Title

A viral reckoning: viruses emerge as essential manipulators of global ecosystems

Journal

Environmental Microbiology Reports, 11(1)

ISSN

1758-2229

Authors

Roux, Simon

Brum, Jennifer R

Publication Date

2019-02-01

DOI

10.1111/1758-2229.12700

Peer reviewed

A viral reckoning: Viruses emerge as essential manipulators of global ecosystems

Simon Roux[1] & Jennifer R. Brum[2]

[1] DOE Joint Genome Institute, Walnut Creek, CA 94598, USA
[2] Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, LA 70808, USA

Viruses are integral components and critical regulators of microbial ecosystems. In terms of numbers alone, virus-like particles seemingly outnumber microbial cells in every ecosystem (Weinbauer and Rassoulzadegan, 2004), with a virus-to-microbe ratio typically ranging from 1 to 100 (Wigington *et al.*, 2016). While challenging to assess, profound ecological and evolutionary impacts of virus-host interactions have nonetheless been uncovered across a broad range of ecosystems, from the bottom of the oceans to bubbling acidic hot springs, coral reefs, and thawing permafrost (Suttle, 2007; Rohwer and Vega Thurber, 2009; Dell'Anno *et al.*, 2015; Williamson *et al.*, 2017; Emerson *et al.*, 2018).

Collectively, these studies highlighted multiple mechanisms by which viruses drive ecological and evolutionary processes in microbial ecosystems (Koskella and Brockhurst, 2014; Breitbart *et al.*, 2018). While the ecological importance of viruses is now undeniable, a thorough assessment of their influence on any microbial system remains elusive. Two of the current challenges are (i) comprehensively exploring and classifying environmental viral diversity, and (ii) establishing host linkages for uncultivated viruses. A number of recent methodological innovations suggest, however, that these hurdles may be overcome sooner rather than later (Mokili *et al.*, 2012; Dang and Sullivan, 2014; Brum and Sullivan, 2015; Sepulveda *et al.*, 2016; Sullivan *et al.*, 2017). Here, based on these latest advances in the field of viral ecology and genomics, we try to imagine how a comprehensive host-resolved mapping of the viral sequence space will enable researchers to address long-standing viral ecology questions in an unprecedented way. We present these as three stories relating how we picture (and/or wish) viral ecology research could be conducted ten years from now.

Story 1. A real-time investigation of virus-host interactions onboard an oceanic research cruise.

The year is 2030. Aboard a research vessel in the Southern Ocean, a group of graduate students is busy debriefing and summarizing the work they have accomplished during their first two weeks of study off the coast of Antarctica. The purpose of this expedition was to monitor how the massive influx of freshwater coming from significant melting of the Antarctic ice sheet influenced microbial communities. Because viral ecology is now routinely taught at the undergraduate and graduate levels, they already know that viruses are important to consider in this context, and that viral communities can be explored by identifying virus genomes within metagenomes. To understand the role of viruses in this changing environment, the students used a robotic sampling device to collect, process, and sequence a high-resolution spatial grid of microbial metagenomes and metatranscriptomes from the coast, obtaining nearly real-time sequencing data.

In anticipation of this large amount of data, the students had previously downloaded the standard Viral Ecology Toolkit on the ship's computing system. This toolkit is a result of intensive, field-wide collaborative efforts to centralize all known virus-related information, and includes programs to (i) extract all virus sequences (DNA and RNA; infecting bacteria, archaea, or eukaryotes), (ii) map and classify these sequences against the global database of publicly available virus genomes, populated with information on each virus host range and ecological distribution, (iii) provide host prediction for the handful of "novel" viruses without a close match in the global database, and (iv) model the biogeochemical effects of viral auxiliary metabolic genes expressed during infection. The group is thus able to quickly obtain a 3D map of viral diversity, host associations, and influences on carbon and nutrient cycling while still on site. They use this information to iteratively refine their

sampling plan and focus on the critical locations they need material from to further study these viruses back in the lab. While on the transit back to warmer climes, after obtaining a relatively comprehensive picture of viral ecology while still in the field, the group of students is already organizing the research paper units for their dissertations, planning the different experiments they want to attempt next, and of course sorting their many pictures of penguins by location, size, and cuteness for social media outreach efforts.

