

# Viruses: unlocking the greatest biodiversity on Earth<sup>1</sup>

Curtis A. Suttle

It is a great pleasure to be here, and I want to thank the organizers for giving me an opportunity to tell you about some of our results and ideas. The last two talks were a great introduction to the focus of my talk, biodiversity; I want to discuss how little we know.

There are three objectives to my talk. First, most of you will not have heard about viruses in the context that I wish to discuss them today; I want to make the case that most of the biodiversity on the planet is actually found in viruses. Second, I want to convince you of the importance of viruses other than as purveyors of disease, and that if it were not for viruses, life on the planet would probably not exist, or at least we would not exist. Third, I wish to show how genomics is the lens through which we can unlock this diversity and potentially make some remarkable discoveries along the way.

This is a symposium about genomics, the power and the promise. The power, of course, is that genomics is allowing us to unlock this diversity. New sequencing technologies are uncovering diversity we never knew existed. The promise is that this diversity has untold riches, that we will be able to use for all kinds of things including understanding how ecosystems function. We heard some nice stories about what diversity represents, but not what is meant by biodiversity. I work in the Beaty Biodiversity Research Centre, and a definition that resonates with me and many of my colleagues is that biodiversity is the totality of genes, species, and ecosystems in a region. In 1992, this definition was agreed upon by the United Nations Environment Program, Global Biodiversity Strategy, after a large round of consultations. Obviously, if there is no genetic diversity, there is no biodiversity. Species richness ultimately stems from the underlying genetic diversity, which underpins ecosystem richness and diversity. Ultimately, the roots of biological diversity and ecosystem function reside in the overall genetic diversity.

In the context of marine biological diversity, our understanding of the oceans has changed markedly. Originally, the seas were seen only as a resource to exploit, whether it was simply as an avenue to transport goods or as a fishery. More recently, the oceans are being viewed as a vast reservoir of potential genetic and biological diversity that can be interrogated to uncover processes and functions.

Perhaps surprising to some, most life in the oceans, by weight, is microbial. In fact, if on one side of a giant balance you could place everything microscopic (the protists, bacteria, archaea, and viruses) and on the other side you placed the things that people can see (the fish, whales, algae, crustacean, zooplankton, etc.), 95% to 98% of the living material, by weight, in the oceans is microscopic. This invisible majority not only produces half the oxygen on the planet, but if we think about biodiversity, ecosystem functioning, or biogeochemical cycles, we need to think about the microbial life in the seas. Almost all of the life in the

oceans, by weight, is prokaryotic (bacteria and archaea), with viruses and protists making up roughly equal amounts of the remaining. The big stuff, the whales and other charismatic megafauna are actually pretty trivial in terms of overall biomass. There are about 200 megatonnes of carbon in viruses in the ocean, which is equal to about 75 million blue whales. The most biomass that was ever in whales, as far as we know, was about 13 megatonnes. These numbers are dwarfed by the carbon in prokaryotes, which is about 5.2 gigatonnes. As I tell my students, whales are great; many are top predators; however, if you are interested in how the planet functions, it is not the whales that are important, it is the microbial life that we cannot see.

If we look at life by numbers, it is different. There are about 10 times as many viruses in the oceans as there are bacteria and about 1000-fold less protists than bacteria. In fact, in a litre of coastal seawater there are more viruses than there are people on the planet. If one takes a drop of seawater and adds a nucleic-acid stain such as Yo-Pro, SYBR Green, or SYBR Gold and looks at it under a microscope, there are a myriad of tiny fluorescent dots that are reminiscent of a clear night sky (Fig. 1). Most of these dots are virus particles. If aliens randomly sampled Earth they would see a planet dominated by microbial life, most of which would be viruses. On average, there are about 10 million viruses and a million bacteria per litre of seawater or freshwater. If we compare the number of viruses in the oceans to the number of stars in the universe, there are about  $10^{23}$  stars in the universe. In contrast, there are about 10 million-fold more viruses in the ocean than there are stars in the universe. If we took the  $10^{30}$  viruses in the oceans and stretched them end-to-end, how far would they go? Assuming, the average length of a virus is about 100 nm, the viruses would stretch  $10^{23}$  m, which is  $10^{20}$  km. If converted to light years, by dividing by  $10^{13}$  km, we end up with 10 million light years. The nearest star is Proxima Centauri, about 4.2 light years away; the Crab Supernova is 1000 light years away; our own galaxy, the Milky Way, is about 150 000 light years across. In fact, all the viruses in the ocean end-to-end would stretch further than the nearest 60 galaxies. This may seem like a trivial calculation, but it is important on a number of levels. For one reason, these viruses are responsible for about Avogadro's number, about  $10^{24}$  infections per second in the ocean. Each one of these events is an opportunity for lateral transfer of genes. Most of the genetic information on Earth probably resides within viruses.

