

Why the ocean virome matters

Expeditions are delivering data wealth about our planet's oceans, including its microbes. Some labs are now diving deep into the ocean virome.

Vivien Marx

Viruses, “even though they are such simple forms of life, seem to have their hands in everything,” says microbiome researcher Chris Bowler from the Institut de Biologie de l'École Normale Supérieure. The ocean microbiome includes bacteria and archaea and many types of single-celled eukaryotes: fungi, protists, phytoplankton, zooplankton. “We also consider viruses to be part of the ocean microbiome,” he and others note in a recent Perspective in *Nature Microbiology*¹. Studies and analyses of marine viruses are revealing the scale of the influence viruses have^{2–7}.

For example, says Bowler, viruses control population numbers of organisms in the ocean they infect, they move genes around, thus “lubricating the ocean,” and they contribute to biogeochemical cycles of the minerals carbon, nitrogen and phosphorus that all life depends on. Bowler is scientific director of the Tara Oceans consortium, which involves many organizations around the world. He was scientific coordinator of the [Tara Oceans](#) expedition, which produced copious data about the ocean. It was supported by the Tara Ocean Foundation, or Fondation Tara Océan, the Tara Oceans consortium and various funders. Bowler was also onboard in Antarctica during the [Tara Mission Microbiomes](#) expedition, which has been underway for two years and wraps up in October 2022.

Because of HIV and COVID-19, the word ‘virus’ seems synonymous lately with disease-causing agents, says James Wainaina, a postdoctoral fellow in the lab of Matthew Sullivan at Ohio State University, a co-author of two recent papers in *Science* on marine RNA viruses and a joint first author of one of them^{2,3}. When they first hear about his marine virome research projects, Wainaina's friends and family worry. But when they hear that these viruses do not infect humans, he says, “everyone is happy and at ease.” His friends and family also ask him about RNA viruses and are “terrified,” says Guillermo Dominguez-Huerta, a former postdoctoral scholar in the Sullivan lab who is now a scientific consultant for the lab, a joint first author of the recent studies and a first author of one of them. In their ocean samples, they checked for coronaviruses



Much about the global oceans is unknown, including its virome. Researchers are exploring the roles of marine viruses in ocean health, biogeochemical cycles and climate. Credit: F. Aurat, Fondation Tara Océan

and found none. It's understandable that people see viruses mainly as threats, but, he says, “viruses are essential parts of natural ecosystems.” They infect organisms such as microbes, which are found, for instance, in soil, in plants, in the human body and in the sea. Just as the human microbiome shapes not just disease but health, the ocean's virome is involved in the ocean's health and in planetary well-being.

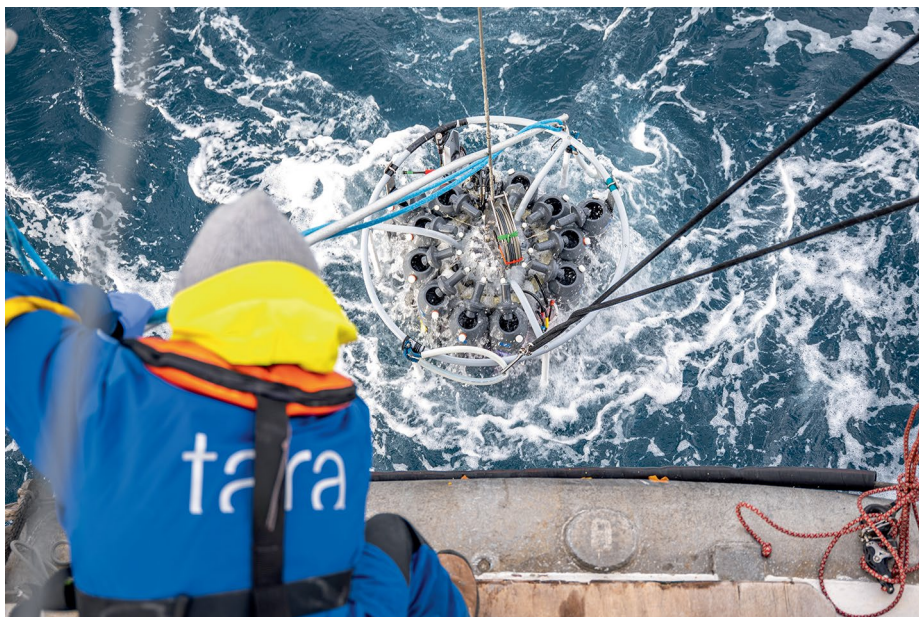
“In environmental microbiology, viruses are overlooked,” says Kathryn Campbell, a member of the Texas A&M University at Galveston lab of marine biologist Jessica Labonté, currently at sea. Campbell and Labonté co-authored a commentary⁴ on one of the *Science* papers³. As oceanic inhabitants, viruses lack the charisma of dolphins, whales or sharks, but “viruses are vital components of any ecosystem, not just aquatic,” says Campbell. Viruses are found wherever their hosts are and play essential roles in shaping communities and driving evolution. “Life itself has evolved tandemly with viruses,” she says. “The two are inseparable.”