Story 2: Taking viruses to task: virus-inspired solutions to local harmful cyanobacterial blooms.

During the same year, 2030, another research team is trying to solve an issue much closer to home. This team focuses on cyanobacterial harmful algal blooms (HABs) occurring in a local freshwater lake. These toxic HABs are not only a human health hazard, they also strongly decrease the surface water quality, hindering the local tourism-based economy. Thanks to progress in satellite monitoring coupled with computational model predictions of HAB dynamics, they already have a good understanding of which environmental conditions trigger these blooms. Hence, they are now in search of solutions for mitigating these HABs before they fully form.

Surveying the current literature, the team quickly realizes that viruses infecting these cyanobacteria (cyanophages) may hold the key to some innovative bloom reduction strategies. They also know that simply releasing phages might have unintended consequences on ecosystem function, so they look instead for alternative "phages-inspired" solutions. Eventually, they plan to test two potential approaches: (i) using phage tail proteins as bait to specifically bind and remove the harmful cyanobacteria from the environment, and (ii) leveraging phage proteins targeting the cyanobacterial peptidoglycan, such as glycoside hydrolases, to be deployed at the onset of a bloom and preclude its development.

Mining the centralized and host-contextualized database of phage genomes now available, the team gathers candidates for tail proteins and glycoside hydrolases specific to their HAB strain in a matter of hours. From there, the team can design and order synthetic constructs for the identified proteins, and within a few weeks are already testing their first concepts for a "HAB-removal phage-inspired tool" under laboratory conditions. This first set of laboratory experiments helps them to determine which approaches could be the most successful. Two designs stand out as promising in these experiments: (i) a polymer nanosheet displaying phage tail proteins binding specifically to HAB cyanobacteria, and (ii) a chimeric protein combining the most efficient cell-binding domain with the most stable catalytic domain of a phage glycoside hydrolase, applied directly to the sample for in situ lysis of HAB cyanobacteria.

The next step is now to test these two designs through experimental mesocosms in their lake, where they plan to apply the treatments and continuously monitor the HAB cyanobacteria, as well as the effects on the microbial community as a whole, using automated sampling. While the team is aware that scaling-up this type of approach is often non-trivial, and expects to go through a few rounds of optimization, they are more convinced than ever that there is a lot to learn from phage in their quest to curb HABs.

Story 3. Back to the future: addressing historical viral ecology questions for ecosystem modeling

At the conclusion of the 2030 International Viral EcoGenomics Conference, as the remaining interactive posters are being turned off and recycled, a group of scientists decides to adjourn to a private room at a nearby gastropub. They want to discuss a new, collaborative project aimed at meaningfully incorporating viruses into predictive models of microbial community structure using the extensive host-resolved mapping of viral sequence space accumulated in the last 30 years. While real-time sequencing and analysis of free virus particles is now automated, there remains a major question to investigate: which of the free viruses will actually find and infect a host cell, and which extracellular viruses will never contact their host, destined to be simply degraded as

dissolved organic matter? This question raises issues first explored during the initial phase of environmental viral ecology in the 1990's, but largely overlooked in the intervening years.

The first topic on the digital board of the meeting room contains only two words: "viral decay". Although focusing on the genome-based exploration of viral diversity was a critical step, viral genomes unfortunately don't inform on key virus biophysical properties, including those that determine how long a virus may remain in the environment. Because this process will be virus- and environment-dependent, the team thus plans on combining "traditional" viral decay experiments with now-routine viral metagenomic analysis under varying environmental conditions. These data will be incorporated into physical models that predict particle transport across changing environmental conditions through time and space, then combined with similar models for the microbial hosts.

