

The Virophage Family *Lavidaviridae*

Matthias G. Fischer^{1*}

¹Department of Biomolecular Mechanisms, Max Planck Institute for Medical Research, Heidelberg, Germany.

*Correspondence: mfischer@mpimf-heidelberg.mpg.de

<https://doi.org/10.21775/cimb.040.001>

Abstract

Double-stranded (ds) DNA viruses of the family *Lavidaviridae*, commonly known as virophages, are a fascinating group of eukaryotic viruses that depend on a coinfecting giant dsDNA virus of the *Mimiviridae* for their propagation. Instead of replicating in the nucleus, virophages multiply in the cytoplasmic virion factory of a coinfecting giant virus inside a phototrophic or heterotrophic protistal host cell. Virophages are parasites of giant viruses and can inhibit their replication, which may lead to increased survival rates of the infected host cell population. The genomes of virophages are 17–33 kilobase pairs (kbp) long and encode 16–34 proteins. Genetic signatures of virophages can be found in metagenomic datasets from various saltwater and freshwater environments around the planet. Most virophages share a set of conserved genes that code for a major and a minor capsid protein, a cysteine protease, a genome-packaging ATPase, and a superfamily 3 helicase, although the genomes are otherwise diverse and variable. *Lavidaviruses* share genes with other mobile genetic elements, suggesting that horizontal gene transfer and recombination have been major forces in shaping these viral genomes. Integrases are occasionally found in virophage genomes and enable these DNA viruses to persist as provirophages in the chromosomes of their viral and cellular hosts. As we watch the genetic diversity of this new viral family unfold through metagenomics, additional isolates are still lacking and critical questions regarding their infection cycle, host range, and ecology remain to be answered.

Introduction

Virophages are a recently discovered class of double-stranded (ds) DNA viruses that have evolved a dependency on complex dsDNA viruses of eukaryotes, so-called giant viruses. The discovery of giant viruses was therefore a prerequisite for the isolation and characterization of virophages (see Fig. 12.1), as shall be reviewed here briefly (See also Reteno et al., 2018).

In 1992, following a pneumonia outbreak in Bradford, England, an intra-amoebal parasite was isolated by the team of T.J. Rowbotham and given the name ‘Bradford coccus’. This microorganism was initially assumed to be a bacterium because of its size and positive Gram-stain reaction; however, attempts to amplify and analyse its 16S ribosomal DNA sequences failed. A decade later in the laboratory of Didier Raoult in Marseille, electron microscopy revealed that Bradford coccus was in fact a giant virus with a fibre-studded, 0.75 µm-diameter particle (La Scola et al., 2003). The serendipitous discovery of *Acanthamoeba polyphaga* mimivirus (family *Mimiviridae*, genus *Mimivirus*, species *Acanthamoeba polyphaga mimivirus*) and the analysis of its 1.2 million base pair dsDNA genome have profoundly changed our view of the viral world (Raoult et al., 2004). This virus exhibited a particle size and genome length that exceeded those of the smallest cellular organisms, blurring the boundary between viruses and cells. The finding that some mimivirus genes were homologous to genes that were previously known to occur only in cellular genomes, in particular four aminoacyl-tRNA synthetases and four translation factors, spurred evolutionary scenarios in which giant viruses were hypothesized to

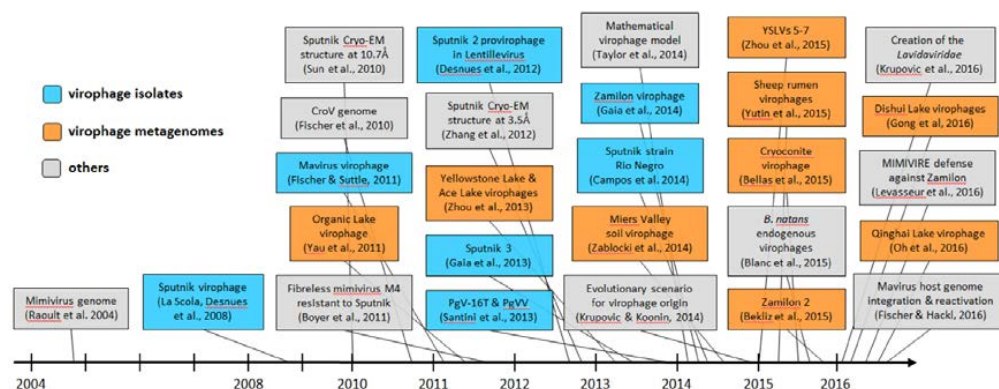


Figure 12.1 Historical time line of virophage-related publications.

be descendants of a cellular ancestor (Boyer *et al.*, 2010; Nasir *et al.*, 2012; Raoult *et al.*, 2004). Such hypotheses, however, have been widely criticized and the origin of giant viruses remains a matter of ongoing debate (Williams *et al.*, 2011; Yutin *et al.*, 2014). Akin to poxviruses (Broyles, 2003), mimivirus and related giant viruses replicate solely in the cytoplasm of their unicellular eukaryotic host (Mutsaers *et al.*, 2010), which is made possible by hundreds of virus-encoded proteins that provide nuclear functions to the virion factory (VF), such as DNA replication and transcription. For instance, the mimivirus-encoded transcription apparatus consists of at least eight DNA-dependent RNA polymerase subunits, a trifunctional mRNA capping enzyme, a polyadenylate polymerase, and several transcription factors.

A targeted search for giant viruses was launched following the characterization of mimivirus and resulted in the isolation of dozens of new viral strains (La Scola *et al.*, 2010). One of the first strains to be isolated was *Acanthamoeba castellanii* mamavirus, a close relative of mimivirus that was recovered from a cooling tower in Paris. Electron microscopy analysis of mamavirus-infected amoebae revealed the presence of a second, smaller icosahedral virus that was named Sputnik (La Scola *et al.*, 2008). Sputnik was unable to infect *Acanthamoeba* cells on its own, and replicated only when the cells were coinfecting with mamavirus. In coinfecting cells, Sputnik colocalized to the mamavirus VF, providing the first evidence that Sputnik uses giant virus-encoded enzymes for its propagation. The

presence of Sputnik also elicited a morphological phenotype in mamavirus, with visible deformations such as partial capsid thickening in many of the newly synthesized virions. The yield of mamavirus progeny from coinfecting cells was reduced by $\approx 70\%$ (La Scola *et al.*, 2008). Sputnik thus acted as a parasite of mimiviruses and the term ‘virophage’ was coined to reflect the relationship as a ‘virus of a virus’. However, the parasitic interaction appears to be restricted to the intracellular phase of the virus life cycle, when the biochemical complexity of giant viruses unfolds in the form of the VF organelle. In the years following the discovery of Sputnik, additional virophages such as mavirus (Fischer and Suttle, 2011) and Zamilon (Gaia *et al.*, 2014) were isolated, and more than a dozen putative virophage genomes were found in metagenomic datasets (Gong *et al.*, 2016; Oh *et al.*, 2016; Yau *et al.*, 2011; Yutin *et al.*, 2015a; Zhou *et al.*, 2013, 2015) (see Table 12.1). These eukaryotic dsDNA viruses are geographically widespread, genetically diverse, and appear to be commonly associated with giant virus-infected protist populations. The interested reader is referred to previously published review articles on this topic (Bekliz *et al.*, 2016; Claverie and Abergel, 2009b; Desnues and Raoult, 2010; Gaia *et al.*, 2013a).

Virion structure

Only few virophage representatives (Sputnik, mavirus, and Zamilon) have been isolated in laboratory culture and are thus amenable to structural studies.

The virophage particles that have been examined so far are 50–75 nm in diameter and possess icosahedral symmetry (Fig. 12.2). A 3.5 Å resolution cryo-electron microscopy (cryo-EM) reconstruction of the 75 nm wide Sputnik particle suggested the lack of a membrane component (Zhang *et al.*, 2012), contradicting earlier reports that assumed a lipid component in the Sputnik particle (Desnues *et al.*, 2012a; Sun *et al.*, 2010). The Sputnik capsid structure is composed of the major capsid protein (MCP) encoded by the V20 gene, and the minor capsid protein (mCP) encoded by the V18/V19 gene. The MCP contains a double jelly-roll fold and forms trimeric capsomers (hexons) with pseudo-hexagonal symmetry that build the 20 faces of the icosahedral particle (Fig. 12.2C), whereas the mCP is a single jelly-roll protein that forms pentameric capsomers (pentons) occupying the 12 vertices. The mature capsid consists of 260 hexons and 12 pentons that are arranged in a lattice with a triangulation (T) number of 27 ($h=3$; $k=3$) (Sun *et al.*, 2010; Zhang *et al.*, 2012).

