



THE GENE WEAVERS

Viruses are often thought of as simple creatures. But their staggering diversity and genetic promiscuity could make them the most creative force in evolution, says **Garry Hamilton**.

When high-school students Joe Gross and Jake Falbo walked into Graham Hatfull's lab at the University of Pittsburgh five years ago, they had no idea what they were getting into. Hoping for a little hands-on scientific experience, the two were instead invited to participate in a mission that would have excited the likes of Linnaeus or Darwin. "I said to them, 'Sure you can do a project'," recalls Hatfull. "Why don't you go out and discover some new viruses?"

It was no joke. Gross and Falbo are among the co-authors of a now heavily cited *Cell* paper for their separate identification of two previously unknown viruses called bacteriophages, 'phages' for short, which infect bacteria¹. But unlike Linnaeus and Darwin, Hatfull's team doesn't have to embark on a far-flung expedition to collect and classify new creatures — they, and other virologists like them, come across them wherever they go. "All you need to do," says Hatfull, "is go and look under a rosebush and you'll find a phage that's unlike anything anyone's ever seen before."

In total, Hatfull and his squadron of phage-hunters have now isolated and sequenced

more than 40 phages, from settings as varied as the grounds of a tuberculosis clinic in India, the monkey house at New York's Bronx Zoo and, yes, even beneath a rosebush in Latrobe, 55 kilometres southeast of their Pittsburgh base. They have discovered that every phage they come across is almost completely different from the next. Findings such as these have prompted a huge change in how biologists see all viruses: the spurt has fuelled the suggestion that viruses have a large part to play in the evolution of life on Earth.

In from the cold

Viruses have long been viewed as nature's outsiders. As parasites that depend on a host cell for survival, they don't seem to have fully earned the stripes of a living organism. But they are actually far more abundant, diverse and complex than once thought. Recent calculations suggest there may be more undiscovered genes in the viral world — most belonging to phages — than in all other life forms combined. And this vast presence and diversity has profound effects on the rest of life. By shuttling genes into and out of their hosts, viruses seem to be a major driving force

in the evolution of higher organisms. Even within our own genome, genes that came from viruses are hard at work. This in turn has led to the realization that viruses probably play a major role in the ecological, biochemical and evolutionary processes that underlie the entire natural world. Rather than being the outsiders, viruses have emerged as perhaps life's most ubiquitous presence.

"I don't think you can look at any system now and leave viruses out of the equation," says Nicholas Mann, a microbiologist at the University of Warwick, UK.

Hints of the vast, unexplored viral world first emerged during the 1980s. Researchers at the University of Bergen in Norway, with the help of a new electron-microscopy technique, showed that viral concentrations in some aquatic habitats were up to 10 million times greater than previous estimates². Viral particles per millilitre ranged from 60,000, deep beneath the Barents Sea, to 254 million, from the surface waters of Germany's Lake Plüsee.

Since then researchers have discovered huge numbers of viruses wherever they have looked, from 2,000 metres below the surface of Earth to the sands of the Sahara Desert, from acidic hot

HIDDEN TALENTS

The discovery that we share the planet with countless viruses has led researchers to explore their strange talents.

What happens when this viral world throws its weight around, for example? By killing bacteria en masse, viruses are thought to play an important role not only in the control of microbial populations, but also in global geochemical processes such as the carbon cycle. Other research has shown that when viruses prey on blooms of marine algae, the volume of dimethyl sulphide released is sufficient to influence the formation of clouds¹².

And viruses seem to have other mysterious abilities. Many viruses carry and use a long list of genes formerly thought to be useful only in cells, such as those needed for

photosynthesis. They probably provide viruses with an evolutionary advantage while they are infecting cells, but it's not always clear how.

At the University of Nebraska in Lincoln, plant virologist James Van Etten has cultured an algae-infecting virus that carries a gene for hyaluronic acid, a molecule used by vertebrates as a joint lubricant, as well as viruses that make chitin, the protein used as a building block in insect exoskeletons¹³. "We know they use these genes," says Van Etten. "But we still have no idea what advantage they gain from having them."