After this initial discussion, the group realizes that modeling both virus decay and transport will not fully solve their problem. The production of new viral particles will also be dependent upon the rate at which they contact their hosts. Thus, the second portion of the board is dedicated to "contact rate". Once calculated using total viral and bacterial abundances, the group now has decades of virus-host specificity information based on cultivation-dependent and -independent studies to draw upon. They plan to utilize this data to incorporate the contact between viral species and the specific host(s) they can infect within the global-scale transport model.

Although tired, the group suddenly realizes they are still thinking of viruses solely as killing machines, so there is yet one more component to include; "viral replication strategy" is added to the board. While lytically replicating viruses were the primary focus of early work in the field, the last decade of lab- and field-work has added considerable information regarding (i) the environmental conditions that select for different virus infection strategies, e.g. chronic, lysogenic, and lytic, and (ii) varying efficiencies for each type of infection in different hosts and environmental contexts. The team thus plans to integrate this data into the overall model, alongside viral decay and virus-host contact rate. As the team exits the gastropub and waits for a self-driving shuttle to take them back to their hotel, they know their model will not be perfect but are confident that it will enhance their ability to predict virus-induced changes brought about as geoengineering strategies are enacted on Earth.

Wishful thinking or realistic path forward?

Some of the stories presented here may seem far-fetched and little more than wishful thinking from scientists eager to witness their field expanding and transforming. Undoubtedly, technological leaps remain difficult to predict, and some or all of the techniques, approaches, and collective knowledge referenced in these stories may not be readily available in the relatively short time-frame of the next 12 years. However, the science imagined for 2030 is grounded on real pilot experiments, projects, and prototypes here in the year 2018.

First, given the growing number of published papers reporting new viruses sequenced from environmental sample or isolates (Pope *et al.*, 2015; Páez-Espino *et al.*, 2016; Roux *et al.*, 2016; Emerson *et al.*, 2018), a global contextualized database for viruses of microbes is certainly on its way. As highlighted in the recently published "Minimum Information on an Uncultivated Virus Genome" framework (Roux *et al.*, 2018), viral genomes will most likely form the backbone of this database. Current efforts include >700,000 complete and partial genomes (Páez-Espino *et al.*, 2016), and should reach 10s of millions of genomes in the coming years, especially once complete genomes can be sequenced from single templates using long read sequencing (Houldcroft *et al.*, 2017). New methods to recover virus particles from low biomass samples are also being developed, hence the virus genome database should cover nearly all type of biomes and ecosystems by 2030.

In the meantime, the development of methods to add information to this genome-centric database is currently at full steam. These include approaches for large scale virus taxonomy classification (Bolduc *et al.*, 2017b; Meier-Kolthoff and Göker, 2017; Nishimura *et al.*, 2017; Aiewsakun and Simmonds, 2018), as well as host linkage for uncultivated viruses either computationally (Edwards *et al.*, 2016; Galiez *et al.*, 2017; Ahlgren *et al.*,

2016) or experimentally (Tadmor *et al.*, 2011; Martínez-García *et al.*, 2014; Deng *et al.*, 2014; Roux *et al.*, 2014; Labonté *et al.*, 2015; Spencer *et al.*, 2016). Beyond virus-specific tools, we anticipate significant improvements in genome annotation capabilities stemming from (i) "multi 'omics approaches" combining transcriptomics, proteomics, and metabolomics studies of individual environments (Franzosa *et al.*, 2015), and (ii) improved functional prediction tools leveraging protein structure constraints and large-scale comparative genomics (Alva *et al.*, 2016; Ovchinnikov *et al.*, 2017). Finally, online platforms enabling high-throughput analysis of user's data are starting to emerge, such as iVirus (Bolduc *et al.*, 2017a) or KBase (Arkin *et al.*, 2018). These, in concert with the progressive establishment of a standardized viral ecogenomics toolkit, should enable every microbiologist to analyze viruses in their system.

