

FALSE-COLOUR ELECTRON MICROGRAPHS (NOT TO SAME SCALE). TOP ROW L-R: SMALLPOX VIRUS; ACIDIANUS BOTTLE-SHAPED VIRUS; ACANTHAMOEBA POLYPHAGA MIMIVIRUS. CENTRE ROW L-R: RABIES VIRUS; T4 BACTERIOPHAGE; ROTAVIRUS. BOTTOM ROW L-R: EBOLA VIRUS; TOBACCO RATTLE VIRUS; HIV-2. SPL. M. HARING ET AL./J. VIROL.; E. GHIGO ET AL./PLOS PATHOG.; FREDERICK A. MURPHY/CDC GLOBAL

Viruses come in all shapes and sizes, such as the giant mimivirus (top right) and the lunar-lander-shaped bacteriophage (centre).

# The incredible diversity of viruses

They're everywhere virologists look, and they're not all bad. Scientists are beginning to identify and classify the nonillions of viruses on the planet and their contributions to global ecosystems.  
**By Amber Dance**



**M**ya Breitbart has hunted novel viruses in African termite mounds, Antarctic seals and water from the Red Sea. But to hit pay dirt, she has only to step into her back garden in Florida. Hanging around her swimming pool are spiny-backed orbweavers (*Gasteracantha cancriformis*) – striking spiders with bulbous white bodies, black speckles and six scarlet spikes that make them look like a piece of medieval weaponry. Even more striking for Breitbart, a viral ecologist at the University of South Florida in St Petersburg, was what was inside. When she and her colleagues collected a few spiders and ground them up, they found two viruses previously unknown to science<sup>1</sup>.

Although we humans have been focused on one particularly nasty virus since early 2020, there are legions of other viruses out there waiting to be discovered. Scientists estimate that there are about  $10^{31}$  individual viral particles inhabiting the oceans alone at any given time – 10 billion times the estimated number of stars in the known Universe.

It's becoming clear that ecosystems and organisms rely on viruses. Tiny but mighty, they have fuelled evolution for millions of years by shuttling genes between hosts. In the oceans, they slice open microorganisms, spilling their contents into the sea and flooding the food web with nutrients. "Without viruses," says Curtis Suttle, a virologist at the University of British Columbia in Vancouver, Canada, "we would not be alive."

There are just 9,110 named species listed by the International Committee on Taxonomy of Viruses (ICTV), but that's obviously a pitiful fraction of the total. In part, that's because officially classifying a virus used to require scientists to culture a virus in its host or host cells – a time-consuming if not impossible process. It's also because the search has been biased towards viruses that cause diseases in humans or organisms we care about, such as farm animals and crop plants. Yet, as the COVID-19 pandemic has reminded us, it's important to understand viruses that might jump from one host to another, threatening us, our animals or our crops.

Over the past ten years, the number of known and named viruses has exploded, owing to advances in the technology for finding them, plus a recent change to the rules for identifying new species, to allow naming without having to culture virus and host. One of the most influential techniques is metagenomics, which allows researchers to sample the genomes in an environment without having to culture individual viruses. Newer technologies, such as single-virus sequencing, are adding even more viruses to the list, including some that are surprisingly common yet remained hidden until now. It's an exciting

time to be doing this kind of research, says Breitbart. "I think, in many ways, now is the time of the virome."

In 2020 alone, the ICTV added 1,044 species to its official list, and thousands more await description and naming. This proliferation of genomes prompted virologists to rethink the way they classify viruses and helped to clarify their evolution. There is strong evidence that viruses emerged multiple times, rather than sprouting from a single origin.

Even so, the true range of the viral world remains mostly uncharted, says Jens Kuhn, a virologist at the US National Institute of Allergy and Infectious Diseases facility at Fort Detrick, Maryland. "We really have absolutely no idea what's out there."

### Here, there and everywhere

All viruses have two things in common: each encases its genome in a protein-based shell, and each relies on its host – be it a person, spider or plant – to reproduce itself. But beyond that general pattern lie endless variations.