Meanwhile, other recent finds have shattered the notion that all viruses are lean and simple. Most notable is the recently sequenced



M. HILL/ALAMY

Viruses thrive everywhere, from polar ice caps to hot springs.

mimivirus, an amoeba-infecting virus whose genome contains 1.2 million base pairs and more than 900 genes¹⁴. The mimivirus's DNA load, more than double that found in the smallest known

bacterial genomes, contains numerous genes long thought to be the exclusive property of cells, including genes for repairing DNA and manufacturing proteins. **G.H.**

springs to polar lakes. In total there are now thought to be some 10^{31} viral particles on the planet — an astronomical figure that one researcher recently described as 250 million light years' worth of viral genes laid end to end.

Then, in the mid-1990s, researchers started to grasp the dazzling variety that exists within this viral multitude. Biologists discovered that the known and cultured viruses represented just a fraction of their respective groups. From one cubic metre of seawater to the next, there was more genetic diversity in viruses than found in any other known group of organisms.

Constant reinvention

More recently, researchers led by microbiologist Forest Rohwer at San Diego State University, California, have developed a technique for extracting and sequencing unknown viral DNA en masse. In a series of studies that examined samples from seawater, marine sediment and human faeces, the researchers discovered that most viral groups had previously been completely undetected³. A kilogram of marine muck was found to contain up to a million different viral genotypes. In the human gut alone there may be as many as many as 1,200 distinct viruses. "Viral diversity is just way different from anything else," says Rohwer. "Every time we sequence it, most of everything we run into is unknown."

Hatfull and his colleague Roger Hendrix have made similar discoveries, with help from a team of high-school students and undergraduates. In the 40 or so phages they have found, roughly half of each genome consists of genes that have never been seen before — not in any of the cellular organisms that have been sequenced to date, not even in any other virus¹. And a similarly high percentage of

unique genes has been found in most other sequenced viral genomes — there are some 450 in the largest virus, the mimivirus (see 'Hidden talents'). This has led to the conclusion that the bulk of nature's genetic information resides in the genomes of viruses. "You could argue that this is nature's greatest genetic experiment," says Hatfull.

And viruses seem to be truly talented experimentalists. Although phages have long been known to cut and paste their genes together from different sources, Hatfull and Hendrix have shown just how rampant this shuffling is. By making a comparison of 14 different phages, the researchers detected spliced-in sequences that were no more than a single gene in length. Other segments were flanked by fragments of genes, as if they had been arbitrarily ripped from one virus and inserted into another. This suggests that, unlike cells, phages can combine bits of DNA even when there is no similarity between the sequences of the different DNA pieces.

Hatfull and Hendrix now argue — and many others concur — that viruses are continually and randomly recombining with whatever DNA they might encounter while infecting a cell, be it phage or host DNA. Thus cells are stewing pots where viruses are continually reinventing themselves thanks to a blind but highly creative process that is not only spitting out novel combinations of genes, but is also generating brand new genes possibly never before seen in nature. Success relies on the huge number of new viruses being created in the world — estimated to be some 10^{24}

per second. "Almost all of this would be junk, useless monsters," says Hendrix. "But it's happening often enough that the few that survive are still a significant number. It's darwinian evolution on a grand scale."

DNA dispersal

Hendrix also points out that the common structure of most phages — a capsule full of DNA and a hollow tail used for injecting that DNA into a cell — would be ideal for facilitating the creative process by requiring excess DNA to act as stuffing that enables the head to maintain its shape. Free from the pressures of selection, such non-essential DNA would have an extended opportunity to evolve into something useful.

The emerging picture suggests viral DNA can spread far and fast. Two years ago Rohwer and his colleagues reported that they had found the same sequence of viral DNA in 49 of 66 widely varied habitats — including the rumen of a cow from Idaho, a hot spring in California, the Antarctic Ocean and mucus from coral growing in the Caribbean Sea⁴. Analysis of 18 of these samples showed that the 533-base-pair sequence differed by no more than three nucleotides. Based on known phage mutation rates, this piece of viral DNA seems to have dispersed sometime in the past 2,000 years⁵.

It is possible that the same host resides in all of these habitats, or that a single virus can infect a wide range of hosts. But Rohwer and others, who have now found similar results for several sequences, believe it is more likely that fragments of viral DNA are being passed

"A viral group is like a super-organism — genetic information is shared between all these different viruses."
— Curtis Suttle

between viruses that share the same host.