In terms of technology, automation and robotics could undeniably provide a significant shift in the scale of biological field sampling. Instruments able to collect and process samples automatically in the field, including those for sequencing, are currently being tested and refined, and would present an unprecedented opportunity for near-continuous monitoring of microbial ecosystem (Ottesen, 2016; Powers *et al.*, 2018). Similarly, synthetic biology approaches are progressing at a fast pace, to the point where sequences of interest can simply be ordered on demand for functional characterization (Smanski *et al.*, 2016; Ziemert *et al.*, 2016). With the throughput and cost of these techniques continuously improving, many groups and laboratories will likely be able to perform quick functional screening of candidate uncharacterized proteins, helping them design innovative biotechnological applications.

Phage therapy, or more generally using viruses to modulate microbial communities, is not a novel idea as it dates from almost a century ago. However, it has been recently revitalized by an improved understanding of phage biology associated with the challenges presented by widespread antibiotic resistance of some bacterial pathogens (Kutter *et al.*, 2015). This type of approach has already been successful beyond human patients, e.g. to protect honeybee hives from *Paenibacillus* infections (Brady *et al.*, 2017). In addition to the "classical" phage therapy using entire infectious phages, a number of methods have been pioneered that use only specific phage components (Fischetti, 2008; Young and Gill, 2015; Ghequire and De Mot, 2015; Fischetti, 2018). Beyond cell lysis, we also anticipate that a number of applications will likely leverage the ability of phages to modify the metabolism and phenotype of their host cell. This is exemplified in the use of *Wolbachia* as a biocontrol agent of mosquito populations, which is predicated on the presence of a specific prophage in the bacteria (Le Page *et al.*, 2017). Finally, the nanosheet referred to in Story 2 already exists (Battigelli *et al.*, 2018), and we can expect further progress in the field of nanomaterials that will enable innovative product design for virus-inspired microbial manipulation.

The major scientific questions highlighted in Story 3 (viral decay, contact rate, and the continuum of viral infection strategies), are long-standing challenging topics. While the first conceptual frameworks were established 20 to 30 years ago (Heldal and Bratbak, 1991; Suttle and Chen, 1992; Murray and Jackson, 1992; Thingstad, 2000), new work and new discoveries are still being made around the same fundamental question today (Dell'Anno *et al.*, 2015; Knowles *et al.*, 2016; Weitz *et al.*, 2017; Köstner *et al.*, 2017). Most of the debate and controversy is rooted in the resolution at which we can observe these phenomenon in nature, i.e. we see "too little from too far". We believe, although maybe optimistically, that we will be able to scrutinize viral communities and virus-host interactions with an unprecedented levels of details and resolution by the ~2030 horizon, allowing us to revisit these fundamental questions. In addition, there is very little doubt that inter- and multi-disciplinary projects will be required to address these issues, combining experimental and *in silico* approaches, and involving experts in microbiology, computational biology, biophysics, statistical modeling, and physical modeling.

Acknowledgments

The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science of the U.S. Department of Energy under contract no. DE-AC02-05CH11231. This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy contract no. DE-AC02-05CH11231. The authors have no conflict of interest to declare.

References

Ahlgren,N. *et al.* (2016) Alignment-free d_2^* oligonucleotide frequency dissimilarity measure improves prediction of hosts from metagenomically-derived viral sequences. *Nucleic Acids Res.* **45**: 39–53.

Aiewsakun,P. and Simmonds,P. (2018) The genomic underpinnings of eukaryotic virus taxonomy : creating a sequence-based framework for family-level virus classification. *Microbiome* **6**: 1–24.

Alva,V. *et al.* (2016) The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res.* **44**: W410–W415.

Arkin,A.P. *et al.* (2018) KBase: The United States department of energy systems biology knowledgebase. *Nat. Biotechnol.* **36**: 566–569.

Battigelli,A. *et al.* (2018) Glycosylated Peptoid Nanosheets as a Multivalent Scaffold for Protein Recognition. *ACS Nano* **12**: 2455–2465.