There are minuscule circoviruses with only two or three genes, and massive mimiviruses that are bigger than some bacteria and carry hundreds of genes. There are lunar-lander-looking phage that infect bacteria and, of course, the killer spiky balls the world is now painfully familiar with. There are viruses that store their genes as DNA, and others that use RNA; there's even a phage that uses an alternative genetic alphabet, replacing the chemical base A in the standard ACGT system with a different molecule, designated Z.

Viruses are so ubiquitous that they can turn up even when scientists aren't looking for



**We really have absolutely no idea what's out there."**

them. Frederik Schulz did not intend to study viruses as he pored over genome sequences from waste water. As a graduate student at the University of Vienna, in 2015 he was using metagenomics to hunt for bacteria. This involves isolating DNA from a whole mix of organisms, chopping it into bits and sequencing all of them. A computer program then assembles the bits into individual genomes; it's like solving hundreds of jigsaw puzzles whose pieces have been jumbled up.

Among the bacterial genomes, Schulz couldn't help but notice a whopper of a virus genome – obvious because it carried genes for a viral shell – with a remarkable 1.57 million base pairs<sup>2</sup>. It turned out to be a giant virus,

part of a group whose members are large in terms of both genome size and absolute size (typically, 200 nanometres or more across). These viruses infect amoebae, algae and other protists, putting them in a position to influence ecosystems both aquatic and terrestrial.

Schulz, now a microbiologist at the US Department of Energy Joint Genome Institute in Berkeley, California, decided to search for related viruses in metagenome data sets. In 2020, in a single paper<sup>3</sup>, he and his colleagues described more than 2,000 genomes from the group that contains giant viruses; before that, just 205 such genomes had been deposited in public databases.

Virologists have also looked inwards to find new species. Viral bioinformatician Luis Camarillo-Guerrero worked with colleagues at the Wellcome Sanger Institute in Hinxton, UK, to analyse metagenomes from the human gut, and built a database containing more than 140,000 kinds of phage. More than half of these were new to science. Their study<sup>4</sup>, published in February, matched others' findings that one of the most common viruses to infect the bacteria in our guts is a group known as crAssphage (named after the cross-assembly software that picked it up in 2014). Despite its abundance, not much is known about how it contributes to our microbiome, says Camarillo-Guerrero, who now works at DNA-sequencing company Illumina in Cambridge, UK.

Metagenomics has turned up a wealth of viruses, but it ignores many, too. RNA viruses aren't sequenced in typical metagenomes, so microbiologist Colin Hill at University College Cork, Ireland, and his colleagues looked for them in databases of RNAs, called metatranscriptomes. Scientists normally use these data to understand the genes in a population that are actively being turned into messenger RNA in to make proteins, but RNA virus genomes can show up, too. Using computational techniques to pull sequences out of the data, the team found 1,015 viral genomes in metatranscriptomes from sludge and water samples<sup>5</sup>. Again, they'd massively increased the number of known viruses with a single paper.

Although it's possible for these techniques to accidentally assemble genomes that aren't real, researchers have quality-control techniques to guard against this. But there are other blind spots. For instance, viral species whose members are very diverse are fiendishly difficult to find because it's hard for computer programs to piece together the disparate sequences.

The alternative is to sequence viral genomes one at a time, as microbiologist Manuel Martinez-Garcia does at the University of Alicante, Spain. He decided to try trickling seawater through a sorting machine to isolate single viruses, amplified their DNA, and got down to sequencing.

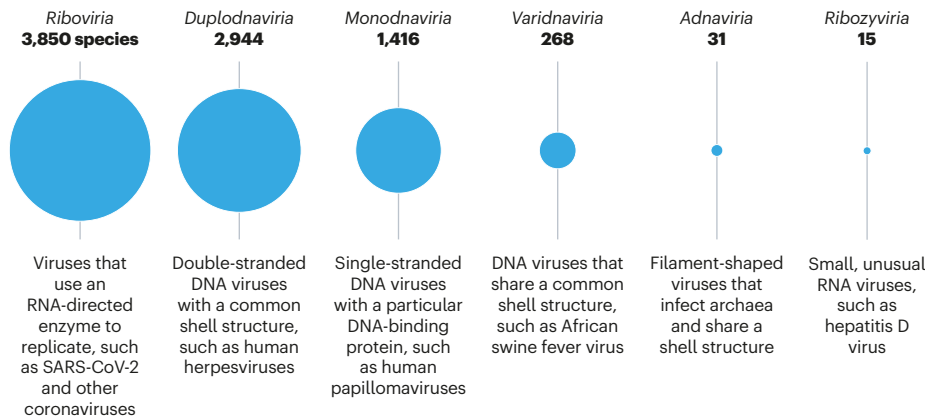
On his first attempt, he found 44 genomes.

