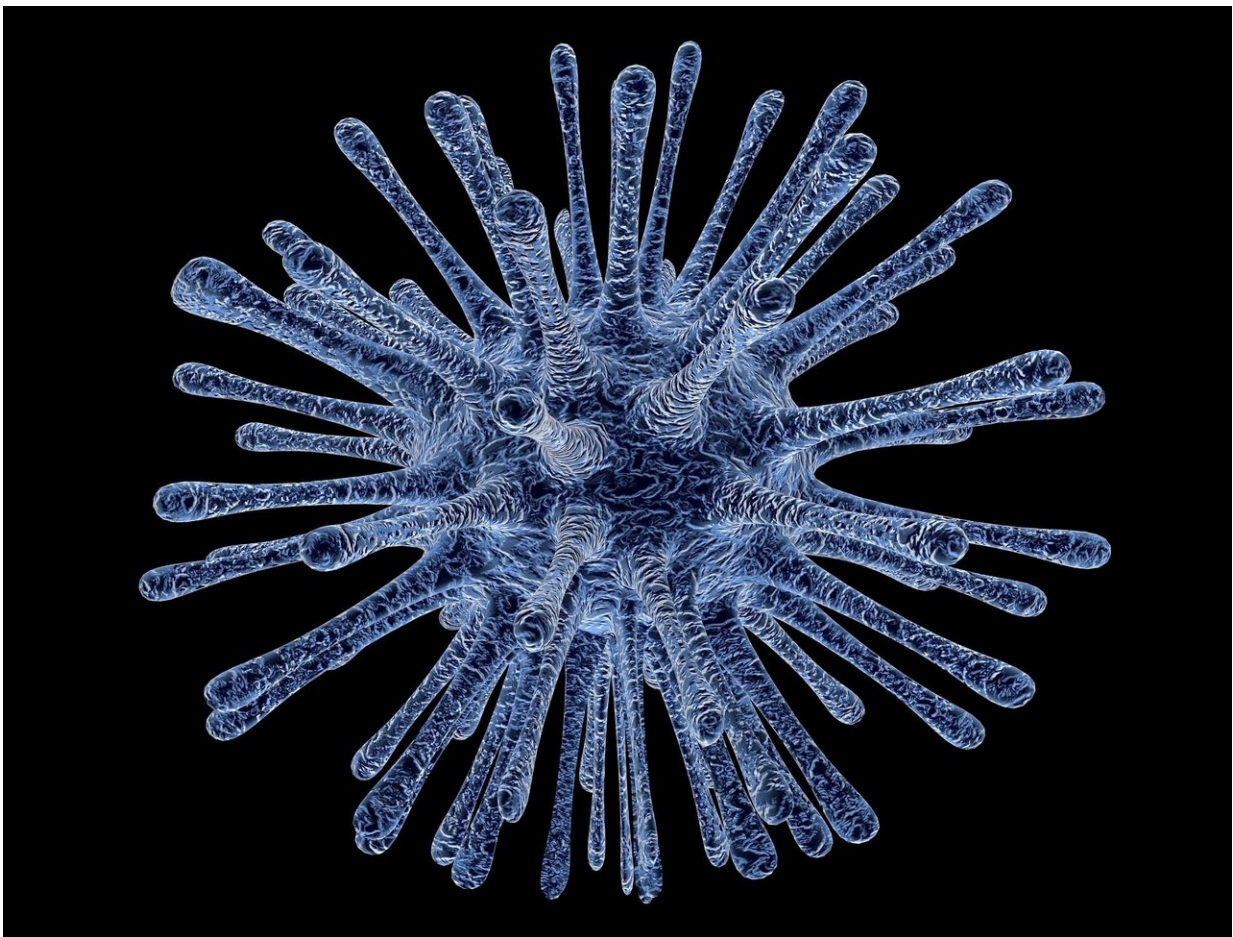


# International consortium offers guidelines, best practices for characterizing uncultivated viruses

December 17 2018

---



Credit: CC0 Public Domain

Microbes in, on and around the planet are said to outnumber the stars in the Milky Way Galaxy. The total number of viruses is expected to vastly exceed even that calculation.