“When you look at a group of viruses, such as the algal viruses, there seems to be a very, very small core of conserved genes,” says Curtis Suttle, a microbiologist at the University of British Columbia in Vancouver, Canada. “The rest is almost like a super-organism — a massive pool of genetic information that’s being shared among all these different viruses.”

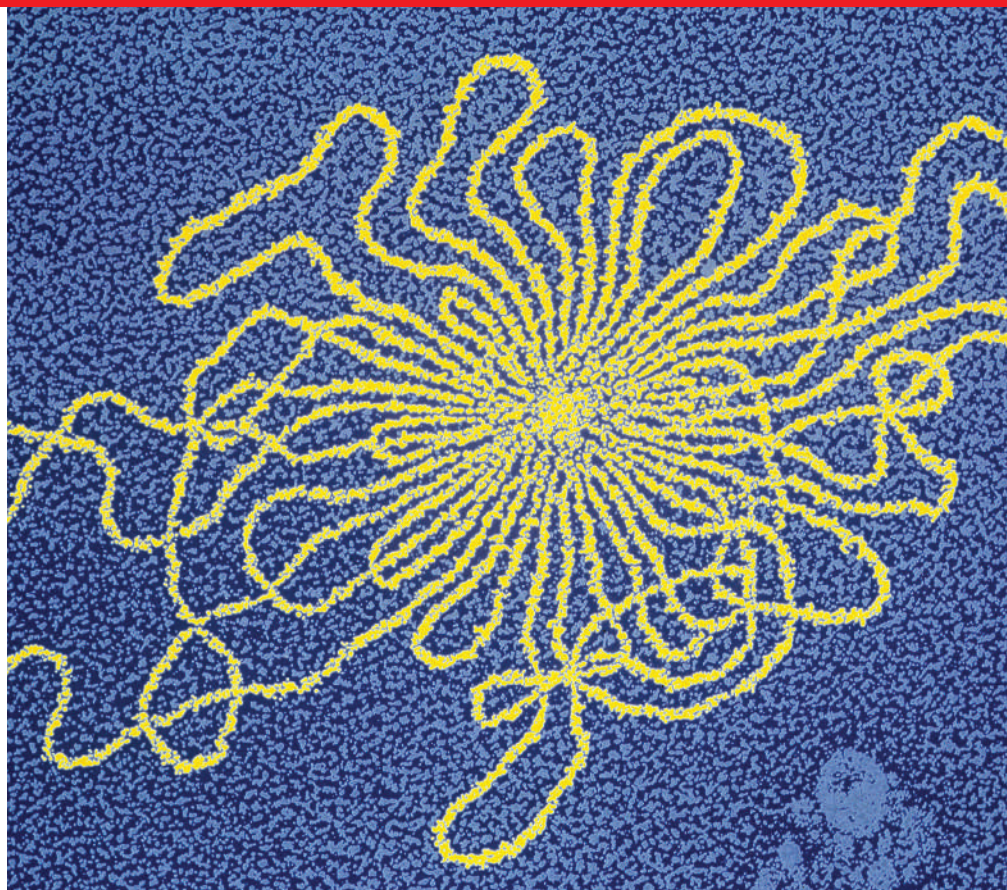
A big question now is the degree to which this super-organism extends its tentacles into host genomes. It has long been known that bacteria use genes acquired from prophages — phages that insert their DNA temporarily or even permanently into the DNA of their host — to gain competitive advantage and exploit new environments. Indeed, work reaching back decades has shown that prophage genes carried by bacteria are responsible for producing the primary toxins associated with diseases such as diphtheria, scarlet fever, food poisoning, botulism and cholera.

Behind bacteria

More recently, whole-genome sequencing has shown that most bacteria harbour an average of two or three prophages. Other studies have shown that phage genes make up the major differences between closely related strains of bacteria. In Japan, for example, researchers carried out whole-genome comparisons between a harmless laboratory strain of *Escherichia coli* and O157:H7, the strain of *E. coli* that has emerged in recent decades as a global health threat. They found that most of what separates the pathogenic strain from its close and harmless relative — almost a million base pairs of DNA — comes from 24 different prophage or prophage-like genetic segments⁶.