The capsid proteins of different virophages show low levels of conservation on the amino acid level (e.g. ~40% for the MCPs of Sputnik and mavirus) (Krupovic *et al.*, 2016). The mCP in particular is highly diverse, which hinders sequence similarity-based identification of novel virophages. Despite the high sequence divergence, all MCPs and mCPs are assumed to adopt the double and single jelly-roll folds, respectively, and preliminary data shows that the three-dimensional structure of the mavirus MCP is very similar to that of Sputnik (D. Born, L. Reuter, U. Mersdorf, M. Mueller, M.G. Fischer, A. Meinhart and J. Reinstein, under review). However, since most of the sequence space occupied by virophages remains unknown, the existence of different capsid architectures cannot be excluded. For instance, the virophage genomes assembled from a sheep rumen metagenome apparently lack a penton gene (Yutin *et al.*, 2015a). In addition to the main capsid components (MCP and mCP), other virophage-encoded proteins may be present in the mature virion, such as the mavirus MV13

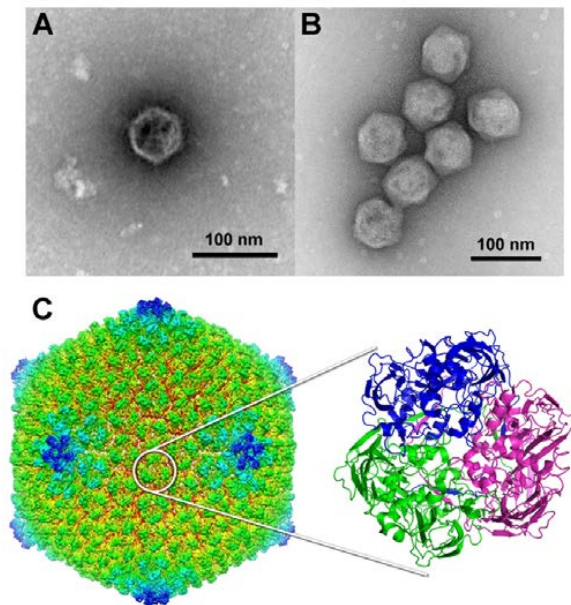


Figure 12.2 The structure of lavidavirus capsids. A) Negative stain electron micrograph of a Sputnik particle (courtesy of M. Gaia and B. La Scola, Univ. Aix-Marseille, France). B) Negative stain EM image of CsCl gradient-purified mavirus particles (U. Mersdorf and M. Fischer, Max Planck Institute for Medical Research, Germany). C) Cryo-EM reconstruction of the Sputnik virion at 3.5 Å resolution with a magnified major capsid protein trimer (PDB entry 3J26). Modified from (Zhang *et al.*, 2012).

protein, which could have lipase activity (Fischer *et al.*, 2014).

Virophage genomics

All virophages described so far possess dsDNA genomes ranging in size from 17 to more than 30 kbp. Fig. 12.3 depicts the genome organization of virophages, and their properties are listed in Table 12.1. Cultured lavidaviruses (Sputnik, mavirus, Zamilon) have circular genomes while those that were assembled from metagenomic sequences are either circular, linear, or have unknown genome topologies (Gong *et al.*, 2016; Oh *et al.*, 2016; Yau *et al.*, 2011; Yutin *et al.*, 2015a; Zhou *et al.*, 2013, 2015). The G+C content of virophage genomes tends to fall between 27% and 37%, with the exception of YSLV 5 (51.1% G+C). Similarly low G+C contents are found in the virophage-associated giant

viruses, e.g. mimivirus (28% G+C), CroV (23%), PgV-16T (32% G+C), and OLPV-1 (30% G+C) (Fischer *et al.*, 2010; Raoult *et al.*, 2004; Santini *et al.*, 2013; Yau *et al.*, 2011). In contrast, the G+C contents of the nuclear genomes in the eukaryotic hosts are typically higher, e.g. 59% for *Acanthamoeba polyphaga* (GenBank Acc. #CDFK000000000) and 65–70% for *Cafeteria roenbergensis* (T. Hackl and M. Fischer, unpublished data). This may indicate that virophage and giant virus genomes are subject to similar evolutionary forces that favour high A+T contents.

Virophage genomes contain from 16 to 34 protein-coding sequences, which are for the most part spaced in a non-overlapping manner, similar to other dsDNA viruses and bacterial genomes (Fig. 12.3). The Organic Lake virophage (OLV) has five annotated ORFs that are nested within larger predicted coding sequences (CDSs) (Yau *et al.*,

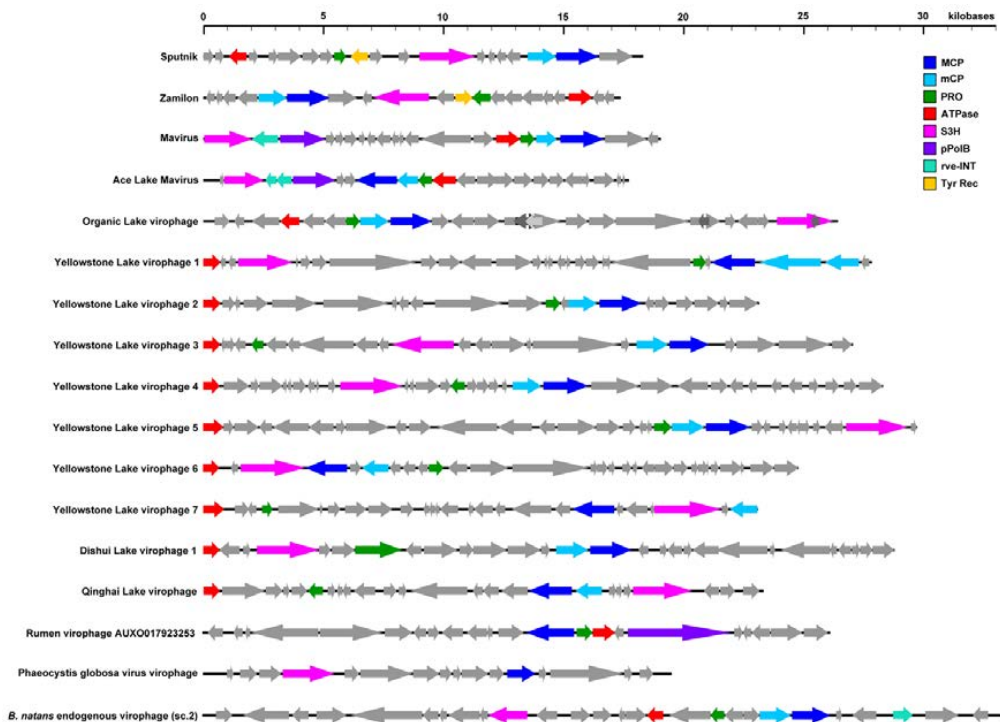


Figure 12.3 Comparative genomics of virophages. Complete and near-complete virophage genomes are drawn to scale and their open reading frames are indicated by arrows. Selected genes are shown in colour. MCP: major capsid protein, mCP: minor capsid protein, PRO: cysteine protease, ATPase: FtsK-HerA-family genome packaging ATPase, Prim/HEL: primase/superfamily 3 helicase, pPolB: protein-primed DNA polymerase B, rve-INT: retrovirus-family integrase, Tyr Rec: tyrosine recombinase-family integrase.

2011). There is no indication that they encode proteins, however, since one of those ORFs (OLV26) has no significant similarity to sequences in public databases, and the other four ORFs (OLV14-16, 21) contain highly repetitive amino acid patterns, which may give false-positive homologues in BLAST searches.

Conserved genes

Nearly all virophage genomes encode a structural module that consists of a major capsid protein (MCP), a minor capsid or penton protein (mCP), a cysteine protease (PRO), and an FtsK-HerA-like ATPase (coloured in Fig. 12.3). These proteins are involved in virion formation. The highly conserved MCP–mCP gene pair is found in all virophages including the virophage-like element PgVV and also in polinton-like viruses (PLVs), which encode poorly conserved versions of these capsid genes (Yutin *et al.*, 2015b). These two genes are syntenic in most virophages (Fig. 12.3), indicative of their importance. The MCP adopts a double jelly-roll fold, whereas the mCP gene encodes the single jelly-roll penton protein of virophage capsids. Yellowstone Lake virophage (YSLV) 1 encodes two versions of the mCP, coding sequences (CDSs) 26 and 27, of which CDS 26 contains a 390 aa long insertion that is missing in other virophage mCPs. Because mCP pentamers are located at the vertices of the icosahedral capsid, they are most likely to interact with host components (e.g. membrane proteins), which may explain their high sequence variability due to adaptation to new hosts. The virophage-encoded PRO bears similarity to maturation proteases from adenoviruses and PRD1, and this protease is most likely required for virion morphogenesis (D. Born, L. Reuter, U. Mersdorf, M. Mueller, M.G. Fischer, A. Meinhart and J. Reinstein, under review). The cryo-EM reconstruction of the Sputnik virion revealed that the assembled MCP was 87 aa shorter than predicted from the DNA sequence (Zhang *et al.*, 2012). Therefore, the C-terminus must be cleaved off at a diglycine motif that is conserved across all virophages except for PgVV and the rumen virophages (RVPs). PRO is thus responsible for C-terminal processing of the MCP in at least Sputnik and mavirus. The last conserved virophage gene is a predicted genome-translocating ATPase that belongs to the FtsK-HerA superfamily.

