


RESEARCH

Open Access



Diversity, evolution, and classification of virophages uncovered through global metagenomics

David Paez-Espino^{1*†} , Jinglie Zhou^{1†}, Simon Roux¹, Stephen Nayfach¹, Georgios A. Pavlopoulos^{1,2}, Frederik Schulz¹, Katherine D. McMahon³, David Walsh⁴, Tanja Woyke¹, Natalia N. Ivanova¹, Emiley A. Eloie-Fadros¹, Susannah G. Tringe^{1*} and Nikos C. Kyrpides^{1*}

Abstract

Background: Virophages are small viruses with double-stranded DNA genomes that replicate along with giant viruses and co-infect eukaryotic cells. Due to the paucity of virophage reference genomes, a collective understanding of the global virophage diversity, distribution, and evolution is lacking.

Results: Here we screened a public collection of over 14,000 metagenomes using the virophage-specific major capsid protein (MCP) as “bait.” We identified 44,221 assembled virophage sequences, of which 328 represent high-quality (complete or near-complete) genomes from diverse habitats including the human gut, plant rhizosphere, and terrestrial subsurface. Comparative genomic analysis confirmed the presence of four core genes in a conserved block. We used these genes to establish a revised virophage classification including 27 clades with consistent genome length, gene content, and habitat distribution. Moreover, for eight high-quality virophage genomes, we computationally predicted putative eukaryotic virus hosts.

Conclusion: Overall, our approach has increased the number of known virophage genomes by 10-fold and revealed patterns of genome evolution and global virophage distribution. We anticipate that the expanded diversity presented here will provide the backbone for further virophage studies.

Keywords: Metagenomics, Virophage, Major capsid protein (MCP), Virophage classification, Virophage-NCLDV interactions, Global distribution

Background

Virophages are a group of circular double-stranded DNA (dsDNA) viruses taxonomically classified within the *Lavidaviridae* family [1]. They **co-infect unicellular eukaryotic hosts with members of the *Mimiviridae* family, a group of nucleocytoplasmic large DNA viruses (NCLDV)** [2–4]. By siphoning off resources within the giant virus factory, virophage replication reduces the number of giant virus progeny, thereby increasing host survival [5].

Since 2008, when virophages were discovered in a water-cooling tower (virophage *Sputnik*) [5], genome sequences have been obtained for five cultured isolates: *Sputnik2* (from lens liquid), *Sputnik3* (from soil), *Mavirus* (from coastal waters), *Zamilon* (from soil), and *Zamilon2* (from a bioreactor) [5–9]. These five virophages have been classified into two genera: *Sputnikvirus* (including *Sputnik* and *Zamilon* genomes) and *Mavirus* [10]. All five reference isolated genomes lack an envelope, form small icosahedral capsids (diameter of 35–74 nm), and have genomes ranging from 17 to 19 kb in length [11].

The rate of discovery of new virophages recently took a big leap due to the recovery of 20 virophage genomes from metagenomes. Of these, 18 were identified in diverse lake microbiomes (from Antarctica [12, 13], China [14, 15], and the US [13, 16, 17]) and the remaining two

* Correspondence: adpaezespino@lbl.gov; sgtringe@lbl.gov; nc kyrpides@lbl.gov

[†]David Paez-Espino and Jinglie Zhou contributed equally to this work.

¹Department of Energy, Joint Genome Institute, 2800 Mitchell Dr., Walnut Creek 94598, USA

Full list of author information is available at the end of the article

