

APRIL 18, 2003

An Unexplored Genomic Terrain in a Handful of Dirt

Howard Hughes Medical Institute researchers and their colleagues have literally unearthed a treasure trove of genomic information from ten newly identified viruses found in the monkey pit at the Bronx Zoo and other locations. The viruses are called mycobacteriophages and they infect a range of bacteria, including those that cause tuberculosis and leprosy.

The studies, which were undertaken with the aid of high school students from Pennsylvania and New York, have also uncovered evidence supporting the theory that mycobacteriophages undergo constant random genetic mixing in the wild. These free-flowing associations produce a mélange of recombinant viruses, with the weaker strains weeded out in survival-of-the-fittest competitions.

"Before, we were tempted to group microbial genes into those that are bacterial and those that are phage genes. Now, we might rethink that classification as a continuum of genes that are predominantly bacterial or phage."

— **Graham Hatfull**

The researchers reported their findings in the April 18, 2003, issue of the journal *Cell*. The research team included HHMI Professor Graham Hatfull at the University of Pittsburgh and HHMI investigator William Jacobs at the Albert Einstein College of Medicine.

The ten soil-dwelling mycobacteriophages selected for genomic sequencing came from zoos, backyards, and even the soil outside a tuberculosis sanitarium in India. Despite their humble surroundings, the phages may have quite a story to tell. "Phages represent the most abundant life form in the biosphere, with an estimated 10^{31} phage particles in the world," said Hatfull. "Our data indicate not only that this amazingly diverse population constitutes the largest unexplored reservoir of sequence information, but one that is extremely dynamic, swapping genes with many other organisms."

Hatfull and his colleagues enlisted the help of high school biology students taught by Jacobs's sister, Debbie Jacobs-Sera, who is a co-author of the article in *Cell*. “One student isolated a phage from the rosebush in her front yard and from a nearby barnyard; and another student found a phage in soil from the Bronx Zoo monkey pit,” Jacobs said. Jacob Falbo and Joseph Gross, two students from Jacobs-Sera's high school class in Latrobe, Pennsylvania, worked with Hatfull's group in Pittsburgh, while Lauren Keenan, a high school student from Pelham, New York, worked with Jacobs. Vanaja Kumar of the Tuberculosis Research Center in Chennai, India also collaborated on the work. All are listed as co-authors on the article in *Cell*.

Until now, scientists had only characterized the complete genomes of four mycobacteriophages. The genomic sequences of the ten newly isolated phages may offer important information to researchers, said Hatfull. “Of those phages that we had genomic sequence information for, the morphologies were rather similar to those that we'd seen previously,” he said. “However, when viewed under the electron microscope, the morphologies of these newly isolated phages were quite a bit more varied than we had imagined. These surprising findings should give us new information about the relationship between genomic information and viral morphologies.”

Hatfull's team turned up some interesting information when comparing the genomes of the ten phages. “We were somewhat surprised by the variety of genome lengths,” he said. “The previously sequenced mycobacteriophage genomes were about in the same range, but these new ones varied as much as threefold, and they didn't appear to fall into any discrete groups.” His group also found that the length of the phage genomes appeared to be statistically correlated with the percentage of the DNA bases guanine and cytosine—a mystery that will take some time to figure out.

The new phages exhibited an extraordinary genomic diversity, said Hatfull. “Some of these new phages are clearly very different in terms of their sequence information from any of the other phages that we've isolated,” said Hatfull. “That suggests that if we extrapolate to the huge group of phages as a whole, they are more diverse than we ever imagined they could be.”

The genetic diversity of the more than 1600 phage genes the researchers identified is just as eye-opening to Jacobs. “What's amazing is that we now have 14 phages sequenced, and there is only one gene that is common between two of the phages. And fifty percent of the genes in this total dataset aren't in any genomic database,” he said.

Hatfull added that the known genes they identified in the phages also presented surprises. “We found a significant number of genes that are related to genes previously seen in bacterial genomes or in other genomes, that hadn't previously been seen in phages,” he said. “Before, we were tempted to group microbial genes into those that are bacterial and those that are phage genes. Now, we might rethink that classification as a continuum of genes that are predominantly bacterial or phage.”

The researchers' findings also suggest how phages mix their genes to produce the widely varied "mosaic" structure of their genomes. One theory posits that the genomes possess specific "linker" sequences at which genes are snipped apart for recombination. However Jacobs, Hatfull and their colleagues found more persuasive genomic evidence for the "illegitimate recombination" theory, in which the phage genes undergo nearly random cutting and splicing, with the resulting phage genomes succeeding or failing based on their competitive evolutionary advantage.

"Most of the time you get genomic trash, garbage that is destined to go nowhere," Hatfull said. "But natural selection can choose the small number of recombinant genomes that are viable, that have the genes required for growth and give the phage a genome of appropriate size."

According to Jacobs, the phages may be able to teach scientists a thing or two about evolution. "Typically, in cells of higher organisms, cell division involves recombination of similar, or homologous, chromosomes. But phages just put genes together randomly, which makes them a fascinating model of evolution," he said.

The studies also raise intriguing questions about the influence that bacteriophages have on their bacterial hosts. "We had known that bacteriophages can be intimately involved in the pathogenesis of their bacterial hosts, such as the infamous pathogenic *E. coli*," said Hatfull. "But in these phages, a number of genes cropped up that raise the question of phage involvement with mycobacterial pathogenesis." Such genes include those that code for proteins that trigger immune responses in tuberculosis and leprosy, he said.

Another phage gene identified by the researchers resembles a human gene called *Ro* that is involved in the autoimmune disease lupus. The scientists detected sequences in one of the phages that resemble "alarm clock" sequences that phages can use to activate dormant bacteria, increasing the phage's chance of reproduction. The presence of alarm clock sequences raises the question of whether these phages are mimicking mechanisms usually involved in the control and activation of tuberculosis latency, said Hatfull.