These NTPases usually form hexamers at a unique vertex of the viral capsid and pump the genome through the central pore of this hexamer, a process that is powered by ATP hydrolysis (Burroughs *et al.*, 2007). The mavirus ATPase differs from those of all other virophages by a unique 54 aa insertion between Walker A and Walker B motifs.

Most virophage genomes encode a superfamily 3 helicase (S3H) domain (Fig. 12.3), but the genetic relationships among these genes are complicated. Virophage S3H genes are apparently polyphyletic, and the versions found in Sputnik, Zamilon, OLV, YSLV 1, YSLV 6, QLV, and PgVV are coupled to an amino-terminal domain with predicted primase-polymerase activity (TVpol, related to bacterial DNA polymerase I enzymes) (Iyer *et al.*, 2008; Yutin *et al.*, 2013). The S3H may thus have adopted different functions in different virophages, e.g. as a helicase in mavirus (which encodes a separate DNA polymerase), and as the primary DNA replication enzyme in Sputnik. Short zinc ribbon domains are also frequently found in virophages, where they can be fused with GIY-YIG endonucleases (e.g. mavirus, OLV), or occur separately (e.g. Sputnik) (Yutin *et al.*, 2013).

Non-conserved genes

Many of the genes found in virophage genomes exhibit diverse evolutionary affiliations, including eukaryotes, bacteria, different viral families, and diverse mobile genetic elements (Fischer and Suttle, 2011; Krupovic and Koonin, 2015; La Scola *et al.*, 2008; Yau *et al.*, 2011; Yutin *et al.*, 2013). Examples are the bacteriophage lambda-type tyrosine recombinase encoded by Sputnik or the retroviral integrase encoded by mavirus. Although virophages are genetically distinct from their associated host viruses, they encode one to several genes with clear affiliations to giant DNA viruses, and these genes often contain repetitive motifs. The genomes of Sputnik, Zamilon, and predicted algae-infecting virophages contain collagen-like repeat proteins that are also prominent in genomes of *Mimiviridae* members (Gaia *et al.*, 2014; La Scola *et al.*, 2008; Suhre, 2005; Yau *et al.*, 2011). Sputnik ORF V12 codes for a low-complexity protein that is similar to mimivirus ORF R546; and Sputnik ORFs V4, V14, V16, and V17 display similarities to proteins encoded mainly by mousmouviruses. The most

Table 12.1 Properties of partial and complete virophage genomes

Virophage	Length (bp)	% GC	Topology	ORFs	Cultured or sequenced	Associated giant virus(es)	Host organism	GenBank accession no.	Reference
Sputnik	18,342	27.04	Circular	21	In culture	Mimi-, Mournou-, Megaviruses	<i>Acanthamoeba</i> sp.	EU606015.1	La Scola <i>et al.</i> (2008)
Sputnik 2	18,338	27.04	Circular	21	In culture	Mimi-, Mournou-, Megaviruses	<i>Acanthamoeba</i> sp.	JN603369.1	Desnues <i>et al.</i> (2012b)
Sputnik 3	18,338	27.04	Circular	21	In culture	Mimi-, Mournou-, Megaviruses	<i>Acanthamoeba</i> sp.	JN603370.1	Gaia <i>et al.</i> (2013b)
Sputnik strain Rio Negro	1788 (partial)	33.45	Unknown	Unknown	In culture	Samba virus (SMBV)	<i>Acanthamoeba</i> sp.	KJ183141.1	Campos <i>et al.</i> (2014)
Zamilon	17,276	29.67	Circular	20	In culture		<i>Acanthamoeba</i> sp.	HG531932.1	Gaia <i>et al.</i> (2014)
Zamilon 2	6717 (partial)	32	Unknown	15	Sequence only	Unknown	Unknown	N/A	Bekiz <i>et al.</i> (2015)
Mavirus (exogenous)	19,063	30.26	Circular	20	In culture	Cafeteria roenbergensis virus (CroV)	<i>Cafeteria roenbergensis</i>	HQ712116.1	Fischer and Suttle (2011)
Mavirus (endogenous)	20,190	30.67	Integrated	20	In culture	Cafeteria roenbergensis virus (CroV)	<i>Cafeteria roenbergensis</i>	KU052222.1	Fischer and Hackl (2016)
Ace Lake Mavirus (ALM)	17,767	26.73	Unknown	22	Sequence only	Unknown	Unknown	KC556923.1	Zhou <i>et al.</i> (2013)
Organic Lake virophage (OLV)	26,421	36.51	Unknown	26	Sequence only	Unknown	Unknown	HQ704801.1	Yau <i>et al.</i> (2011)
Yellowstone Lake virophage (YSLV) 1	27,851	33.37	Unknown	28	Sequence only	Unknown	Unknown	KC556924.1	Zhou <i>et al.</i> (2013)

YSLV2	23,184	33.16	Unknown	21	Sequence only	Unknown	Unknown	KC556925.1	Zhou <i>et al.</i> (2013)
YSLV3	27,050	35.86	Unknown	23	Sequence only	Unknown	Unknown	KC556926.1	Zhou <i>et al.</i> (2013)
YSLV4	28,306	37.07	Unknown	34	Sequence only	Unknown	Unknown	KC556922.1	Zhou <i>et al.</i> (2013)
YSLV5	29,767	51.10	Unknown	32	Sequence only	Unknown	Unknown	KM502589.1	Zhou <i>et al.</i> (2015)
YSLV6	24,837	26.84	Unknown	29	Sequence only	Unknown	Unknown	KM502590.1	Zhou <i>et al.</i> (2015)
YSLV7	23,193	27.33	Unknown	26	Sequence only	Unknown	Unknown	KM502591.1	Zhou <i>et al.</i> (2015)
Dishui Lake virophage (DSLVI)	28,788	43.20	Circular	28	Sequence only	Unknown	Unknown	KT894027.1	Gong <i>et al.</i> (2016)
Qinghai Lake virophage (QLV)	23,379	33.17	Circular	25	Sequence only	Unknown	Unknown	KJ854379.1	Oh <i>et al.</i> (2016)
Rumen virophages (RVP)	varies, longest: 26,209	varies, longest: 34.64	Linear with TIRs	varies, longest: 22	Sequence only	Unknown	Unknown	longest AUXO017923253	Yutin <i>et al.</i> (2015a)
Cryoconite virophage (CrV)	12,572 (partial)	32.80	unknown	15	Sequence only	Unknown	Unknown	LDNN01000025.1	Bellas <i>et al.</i> (2015)
Miers Valley soil virophage (MVSV)	1035 (partial)	34.69	unknown	unknown	sequence only	unknown	unknown	CZR14427.1	Zablocki <i>et al.</i> (2014)
Phaeocystis globosa virus virophage (PgVV)	19,527	35.78	linear or integrated	16	in culture	Phaeocystis globosa virus 16T (PgV-16T)	Phaeocystis globosa	NC_021333	Santini <i>et al.</i> (2013)
<i>Bigeloviella natans</i> virophages (BnVP)	varies, up to 33.3 kb	varies, ≈36.4	integrated	varies, up to 27	sequence only	unknown	<i>Bigeloviella natans</i>	Assembly GCA_000320545	Blanc <i>et al.</i> (2015)

frequent repetitive element in the CroV genome is called FNIP or IP22 (Fischer *et al.*, 2010; O'Day *et al.*, 2006) and occupies more than 10% of the 692 kbp genome (Hackl and Fischer, in preparation), and the same repeat is also found in mavirus ORF MV20. These observations suggest that DNA or protein repeats may be involved in virophage–giant virus interactions. Repetitive proteins are also found in the genomes of uncultivated virophages, e.g. the Kelch, ankyrin, and adhesin domain-containing proteins in the endogenous *Bigelowiella natans* virophages (Blanc *et al.*, 2015).

Diversity and taxonomy of virophages

After the isolation and characterization of Sputnik and mavirus, it became apparent that these viruses were the first representatives of a previously unknown group of dsDNA viruses. Classification of virophages was initially complicated by their requirement for coinfecting giant dsDNA viruses for their propagation. This property, which virophages share with satellite viruses, led to the initial classification of Sputnik as a satellite virus, placed within the obscure category of 'subviral agents' in the Ninth Report of the International Committee on Taxonomy of Viruses (ICTV) (King *et al.*, 2011). The distinction between satellite viruses and virophages has caused some confusion and debate (Desnues and Raoult, 2012; Fischer, 2011; Krupovic and Cvirkaite-Krupovic, 2011) and shall be discussed here briefly.

