

The virome hunters

Ambitious efforts to catalog viruses across the globe may facilitate our understanding of viral communities and ecology, boost infectious disease diagnostics and surveillance, and spur new therapeutics. Charles Schmidt investigates.

In July, scientists from UC Davis and Columbia University announced they had isolated a new species of the Ebola virus from bats roosting inside houses in Sierra Leone. Dubbed Bombali after the district where the bats were captured, this new species is the first Ebola virus to have its initial identification in an animal host rather than from a sick person. According to Tracey Goldstein, associate director of the One Health Institute at the University of California, Davis, who led the team behind the research, it isn't yet clear whether Bombali can infect people in the field, although it has been shown to infect cultured human cells¹.

The discovery of Bombali is notable for another reason: it was detected as a result of sequencing the entire virome of bats that had tested positive for Ebola in a consensus PCR-based assay¹. This new approach to virology, which takes advantage of high-throughput genomic technologies like next generation sequencing, is a novel adjunct to other approaches for identifying emerging viral pathogens before they 'spill over' into humans.

These are early days. "We know of only a minuscule fraction of the viruses out there,

and our questions about the viral world are profound," says Edward Holmes, a virologist and professor at the University of Sydney in Australia. Along with new species, investigators are turning up vast stretches of what they call dark matter—viral sequences unlike any seen previously. They're using sophisticated bioinformatics to characterize viral RNA and DNA and its various functions, and findings have shown already that viruses can play essential as well as harmful roles in human health. Ideally, virome research will lead to biomedical payoffs, such as new therapies, vaccines, and opportunities to head off new disease outbreaks.

A new high-throughput era

During the 2000s, genomic sequencing combined with advances in high-resolution microscopy ushered in the modern era of virome research. The first uncultured viral genome was sequenced in 2002 by Forest Rowher at San Diego State University from seawater samples collected off the California coast². More than 65% of the viral sequences in those samples had never been seen before, reflecting how viral

diversity was—and still is to this day—mostly uncharacterized. Scientists have since expanded their analyses into other many environments, as well as animal and human viromes. Indeed, a pure metagenomic analysis of human fecal samples revealed a previously unknown virus that represents a large part of the dark matter—as much as 90%—of the human gut virome. Dubbed the crAssphage by Robert Edwards and collaborators from San Diego State because it was pieced together by tool they invented called cross assembly analysis (although its origin in stool seems to have been in the minds of the researchers), it was called "one of the most striking feats of metagenomics at that time" by Eugene Koonin at US National Center for Biotechnology Information (NCBI). Koonin's group collaborated with Edwards' to characterize this family of phage and annotate some of the 80% of the 100kb genome that didn't align with known viral known viral proteins³.

Characterizing viromes, however, is complicated by the lack of a shared genetic marker among viruses analogous to the 16S ribosomal RNA gene in bacteria. All bacteria contain a version of that gene, allowing scientists to identify a particular species on the basis of its unique 16S signature. Viral identification relies instead on multiple markers associated with different taxonomic groups and on the way the sequences match up with those from known viruses in genomic repositories, such as the NCBI's Genome database.

David Paez Espino, a bioinformaticist at the US Department of Energy's Joint Genome

Institute (JGI) in Walnut Creek, California, explains that scientists can isolate a viral fraction in a sample by filtering it or by extracting and sequencing the entire microbial nucleic acid content. Applying metagenomic methods will home in on the genomes of DNA viruses alone, whereas metatranscriptomic methods will reveal the sequences of both viral DNA and viral RNA. Paez Espino explains that the analytical approaches are continually evolving, but as metagenomic methods came first—and DNA is inherently more stable than RNA—viral DNA sequences still predominate in microbial databases. But metatranscriptomic analyses are taking hold at places like the JGI, according to Paez Espino, because they provide so much more information about the virome—not just sequences, but also expression patterns. Moreover, scientists are motivated to sequence RNA viruses because they account for roughly half the entire viral world. “Most of big infectious diseases, such as Zika, Ebola and influenza, are caused by RNA viruses,” Paez Espino adds.