“I am just amazed by the sheer amount of evidence for viral diversity we keep discovering,” says microbiome

researcher Shinichi Sunagawa of the ETH Zurich. With the data researchers currently have at hand, he is excited about the prospects of what else can be discovered in terms of viral ecology and evolution in the years to come. “Seeing the ocean as the place where life originated and given the enormous amount of environmental variability, I like imagining that we will still learn a lot of new biology from ocean viruses.”

Finding and counting viruses

There are many places to computationally trawl for microbial and viral data about the oceans; [a second story](#) in this issue describes some of these resources, as well as some other aspects of marine virus research. Viruses in the sea infect around 20–40% of the bacteria in the ocean per day and when the infected cells are lysed, this releases carbon and other nutrients into the food web⁵. Viruses can shape microbial communities, given the way they transfer genes to host cells and also ‘steal’ genes to manipulate host physiology, all of which affects biogeochemical cycling. Gene-to-ecosystem modeling has led many



Viruses are captured from water samples taken in the upper layers of the ocean, the epipelagic waters, and the deeper ones down to 1,000 meters. Organisms are filtered and nucleic acids extracted and sequenced to yield metatranscriptomic and metatranscriptomic data. Credit: M. Bardy, Fondation Tara Océan

labs to focus on the roles viruses play in biogeochemical cycles and climate.

A mouthful of coastal ocean water might have as many as 30 million viruses that can be counted with visual methods, says Dominguez-Huerta. This estimate is based on the number of known DNA viruses, which is perhaps around one-half of all DNA viruses. And then there are the numerous, elusive RNA viruses. To be analyzed, viruses have to first be captured, from both the upper layers of the ocean, the epipelagic waters, and the deeper ones down to 1,000 meters. During Tara Oceans, scientists collected samples with, for example, 250-kilogram rosettes of Niskin bottles that collect water at various depths and also carry sensors to collect data such as temperature, and there were pumps and plankton nets, too⁶. After the organisms are filtered and separated by size, nucleic acids are extracted and sequenced, yielding metagenomic and metatranscriptomic data. Illumina sequencing was used to obtain deep sequence data to use in identifying the plankton species in these samples.

Epifluorescence is used to visually tally double-stranded (ds) DNA viruses, which are prokaryotic viruses; they infect bacteria and archaea, says Dominguez-Huerta. But it is much more challenging to count other important viruses that have different types of genomes—single-stranded DNA or RNA. In their recent work, the team didn't count actual virus particles but rather

tallied nucleotide sequences derived from RNA viruses, such as genomic sequence, replication intermediates and more. The team identified 5,500 marine virus species in plankton samples from the Tara Oceans Consortium. This information stems from screening metatranscriptomes of marine plankton, which viruses infect.

What has impressed him about this work on marine viruses² from the Sullivan lab and colleagues, says Huahua Jian, who leads a deep-sea virus research group at Shanghai Jiao Tong University, is how the group has reliably identified RNA virus genomes from “the vast and messy RNA sequences.” The efforts, he says “will definitely promote the establishment of standardized methods of analysis across the research community,” which will make it easier to integrate findings and data from different labs.

Ocean metabolism

Kyoto University microbiome researcher Hiroyuki Ogata says that the recent work^{2,3} further connects RNA viruses and the carbon pump, which affects the Earth's biogeochemical cycles and thus its climate. And it sheds light on the diversity, evolution and ecology of RNA viruses, which has not previously been possible through applying the techniques of traditional DNA-based metagenomics. The team found many new lineages at the phylum-level by using “highly sensitive” computational approaches.

It's possible to assess the ecosystem impact of viruses by inferring auxiliary metabolic genes (AMGs). AMGs hint at the ways RNA viruses manipulate the physiology of their hosts as they seek to maximize production of more virus through the host. As Jian explains, labs have identified a variety of AMGs that are encoded by DNA viruses and, he says, it's “well-recognized” that AMGs probably play a role in marine ecosystems. It was unknown if AMGs could be found in RNA viruses, which the recent *Science* paper² has now established, he says. Jian sees this work as providing “a very important foundational dataset” for exploring questions connected to AMGs. “In my opinion, if more long-sequence or complete marine RNA virus genomes can be obtained in the future, and they can be further connected with specific hosts, it will greatly promote the understanding of the ecological impact of RNA viruses in the oceans.”