Virophages and satellite viruses

Virophages as well as satellite viruses can exert a negative effect on their associated giant viruses, or helper viruses, respectively. This effect varies significantly depending on the virus and multiple infection parameters. Satellite viruses also share some other aspects with virophages, such as sub-cellular localization or overlap with their helper viruses regarding polyadenylation sites (Krupovic and Cvirkaite-Krupovic, 2011). On the other hand, virophages are genetically distinct from satellite viruses. While the latter have short ssRNA genomes (satellite viruses infecting plants and arthropods) or short ssDNA genomes (adeno-associated viruses) (King *et al.*, 2011), virophages possess considerably

longer dsDNA genomes encoding genes that are conserved in the PRD1–adenovirus lineage. Unlike satellite viruses and owing to the greater complexity of virophage particles, virophages encode multiple morphogenesis proteins (two capsid proteins, maturation protease, DNA-packaging ATPase) and DNA replication proteins. As discussed in more detail below, virophages probably depend on the giant virus-provided transcription apparatus in a way that is similar to how other small dsDNA viruses depend on the cellular transcription system.

Thus, while several phenomena are shared between virophages and satellite viruses (dependency on another virus, variable hypovirulence), their larger particle size and coding potential, genetic and infection properties, and relatedness to viruses of the PRD1–adenovirus lineage clearly set virophages apart from those much simpler and genetically distinct groups of satellite viruses. Virophages can be viewed as eukaryotic dsDNA viruses that became evolutionarily adapted to giant dsDNA viruses, and that use the viral transcription enzymes in the cytoplasm instead of accessing the host nucleus for cellular transcription.

The *Lavidaviridae*

The family *Lavidaviridae* was created to reflect the unique properties of virophages, which distinguish them from other groups of viruses including satellite viruses (Krupovic *et al.*, 2016). 'Lavida-' is an acronym for 'large virus dependent or associated' and highlights the most striking feature of these viruses, i.e. their co-dependency on giant dsDNA viruses. This family currently contains two genera, *Sputnikvirus* and *Mavirus*, which were created based on shared subsets of genes that are specific to each genus, in addition to phylogenetic analysis of genes that are conserved in all virophages.

The topologies of the resulting trees vary somewhat depending on the chosen marker gene (Fig. 12.4). The most highly conserved gene and thus arguably the best phylogenetic marker is the major capsid protein (Fig. 12.4A), but the ATPase is also conserved enough to be useful for phylogenetic studies (Fig. 12.4B). Both marker genes resolve the two genera with high branch support, and additional well-supported clusters can be observed, as discussed below. Given the rapid accumulation of virophage genomes from metagenomes, which

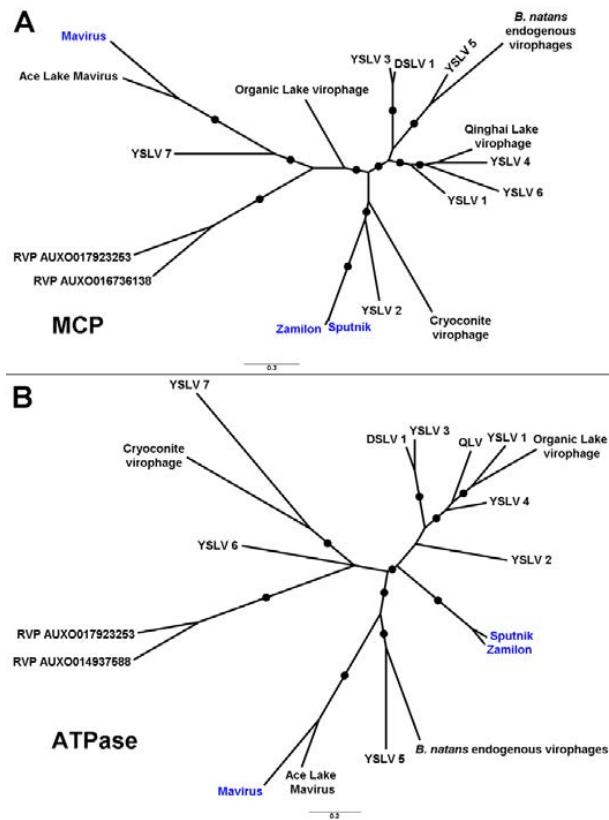


Figure 12.4 Phylogenetic relationships among virophages. Major capsid protein (A) and FtsK-HerA-family ATPase (B) sequences were aligned with PROMALS3D (Pei *et al.*, 2008), and the manually edited alignment was used to create an unrooted Bayesian phylogenetic tree using MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003) with 1 million generations and a burn-in of 1000. Branches with posterior probabilities less than 0.5 were collapsed; those with posterior probabilities higher than 0.90 are marked by black dots. Cultured virophages are printed in blue.

leads to a better resolution of the phylogenetic reconstructions, additional *Lavidaviridae* genera will likely be added in the near future.

The genus *Sputnikvirus*

Sputnik, the first virophage to be discovered, was isolated from water of a cooling tower near Paris, France (La Scola *et al.*, 2008). The same sample also contained a new mimivirus strain named mamavirus with a 1,191,693 bp dsDNA genome. Based on transmission electron microscopy, the Sputnik particles were initially reported to be 50 nm wide, in contrast to cryo-EM studies that were conducted later and reported a capsid diameter of 74 nm (Sun *et al.*, 2010). After Sputnik virions were purified

from mamavirus by filtration through 0.2 μ m pore-size filters and inoculated on amoebae, no Sputnik multiplication was observed. This virus instead only replicated when the amoebae were coinfecting with mamavirus. Inspection of coinfecting amoebal cells by electron microscopy further revealed that Sputnik and mamavirus particles emerged from the same cytoplasmic virion factory (Fig. 12.5). Sputnik multiplied faster than mamavirus and progeny Sputnik virions appeared at distinct locations of the virion factory. In addition, it was noted that Sputnik had a negative effect on mamavirus production, as the yield of the latter was reduced by ~70% and abnormal giant virus capsids were observed in the presence of Sputnik (La Scola *et al.*, 2008). This

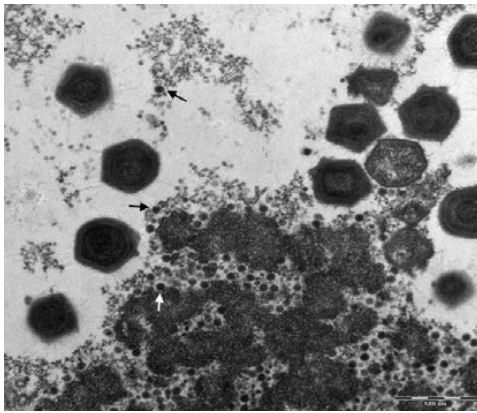


Figure 12.5 Sputnik particles (arrows) produced in *Acanthamoeba polyphaga* cytoplasmic mimivirus factories. This thin-section transmission electron micrograph is courtesy of M. Gaia and B. La Scola, Univ. Aix-Marseille, France.

led to the classification of Sputnik as a parasite of mamavirus and the introduction of the term ‘virophage’, as a virus infecting another virus.

The Sputnik genome consists of a circular dsDNA molecule 18,343 bp in length with an A+T content of 73%, and encodes 21 predicted protein-coding sequences (CDSs). Two other Sputnik strains have been isolated and genetically characterized. Sputnik 2 was found associated with lentilivirus, a mimivirus strain isolated from lens fluid (Desnues *et al.*, 2012b; La Scola *et al.*, 2010) and Sputnik 3 was recovered via a reporter system from a soil sample near Marseille, France during a screen for new virophages (Gaia *et al.*, 2013b). The Sputnik 2 and 3 genomes are four base pairs shorter than Sputnik 1, which results from single base pair deletions at positions 877, 7949, 7958, and 12936. Initially, the Sputnik 1 genome displayed another additional base at position 14047. This, however, was found to be a sequencing artifact that resulted in a frame shift artificially separating ORFs 18 and 19 (Gaia *et al.*, 2013b; Zhang *et al.*, 2012). The corrected Sputnik 1 sequence encodes the minor capsid protein V18/19 in a single ORF, which is not only in agreement with the gene topology in Sputnik 2 and 3, but also enables full-length sequence alignments with homologous mCP genes in other virophages. The genome of Sputnik 3 further differs from Sputnik 1 and 2 by two T to A transversions at

positions 8986 and 8991 (located in the intergenic region between V12 and V13) and two A to G transitions at positions 16098 and 17666, resulting in a A464T change in the MCP V20 and a D371N change in V21, respectively. Due to their nearly identical genomes, Sputniks 1, 2 and 3 are closely related strains.