Simon Roux, a research scientist at the JGI’s facility in Berkeley, California, adds that analytical methods used in virome research each have their inherent limitations. For instance, scientists still can’t purify the viral content in a given sample completely. Some may be lost during filtration, for example. And the short reads that one gets from sequencing often have uncertain origins: they could be viral or derived from some other microbe. To confirm their sources, scientists stitch the reads together into ‘contigs’, or longer sequences that may have recognizable functions. According to Roux, that process requires bioinformatic algorithms that look for features unique to viruses that other microbes don’t share. For instance, newly formed viruses—but not other microbes—come wrapped in a protein capsid that can give away their identity. And microbial sequences that look completely unlike anything seen previously are often assumed to be viral simply because they are so novel. “Say you’ve gotten ten genes in your contig, and one looks like it encodes for a capsid, and the other nine genes are totally new and we have no idea what they’re doing,” Roux says. “That’s what you’d expect with a new virus genome.”

Although bacterial genomes greatly outnumbered their viral counterparts in microbial databases until recently, the number of published viral sequences is rapidly growing (Fig. 1). In 2016, JGI scientists unveiled roughly 125,000 partial and complete DNA virus genomes from samples taken around the world, including oceans, freshwater systems, soils, plants, animals and humans⁴, by mining the Department of Energy’s Integrated

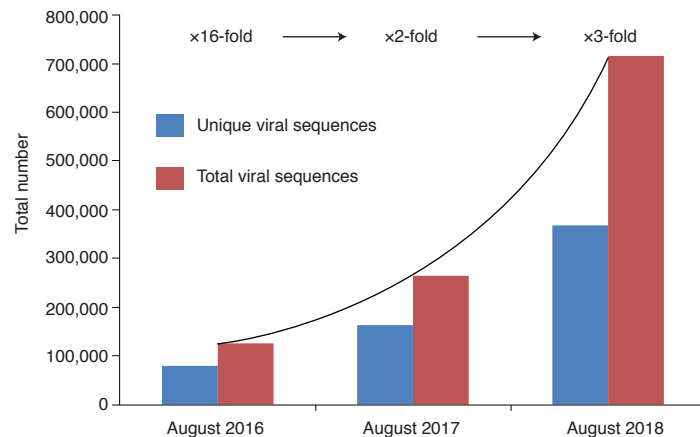


Figure 1 Growth rate of virus identification and microbial host prediction. Growth over time in the total and unique number of viral sequences in the January 2018 release of the IMG database. The first data point represents a 16-fold increase in the number of species in comparison to the number of previously identified viruses. Subsequent points are twofold and threefold higher, respectively (image by David Paez Espino, JGI).

Microbial Genomes (IMG) and Microbiomes database. Those viral contigs currently add up to 750,000 total sequences (428,000 unique) owing to the continuous addition of new samples submitted by the JGI and the NCBI and are deposited into the IMG/VR, a database of cultured and uncultured DNA viruses and retroviruses.

Into the wild

Goldstein’s team and many other groups are leveraging virome research with the aim of intercepting disease pandemics before they occur. The new Bombali finding is only the latest to emerge from the PREDICT project, a global effort to discover new viral threats in wildlife with the potential to spill over into human populations. Funded by the US Agency for International Development and based at the University of California, Davis, PREDICT—which launched in 2009—relies on PCR and next-generation sequencing to characterize viromes in bats, rodents and primates, three taxonomic groups that account for a high proportion of zoonotic viral diseases. According to Goldstein, PREDICT scientists have sampled more than 70,000 animals and people in over 30 countries at high-risk interfaces and, using consensus PCR, have discovered close to 1,000 new virus species. In bats, for example, which are now believed to be the evolutionary source of viruses such as SARS and MERS, the group detected sequences from 91 coronaviruses in 20 countries⁵. To further characterize those viruses, PREDICT researchers will isolate them or, if need be, synthesize them to see how they behave and replicate in a cellular host. However, given the potential danger, subsequent step in the research will have to take place only in level 4 biosafety facilities.

Goldstein says that after a new viral threat is identified, the hard work of determining pathogenicity begins. In her own laboratory, Goldstein starts by evaluating whether proteins from a particular virus bind to human cell receptors, as was demonstrated with Bombali. If they do, then researchers will try to grow the virus in cell culture, test whether it causes disease symptoms in experimental animals, and look for antibodies against the virus in people who live near where it was discovered.