# Feature

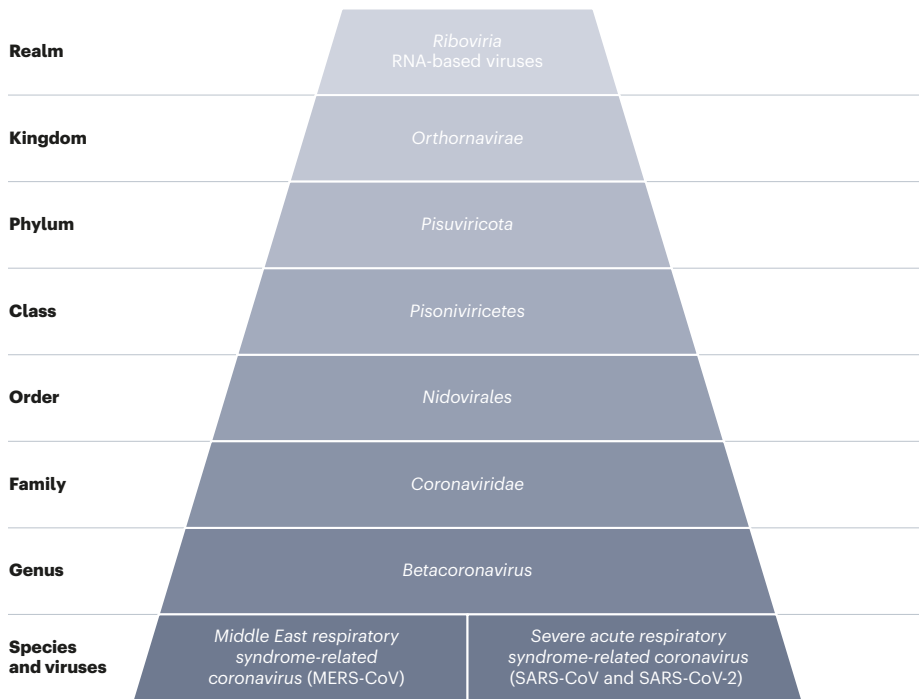
## VIRUS REALMS

Unlike cell-based life forms, viruses don't have a single common ancestor, which makes it impossible to create a phylogenetic tree. Instead, as of 2021, the International Committee on Taxonomy of Viruses recognizes six realms, defined by similarities in their members' genes and proteins, each arising from a different common ancestor.

### Realms by number of species



### The realm of SARS-CoV-2 and its relatives



One turned out to represent some of the most abundant viruses in the ocean<sup>6</sup>. This virus is so diverse – its genetic jigsaw pieces so varied from one virus particle to the next – that its genome had never popped up in metagenomics studies. The team calls it 37-F6, for its location on the original laboratory dish, but Martinez-Garcia jokes that, given its ability to hide in plain sight, it should have been named 007, after fictional superspy James Bond.

### Virus family trees

The James Bond of ocean viruses lacks an official Latin species name, and so do most of the thousands of viral genomes discovered by metagenomics over the past decade. Those sequences presented the ICTV with a dilemma: is a genome enough to name a virus? Until

2016, proposing a new virus or taxonomic group to the ICTV required scientists to have that virus and its host in culture, with rare exceptions. But that year, after a contentious but cordial debate, virologists agreed that a genome was sufficient<sup>7</sup>.

Proposals for new viruses and groups poured in. But the evolutionary relationships between these viruses were often unclear. Virologists usually categorize viruses on the basis of their shapes (long and thin, say, or a head with a tail) or their genomes (DNA or RNA, single- or double-stranded), but this says surprisingly little about shared ancestry. For example, viruses with double-stranded DNA genomes seem to have arisen on at least four separate occasions.