Another Sputnik isolate is the Rio Negro virophage (RNV) that was found in the Rio Negro River of the Brazilian Amazon (Campos *et al.*, 2014). RNV is associated with a mimivirus variant called Samba virus. The RNV particles were reported to be only 35 nm in diameter, even though the partial major capsid gene sequence was 100% identical to the Sputnik MCP gene, which makes it likely that the true capsid size of RNV is comparable to that of Sputnik.

The Zamilon virophage was isolated from Tunisian soil together with its host virus, the Mont1 mimivirus (Boughalmi *et al.*, 2013; Gaia *et al.*, 2014). Zamilon (Arabic for ‘neighbour’) is a relative of the Sputnik virophages, and its circular 17,276-bp-long genome shares 76% nucleotide identity with Sputnik. The Zamilon genome encodes 20 predicted proteins, 17 of which have homologues in Sputnik. In contrast to Sputnik, Zamilon does not replicate with all members of the *Mimivirus* genus. Zamilon replication was only detected during coinfection with viruses of lineage B (moumouvirus) and lineage C (megavirus), but not with lineage A members (mimivirus) (Gaia *et al.*, 2014). In addition, Zamilon differs from Sputnik by not inhibiting the replication of its coinfecting giant virus.

The genus *Mavirus*

Mavirus

The second described virophage was mavirus (meaning, maverick-related virus; mavericks are large transposons found in eukaryotic genomes that are also called polintons from the presence of DNA polymerase and integrase genes; see ‘Evolutionary connections of virophages’ below), a parasite of *Cafeteria roenbergensis* virus (CroV). The exact geographic origin of mavirus is not clear. CroV was isolated from coastal waters of the Gulf of Mexico near Aransas Pass, Texas, USA, in 1989 (Garza and Suttle, 1995). The flagellate host strain, on which CroV was isolated, originated from the Pacific coast near Yaquina Bay, Oregon, USA

(Gonzalez and Suttle, 1993). After mavirus was discovered in CroV-infected flagellate cultures (Fischer and Suttle, 2011), it was assumed that mavirus had been co-isolated with CroV from the same Gulf of Mexico water sample and that it had been propagated alongside CroV until the virophage was identified in 2009. Recent findings of endogenous mavirus genomes integrated in flagellate chromosomal DNA (Fischer and Hackl, 2016), however, have opened up an alternative explanation, according to which some cells of the Oregon host strain could have harboured a mavirus provirophage that was subsequently reactivated upon contact with CroV.

The mavirus genome is a circular 19,063-bp-long molecule that is able to linearize for integration. The terminal inverted repeats (TIRs) of exogenous mavirus genomes are ≈ 50 bp long (Fischer and Suttle, 2011), whereas the TIRs of the endogenous mavirus genome are 615/616 bp long, resulting in a total length of 20,190 bp for the provirophage genome (Fischer and Hackl, 2016). Mavirus encodes 20 predicted proteins and has a GC content of 30%, which is similar to the properties of Sputnik and Zamilon.

The particles of mavirus are icosahedral in shape and have a diameter of 60–65 nm in thin-section transmission electron micrographs, and 70–75 nm in negative stain electron micrographs (Fig. 12.2B). The burst size of mavirus is roughly ten times higher than that of CroV, the latter as measured during a mavirus-free CroV infection of *C. roenbergensis*.

Ace Lake mavirus

Although known only from metagenomic sequences, the genome of Ace Lake Mavirus (ALM) clearly falls within the genus *Mavirus*. The ALM genome was assembled from short metagenomic reads that originated from the Antarctic Ace Lake (Zhou *et al.*, 2013). The genome sequence is not complete and may contain sequence errors due to erroneous assembly and low sequence coverage. Despite these potential problems, the ALM genome provided a first glance at sequence diversity within the mavirus subgroup of virophages. ALM and mavirus share 13 of their 22 and 20 predicted genes, respectively. The gene order is also

conserved; however, an inversion affecting seven genes occurred in one of the genomes. In addition to the four highly conserved morphogenesis genes, mavirus and ALM encode homologues for the predicted pPolB, rve-INT, S3H, as well as six genes of unknown function.

ALM ORF14 probably starts at the second ATG start codon, as suggested by alignments with its homologue mavirus ORF12 and the presence of a putative promoter motif directly upstream of the second ATG site, whereas no such promoter motif is present in the immediate upstream region of the first start codon. Like mavirus, ALM encodes an rve-INT. The ALM homologue is annotated in two adjacent reading frames and at present it is not known whether this gene is split into two separate ORFs, or whether an artificial frameshift was introduced during assembly of the metagenomic reads.

Additional virophages from metagenomes

Virophages appear to be distributed world-wide and several groups of virophage-like genomes have been discovered in metagenomes in recent years (Figs. 12.1, 12.3 and 12.6). This was made possible by seeding BLAST search with genes from fully sequenced genomes of the few virophage isolates that are currently available. Interestingly, whereas Sputnik originated from a freshwater cooling tower, Zamilon from soil, and mavirus from an oceanic environment, most uncultured virophages were found in lakes, suggesting that lake ecosystems provide conditions that allow certain virophage populations to thrive.

The Organic Lake virophage

The first virophage genome assembled from a metagenome was reported in 2011 (Yau *et al.*, 2011). A genomic study of the microbial community in East Antarctica's hypersaline Organic Lake revealed the presence of a virophage-like genome. The Organic Lake Virophage (OLV) is probably associated with NCLDV of the extended *Mimiviridae* family that infects a photosynthetic host². Its genome was assembled into a 26,421 bp long sequence, encoding 24 ORFs. The phylogenetic position of OLV depends on the gene used (Fig.

² The same publication reports the partial assembly of two giant virus genomes, named Organic Lake phycodnavirus (OLPV) 1 and 2 (Yau *et al.*, 2011). Despite the name, the OLPVs are not part of the *Phycodnaviridae*, but phylogenetically more closely related to the *Mimiviridae*.



Figure 12.6 Geographic origin of published virophage genomes.

12.4), but based on MCP phylogeny, there are currently no known close relatives of OLV.

The Yellowstone Lake virophages and Ace Lake mavirus

In 2013, Zhou *et al.* reported the assembly of four virophage genomes (YSLV1–4) from a metagenome of Yellowstone Lake water samples (Zhou *et al.*, 2013). These findings were expanded by the same group of researchers with the publication of three additional Yellowstone Lake virophage sequences (YSLV5–7) 2 years later (Zhou *et al.*, 2015). The YSLV genomes are 23–30 kb long and appear to belong to different subgroups, as shown by phylogenetic reconstruction (Fig. 12.4). In particular, YSLV5 and YSLV7 do not cluster with other YSLVs. In their 2013 publication, the authors also described the genome of Ace Lake mavirus (ALM), assembled from the saline meromictic Ace Lake in the Vestfold Hills of Antarctica (Zhou *et al.*, 2013). Interestingly, this lake is in close geographic proximity to Organic Lake, where OLV was detected (Yau *et al.*, 2011).

The Dishui Lake and Qinghai Lake virophages

In 2016, virophage genomes were reconstructed from metagenomic data of two Asian freshwater lakes, one in Shanghai and one in Tibet. The Dishui

Lake virophages (DSLVs) occur in an artificial freshwater lake in Shanghai, China, and are closely related to YSLV 3 (Fig. 12.4) (Gong *et al.*, 2016). They are presumed to replicate in combination with algae-infecting giant viruses. Another population of virophages was found in a surface water metagenome of Lake Qinghai, in the mountains of Tibet (Oh *et al.*, 2016). The Qinghai Lake virophages (QLVs) are most closely related to YSLVs 1 + 4 and could infect phototrophic protists.

Virophages from a sheep rumen metagenome

A novel lineage of virophages was identified in the metagenome of sheep rumen (Yutin *et al.*, 2015a). The longest of these was 26,209 bp long and encoded 22 ORFs. The rumen virophage (RVP) genomes appear to be a hybrid between bona fide virophages (with genes for MCP, ATPase, PRO) and polintons (mavericks, encoding a protein-primed DNA polymerase B). Such a scenario has also been proposed for the origin of mavirus (Yutin *et al.*, 2013); however, RVPs and mavirus are phylogenetically distinct based on pPolB and MCP analyses and RVPs branch separately from all other virophages (Yutin *et al.*, 2015a) (Fig. 12.4). So far, no minor capsid protein has been found in RVP genomes, suggesting that their capsid architecture differs from that of Sputnik and mavirus.