Several other large-scale efforts are now expanding on PREDICT. The Global Virome Project (GVP) is setting out to discover roughly 1.2 million new zoonotic viruses in animals over the next ten years⁶. The project will depend heavily on the development and use of low-cost sequencing tools designed for developing countries. Peter Daszak, president of the New York-based EcoHealth Alliance, a nonprofit organization that works on global infectious disease issues, is among those directing the project. The GVP’s goal, he says, is to go from being reactive to proactive in the way health officials confront zoonotic pathogens. Should the GVP succeed, he adds, then it will have accumulated a database of all the high-risk viruses that threaten human populations.

Run by the US Department of Defense, the PREEMPT project has a complementary focus on the biological mechanisms underlying viral spillover into humans. According to Jim Gimlett, a program manager in the Department of Defense’s Defense Advanced Research Projects Agency, PREEMPT focuses on known classes of dangerous viruses, such as Ebola, Lassa fever, Rift Valley fever and avian influenza. Project scientists are modeling viral evolution and zoonotic potential, “and we’re testing scalable methods to prevent viral

species from jumping to humans in the first place,” Gimlett says.

Still, the notion that scientists could prevent new zoonotic pandemics strikes some scientists as unrealistic. Holmes, for instance, argues such efforts amount to “an absolute waste of time” because new disease outbreaks occur infrequently relative to the virome’s immensity. “If the goal is to better understand viral diversity, evolution and ecology, then great—that’s what we should do,” he says. “But there are just an enormous number of viruses in wildlife. And to try to predict which of them will emerge in humans is totally infeasible. It’s using rare data to predict rare events, and that just won’t work.”

The GVP’s Daszak responds that skeptics were once similarly dubious about weather prediction. But as meteorological data accumulated with time, he says, weather reports became increasingly reliable, and the same might prove true for predicting new outbreaks of disease. “More data to get to better predictions is exactly what we’re trying to provide,” he says. “Because if we don’t understand what’s going on out there, then we’re just stuck in the same situation: discovering viruses the hard way, which is what we want to avoid.”

The human virome

Apart from cataloging new viruses in the environment, researchers are also trying to understand what viruses are doing in their hosts and where—in what exact cell types—they are doing it (Table 1). The human virome differs from one person to the next, but within an individual it is remarkably stable over time. The viral composition in the gut, for instance, which is dominated by temperate DNA bacteriophages (which can shift between being temperate, or dormant, and replicating, or lytic), not only varies with age, health status,

geography and especially diet, but also has a major influence on the gut’s bacterial make-up. However, Frederic Bushman and colleagues at the University of Pennsylvania’s Perelman School of Medicine found that in stool samples isolated from the same person repeatedly for two-and-a-half years, 80% of the viral sequences were unchanged⁷. Temperate phages were stable, but lytic phages—which battle constantly with bacterial defenses—evolved rapidly: “It was as if some of them had become completely different species,” Bushman says.

A controversial finding is that some viruses can be detected that reside in body fluids—body fluids that were once considered sterile, such as healthy blood and cerebrospinal fluid. Amalio Telenti, a computational biologist at Scripps Research in La Jolla, California, had detected what he suspected were blood-borne viral fragments in healthy people years ago with PCR. Newer sequencing tools, he says, provided an opportunity to investigate whether these fragments were indeed viral, as opposed to bacterial, human “or just low-quality reads.” Others had found that 95% of DNA sequences in blood were from human cells, and they typically ignored the rest. Telenti wanted to characterize those residual sequences, since the presence of pathogenic viruses in transfusion blood products (for example, red cells, platelets and plasma) constitutes a major hazard, particularly for immunocompromised patients. So, with funding from Human Longevity in San Diego, a company cofounded by J. Craig Venter (who was a coauthor on the subsequent publication but has since retired), he led a team that sequenced blood samples from more than 8,000 healthy participants in a large-scale investigation of the whole human genome. The findings revealed 94 different phages and eukaryotic viruses⁸. Although Telenti says

some were likely introduced as contaminants in sequence reagents, he posits that the others, which include species from the Herpesviridae (herpesviruses) and Anelloviridae (for example, the torque teno virus), may reside permanently in healthy blood. “Sometimes viruses will attack us, but we also have to embrace the fact there’s probably much more symbiosis than we anticipated,” Telenti says. “Everyone seems to have a herpesvirus, so you have to wonder why they might be useful.” One possibility, Telenti adds, is that persistent, latent herpesviruses have roles in modulating and “educating the immune system.”

