



# A Viral Ecogenomics Framework To Uncover the Secrets of Nature's "Microbe Whisperers"

 Simon Roux<sup>a</sup>

<sup>a</sup>DOE Joint Genome Institute, Walnut Creek, California, USA

**ABSTRACT** Microbes drive critical ecosystem functions and affect global nutrient cycling along with human health and disease. They do so under strong constraints exerted by viruses, which shape microbial communities' structure and shift host cell metabolism during infection. While the majority of viruses and their associated impacts remain poorly characterized, a number of mechanisms by which viruses alter microbial cells and ecosystems have already been revealed. Here I outline how a comprehensive host-resolved mapping of viral sequence space will enable a thorough characterization of virus-encoded mechanisms for microbial manipulation. With soon-to-be millions of virus genomes obtained from metagenomes, one of the major challenges resides in the development of methods for high-throughput and high-resolution virus-host pairing, before multi-omics approaches can be leveraged to fully decipher virus-host dynamics in nature. Beyond novel fundamental biological knowledge, these studies will likely provide new molecular tools enabling a precise engineering of microbial cells and communities.

**KEYWORDS** ecogenomics, metagenomics, phage, virus

Microbes are now recognized to play key roles in all of Earth's ecosystems, driving nutrient and energy transfers, and directly influencing human health and disease (1). These microbial processes are strongly constrained by viruses, which are globally abundant and infect all cellular life forms. Because of technical challenges limiting our ability to cultivate viruses infecting these microbes and, until recently, to even survey their diversity in nature, the vast majority of viruses and their associated impacts remain poorly understood. Nevertheless, a number of mechanisms by which viruses profoundly influence microbial ecosystems have already been revealed or predicted (2, 3). The most intuitive impact of viral infection is linked to virus-induced mortality which can trigger large-scale reshuffling of microbial communities, although the associated ecological and evolutionary drivers are not fully characterized yet. However, viruses can also modify host cell metabolism and alter host cell fitness during infection.

Modification of host cell properties during viral infection has been primarily observed and studied for temperate bacteriophages integrated into the host's genome (prophages) (3, 4). Prophages were found to alter their host cell features through gene disruption (when integrated), protection from infection by similar and/or unrelated phages (superinfection exclusion), and introduction of new genes either through horizontal transfer or directly expressed from the prophage (lysogenic conversion). The latter notably includes virus-borne toxin genes leading to increased virulence for major bacterial pathogens such as *Vibrio cholerae*, *Clostridium botulinum*, or *Escherichia coli* (5). Meanwhile, the protection provided by phages against additional infections has consequences not just for individual host cells but also for the whole community dynamics by altering virus/host ratios and global infection efficiency (6). Beyond temperate phages, fitness-altering genes have been discovered in lytic viruses, i.e., viruses whose infection cycle does not include a latent phase but who instead directly replicate their

**Citation** Roux S. 2019. A viral ecogenomics framework to uncover the secrets of nature's "microbe whisperers." mSystems 4:e00111-19. <https://doi.org/10.1128/mSystems.00111-19>.

**Copyright** © 2019 Roux. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to [sroux@lbl.gov](mailto:sroux@lbl.gov).

Conflict of Interest Disclosures: S.R. has nothing to disclose.

mSystems® vol. 4, no. 3, is a special issue sponsored by Illumina.

 Understanding the myriad ways by which viruses alter (more or less subtly) microbial cells and communities in nature is key for modeling microbial ecosystems and a potential source of new approaches to manipulate microbial consortia.

**Received** 15 February 2019

**Accepted** 19 March 2019

**Published** 14 May 2019

genome and produce new virus particles upon entry in the host cell. Photosystem genes carried in lytic cyanophages were the first example of host cell reprogramming by lytic viruses to be comprehensively characterized (7), but other genes potentially regulating nitrogen, sulfur, phosphorus, and central carbon metabolisms have since been identified. Collectively, these studies suggest that viruses, both temperate and lytic, manipulate a broad range of host cell features.

Because exploring environmental viral communities has remained a challenging task until very recently, our collective understanding of the molecular mechanisms by which viruses can influence microbial cells is still largely partial, and known examples are mostly derived from a few well-studied cultivated model systems. Interrogating the vast diversity of uncultivated viruses holds tremendous potential for the discovery of entirely novel avenues for microbial manipulation, including strategies tailored to specific hosts and environmental conditions. Fortunately, this type of approach might become viable in a not-too-distant future considering the current pace of progress and improvements in the field of viral ecogenomics. Here I outline the major steps and remaining challenges toward a comprehensive characterization of viral diversity and virus-host interactions in nature and the potential use of this knowledge for manipulation of microbial communities.