Other virophage sightings

A metagenomic study of viromes from cryoconite holes in Greenland yielded a partial virophage genome (Bellas *et al.*, 2015). Cryoconite ('ice dust') holes are cylindrical depressions at the surface of glaciers which are filled with dark coloured sediments. The cryoconite virophage (CryV) partial genome is 12,572 bp long and includes the four morphogenesis genes (MCP, mCP, PRO, ATPase). Virophage genome fragments were also identified at the opposite pole, from surface soil of the dry Miers Valley in Eastern Antarctica (Zablocki *et al.*, 2014). Gene fragments of this Miers Valley soil virophage (MVSV) suggested that viruses of the Sputnik clade exist in Antarctica, as well as mimiviruses and phycodnaviruses, whose genomic signatures were also found in the same study. Antarctica therefore harbours a diverse virophage community in its aquatic and terrestrial environments, as evidenced by OLV, ALM, and MVSV. Not surprisingly, genetic signatures of giant viruses are also present in Antarctica (Kerepesi and Grolmusz, 2017; Yau *et al.*, 2011). The endogenous virophage-like elements found in the nuclear genome of *Bigelowiella natans* (Blanc *et al.*, 2015) are discussed later in this review. A recent study identified 25 uncultivated virophage populations using time-series metagenomes from two North-American freshwater lakes (Roux *et al.*, 2017). This report will help refine the taxonomy of virophages and confirms that most of their genetic diversity remains to be described.

The virophage-like element PgVV and the polinton-like viruses

The large DNA virus PgV-16T belongs to the extended *Mimiviridae* family and infects the bloom-forming microalga *Phaeocystis globosa* (Santini *et al.*, 2013). During assembly of the 459,984-bp-long, linear dsDNA PgV-16T genome, an additional contig was found. This 19,527 bp long linear dsDNA molecule with 1 kb terminal inverted repeats contained 16 predicted CDSs that are all located on the same strand. Based on the similarity of three of these CDSs to genes in OLV and mavirus, this genetic element was termed PgV-associated virophage (PgVV). However, no 50–80 nm sized capsids were found and the authors proposed that PgVV replicates either as a linear plasmid or as a provirophage that is

integrated in the PgV-16T genome (Santini *et al.*, 2013). Although the PgVV genome was initially reported to lack any recognizable capsid genes, ORF PgVV_00012 likely encodes a distant version of a double-jelly roll major capsid protein and ORF PgVV_00010 could encode a minor capsid protein (Krupovic *et al.*, 2014). At the time of discovery, PgVV was most closely related to mavirus and OLV. Based on the identification of polinton-like viruses (PLVs) in metagenomic datasets using the PgVV-predicted MCP as bait, however, PgVV is considered to be a PLV rather than a virophage (Yutin *et al.*, 2015b). The distinction is justified by the lack of a cysteine protease that is conserved in virophages, as well as by the distinct versions of MCP, mCP, and ATPase genes found in PgVV and PLVs. (Krupovic *et al.*, 2014; Yutin *et al.*, 2015b).

Evolutionary connections of virophages

Virophages may have evolved from an ancestral virus of the PRD1–adenovirus lineage by multiple recombination and gene replacement events (Krupovic and Koonin, 2015). Genome analysis of the first virophage, Sputnik, suggested mixed origins for its genes which exhibited links to viruses infecting archaea, bacteria, and eukaryotes (La Scola *et al.*, 2008). Surprisingly, seven of the 20 genes from the second described virophage, mavirus, had their closest homologues in one particular group of genetic elements, the maverick/polinton elements (MPEs) (Fischer and Suttle, 2011). The MPEs were originally described as a class of DNA transposons widespread in eukaryotic genomes with highly variable copy numbers (Jordan *et al.*, 2004; Kapitonov and Jurka, 2006; Pritham *et al.*, 2007). They are 15–20 kb large and contain conserved rve-INT integrase and pPolB DNA polymerase genes in addition to other virus-affiliated genes such as an FtsK/HerA-like genome packaging ATPase and an adenovirus-like protease. Recently, it was shown that MPEs also code for putative mCP and MCP genes (Krupovic *et al.*, 2014), which resulted in a shift of perception, and capsid-encoding MPEs are now regarded as endogenous viruses, called polintoviruses (Krupovic and Koonin, 2015), rather than as transposons. Roughly one-third of the genome of the human parasite *Trichomonas*

vaginalis consists of MPEs, indicating that at least in some eukaryotic lineages these elements are able to spread intragenomically.

MPEs and virophages of the genus *Mavirus* share seven homologous proteins: S3H, rve-INT, pPolB, FtsK/HerA-type ATPase, PRO, mCP and MCP. In addition, virophages and MPEs have similar genome length (~20 kb), overlapping host range (mainly protists, although MPEs are also found in animals), and the mavirus genome exhibits terminal inverted repeats similar to those found in MPEs. The genes shared by mavirus and MPEs have thus evolved from a common ancestor; however, the nature of this ancestor is unclear. One hypothesis states that ancestral mavirus-like virophages integrated into a eukaryotic host genome, where they conferred resistance to the host against infection by giant viruses and were positively selected for in the host cell population (Fischer and Suttle, 2011). An alternative hypothesis, which is mainly based on tree topologies in phylogenetic reconstructions of conserved MPE/virophage proteins, suggests that virophages evolved from ancestral MPEs through multiple gene transfers and rearrangements and that they subsequently became dependent on giant DNA viruses (Krupovic and Koonin, 2015; Yutin *et al.*, 2013).

Recent data mining approaches in metagenomic sequences have revealed several additional groups of viral elements with ties to virophages and MPEs. The RVP virophages found in a sheep rumen metagenome share with MPEs the pPolB, ATPase, PRO and MCP genes, but lack the rve-INT and mCP. In phylogenetic trees based on multiple sequence alignments of MCP, the RVP virophages cluster as a monophyletic group, distinct from Sputnik-, mavirus-, and OLV-like virophages (Yutin *et al.*, 2015a) (Fig. 12.4). Another group of related viruses discovered in metagenomes are the polinton-like viruses (PLVs) (Yutin *et al.*, 2015b), which share a highly diverged version of the double-jelly roll MCP with PgVV. The PLVs are 18–28 kb in length, encode ATPase, mCP, and MCP (but no PRO) and some of them contain long TIRs. Notably, PLVs were also found in the genomes of photosynthetic protists, which suggests – together with the endogenous *B. natans* virophages – that virophage integration into eukaryotic genomes may be a common phenomenon (Blanc *et al.*, 2015; Fischer, 2015; Yutin *et al.*, 2015b). For a more detailed

overview of the evolutionary connections between virophages, MPEs, PLVs, and a class of mimivirus-associated parasitic linear DNA elements called transpovirons, see Koonin and Krupovic (2017) and Krupovic and Koonin (2016).

The virophage infection cycle

Owing to the paucity of culture-based virophage systems, many details of their replication cycle remain unknown. Studies on the *Acanthamoeba*–mimivirus–Sputnik and *Cafeteria*–CroV–mavirus systems, however, have given us a vague idea about the intracellular events that occur during virophage infection (Fig. 12.7). The most striking feature about these viruses is that, despite their respectable size and coding potential, they cannot reproduce inside a host cell without a suitable coinfecting giant virus. Hence, the replication cycle of virophages is tightly linked to that of their host viruses.

Virion entry

Virions enter the host cell either by endocytosis, or as a composite with their host virus by phagocytosis. An example for the former entry mode is mavirus, which attaches to the host cell surface via an unknown receptor (Fig. 12.8B). The membrane at the attachment site then starts to invaginate and the virus is endocytosed (Fig. 12.8C). Inside infected host cells, mavirus particles can be observed within coated (Fig. 12.8A and D) as well as uncoated vesicles (Fig. 12.8E) by electron microscopy, which implies that virions are internalized by clathrin-mediated endocytosis (Fischer and Suttle, 2011).

In contrast, Sputnik has never been observed to enter the amoeba independently of a coinfecting virus of the *Mimiviridae* family. Sputnik is able to adhere to the 1250 Å long, heavily glycosylated surface fibres that cover the mimivirus capsid (Piacente *et al.*, 2012; Xiao *et al.*, 2009). This interaction may be mediated by the ~100 Å long mushroom-like surface fibres that decorate the Sputnik virion (Sun *et al.*, 2010). The virophage–giant virus composite is then phagocytosed by the amoeba and Sputnik particles entangled within mimivirus surface fibres are occasionally observed by electron microscopy (Desnues and Raoult, 2010). In particular, the mimivirus R135 gene product, a glycosylated fibre protein related to GMC-type oxidoreductases that elicits an antigenic response (Klose *et al.*, 2015;

