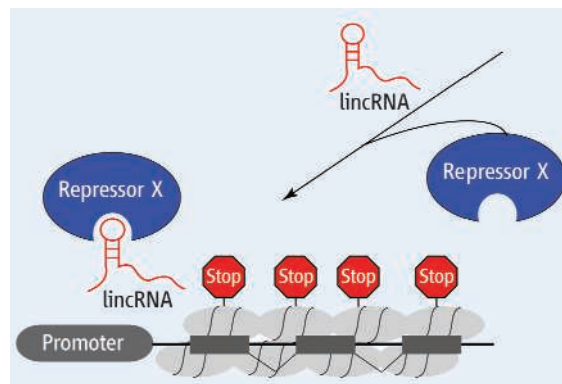
there are about 1000 more. Not everyone agrees that lincRNAs represent a true subclass of noncoding RNAs, as they worry that some of the DNA sequence encoding long noncoding RNAs may extend into genes, blurring the definition of "intervening." "Making a new class may be premature," Gingeras says.

Guttman, Rinn, and their colleagues have also looked for patterns of coexpression between protein-coding genes and lincRNAs in 21 tissues. They found quite a few gene-lincRNA overlaps, from which they concluded that, broadly speaking, lincRNAs are involved in the regulation of the cell cycle, immunity, and stem cell differentiation, Guttman reported. For example, 39 associ-

ated with the tumor suppressor gene *p53*. One "is directly regulated by *p53*," Rinn reported at the meeting. He thinks that particular lincRNA acts as a global repressor of the *p53* pathway, because 1000 genes increased their expression when he disabled either *p53* or that particular lincRNA.

The story sounded vaguely familiar, for that is how HOTAIR and XIST seem to work. So Rinn and Guttman did an experiment to assess how many lincRNAs bind to polycomb protein complexes. These complexes remodel chromatin, reconfiguring this DNA-protein matrix to shut out transcription factors and silence certain genes. They found that almost 25% of the lincRNAs latch onto these complexes in one cell type or another. Furthermore, the RNAs were required for gene silencing. In total, 38% of the lincRNAs are tied in with one of four chromatin-remodeling complexes, Rinn reported. "We think RNA is a scaffold to [help] bring in the right proteins," says Rinn.

"Those results are very exciting," says David Haussler, a bioinformaticist at the University of California, Santa Cruz. "This is far from being a mature scientific story, but there are tantalizing hints [of a repressive function]." Adds cancer geneticist Victor Velculescu of Johns Hopkins Kimmel Cancer Center in Baltimore, Maryland: "It reinforces the fact that these lincRNAs are likely to be important physiologically." **—ELIZABETH PENNISI**



**Gene stopper?** LincRNAs may bind and lead repressor proteins to promoters so as to silence genes.