The view that some viruses help to keep the immune system nimble and responsive by stimulating low-level reactions—even as the immune system regulates viral behavior to keep illness in check—is gaining support. Mounting evidence suggests that phages, for instance, contribute to normal gut functioning by pruning the commensal bacteria we ordinarily live with, as well as by killing off bacterial pathogens⁹.

However, changes in phage distribution have also been linked with disease. Herbert Virgin, now executive vice president and chief science officer at San Francisco-based Vir Biotechnology (headed by George Scangos), was the first to associate the phage virome with human illness. While at the Washington University in St. Louis, Missouri, he and his colleagues reported in 2015 that people with ulcerative colitis and Crohn’s disease have elevated numbers of Caudovirales phages in their gut. The increased abundance of these viruses led him to speculate that an unbalanced phage virome contributes to these illnesses, and possibly to others as well¹⁰.

Eukaryotic viruses that infect human cells are rare in the gut by comparison, but knowledge is also increasing about their role

Table 1 Known virotypes according to culturomics and metagenomics

Virus type	Genome type	Environment	Associated disease
Eukaryotic virus			
Rotavirus, Astrovirus, Calicivirus, Norovirus, hepatitis E virus, Coronavirus, Torovirus, Adenovirus (serotypes 40,41)	All RNA except adenovirus (DNA)	Human small bowel and colon	Gastroenteritis
Adenoviridae, Picornaviridae and Reoviridae (genus Enterovirus)	RNA	Human intestine	Unknown
Plant-derived virus			
Pepper mild mottle virus (PMMV), oat blue dwarf virus, grapevine asteroid mosaic-associated virus, maize chlorotic mottle virus, oat chlorotic stunt virus, <i>Panicum</i> mosaic virus, tobacco mosaic virus	RNA	Plants and human feces	Pathogenic for plants Nonpathogenic for humans
Giant virus (>300 kb)			
Mimiviridae, Mamaviridae, Marseilleviridae, Poxviridae, Iridoviridae, Ascoviridae, Phycodnaviridae, Asfaviridae	DNA	Human fecal protists; amoebae in lakes, rivers and seawater	Pneumonitis, childhood diarrhea (Mimiviridae only)
Prophages			
Myoviridae, Siphoviridae, Podoviridae, Tectiviridae, Leviviridae, Inoviridae	dsDNA	Human feces	Unknown
Virus (<145 kb)			
Microviridae (Microvirus, Gokushovirinae, Alpavirinae, Pichovirinae)	ssDNA	Seawater, human gut bacteria	Unknown

dsDNA, double-stranded DNA; ssDNA, single-stranded DNA (modified from Scarpellini, E. *et al. Dig. Liver Dis.* 47, 1007–1012, 2015).

in health and disease. Virgin and his academic collaborators also reported that they had detected several eukaryotic viruses—including species from the Circoviridae, Anelloviridae, and Picobirnaviridae—in stool samples from healthy children¹¹. In his view, the evidence points to these viruses as members of a normal gut virome, even though they can also make people sick. Furthermore, Virgin discovered that viral diversity was remarkably low in the stool of children with type 1 diabetes, whereas stool samples from healthy children were enriched for eukaryotic Circoviridae viruses that seemed to protect against the disease.

Translating the virome

Scott Plevy, a gastroenterologist and expert on inflammatory bowel disease (IBD) at Janssen Research & Development in Raritan, New Jersey, emphasizes that the virome offers a wealth of untapped diagnostic and therapeutic opportunities. He's now exploring how phage composition varies in response to new IBD treatments. That research was prompted by findings showing that the gut's bacterial composition varies in tandem with IBD flareups and remissions. Plevy describes those bacterial changes as merely “the tip of the iceberg in terms of the diagnostic information we can extract from the gut.” Corresponding phage changes, he says, could offer even deeper insights into how patients respond to therapeutic interventions. And ideally, that could pave the way toward phage therapy—using phages as medical treatments to kill the bacteria that might cause or exacerbate IBD.

