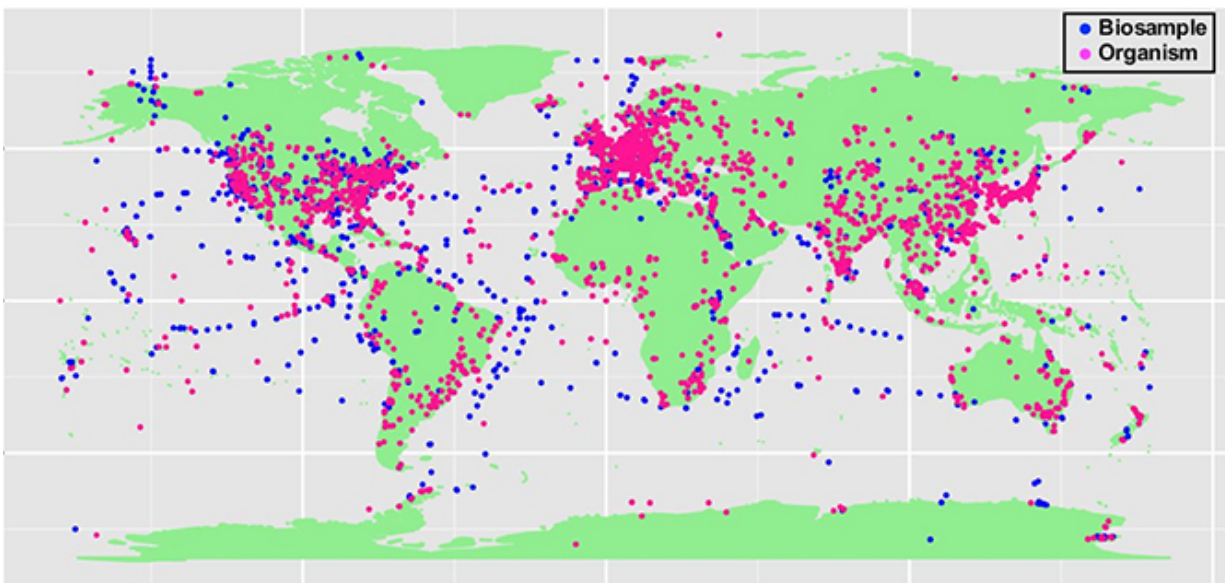


New database of DNA viruses and retroviruses debuts

March 6 2017



Geographic distribution of biosamples and organisms encompassed by the Genomes OnLine Database. The location where the organism was isolated is marked with pink dots; the location of the biosample is marked with blue dots. Credit: US Department of Energy

There are more microbes in, on, and around the planet than there are stars in the Milky Way. Microbes affect food production; air quality; natural breakdown of plants, trees and biomass; soil quality for agriculture; and much more. To work with these microbes, scientists need to learn more about how microbes and viruses interact. Viruses

influence microbes' abilities to work. Scientists at the U.S. Department of Energy (DOE) Joint Genome Institute built the largest publicly available database for viruses. This single effort increases the number of known viral genes by a factor of 16. Further, in a series of four articles published in *Nucleic Acids Research*, DOE Joint Genome Institute researchers report on the latest updates to several other publicly accessible databases and computational tools. These databases and tools will benefit the global community of microbial researchers.

Microbes play key roles in the planet's biogeochemical cycles. Viruses, thought to outnumber [microbes](#) by 10 fold, can alter those cycles. Advances in sequencing technologies have generated vast amounts of data about the viruses. Working with the data requires tools to manage and interpret it. The updated database analytical tools offer details on microbes, specifically microbial genomics, and viruses. A new database, available to the public, enables analyses that can help scientists answer questions about the viruses that infect microbes, an interaction that affects countless natural processes and may provide tools for speeding up beneficial (or slowing down harmful) microbial actions. The insights gained from this work are relevant to DOE missions in bioenergy and environment.

Providing high-quality, publicly accessible sequence data goes hand-in-hand with developing and maintaining the databases and tools that the research community can harness to help answer scientific questions. In a recent series of articles published in *Nucleic Acids Research*, researchers at DOE's Joint Genome Institute, a national scientific user facility, describe a database called Integrated Microbial Genomes with Virus Samples (IMG/VR). IMG/VR is a comprehensive computational platform integrating all the sequences in the database with associated metadata and analytical tools.

IMG/VR follows on the heels of a recent DOE Joint Genome Institute

viral diversity report in Nature. Additional articles in the same issue describe updates to several publicly accessible, interactive databases since the last set of reports published in 2014. For example, as of July 2016, there were 47,516 archaeal, bacterial, and eukaryotic genomes in the IMG with Microbiome Samples (IMG/M) system, with researchers noting that number "represents an over 300% increase since September 2013." IMG/M contains annotated DNA and RNA sequence data of archaeal, bacterial, eukaryotic, and viral genomes from cultured organisms; single cell genomes (SCG) and genomes from metagenomes from uncultured archaea, bacteria, and [viruses](#); and metagenomes from environmental, host-associated, and engineered microbiome samples.

Another paper concerns the Genomes OnLine Database (GOLD), a manually curated data management system that catalogs sequencing projects with associated metadata from around the world. In the current version of GOLD (v.6), all projects are organized based on a four-level classification system in the form of a study, organism (for isolates) or biosample (for environmental samples), sequencing project, and analysis project. A fourth paper focuses on the IMG Atlas of Biosynthetic gene Clusters (IMG-ABC). Launched in 2015, IMG-ABC enables researchers to search for biosynthetic gene clusters and secondary metabolites. Their latest update now incorporates ClusterScout, a tool for targeted identification of custom biosynthetic gene clusters across several thousand isolate microbial genomes, as well as a new search capability.

More information: I-Min A. Chen et al. IMG/M: integrated genome and metagenome comparative data analysis system, *Nucleic Acids Research* (2017). [DOI: 10.1093/nar/gkw929](https://doi.org/10.1093/nar/gkw929)

Supratim Mukherjee et al. Genomes OnLine Database (GOLD) v.6: data updates and feature enhancements, *Nucleic Acids Research* (2017). [DOI: 10.1093/nar/gkw992](https://doi.org/10.1093/nar/gkw992)

David Paez-Espino et al. IMG/VR: a database of cultured and uncultured DNA Viruses and retroviruses, *Nucleic Acids Research* (2017). [DOI: 10.1093/nar/gkw1030](https://doi.org/10.1093/nar/gkw1030)

Provided by US Department of Energy

Citation: New database of DNA viruses and retroviruses debuts (2017, March 6) retrieved 18 April 2024 from <https://phys.org/news/2017-03-database-dna-viruses-retroviruses-debuts.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.