### A GLOBAL MAP OF EARTH'S VIRUSES

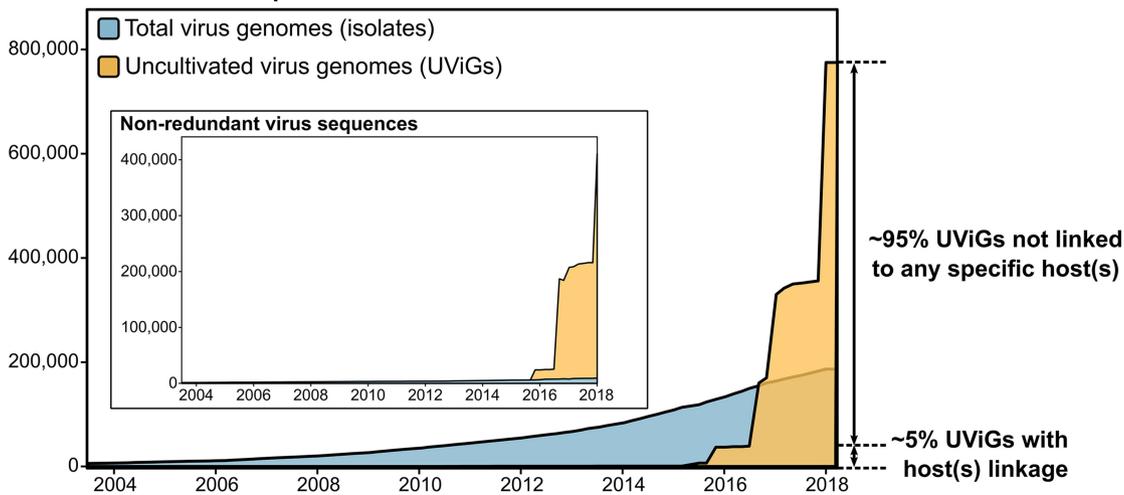
In order to fully understand and leverage the impact of viruses on microbial cells, a large-scale mapping of viral sequence space covering most (if not all) of viral diversity must first be achieved. This is the topic that saw the most remarkable improvement in the last few years, especially compared to the exploration of microbial diversity, mainly due to a shift in the methods used. The vast diversity of uncultivated microbes was initially revealed through 16S and 18S amplicon sequencing (8); however, this approach cannot be applied to viruses as they lack a universally conserved gene. Hence, while amplicon-based surveys of microbial communities were comprehensive and revealed large portions of novel diversity, their viral counterparts focused on small groups of known viruses. Uncultivated organisms are, however, increasingly studied through *de novo* genome assembly from shotgun metagenomes (9). In contrast to amplicon sequencing, viruses are often more easily amenable to genome assembly from metagenomes than microbes, primarily because viruses tend to have smaller genomes, a lower repeat content, and higher levels of differentiation between genotypes (10).

Pragmatically, this means that the advent of genome assembly from metagenomes has been particularly impactful for the exploration of viral diversity (Fig. 1). Currently, the largest database of uncultivated virus genomes (IMG/VR; <https://img.jgi.doe.gov/cgi-bin/vr/main.cgi>) is composed of >700,000 uncultivated genome fragments, the overwhelming majority collected in the last 2 years alone (11). Among these, >25,000 were estimated to represent near-complete or complete genomes ( $\geq 90\%$  complete, "high-quality viral genomes" sensu [12]). This means that metagenome assemblies have already yielded more than twice the total number of (near-)complete genomes obtained from isolates. In addition, complementary methods already exist for viruses recalcitrant to assembly from metagenomes, including viral single-amplified genomes (13) and single-template long-read sequencing (14). The pace at which data accumulate and new complementary methods are developed suggests that an exhaustive mapping of virus genome diversity is on the horizon.

### ESTABLISHING HIGH-RESOLUTION VIRUS-HOST LINKAGES

One of the main pitfalls of genome assembly from metagenomes is that, in contrast to viral isolates, uncultivated viruses are not immediately associated with a host. This host linkage information is critical to identify and interpret virus impacts on microbial ecosystems. Currently, most of this virus-host linkage is based on computational predictions using methods still being actively developed and improved, including identification of horizontal gene transfers between virus and hosts, sequence similarity between host CRISPR spacers and viral genomes, and detection of nucleotide compo-

**Total number of virus sequences**



**FIG 1** Size of virus genome databases over time. The total number of genomes from isolates was based on queries to the NCBI nucleotide database portal, while the number of uncultivated virus genomes (UViGs) was estimated by compiling data from the literature and from the IMG/VR database, as in reference 12. The inset displays the same data after dereplication of the genomes, i.e., one per species for NCBI nucleotide database, and clustering based on pairwise nucleotide similarity for UViGs using cutoffs of 95% average nucleotide identity and 85% alignment fraction (12). Host linkage information is from the IMG/VR database and was derived from matches between UViGs and isolate viruses, prophages, and CRISPR spacers.

sition (kmer) similarity between host and virus genomes (15). While these approaches are promising, they remain limited in scope and resolution. The vast majority of uncultivated virus sequences remain without any host (Fig. 1), and *in silico* predictions typically provide information on host taxonomy at the family or genus ranks, more rarely down to species, while virus-host interactions tend to be strain specific.