Bolduc,B. *et al.* (2017a) iVirus: facilitating new insights in viral ecology with software and community data sets imbedded in a cyberinfrastructure. *ISME J.* **11**: 7–14.

Bolduc,B. *et al.* (2017b) vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ* **5**: e3243.

Brady,T.S. *et al.* (2017) Bacteriophages as an alternative to conventional antibiotic use for the prevention or treatment of Paenibacillus larvae in honeybee hives. *J. Invertebr. Pathol.* **150**: 94–100.

Breitbart,M. *et al.* (2018) Phage puppet masters of the marine microbial realm. *Nat. Microbiol.* **3**: 754–766.

Brum,J.R. and Sullivan,M.B. (2015) Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat. Rev. Microbiol.* **13**: 1–13.

Dang,V.T. and Sullivan,M.B. (2014) Emerging methods to study viral infection at the single-cell level. *Front. Microbiol.* **5**: 724.

Dell'Anno,A. *et al.* (2015) Virus decomposition provides an important contribution to benthic deep-sea ecosystem functioning. *Proc. Natl. Acad. Sci.* **112**: E2014–E2019.

Deng,L. *et al.* (2014) Viral tagging reveals discrete populations in Synechococcus viral genome sequence space. *Nature* **513**: 242–45.

Edwards,R.A. *et al.* (2016) Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol. Rev.* **40**: 258–272.

Emerson,J.B. *et al.* (2018) Host-linked soil viral ecology along a permafrost thaw gradient. *Nat. Microbiol.* **3**: 870–880.

Fischetti,V.A. (2008) Bacteriophage lysins as effective antibacterials. *Curr. Opin. Microbiol.* **11**: 393–400.

Fischetti,V.A. (2018) Development of phage lysins as novel therapeutics: A historical perspective. *Viruses* **10**:.

Franzosa,E.A. *et al.* (2015) Sequencing and beyond: Integrating molecular "omics" for microbial community profiling. *Nat. Rev. Microbiol.* **13**: 360–372.

Galiez,C. *et al.* (2017) WIsH: who is the host? Predicting prokaryotic hosts from metagenomic phage contigs. *Bioinformatics* **33**: 3113–14.

Ghequire,M.G.K. and De Mot,R. (2015) The Tailocin Tale: Peeling off Phage Tails. *Trends Microbiol.* **23**: 587–590.

Heldal,M. and Bratbak,G. (1991) Production and decay of viruses in aquatic environments. *Mar. Ecol. Prog. Ser.* **72**: 205–212.

Houldcroft,C.J. *et al.* (2017) Clinical and biological insights from viral genome sequencing. *Nat. Rev. Microbiol.* **15**: 183–192.

Knowles,B. *et al.* (2016) Lytic to temperate switching of viral communities. *Nature* **531**: 533–537.

Koskella,B. and Brockhurst,M.A. (2014) Bacteria-phage coevolution as a driver of ecological and evolutionary processes in microbial communities. *FEMS Microbiol. Rev.* **38**: 916–931.

Köstner,N. *et al.* (2017) High viral abundance as a consequence of low viral decay in the Baltic Sea redoxcline. *PLoS One* **12**: 1–21.

Kutter,E.M. *et al.* (2015) Re-establishing a place for phage therapy in western medicine. *Future Microbiol.* **10**: 685–688.

Labonté,J.M. *et al.* (2015) Single cell genomics-based analysis of virus-host interactions in marine surface bacterioplankton. *ISME J.* **9**: 2386–99.

Martínez-García,M. *et al.* (2014) Unveiling viral–host interactions within the 'microbial dark matter.' *Nat. Commun.* **5**: 1–8.

Meier-Kolthoff,J.P. and Göker,M. (2017) VICTOR: genome-based phylogeny and classification of prokaryotic viruses. *Bioinformatics* **33**: 3396–3404.

Mokili,J.L. *et al.* (2012) Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* **2**: 63–77.