To tease out AMGs, the scientists used a variety of tools, such as viral identification software for both DNA and RNA viruses, says Wainaina. The ones for DNA viruses are available on [Cyverse](#), and the protocols for the tools from the Sullivan lab are on [protocols.io](#). One method for RNA viruses is in progress and will be soon available on Cyverse, he says. DNA viral identification tools include [VirSorter2](#), a pipeline for identifying viral sequence from metagenomics data, and the protocol for using this and other tools are also on [protocols.io](#). To identify AMGs from viral sequence that had been identified through VirSorter, the team used [DRAM-v](#), a software tool from the lab of microbiome researcher Kelly Wrighton at Colorado State University. Her group had created [Distilled and Refined Annotation of Metabolism \(DRAM\)](#), a framework to resolve metabolic information from microbial data. The companion tool [DRAM-v](#) is for viruses and can be applied to metagenomic data sets for annotating metagenomics-based assembled genomes, for example through the Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway database, and to contiguous viral sequences identified by VirSorter.

The hunt for AMGs is one instance in which the team needed to determine in each case whether a sequence was likely ‘stolen’ from host cells, says Dominguez-Huerta. RNA viral genomes are less than 40 kilobases long and usually have complicated genomic organization, both in a structural genomics sense related to the physical arrangement of genes along the viral genome and in a functional sense in terms of transcription and translation: there are

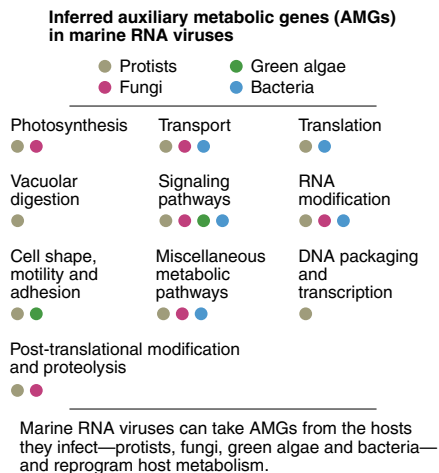
overlapping genes, frameshifts and more, all of which makes this kind of annotation difficult. And sometimes information in the annotation databases is wrong and indicates that a match is cellular when it is in fact viral. Thus, to find AMG, “we don’t have a defined clean methodology automated in a pipeline yet,” he says. It remains a time-consuming task. Assigning putative function to the protein sequences encoded by AMGs also involves checking the literature and comparing different annotation sources.

Dominguez-Huerta says he and the team were glad they could assemble AMG functionalities to suggest the range of ways in which RNA viruses manipulate the metabolisms of their hosts—from photosynthesis to central carbon metabolism to vacuolar digestion and RNA repair. This overview let them see how some AMGs are repeated across different viruses across the oceans. Finding AMGs in long-read sequence is what he calls a “fire test” for the lab. To avoid ‘false AMGs’ from unreliable matches, they use [BLASTP](#), the Basic Alignment Search Tool that compares a protein query sequence to a protein database.

“I am fascinated by the ability of viruses to metabolic reprogram not only their hosts but more importantly at the ecosystem level,” says Wainaina. It is probable that the AMGs the team identified “are a central cog in microbial metabolism networks.” Current and future modeling efforts will hopefully provide insights into the ecosystem roles of viruses—both DNA viruses and RNA viruses—and on a global scale both within the ocean ecosystem and beyond.

Host inference is challenging, says Dominguez-Huerta, because, for example, viruses with RNA genomes do not share genetic information with their host genomic DNA the way dsDNA viruses do when they infect bacteria. That means there is no clear signal to be derived from the host genome to help one guess the possible host. But sometimes RNA viruses do integrate into host genomes, and those, likely more accidental, events were sufficient for the scientists to capture some signal to infer hosts. “We also performed statistical co-occurrence analytics using abundances to infer the hosts with certain success,” he says.

Unlike dsDNA viruses, RNA viruses infect mostly eukaryotes, from protists and fungi to invertebrates and fish larvae; only a minority infect bacteria. Overall, the team has been able to capture “a picture of dsDNA viruses infecting prokaryotes and RNA viruses infecting eukaryotes in the oceans, complementing each other in their marine hosts,” says Dominguez-Huerta.