One of the major challenges faced by the field of viral ecogenomics is thus to develop experimental approaches providing high-throughput, high-resolution host linkage for uncultivated viruses. These approaches will most likely rely on high-throughput sequencing of single cells and/or targeted host (sub)populations, sorted and selected via a combination of, e.g., microfluidic separation, microdroplet encapsulation, and/or flow cytometry cell sorting. The codetection of virus and host at the individual cell level should then provide unique insights into strain-resolved host range and infection rates for individual virus genotypes in nature, without cultivation of either the microbial host or the virus. Given that pilot experiments are already ongoing, it seems reasonable to hope for some of these methods to be established and broadly available at some point in the next decade. The data obtained by applying these methods across a large range of biomes and ecological conditions will then enable the construction of a global host-resolved mapping of the virus sequence space.

**VIRUS-INSPIRED SOLUTIONS FOR MICROBE MANIPULATION**

Finally, a global host-contextualized mapping of the virosphere would serve as a foundation for multi-omics studies aiming at deciphering virus-host interactions and dynamics in nature. Sample sets such as time series, patient cohorts, and incubations under controlled conditions, either in the laboratory or *in situ*, will be especially relevant in this context. Beyond a treasure trove of fundamental biological discoveries about ecological and evolutionary drivers of virus-host dynamics, these studies would also benefit every approach and biotechnological application leveraging microbes and microbial processes. Specifically, a comprehensive understanding of the broad range of virus-host interactions ongoing in nature would provide a fantastic resource that could be mined for new ways to precisely and predictably engineer microbial cells, including for targeted modification of specific populations within a complex microbial community.

Already, results from isolated viruses can provide a glimpse at the type of mecha-

nisms and solutions for microbial manipulations that could be inspired by or derived from viruses. The first and most obvious application of viruses would be to kill specific microbial strains such as contaminants, either through active viruses, i.e., “phage therapy,” or by using only specific components of virus particles, e.g., tailocins. Viruses can also contain many key genes able to redirect entire metabolic pathways or to fundamentally alter host behavior and fitness, including variants of key cellular enzymes displaying unique properties as well as highly efficient transcriptional and translational regulators. Hence, a global and host-contextualized virus genome database could be used to identify a candidate gene(s) for modulating specific metabolisms or expressing artificial constructs in any target microbe. Finally, some viruses are particularly amenable to genome engineering and could thus be repurposed as custom-targeted delivery mechanisms. Members of the *Inoviridae* family, i.e., filamentous phages, are particularly interesting in this framework because they have a small genome which can be expressed in a nonhost cell and relatively simple virions whose properties could likely be modified by mutating only one or a couple of genes.

## CONCLUSION

Right now is an exciting time to be part of (or join) the young and highly dynamic field of virus ecogenomics. A lot of the infrastructure and conceptual framework is being built (12), and data pile up at an incredible rate which constantly enables new analyses at unprecedented scales. While fantastic discoveries have already been made, what is maybe even more impressive is the incredible potential of virus ecogenomics approaches to achieve a more detailed characterization of virus impacts on microbial ecosystems, of their role(s) in the evolution of life on earth, and of their potential for improving microbially based biotechnological applications. While the path and prospects outlined here will require important methodological and conceptual breakthroughs, it is nonetheless difficult not to feel a sense of awe (or even dizziness) when contemplating the extent of novel viruses that remain to be unveiled and all their intricate microbial manipulation mechanisms still to be discovered and leveraged.

## ACKNOWLEDGMENTS

I thank Nicole Webster for the invitation to contribute to this special issue and colleagues at the JGI (especially in the Environmental Genomics group), in the Biosciences Area of Lawrence Berkeley National Laboratory, and in viral ecogenomics laboratories around the world for the numerous discussions on the topics discussed here that fueled this perspective.

This work was conducted at the U.S. Department of Energy Joint Genome Institute and supported by the Office of Science of the U.S. Department of Energy and the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory, under contract no. DE-AC02-05CH11231.