Murray,A.G. and Jackson,G. a (1992) Viral dynamics : a model of the effects of size , shape , motion and abundance of single-celled planktonic organisms and other particles. *Mar. Ecol. Prog. Ser.* **89**: 103–116.

Nishimura,Y. *et al.* (2017) ViPTree: The viral proteomic tree server. *Bioinformatics* **33**: 2379–2380.

Ottesen,E.A. (2016) Probing the living ocean with ecogenomic sensors. *Curr. Opin. Microbiol.* **31**: 132–139.

Ovchinnikov,S. *et al.* (2017) Protein structure determination using metagenome sequence data. *Science* **355**: 294–298.

Páez-Espino,D. *et al.* (2016) IMG/VR: a database of cultured and uncultured DNA Viruses and retroviruses. *Nucleic Acids Res.* **45**: D457–D465.

Le Page,D.P. *et al.* (2017) Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. *Nature* **543**: 243–247.

Pope,W.H. *et al.* (2015) Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity. *Elife* **4**: e06416.

Powers,C.W. *et al.* (2018) Coordinated Sampling of microorganisms over freshwater and saltwater environments using an unmanned surface vehicle (USV) and a small unmanned aircraft system (sUAS). *Front. Microbiol.* **9**: 1–12.

Rohwer,F. and Vega Thurber,R. (2009) Viruses manipulate the marine environment. *Nature* **459**: 207–212.

Roux,S. *et al.* (2016) Ecogenomics and potential biogeochemical impacts of uncultivated globally abundant ocean viruses. *Nature* **537**: 689–93.

Roux,S. *et al.* (2014) Ecology and evolution of viruses infecting uncultivated SUP05 bacteria as revealed by single-cell- and meta- genomics. *Elife* **3**: e03125.

Roux,S. *et al.* (2018) Minimum Information about an Uncultivated Virus Genome (MIUViG). *Nat. Biotechnol.* **in press**:

Sepulveda,B.P. *et al.* (2016) Marine phage genomics : the tip of the iceberg. *FEMS Microbiol. Lett.* **363**: fnw158.

Smanski,M.J. *et al.* (2016) Synthetic biology to access and expand nature's chemical diversity. *Nat. Rev. Microbiol.* **14**: 135–149.

Spencer,S.J. *et al.* (2016) Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers. *ISME J.* **10**: 427–436.

Sullivan,M.B. *et al.* (2017) Viral ecology comes of age. *Environ. Microbiol. Rep.* **9**: 33–35.

Suttle,C.A. (2007) Marine viruses--major players in the global ecosystem. *Nat. Rev. Microbiol.* **5**: 801–812.

Suttle,C.A. and Chen,F. (1992) Mechanisms and rates of decay of marine viruses in seawater. *Appl. Environ. Microbiol.* **58**: 3721–3729.

Tadmor,A.D. *et al.* (2011) Probing Individual Environmental Bacteria for Viruses by Using Microfluidic Digital PCR. *Science* **333**: 58–62.

Thingstad,T.F. (2000) Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol. Oceanogr.* **45**: 1320–1328.

Weinbauer,M.G. and Rassoulzadegan,F. (2004) Are viruses driving microbial diversification and diversity? *Environ. Microbiol.* **6**: 1–11.

Weitz,J.S. *et al.* (2017) Lysis, lysogeny and virus-microbe ratios. *Nature* **549**: E1–E3.

Wigington,C.H. *et al.* (2016) Re-examining the relationship between virus and microbial cell abundances. *Nat. Microbiol.* **1**:.

285 Williamson,K.E. *et al.* (2017) Viruses in Soil Ecosystems: An Unknown Quantity Within an Unexplored Territory. *Annu. Rev. Virol.* **4**: 201–219.

Young,R. and Gill,J.J. (2015) Phage therapy redux — what is to be done? *Science* **350**: 1164–4.

Ziemert,N. *et al.* (2016) The evolution of genome mining in microbes-a review. *Nat. Prod. Rep.* **33**: 988–1005.