While many viruses remain unknown and uncultivated, advances in [genome sequencing](#) and analyses have allowed researchers to identify more than 750,000 uncultivated virus genomes from metagenomic and metatranscriptomic data sets. In IMG/VR, a database for virus sequences established and maintained by researchers at the U.S. Department of Energy (DOE) Joint Genome Institute (JGI), a DOE Office of Science User Facility, the viral diversity available has tripled within a single year.

As more and more researchers continue to assemble new genome sequences of uncultivated viruses, JGI researchers led a community effort to develop guidelines and best practices for defining virus data quality. In a report published December 17, 2018, in *Nature Biotechnology*, JGI partnered with a number of virus experts; as well as representatives from the Genomic Standards Consortium (GSC), an open-membership working body that engages the [research community](#) in the standards development process; and the International Committee on Taxonomy of Viruses, the premier authority on the official taxonomy of viruses which is currently re-evaluating virus classification based on sequence-based information.

## Guidelines for Quality and Analyses

"Viruses are critical components of every microbial ecosystem. The JGI is especially interested in developing standards for virus genomes because we generate much of these data ourselves," said JGI research scientist and first author Simon Roux. "We are part of a small group of researchers who have scrutinized these data at length, have seen the metrics, and can provide guidance to help determine data quality. Additionally, in this paper, we've tried to provide not just standards, but

also outline what type of analyses can be performed on these data, to help researchers who want to characterize their own novel viruses."

Cultured viruses already have their own data quality standards, but these cannot be directly applied to uncultured viruses, whose sequences are often incomplete and for which some properties can only be predicted indirectly using computational approaches.

"The uncultivated virus genome community has come together to define what is important to report and valuable to the research community," said GSC President Lynn Schriml of the Institute of Genome Sciences at University of Maryland School of Medicine. The GSC includes representatives from the National Center for Biotechnology Information (NCBI), the European Bioinformatics Institute, and the DNA Data Bank of Japan (DDBJ), who also collaborated on this article.

## **Categories of Virus Genome Quality**

In the paper, Roux and his colleagues outlined the minimum amount of information for an uncultivated virus genome, including the source, methods of identification of the virus genome, and data quality. The JGI has previously developed standards for the minimum metadata to be reported with single amplified genomes (SAGs) and metagenome-assembled genomes (MAGs) submitted to public databases.

"The tremendous growth of virus sequence data, and microbiome data in general, necessitates robust standards and data quality metrics to allow the research community to leverage this data for comparative analyses," said JGI Metagenome Program head and study senior author Emiley Eloë-Fadrosch. "By establishing and promoting '[best practices](#),' the research community has the opportunity to break down barriers of data accessibility and reusability, thereby amplifying the research beyond the initial project scope."

The team proposed three categories of genome quality. "Genome fragments" are comprised of single or multiple fragments that are predicted to be less than 90 percent complete, or have no estimated genome size, and are minimally annotated. A "high-quality draft genome" is estimated to represent 90 percent or more of the complete expected genome sequence, in fragment(s) where any gaps span mostly repetitive regions. Finally, a "finished genome" would include both a complete genome comprised of a single contiguous sequence without gaps, and extensive annotation.

"If you're going to build a standard," Schriml noted, "it is essential to discuss what should be represented with the research community, taxonomists and database providers and to integrate these data needs into the standard." Schriml added journals have also started endorsing the application of the GSC's "Minimum Information about any (X) Sequence (MIxS)" guidelines, the umbrella under which the uncultivated [virus genome](#) standards and other similar community efforts reside. The GSC tracks the adoption of these standards developed over the past decade using records uploaded to the BioSample database. These records reflect individual samples collected, sequenced and annotated, and Schriml said that nearly 450,000 BioSample records currently reference MIxS guidelines, up from 326,000 records tracked in the spring.

**More information:** Minimum Information about an Uncultivated Virus Genome (MIUViG) *Nature Biotechnology* (2018). [DOI: 10.1038/nbt.4306](https://doi.org/10.1038/nbt.4306)

Provided by Lawrence Berkeley National Laboratory

Citation: International consortium offers guidelines, best practices for characterizing uncultivated viruses (2018, December 17) retrieved 17 July 2024 from

<https://phys.org/news/2018-12-international-consortium-guidelines-characterizing-uncultivated.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.