## REFERENCES

1. Falkowski PG, Fenchel T, Delong EF. 2008. The microbial engines that drive earth's biogeochemical cycles. *Science* 320:1034–1039. <https://doi.org/10.1126/science.1153213>.
2. Breitbart M, Bonnain C, Malki K, Sawaya NA. 2018. Phage puppet masters of the marine microbial realm. *Nat Microbiol* 3:754–766. <https://doi.org/10.1038/s41564-018-0166-y>.
3. Argov T, Azulay G, Pasechnek A, Stadnyuk O, Ran-Sapir S, Borovok I, Sigal N, Herskovits AA. 2017. Temperate bacteriophages as regulators of host behavior. *Curr Opin Microbiol* 38:81–87. <https://doi.org/10.1016/j.mib.2017.05.002>.
4. Bondy-Denomy J, Davidson AR. 2014. When a virus is not a parasite: the beneficial effects of prophages on bacterial fitness. *J Microbiol* 52: 235–242. <https://doi.org/10.1007/s12275-014-4083-3>.
5. Brüssow H, Canchaya C, Hardt W. 2004. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 68:560–602. <https://doi.org/10.1128/MMBR.68.3.560-602.2004>.
6. Seed KD. 2015. Battling phages: how bacteria defend against viral attack. *PLoS Pathog* 11:e1004847. <https://doi.org/10.1371/journal.ppat.1004847>.
7. Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, Kettler G, Sullivan MB, Steen R, Hess WR, Church GM, Chisholm SW. 2007. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* 449:83–86. <https://doi.org/10.1038/nature06130>.
8. Hugenholtz P, Goebel BM, Pace NR. 1998. Impact of culture independent studies on the emerging phylogenetic view of bacterial diversity. *J Bacteriol* 180:4765–4774.
9. Parks DH, Rinke C, Chuvochina M, Chaumeil P, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-

- assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>.
10. Roux S, Emerson JB, Eloe-Fadrosh EA, Sullivan MB. 2017. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ* 5:e3817. <https://doi.org/10.7717/peerj.3817>.
  11. Paez-Espino D, Roux S, Chen I-M, Palaniappan K, Ratner A, Chu K, Hunt-emann M, Reddy TBK, Pons JC, Llabrés M, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. 2019. IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Res* 47(D1):D678–D686. <https://doi.org/10.1093/nar/gky1127>.
  12. Roux S, Adriaenssens EM, Dutilh BE, Koonin EV, Kropinski AM, Krupović M, Kuhn JH, Lavigne R, Brister JR, Varsani A, Amid C, Aziz RK, Bordenstein SR, Bork P, Breitbart M, Cochrane G, Daly RA, Desnues C, Duhaime MB, Emerson JB, Enault F, Fuhrman JA, Hingamp P, Hugenholtz P, Hurwitz BL, Ivanova NN, Labonté JM, Lee K-B, Malmstrom RR, Martinez-Garcia M, Karsch-Mizrachi I, Ogata H, Paez-Espino D, Petit M-A, Putonti C, Rattei T, Reyes A, Rodriguez-Valera F, Rosario K, Schriml L, Schulz F, Steward GF, Sullivan MB, Sunagawa S, Suttle CA, Temperton B, Tringe SG, Vega-Thurber R, Webster NS, Whiteson KL, Wilhelm SW, Wommack KE, Woyke T, Wrighton KC, Yilmaz P, Yoshida T, Young MJ, Yutin N, Zeigler Allen L, Kyrpides NC, Eloe-Fadrosh EA. 2018. Minimum information about an uncultivated virus genome (MIUViG). *Nat Biotechnol* 37:29–37. <https://doi.org/10.1038/nbt.4306>.
  13. Martinez-Hernandez F, Fornas O, Lluesma M, Bolduc B, Cruz M, Martinez Martinez J, Anton J, Gasol J, Rosselli R, Rodriguez-Valera R, Sullivan MB, Acinas SG, Martinez-Garcia M. 2017. Single-virus genomics reveals hidden cosmopolitan and abundant viruses. *Nat Commun* 8:15892. <https://doi.org/10.1038/ncomms15892>.
  14. Warwick-Dugdale J, Solonenko N, Moore K, Chittick L, Gregory AC, Allen MJ, Sullivan MB, Temperton B. 2018. Long-read viral metagenomics enables capture of abundant and microdiverse viral populations and their niche-defining genomic islands. *bioRxiv* <https://doi.org/10.1101/345041>.
  15. Edwards RA, McNair K, Faust K, Raes J, Dutilh BE. 2016. Computational approaches to predict bacteriophage-host relationships. *FEMS Microbiol Rev* 40:258–272. <https://doi.org/10.1093/femsre/fuv048>.