The ecosystem impact of marine viruses can be assessed by inferring auxiliary metabolic genes (AMGs). These AMGs hint at how viruses kidnap genes to manipulate host physiology. Recently, a group of scientists inferred AMGs for marine RNA viruses. Credit: Sullivan lab, Ohio State University, ref. ²; Thomas Philipps, Springer Nature.

The fact that the scientists can infer “that RNA viruses can steal genes from the host,” in the form of AMGs, to then reprogram host metabolism matters not only as scientists complete the picture of how viruses directly tune the activity of hosts during infection, but also in regard to how this influences biogeochemical cycles, he says. “We think that these AMGs are incorporated into the RNA virus genomes from cellular mRNA transcripts by non-homologous recombination,” he says. This gives, in his view, a new picture of RNA viruses, which, despite their small genome sizes, can squeeze in protein-coding genes. Such proteins could be sufficient to boost the production of virus particles per infected cell, perhaps increasing viral fitness in the difficult conditions of the oligotrophic open ocean and letting the viruses better propagate in the environment.

More generally, says Dominguez-Huerta, capturing RNA from ocean samples is difficult, because RNA is physically fragile and degrades rapidly. When digging into metatranscriptomic data, which include the RNA from plankton and RNA from other organisms, less than 1% of this RNA is likely to be viral RNA, he says. Previously, some labs have first purified RNA from samples, enriched it for replicating RNA viruses and then applied a method called dsRNA-seq to recover dsRNA virus sequence and replicate sequences from single-stranded RNA viruses. For future ocean RNA virus projects, he says that

the lab is currently working on a wet-lab method to purify RNA virus particles from seawater to solve the challenges of obtaining viral RNA for analysis.

A missing link

RNA viruses evolve quickly, and virus sequences are quite divergent. This was true for one gene that the Sullivan lab and colleagues focused their analysis on in their work^{2,3}. It’s the gene encoding RNA-directed RNA polymerase (*RdRp*), which is present in all Orthornavirae RNA viruses, says Wainaina. Because *RdRp* is so widely distributed, they used it to “fish” RNA viruses from the wealth of RNA sequences that metatranscriptomes comprise, says Dominguez Huerta.

The *RdRp* sequence has diverged, which makes sequence alignment difficult, especially across a large set of taxonomic levels of phyla and sub-phyla, and thus it’s hard, says Wainaina, “to establish who is related to whom.” Network analysis, he says, helps to provide “reasonable virus clusters” that can be fed into the phylogenetic analysis pipeline. A hidden Markov model (HMM), a type of statistical model, let the team search for signals from *RdRp* across their large data sets. In this way, they could “uncover novel and cool viruses” beyond those that are known or that could have been found had the team relied on a purely database approach. “It is more of a data-driven search rather a database-dependent exploration of the sequence space,” he says. *RdRp* was probably present in the RNA–Peptide World, the putative world that predated the evolution of the first cells. The *RdRp* sequence of the suggested novel phylum ‘Taraviricota,’ which, in addition to being very divergent, is predicted to have a three-dimensional protein structure resembling those of reverse transcriptases. This, says Wainaina, could suggest that the *RdRp* of taraviricots is the missing link between *RdRps* and reverse transcriptases in the context of ancient precellular life. “So we needed to improve our power of detection to the maximum level,” he says. This is where their strategy for identifying viral *RdRp* gene sequences based on searching and automated updating of HMMs went to work.

The pipeline starts, says Dominguez-Huerta, with a collection of *RdRp* HMMs and reference sequences to query sequences found across the oceans. The best *RdRp* sequences are extracted and incorporated to build HMMs with increased detection power. A total of ten search and update cycles saturate the search for divergent *RdRp* sequences. The team found protein sequences that were very divergent. Using HMM clustering, they found the ones



On board *Tara*, Xiomara Franchesa Garcia Diaz studies microplastics in a sample from Tara Mission Microbiomes. Data from this expedition will include imaging data, long-read sequence data and genome-wide chromatin organization data obtained by Hi-C analysis. Credit: M. Bardy, Fondation Tara Océan

that could not be clustered with previously reported viral *RdRp* genes. The clustering indicated that they had found five additional phyla of RNA viruses within the kingdom of Orthornavirae, which contains all RNA viruses that use RdRp to replicate.