The original ICTV viral classification, which

is entirely separate from the tree of cellular life, included only the lower rungs of the evolutionary hierarchy, from species and genus up to the order level – a tier equivalent to primates or trees with cones in the classification of multicellular life. There were no higher levels. And many viral families floated alone, with no links to other kinds of virus. So in 2018, the ICTV added higher-order levels: classes, phyla and kingdoms<sup>8</sup>.

At the very top, it invented 'realms', intended as counterparts to the 'domains' of cellular life – Bacteria, Archaea and Eukaryota – but using a different word to differentiate between the two trees. (Several years ago, some scientists suggested that certain viruses might fit into the cell-based evolutionary tree, but that idea has not gained widespread favour.)

The ICTV outlined the branches of the tree, and grouped RNA-based viruses into a realm called *Riboviria*. SARS-CoV-2 and other coronaviruses, which have single-stranded RNA genomes, are part of this realm. But then it was up to the broader community of virologists to propose further taxonomic groups. As it happened, Eugene Koonin, an evolutionary biologist at the National Center for Biotechnology Information in Bethesda, Maryland, had assembled a team to analyse all the viral genomes, as well as the latest research on viral proteins, to create a first-draft taxonomy<sup>9</sup>.

They reorganized *Riboviria* and proposed three more realms (see 'Virus realms'). There was some quibbling over the details, Koonin says, but the taxonomy was ratified without much trouble by ICTV members in 2020. Two further realms got the green light in 2021, but the original four realms will probably remain the largest, he says. Eventually, Koonin speculates, the realms might number up to 25.

That number supports many scientists' suspicion that there's no one common ancestor for virus-kind. "There is no single root for all viruses," says Koonin. "It simply does not exist." That means that viruses probably arose several times in the history of life on Earth – and there's no reason to think such emergence can't happen again. "The *de novo* origin of new viruses, it's still ongoing," says Mart Krupovic, a virologist at the Pasteur Institute in Paris who was involved in both the ICTV decisions and Koonin's taxonomy team.

As to how the realms arose, virologists have several ideas. Perhaps they descended from independent genetic elements at the dawn of life on Earth, before cells even took shape. Maybe they escaped or 'devolved' from whole cells, ditching most of the cellular machinery for a minimal lifestyle. Koonin and Krupovic favour a hybrid hypothesis in which those primordial genetic elements stole genes from cellular life to build their virus particles. Because there are multiple origins for viruses, it's possible there are multiple ways they've originated, says Kuhn, who also served on the

SOURCE: ICTV (TALK.ICTVONLINE.ORG/TAXONOMY); ICTV CORONAVIRIDAE STUDY GROUP; NATURE MICROBIOL., 5, 536–544 (2020)



ICTV committee and worked on the new taxonomy proposal.

Thus, although the viral and cellular trees of life are distinct, the branches touch, and genes pass between the two. Whether viruses count as being ‘alive’ depends on your personal definition of life. Many researchers do not consider them to be living things, but others disagree. “I tend to believe that they are living,” says Hiroyuki Ogata, a bioinformatician working on viruses at Kyoto University in Japan. “They are evolving, they have genetic material composed of DNA and RNA, and they are very important in the evolution of all life.”

The current classification is widely recognized as just the first attempt, and some virologists say it’s a bit of a mess. A score of families still lack links to any realm. “The good point is, we are trying to put some order in that mess,” says Martinez-Garcia.

## World changers

With the total mass of viruses on Earth equivalent to that of 75 million blue whales, scientists are certain they make a difference to food webs, ecosystems and even the planet’s atmosphere. The accelerating discovery of new viruses “has revealed a watershed of new ways viruses directly impact ecosystems”, says Matthew Sullivan, an environmental virologist at Ohio State University in Columbus. But scientists are still struggling to quantify how much of an impact they have.

“We don’t have a very simple story around here at the moment,” says Ogata. In the ocean, viruses can burst out of their microbial hosts, releasing carbon to be recycled by others that eat the host’s innards and then produce carbon dioxide. But, more recently, scientists have also come to appreciate that popped cells often clump together and sink to the bottom of the ocean, sequestering carbon away from the atmosphere.