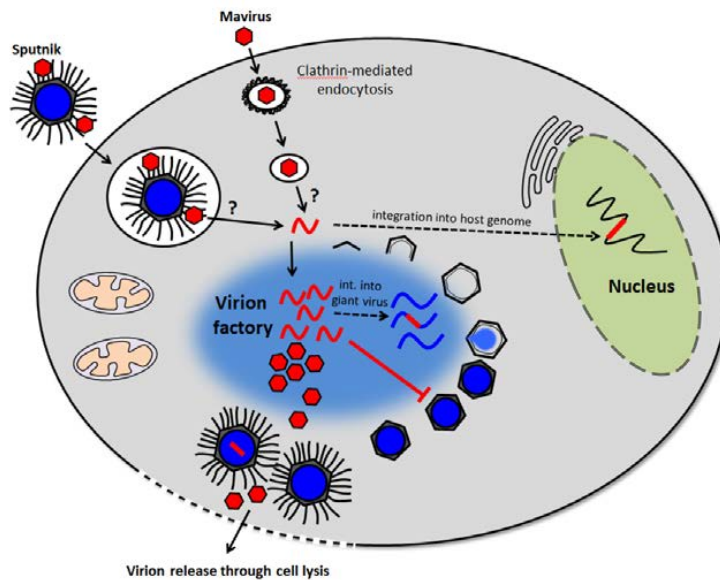


Figure 12.7 Simplified diagram depicting the infection cycle of lavidaviruses. Virophages enter the cell either via phagocytosis as a composite with a giant virus (Sputnik), or independently via endocytosis (mavirus). The mechanism by which they escape the vesicle and target the cytoplasmic virion factory of their giant viruses is unknown. Virophages occasionally integrate their genomes into the nuclear host genome or 'int.' into giant virus genomes. The virophage genome is replicated in the virion factory during the late phase of giant virus infection; virophage proteins are synthesized and progeny virions are assembled. During any of these processes, giant virus replication may be impaired. Finally, virophage and giant virus particles are released upon cell lysis.

Pelletier *et al.*, 2009), appears to have a high affinity for Sputnik particles, as this protein was identified by mass spectrometry in a purified Sputnik preparation (La Scola *et al.*, 2008). Additional support for this hypothesis stems from the analysis of a mimivirus deletion mutant, which lacks the surface fibres due to the deletion of 155 genes. This bald form of mimivirus, termed M4, no longer supports Sputnik replication (Boyer *et al.*, 2011). The R135 gene is among the genes that are deleted in mimivirus M4.

The composite entry model of Sputnik and mimivirus via phagocytosis creates the need for a mechanism that would allow Sputnik particles to escape from the phagosome. Uncoating of mimivirus particles in the phagosome is initiated by opening of the stargate portal, a unique structural feature of mimiviruses (Reteno *et al.*, 2018). The open stargate exposes the inner viral membrane, which subsequently fuses with the phagosomal membrane, thereby creating a gateway to the cytoplasm for the viral core (Zauberman *et al.*, 2008). This uncoating mechanism, however, grants

cytoplasmic access only to molecules that are located inside the mimivirus membrane. Any other particles inside the phagosome, such as Sputnik capsids, would have to use an alternative escape route. For this reason, a mimivirus-independent entry pathway of Sputnik cannot be excluded, and the exact mechanism of viral genome delivery from the particle into the host cytoplasm remains to be determined. Cryo-EM studies of empty Sputnik virions suggested that low pH conditions (pH ~5.5) can trigger the dissociation of some of the penton structures, which would create a portal through which the viral DNA could exit the virion (Zhang *et al.*, 2012). Capsid dismantling of Sputnik may thus resemble that of adenoviruses, which are known to lose penton bases in acidified endosomes during cell entry (Greber *et al.*, 1993). The uncoating process of mavirus, by contrast, is not known.

As suggested by mathematical modelling, the two different entry modes may have consequences for long-term population dynamics of viruses and hosts (Taylor *et al.*, 2014).

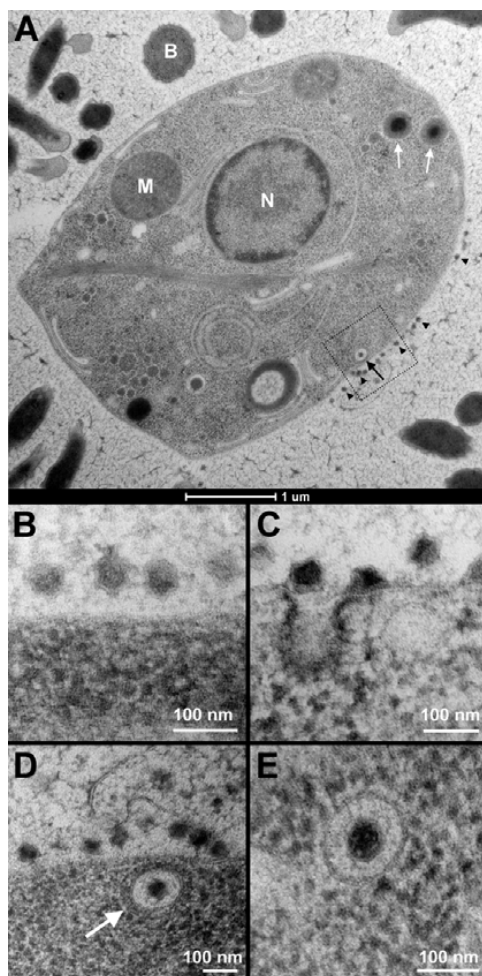


Figure 12.8 Entry of mavirus into *C. roenbergensis*. (A) Host cell coinfecting with at least two CroV cores (white arrows) and a mavirus particle (black arrow). Several other mavirus particles are attached to the cell surface (arrowheads). The boxed area is magnified in D). B: bacterium, M: mitochondrion, N: nucleus. (B) Mavirus particles aligned on the cytoplasmic membrane. (C) Mavirus particles triggering the invagination of the cytoplasmic membrane, creating a coated pit. (D) Mavirus particle inside a coated vesicle. Other virions are visible outside the cell. (E) Mavirus particle inside an uncoated vesicle in the host cytoplasm. All images by M. Fischer, C. Suttle, and the Bioimaging Facility at the University of British Columbia, Canada. Images C) – E) modified from Fischer and Suttle, 2011.

Virus factories

Following entry, the virophage genome is targeted to the developing virion factory of the coinfecting giant virus. After an eclipse phase, which lasts 2–4 h for Sputnik, newly synthesized virophage particles are visible within the virion factory (Figs. 12.5 and 12.9). Sputnik particle production usually begins at one pole of the mimivirus factory (La Scola *et al.*, 2008). The phenotype of mavirus-infected CroV factories varies, depending on parameters such as cell line, multiplicity of infection (MOI), and whether a coinfection with external mavirus particles or a reactivation of endogenous mavirus takes place. Occasionally, factories producing both CroV and mavirus particles can be observed (Fig. 12.9A). More frequently, however, and especially during coinfection with high MOIs of mavirus, the CroV factories will develop to occupy a large part of the cytoplasm, but no CroV particle production occurs. Instead, nested electron-dense subfactories synthesizing mavirus virions can be seen (Fig. 12.9B).

Coinfection dependence

Sputnik, Zamilon, mavirus, and presumably other virophages to be isolated in the future cannot replicate in their host cells without coinfection of their respective host viruses. At first glance, the gene content of virophages does not offer any obvious explanation for such a dependency. The conserved morphogenetic gene module consists of MCP, mCP, PRO and ATPase, and virophages are thus presumably able to synthesize their own capsids. Virophage genomes also code for DNA replication proteins, which should enable them to copy their genetic information. Therefore, virophages appear to be self-sufficient for DNA replication and particle assembly.

Several observations, in contrast, suggest that transcription is the process for which virophages depend on their coinfecting host viruses. First, the computational analysis of the mavirus genome revealed that all 20 genes contain a conserved sequence motif in their immediate 5' upstream region (Fischer and Suttle, 2011). This motif consists of a 'TCTA' core which is flanked by AT-rich sequences and is located approximately 14 nt upstream of the start codon. A highly similar conserved sequence motif had been found earlier at the same upstream position of those CroV genes that are expressed late during infection (Fischer *et al.*,

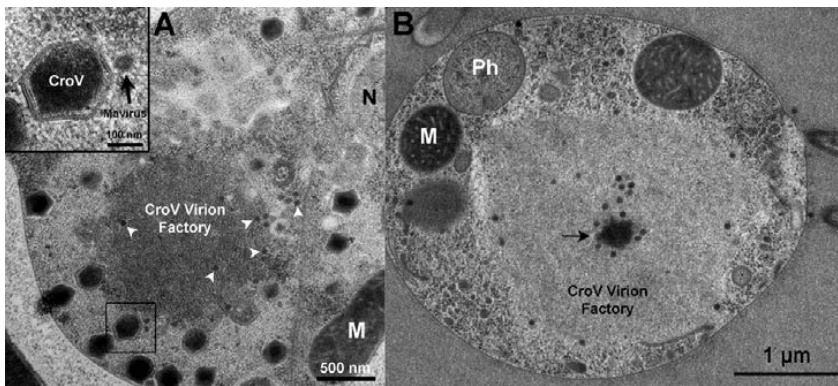


Figure 12.9 Variable CroV phenotypes during coinfection with mavirus. A) *C. roenbergensis* cell displaying a virion factory that produces both CroV and mavirus particles. Modified from (Fischer and Suttle, 2011). The inset shows a magnification of the boxed area. Mavirus particles are marked by white arrowheads (large image) or a black arrow (inset). B) *C. roenbergensis* cell featuring a large virion factory that produces no CroV virions, but contains a nested factory (arrow) that produces mavirus particles. Thin-section EM image by U. Mersdorf and M. Fischer, Max Planck Institute for Medical Research, Germany. M: mitochondrion, N: nucleus, Ph: phagosome.