A more recent genome comparison between pathogenic and non-pathogenic isolates of *Neisseria meningitidis*, a normally harmless nasal-dwelling bacterium that is also known to cause meningitis, found only one noticeable difference between the two groups: a cluster of prophage genes that turned up in 29 of the disease-causing isolates but only 2 of the 20 benign bacteria⁷. This adds to a list of similar studies, such as one involving *Streptococcus pyogenes*, a resident of the skin and mouth that can cause a range of ailments including scarlet fever, rheumatic fever and toxic shock syndrome. In the study, researchers demonstrated how strains associated with three different disease processes each have their own distinct cocktail of phage-encoded toxins⁸.

But pathogenicity is just one trait that can affect bacterial evolution; scientists are beginning to find evidence that viral influence on the evolution of life may be a more general phenomenon. The bacterium *Pseudomonas aeruginosa*, for example, kills its competitors



Genetic cradle: this bacteriophage's DNA (yellow) could be a breeding ground for new genes.

with compounds produced by two modified phage genes⁹.

Scientists don't actually have to go as far as deadly bacteria, or even the nearest rosebush, to get new viral DNA. Human beings are full of it. Retroviruses are a type of virus that specialize in attacking animal cells. And roughly 8% of our genome consists of DNA copies of these RNA-based viruses that have incorporated themselves into our genetic make-up. One example is a protein used by viruses when they attach themselves to a host cell at the start of infection: it has recently been demonstrated to play an active role in binding together cells during development of the placenta¹⁰.

Evolutionary thoughts

Phage DNA has also made it into the human genome, despite the fact that phages target bacteria. The source was our mitochondria, the energy-producing capsules found in our cells and thought to be the descendants of a formerly free-living bacterium. During evolution, phage genes from the mitochondrial genome have transferred into our main genome, housed in the cell's nucleus. There, they help copy and express the few bacterial genes still present in the mitochondrial genome. Recently, this same phage DNA has been spotted in various modern bacteria belonging to the group from which mitochon-

dria are thought to have descended¹¹. That supports the idea that the gene made its way from virus to bacterium to cell nucleus, where it now plays a key role in the molecular circuitry that drives all multicelled organisms.

Some researchers now believe that viruses have been instrumental in assembling the various molecular components that define the cell types associated with life's three domains — bacteria, archaea and eukaryotes. Luis Villarreal, director of the Center for Virus Research at the University of California, Irvine, even argues that a large portion of what distinguishes humans from chimps is viral DNA.

“I think it's become apparent that viruses are involved everywhere,” says Villarreal. “I would argue they are the most creative genetic entities that we know.”

Garry Hamilton is a science writer based in Seattle, Washington.

1. Pedulla, M. L. *et al. Cell* **18**, 171–182 (2003).
2. Bergh, Ø. *et al. Nature* **340**, 467–468 (1989).
3. Breitbart, M. *et al. Proc. Biol. Sci.* **271**, 565–574 (2004).
4. Breitbart, M. *et al. FEMS Microbiol. Lett.* **236**, 249–256 (2004).
5. Breitbart, M. & Rohwer, F. *Trends Microbiol.* **13**, 278–284 (2005).
6. Hayashi, T. *et al. DNA Res.* **8**, 11–22 (2001).
7. Bille, E. *et al. J. Exp. Med.* **201**, 1905–1913 (2005).
8. Beres, S. B. *et al. Proc. Natl Acad. Sci. USA* **99**, 10078–10083 (2002).
9. Nakayama, K. *et al. Mol. Microbiol.* **38**, 213–231 (2000).
10. Mallet, F. *Proc. Natl Acad. Sci. USA* **101**, 1731–1736 (2004).
11. Filee, J. & Forterre, P. *Trends Microbiol.* **13**, 510–513 (2005).
12. Fuhrman, J. A. *Nature* **399**, 541–548 (1999).
13. De Angelis, P. L., Jing, W., Graves, M. V., Burbank, D. E. & Van Etten, J. L. *Science* **278**, 1800–1803 (1997).
14. Raoult, D. *et al. Science* **306**, 1344–1350 (2005).

“Viral diversity is just way different from anything else. Every time we sequence it, we run into the unknown.”
—Forest Rohwer