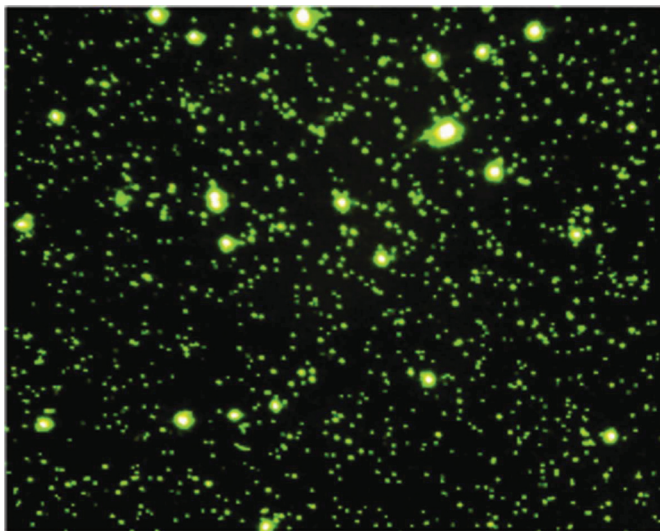
The number of viruses in the sea is enormous, but why does anybody care? As mentioned before, 95% to 98% of the biomass in the ocean is in microbes, which produce about half of the oxygen on the planet. Moreover, viruses kill about 20% of the living material in the ocean every day; hence, viruses are incredibly important in driving global cycles. Earlier concepts that the flow of the ocean's living resources goes from phytoplankton to zooplankton to fish have to be modified to incorporate the viral shunt, which

C.A. Suttle, Associate Dean of Science and Professor, Departments of Earth, Ocean and Atmospheric Sciences, Microbiology & Immunology, and Botany, Senior Fellow Canadian Institute for Advanced Research, University of British Columbia, Vancouver, BC V6T 1Z4, Canada.

E-mail for correspondence: [suttle@science.ubc.ca](mailto:suttle@science.ubc.ca).

<sup>1</sup>This article is part of a Special Commemorative Issue marking the one-year anniversary of "Genomics: The Power and the Promise".

**Fig. 1.** Epifluorescence micrograph of a seawater sample stained with Yo-Pro. On average there are more than 10 million virus particles per millilitre in coastal seawater. If lined up end-to-end, viruses in the ocean would stretch  $10^7$  light years, or further than the nearest 60 galaxies. Collectively, ocean viruses contain perhaps the largest reservoir of unexplored biological and genetic diversity on Earth.



moves living material into particulate and dissolved and organic matter that ultimately cycles back through bacteria and produces  $\text{CO}_2$ . Therefore, our understanding of pathways of nutrient and energy cycling on the planet has changed as a result of what viruses are doing. If viruses are removed from the oceans, photosynthesis decreases because virus-mediated nutrient cycling is less. Hopefully, I have made the case that viruses are important other than just as purveyors of disease in humans, animals, and plants.

If diversity is considered in the context of ribosomal RNA sequences, the Tree of Life is divided into three big branches. For about 3 billion years, life was confined to two of these branches, the bacteria and archaea, before microbial eukaryotes, the protists, arose. This represents the diversity of life on the planet today. One might ask, Where are the higher organisms? A representative higher organism would be a mushroom, and as we carry most of the same genetic information as mushrooms, on an evolutionary tree humans reside right beside mushrooms. The rest of the Tree, literally everything more distantly related, is microscopic. Therefore, macroscopic organisms are a small fraction of the existing genetic diversity on Earth. Most genetic novelty is encompassed in things that are too small to see, whose ancestors gave rise to enormous genetic innovation over billions of years. This includes viruses, which are so distant to other life forms that they are not included on the Tree of Life. The emerging consensus is that viruses arose before cellular life, and both viruses and cells arose from the same pool of genetic information. The origin of cellular and viral life is unknown, but it is increasingly clear that most viruses did not arise from cells.

In my laboratory the diversity of viruses is studied by collecting large volumes of seawater from which viruses are concentrated and purified. We now have a library of about 2000 concentrated virus communities from around the world. What have these samples taught us about the genetic diversity of viruses in the oceans? In collaboration with Forest Rohwer's group at San Diego State University, we used high-throughput sequencing and metagenomic analyses to examine 56 viral communities from the Arctic Ocean, 85 from the Pacific, 45 from the Gulf of Mexico, and 1 from the Atlantic near Bermuda. Remarkably, about 95% of the DNA

sequences had no obvious homologs in any database, meaning that the functions of the proteins that these sequences code for cannot be inferred. Even now, with longer sequences and much larger databases for comparison, about 70% to 80% of the coding sequences from natural marine viral communities do not have obvious homologs. The genetic diversity of marine viruses is enormous, and we are a long way from closing the sequence space. When Alex Culley and Andrew Lang were in my laboratory they asked similar questions about RNA viruses. Again, 60% to 80% of sequences had no obvious similarity to sequences in other databases. Moreover, there was almost no sequence overlap between samples that were taken about 100 km apart, even in the families of viruses that were represented. A third example from my laboratory is the work of Jessica Labonté, who examined the diversity of single-stranded DNA viruses. She found that 90% of the sequences had no matches in other databases, while the other 10% encompassed the known families of single-stranded DNA viruses. Through a series of sophisticated network and genomic analyses, she was able to close more than 600 viral genomes that represented 125 distinct evolutionary groups that had no relationship to each other or any known group of single-stranded DNA viruses. There are eight recognized families of single-stranded DNA viruses; her results may have uncovered more than 100 new families. Clearly, the genetic diversity in viruses is enormous and largely unknown.