On land, thawing permafrost is a major source of carbon, says Sullivan, and viruses seem to be instrumental in carbon release from microbes in that environment. In 2018, he and his colleagues described 1,907 viral genomes and fragments collected from thawing permafrost in Sweden, including genes for proteins that might influence how carbon compounds break down and, potentially, become greenhouse gases<sup>10</sup>.

Viruses can also influence other organisms by stirring up their genomes. For example, when viruses transfer antibiotic-resistance genes from one bacterium to another, drug-resistant strains can take over. Over time, this kind of transfer can create major evolutionary shifts in a population, says Camarillo-Guerrero. And not just in bacteria – an estimated 8% of human DNA is of viral origin. For example, our mammalian ancestors acquired a gene essential for placental development from a virus.

For many questions about viral lifestyles,



Studies of the spiny-backed orbweaver found two viruses previously unknown to science.

scientists will need more than just genomes. They will need to find the virus’s hosts. A virus itself might carry clues: it could be toting a recognizable bit of host genetic material in its own genome, for example.

Martinez-Garcia and his colleagues used single-cell genomics to identify the microbes that contained the newly discovered 37-F6 virus. The host, too, is one of the most abundant and diverse organisms in the sea, a bacterium known as *Pelagibacter*<sup>11</sup>. In some waters, *Pelagibacter* makes up half the cells

“**In many ways, now is the time of the virome.”**

present. If just this one type of virus were to suddenly disappear, says Martinez-Garcia, ocean life would be thrown wildly off balance.

To understand a virus’s full impact, scientists need to work out how it changes its host, says Alexandra Worden, an evolutionary ecologist at the GEOMAR Helmholtz Centre for Ocean Research in Kiel, Germany. She’s studying giant viruses that carry genes for light-harvesting proteins called rhodopsins. Theoretically, these genes could be useful to the hosts – for purposes such as energy transfer or signalling – but the sequences can’t confirm that. To find out what’s going on with these rhodopsin genes, Worden plans to culture the host and virus together, and study how the pair function in the combined, ‘virocell’ state. “Cell biology is the only way you can

say what that true role is, how does this really affect the carbon cycle,” she says.

Back in Florida, Breitbart hasn’t cultured her spider viruses, but she’s learnt some more about them. The pair of viruses belong to a category Breitbart calls mind-boggling for their tiny, circular genomes, encoding just one gene for their protein coat and one for their replication protein. One of the viruses is found only in the spider’s body, never its legs, so she thinks it’s actually infecting some creature the spider eats. The other virus is found throughout the spider’s body, and in its eggs and spiderlings, so she thinks it’s transmitted from parent to offspring<sup>12</sup>. It doesn’t seem to be doing them any harm, as far as Breitbart can tell.

With viruses, “finding them’s actually the easy part”, she says. Picking apart how viruses influence host life cycles and ecology is much trickier. But first, virologists must answer one of the toughest questions of all, Breitbart says: “How do you pick which one to study?”

**Amber Dance** is a science journalist in Los Angeles, California.

- Rosario, K. et al. *PeerJ* **6**, e5761 (2018).
- Schulz, F. et al. *Science* **356**, 85–85 (2017).
- Schulz, F. et al. *Nature* **578**, 432–436 (2020).
- Camarillo-Guerrero, L. F., Almeida, A., Rangel-Pineros, G., Finn, R. D. & Lawley, T. D. *Cell* **184**, 1098–1109.e9 (2021).
- Callanan, J. et al. *Sci. Adv.* **6**, eaay591 (2020).
- Martinez-Hernandez, F. et al. *Nature Commun.* **8**, 15892 (2017).
- Simmonds, P. et al. *Nature Rev. Microbiol.* **15**, 161–168 (2017).
- International Committee on Taxonomy of Viruses Executive Committee. *Nature Microbiol.* **5**, 668–674 (2020).
- Koonin, E. V. et al. *Microbiol. Mol. Biol. Rev.* **84**, e00061-19 (2020).
- Emerson, J. B. et al. *Nature Microbiol.* **3**, 870–770 (2018).
- Martinez-Hernandez, F. et al. *ISME J.* **13**, 232–236 (2019).
- Rosario, K., Mettel, K. A., Greco, A. M. & Breitbart, M. *J. Gen. Virol.* **100**, 1253–1265 (2019).