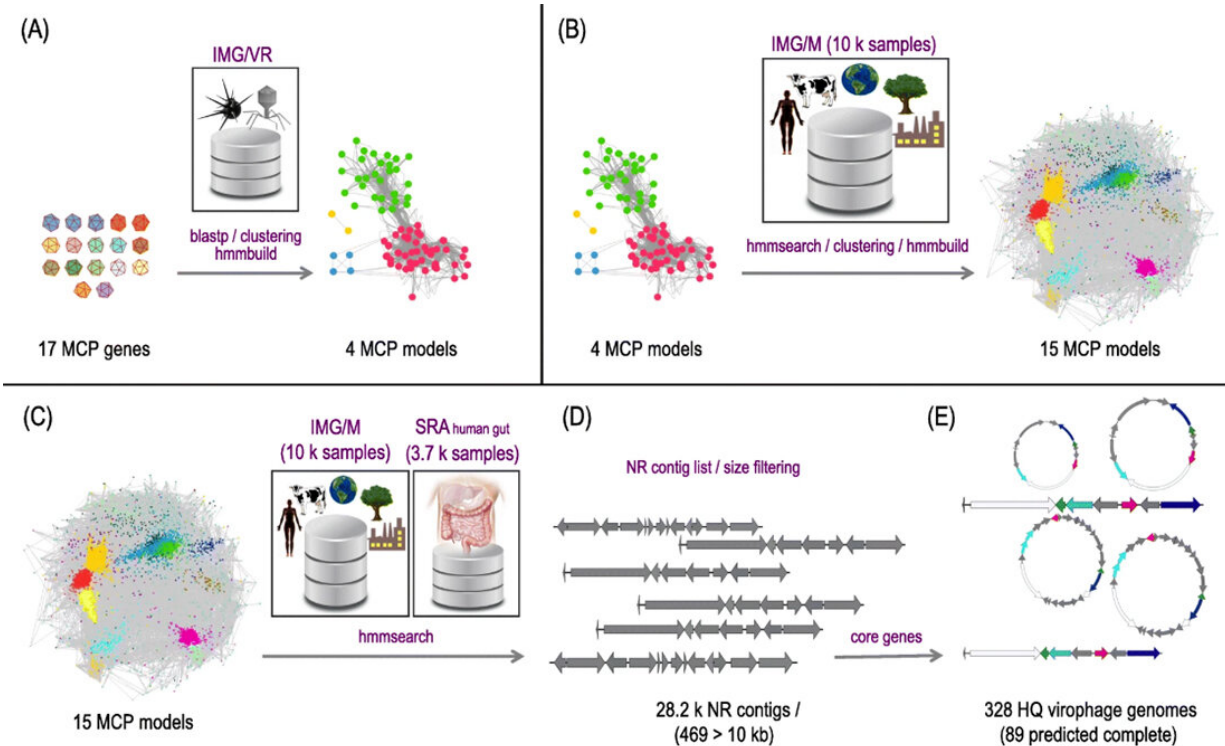


Expanding virophage diversity

December 24 2019



Virophage discovery pipeline. (A) MCP amino acid sequences from reference isolated genomes and published metagenomic contigs were queried against the IMG/VR database with stringent e value cutoffs. All homologous sequences detected were then clustered together to build four independent MCP profiles. (B) The resulting four MCP models were used to recruit additional homologous sequences from the entire IMG/M system. All new sequences were clustered, and models were built creating a final set of 15 unique MCP HMMs. (C) These 15 unique MCP HMMs were then used to search two different databases for homologous sequences: the IMG/M system and a custom assembled human gut database containing 3771 samples from NCBI’s Sequence Read Archive (SRA). (D) The resulting set of 28,294 non-redundant (NR) sequences with stringent e

value cutoffs was filtered by size and (E) by the presence of the four core virophage genes (high-quality genomes; HQ virophages). Finally, completeness of novel metagenomic virophage genomes was predicted based on circularity or presence of inverted terminal repeats (ITR). Credit: Figure from Paez-Espino et al. *Microbiome* (2019) 7:157 <https://doi.org/10.1186/s40168-019-0768-5>

Virophages are small viruses with double-stranded DNA genomes that co-infect eukaryotic cells along with giant viruses. Almost all known virophage genomes share only four genes in common: major and minor capsid proteins (MCP and mCP, respectively), ATPase involved in DNA packaging, and PRO, a cysteine protease involved in capsid maturation.

Recently reported in *Microbiome*, researchers from the US Department of Energy (DOE) Joint Genome Institute (JGI), a DOE Office of Science facility, have increased the number of known high quality [virophage](#) genome sequences 10-fold through computational approaches. By mining more than 14,000 publicly available metagenomic datasets in JGI's Integrated Microbial Genomes & Microbiomes (IMG/M) data suite—which includes IMG/VR (for Virus)—for the virophage marker gene MCP, they were able to identify 44,221 total virophage partial sequences, including over 28,000 unique MCP sequences.

Further analysis led to the identification of 328 "high quality" (based on completeness) diverse new virophage genomes containing all four core genes. These virophages were found in diverse habitats including the air, plant rhizosphere, wastewater, and even animal and human (for the first time) gut and taxonomically classified into 27 distinct clades (17 of them without previously known representatives). Of these, 89 contigs were considered to be complete genomes, and their discovery has extended the possible virophage genome size range from 13.8-29.3 kilobases (Kb) to 10.9-42.3 Kb. Additionally, the gene counts have similarly gone up

from 13-25 to 12-39. "Overall," the team concluded, "we provide a global analysis of the diversity, distribution, and evolution of virophages."

More information: David Paez-Espino et al. Diversity, evolution, and classification of virophages uncovered through global metagenomics, *Microbiome* (2019). [DOI: 10.1186/s40168-019-0768-5](https://doi.org/10.1186/s40168-019-0768-5)

Provided by DOE/Joint Genome Institute

Citation: Expanding virophage diversity (2019, December 24) retrieved 23 June 2024 from <https://phys.org/news/2019-12-virophage-diversity.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.