How to go from association to causation is “the million dollar question,” says David Wang

of Washington University in St. Louis. The field of virome research is a decade behind the study of the microbiome, according to Wang, and many of the tools that are being used to do functional testing of elements of the microbiome (fecal transplantation, gnotobiotic mice, and even antibiotics) are simply not available for the virome. “The first [thing we have to do] is to develop culture systems for any virus that we find an association for. If we don't have a culture system, we can't begin to try to do infection experiments in different settings,” he says. Wang has developed one of the few systems to date for culturing a novel virus detected in human stool. In a 2017 paper, his group described a Caco-2 culture system for astrovirus VA1/HMO-C, which is prevalent in human encephalitis¹². But, he points out, this is only the first step. Once it is established that a novel virus infects human cells, animal models will have to be established to study pathogenicity. According to Wang, until these systems are developed—and they are both challenging to implement and difficult to get funding for—the field will be limited to association studies.

Aleks Radovic-Moreno, vice president at PureTech Health, which is affiliated with the microbiome company Vedanta Biosciences, agrees that until these tools are available, companies will sit on the sidelines. “Solving these two challenges will be necessary for the field to accelerate,” he says. And it's going to take a dedicated effort, not something a microbiome company can do by having one or two scientists looking at viruses, according to Radovic-Moreno. Nonetheless, he believes that we are closer than ever before to

achieving this reality. “We will have to properly identify the ‘killer’ application for human virome-based therapeutics—one where the risk of infecting people with a virus is worth the risk. This is not straightforward, since there are multiple ethical and safety concerns,” he says.

Vir Biotechnology's Virgin also cautions that we're still “substantially behind in drawing statistical associations between the virome and disease.” And the road to turning these associations into treatments will be even longer. If the aim is to design approaches that would have an immediate therapeutic effect, he says, “we are further behind still in our ability to manipulate the virome in a predictable manner.”

But after spending several decades doing seminal research on viruses at Washington University before joining Vir, he says he's optimistic about the field's emerging prospects. “I think there's enormous potential,” he says.

Charles Schmidt, Portland, Maine

1. Goldstein, T. *et al. Nat. Microbiol.* <https://doi.org/10.1038/s41564-018-0227-2> (2018).
2. Breitbart, M. *et al. Proc. Natl. Acad. Sci. USA* **99**, 14250–14255 (2002).
3. Yutin, N. *et al. Nat. Microbiol.* **3**, 38–46 (2018).
4. Paez-Espino, D. *et al. Nature* **536**, 425–430 (2016).
5. Anthony, S.J. *et al. Virus Evol.* **3**, vex012 (2017).
6. Carroll, D. *et al. Science* **359**, 872–874 (2018).
7. Minot, S. *et al. Proc. Natl. Acad. Sci. USA* **110**, 12450–12455 (2013).
8. Moustafa, A. *et al. PLoS Pathog.* <https://doi.org/10.1371/journal.ppat.1006292> (2017).
9. Metzger, R.N. *et al. Viruses* **10**, 146 (2018).
10. Norman, J.M. *et al. Cell* **160**, 447–460 (2015).
11. Zhao, G. *et al. Proc. Natl. Acad. Sci. USA* **114**, E6166–E6175 (2017).
12. Janowski, A.B. *et al. J. Virol.* <https://doi.org/10.1128/JVI.00740-17> (2017).

Corrected after print 18 December 2018

Correction: The virome hunters

Charles Schmidt

Nat. Biotechnol. **36**, 916–919 (2018); published online 11 October 2018; corrected after print 18 December 2018

In the version of this article initially published, the work of the PREDICT project was described incorrectly, and the associated citation was incorrect. The text that read “PREDICT scientists have sampled more than 70,000 animals and people in over 30 countries with high zoonotic disease risks, and the researchers have reported on the discovery of 1,000 virus species with the potential to infect human beings⁵” has been changed to “PREDICT scientists have sampled more than 70,000 animals and people in over 30 countries at high-risk interfaces and, using consensus PCR, have discovered close to 1,000 new virus species. In bats, for example, which are now believed to be the evolutionary source of viruses such as SARS and MERS, the group detected sequences from 91 coronaviruses in 20 countries⁵.” Ref. 5 has been changed from Olival, K.J. *et al. Nature* **546**, 646–650 (2017) to Anthony, S.J. *et al. Virus Evol.* **3**, vex012 (2017). The errors have been corrected in the HTML and PDF versions of the article.