The sequence data the team used came from different RNA samples and cDNA libraries that were prepped in two different ways. One was set up to avoid RNA molecules with a poly(A) tail, and the other selected for such molecules. The non-poly(A)-tail approach was used for prokaryotic plankton and the second approach for eukaryotic plankton samples. “Despite the biases introduced by these two cDNA library prep strategies, we could capture both types of RNA virus molecules regardless of the library prep method,” says Dominguez-Huerta.

Deep down

Over 30 years ago, says Jian, a team at the University of Bergen found a high abundance of virus in seawater⁸. The latest work is revealing the vast “undiscovered diversity of viruses in the ocean.” And there is much to yet find. Compared to the vastness of the ocean, he says, current sampling covers only a small area of the surface ocean. What awaits discovery is

this huge diversity of marine viruses. The next step, he says, is to understand their patterns and functions and how they affect marine ecosystems.

For now, much understanding stems from inference that is based on sequence and gene annotations of marine viruses. “I am very much looking forward to making more representative marine viruses and their hosts culturable in the laboratory and conducting in-depth analysis,” says Jian. In his work, he explores the ‘life features’ and ecological functions of deep-sea viruses, as well as the deep-time evolution of prokaryotic viruses.

The abundance distribution and biodiversity of viral communities in the ocean down to 1,000 meters deep are relatively well studied, says Jian, but the understanding of viruses in the deep ocean remains poor. His deep-sea-focused team is part of the International Center for Deep Life Investigation (IC-DLI) at Shanghai Jiao Tong University that studies the deep ocean, in particular the diversity and function of microbes.

Previously, he and his team generated an oceanic trench viral genome data set⁹ (OTVGD), which showed “remarkably high novelty of viral communities” from seawater and sediment samples of the Mariana, Yap and Kermadec Trenches. Next for his team is creating OTVGD 2.0 to assess the diversity and function of viruses that are found, for example, in the deepest reaches of hadal trenches. The hadal zone, which comprises ocean trenches such as those of the Pacific, is found below 6,000 meters deep.

Who infects whom

“Who infects whom” in the ocean, says Wainaina, is a question that often emerges in their group and the wider research community. It’s one with ecological and evolutionary importance. “It is still insanely hard to identify all possible viral hosts from metagenomes given the fact we cannot culture a majority of the metagenome-derived viruses,” he says. But efforts are underway to develop tools to solve some of these issues.

Bowler says he is intrigued to learn more about the host ranges of marine viruses.

“Who do they infect and are their host ranges very broad? And why have marine viruses evolved to use AMGs so extensively? This doesn’t seem to be the case for viruses infecting mammals and plants, so what’s different about the ocean?”

Texas A&M Galveston’s Campbell agrees on the importance of research that links viruses to their hosts. “Viruses don’t exist without their hosts,” she says. Thus, understanding who in the ecosystem they’re interacting with and who they are providing AMGs to is, in her view, “a high-priority issue” for gaining ecosystem understanding.

Looking at how prevalent and active the hosts are sheds light on the impact viruses have on the overall ecosystem. This is work Campbell personally hopes to focus on over the course of her career. “It would be incredible to drill into more volcanoes or anywhere in the deep sea and try to make viruses the focus or at least an important component of those types of studies,” she says. “The more places we look the more we’ll find.”

As Kyoto University’s Ogata says, only in the last decade has the importance of the human microbiome for human health become recognized. Although it may be possible to assess the health of Planet Earth without considering viruses, “this is like understanding the human physiology and its relationship with health without seeing the human microbiome,” he says. “I see viruses are at the heart of evolution and ecology.” □

Vivien Marx 

Nature Methods.

 e-mail: v.marx@us.nature.com

Published online: 29 July 2022

<https://doi.org/10.1038/s41592-022-01567-3>

References

1. Tara Ocean Foundation et al. *Nat. Microbiol.* **7**, 937–947 (2022).
2. Dominguez-Huerta, G. et al. *Science* **376**, 1202–1208 (2022).
3. Zayed, A. A. et al. *Science* **376**, 156–162 (2022).
4. Labonté, J. M. & Campbell, K. L. *Science* **376**, 138–139 (2022).
5. Gregory, A. C. et al. *Cell* **177**, 1109–1123.e14 (2019).
6. Sunagawa, S. et al. *Nat. Rev. Microbiol.* **18**, 428–445 (2020).
7. Brum, J. R., Hurwitz, B. L., Schofield, O., Ducklow, H. W. & Sullivan, M. B. *ISME J.* **10**, 437–449 (2016).
8. Bergh, O., Børsheim, K. Y., Bratbak, G. & Haldal, M. *Nature* **340**, 467–468 (1989).
9. Jian, H. et al. *ISME J.* **15**, 3094–3110 (2021).