So,—the promise—can any of this information be translated into application? It is difficult to figure out function from novel genetic information and even more difficult to translate genetic sequence into an application, but the potential is enormous. If marine genetic patents are examined, there are many examples of sequences that have homologs where the function of the coded protein is known. This leads to the potential to realize a wide range of applications including human health, biofuels, bioremediation, and biotechnological tools. The number of sequences that are being patented because the function of the coded protein has some putative benefit has been increasing rapidly, but the genetic potential of marine microbes is completely unrealized.

A different problem is that existing patents are held almost entirely by a few economically developed countries that have the ability to benefit from this genetic information. Most of the patents are held by less than 1% of countries. This brings up the question of who should own the ocean's genetic resources. Are marine genetic resources only for exploitation by the most developed countries or are they a global resource that should be shared? These are complicated questions that are being explored through the United Nations Convention on the Law of the Sea (UNCLOS). It is made more complicated because microbes do not respect territorial boundaries.

Exciting new findings about viruses and their genes also can stem from characterizing unusual virus isolates. An example is CroV, a virus that infects and kills *Cafeteria roenbergensis*, a unicellular phagotrophic zooplankton that primarily eats bacteria. CroV was characterized by Matthias Fischer when he was in my laboratory. It is one of the largest viruses known with 730 000 base pairs of coding sequence. About half the genes have similarity to others, but remarkably the closest homologs are distributed among bacteria, archaea, eukaryotes, and other viruses. This virus has acquired genes from all the domains of life. CroV is one of the most complex viruses known. It has more than 40 proteins associated with DNA replication, 21 with transcription, 13 with translation, and 9 involved with DNA repair.

A remarkable observation is that CroV can be parasitized by a small virus called Mavirus, which is an obligate parasite of CroV. When infected by CroV, the zooplankton transports Mavirus from outside the cell. Mavirus then kills CroV and rescues the cell. The closest relatives of Mavirus are not other viruses but large transposable elements that are found in eukaryotes. This leads to the hypothesis that over evolutionary time some species incorporated

this genetic information into their genomes and became immune to infection by these large viruses. This gave rise to *Maverick* transposable elements, or polintons that persist in eukaryotic genomes. The *Economist* picked up on this story in the context of the International Conference on Biodiversity in Nagoya, Japan, emphasizing that “biodiversity is not just a matter of tigers and whales, or butterflies and trees, or even coral reefs and tuna. It is also about myriad creatures too small to see that live in numbers too large to count in ways too numerous to imagine”. Not only does most of Earth’s genetic and biological diversity lie within viruses, it is reasonable to hypothesize that viruses archive homologs of all the genetic information on the planet; yet, most genes in viruses have no homologs in cellular organisms.

I want to end this talk as it began and consider biodiversity. It will be awhile before Greenpeace can be convinced to mobilize and save the viruses; yet, preserving microbial diversity is a legitimate concern. Caroline Chénard, a graduate student in my laboratory, does her field work on the northern tip of Ellesmere Island, which is about as close to the North Pole as one can get and be on land. This is where the Markham ice shelf was in place for tens of thousands of years at the mouth of a long fjord. It formed a giant ice dam, many meters deep that isolated the water behind

it from the Arctic Ocean, forming a giant freshwater lake. Studies have shown that the microbial communities in these ice-dammed lakes are distinct from microbial communities elsewhere. For example, viruses in the seasonal lakes lying on top of these ice shelves are more similar to the viruses in rice paddies in Southeast Asia than they are to the viruses in the immediately surrounding seawater. Over a short period of time the Markham ice shelf broke off and seawater flooded into the fjord, destroying a lake system that had been there for thousands of years. What was lost and its potential benefits to future generations will never be known. There is a case to be made that even if exploring microbial diversity may not translate into products with commercial potential within five years, this genetic information will be the source of new ideas and new technology that will yield enormous benefits.

I hope I have convinced you that genetic diversity underpins biological diversity, and that marine microbes, particularly viruses, encompass much of the unknown genetic diversity on Earth. Ultimately, the identification and preservation of these genetic resources will lead to future breakthrough technologies.

Finally, I want to thank the people in my laboratory and our collaborators that make the research possible and fun, and the individuals and agencies that recognize the value of basic research in leading to major advances in our understanding.