2010). Second, quantitative real-time PCR analysis of cDNA generated from RNA samples isolated at different time points post infection revealed that mavirus genes become active at the beginning of late phase of CroV infection (≈ 4 h p.i.), whereas no mavirus gene activity was detected before the onset of late phase (K. Fenzl and M. Fischer, unpublished data). For Sputnik, a less obvious promoter motif was identified, which also resembled a putative late promoter motif of mimivirus (Legendre *et al.*, 2010). Third, Sputnik and mimivirus share similar transcription termination signals. In mimivirus, the location of polyadenylation sites appear to be defined by palindromic sequences, leading to the proposition of a ‘hairpin rule’ for mimivirus polyadenylation (Byrne *et al.*, 2009). In the Sputnik genome, 16 such palindromic sequences were identified, 14 of which were located in intergenic regions, which makes them candidates for polyadenylation signals (Claverie and Abergel, 2009a). In summary, these findings strongly suggest that virophages use the giant virus-encoded transcription machinery, and in particular the late phase-specific transcription factors to express their genes.

DNA replication

There are currently no data to indicate that the DNA replication proteins encoded by giant viruses may play a role in replicating virophage genomes.

Members of the *Mavirus* genus encode a predicted protein-primed B-family DNA polymerase, whereas a distinct version of bacterial DNA polymerase I (TVPol) with a predicted primase activity is found in the Sputnik and OLV clades (Iyer *et al.*, 2008; Yutin *et al.*, 2013). Protein-primed family B DNA polymerases are encoded by various parasitic elements such as adenoviruses, phi29-like viruses, tectiviruses, some archaeal viruses, and mavericks/polintons (Redrejo-Rodríguez and Salas, 2014). These polymerases use a terminal protein that is covalently attached to the 5′ ends of the genome as a primer. Due to low sequence conservation, however, a terminal protein for mavirus has not been identified yet. In addition, protein-primed replication requires a linear DNA template and the mavirus genome has circular topology, thus the exact mechanism of DNA replication in viruses of the Mavirus genus remains to be explored.

The protist–virus–virophage triangle

Host range

Lavidaviruses require two distinct hosts for successful replication, a susceptible host cell and a permissive host virus in the form of the cytoplasmic virion factory. Each of the two hosts must fulfil

certain requirements for virophage replication to occur. The host cell must be susceptible to the virophage, i.e. the virophage must be able to enter the cell. In addition, the virophage relies on the host cell for energy, metabolites, protein translation, and other essential systems. As a second host component, the virophage requires the presence of a permissive coinfecting giant virus, which provides the components (presumably transcription proteins) that are necessary for the virophage to complete its replication cycle. Currently, the only known cellular hosts for virophages are phagotrophic protists. With only two such virophage–virus–host systems in culture (*Acanthamoeba* sp. and *Cafeteria* sp.), however, the information on virophage host cell range is extremely low. Most of the environmental virophage genomes that have been assembled from metagenomic sequences, or host-integrated virophage genomes (Blanc *et al.*, 2015; Gong *et al.*, 2016; Oh *et al.*, 2016; Yau *et al.*, 2011; Zhou *et al.*, 2013) are assumed to be associated with giant viruses of the extended *Mimiviridae* family (Fischer, 2016), which mostly infect phototrophic protists. We can predict that it will only be a matter of time until the first algal virophages are isolated and grown in laboratory cultures.

Cellular host range of virophages will mainly be determined by the ability of the virophage to enter the cell. Members of the genus *Sputnikvirus* are proposed to enter the amoebal host by attachment to the external fibres of mimiviruses and via subsequent uptake of the virus–virophage composite by phagocytosis. Since this entry mode presumably does not depend on virophage–host cell contact, cellular host range should be determined solely by the giant virus host requirements, and the host cell might not be able to become resistant to the virophage while remaining susceptible to the giant virus. In contrast, members of the genus *Mavirus* enter their flagellate host cells by receptor-mediated endocytosis, a process that requires specific virus–host interactions at the cell surface. Here, it is conceivable that a cellular host becomes resistant to the virophage by mutations in the cell surface receptor, while remaining susceptible to the giant virus. Evolutionarily, this would probably result

in a selective disadvantage for the host cell, since *Mavirus* can strongly inhibit CroV replication and increase the survival of CroV-infected flagellate populations (Fischer and Hackl, 2016).

The type of entry of virophages may reflect the host specificity of their associated giant viruses. Mimiviruses are assumed to have a broad host range and can replicate in various amoebae (Reteno *et al.*, 2018). Although mimiviruses replicate well in *Acanthamoeba* sp., their main natural host is unknown. Hence, mimivirus-associated virophages may be maladapted to *Acanthamoeba*, which might explain why they are not capable of independent cell entry. *Mavirus*, on the other hand, is able to trigger specific uptake by *C. roenbergensis* cells via clathrin-mediated endocytosis. Its viral host CroV infects members of the genus *Cafeteria*, as well as closely related genera such as *Picophagus* (M. Fischer, unpublished data). Unlike mimiviruses, CroV appears to enter host cells not by phagocytosis, but by docking to the cytoplasmic membrane and releasing only the viral core into the cytoplasm (U. Mersdorf and M. Fischer, unpublished results). This suggests that marine heterotrophic nanoflagellates such as *Cafeteria* are the natural hosts for CroV and *Mavirus*.

In general, virophage replication depends on the availability of cytoplasmic transcription. Therefore, the giant virus specificity of lavidaviruses is likely determined by the ability of the giant virus-encoded transcription machinery to synthesize virophage transcripts. Based on the high similarity of late gene promoters as well as transcription termination motifs between giant viruses and their virophages, it is predicted that this interaction largely depends on the recognition of virophage gene promoters by a giant virus-encoded transcription factor. The *Mimiviridae* is not the only family of large dsDNA viruses that encode their own transcription system, and it is possible that virophage-like parasites also exist for other members of the proposed viral order *Megavirales*³, such as pithoviruses and poxviruses.

Virophages are unlikely to replicate in human cells or cause disease (Desnues *et al.*, 2012b; Parola *et al.*, 2012). The case of Sputnik 2, however, shows that virophages may be part of the human virome.

³ The order *Megavirales* was proposed to replace the informal grouping of ‘Nucleocytoplasmic large DNA viruses’, which includes eukaryotic large dsDNA viruses that are of possible monophyletic origin, with a proper taxon (Colson *et al.*, 2013; Iyer *et al.*, 2001, 2006). However, at the time of writing, the ICTV has not approved this new order, nor does the proposal appear to be scheduled for discussion any time soon.

Sputnik 2 was isolated together with lentillevirus from the contact lens of a keratitis patient, whose symptoms were probably caused by *Acanthamoeba* cells that also happened to harbour the giant virus and the virophage (Desnues *et al.*, 2012b).

Virophage–giant virus interactions

Virophages that encode integrases are able to physically link their genomes to those of other organisms. The tyrosine recombinase found in Sputnik virophages may catalyse the occasional integration into mimivirus genomes. Analysis of paired-end Illumina reads of lentillevirus DNA revealed that Sputnik 2 may also exist as a provirophage in the genome of its host virus (Desnues *et al.*, 2012b). Similarly, the PgVV genome likely integrates at multiple sites within its associated *Phaeocystis globosa* virus 16T (Santini *et al.*, 2013). Persistence as a provirophage in a giant virus genome has clear advantages for the smaller virus genome, as it will automatically coinfect and colocalize with its supporting giant virus.

Even although virophages are known to inhibit the production of giant virus particles during a coinfection, their parasitic effect on the host virus appears to vary considerably. Whereas mavirus is a potent inhibitor of CroV and coinfecting cells often produce no CroV virions at all, the negative impact of Sputnik on mimivirus is less pronounced and viral factories producing both types of capsids are frequently observed. The Zamilon virophage on the other hand has no discernible negative impact on the replication of a coinfecting mimivirus. The reason for this variation is unknown, but the degree of virus-against-virus pathogenicity most certainly affects the cellular and viral host dynamics, and the resulting ecological consequences will differ for different virophages. As is to be expected for a host–parasite system, the host can become resistant to the parasite, or acquire means to control the parasite. Mimivirus thus has been shown to become resistant to Sputnik through large-scale deletions at the genome termini, resulting in a ‘bald’ capsid phenotype where external fibres are no longer present (Boyer *et al.*, 2011). This may prevent Sputnik from entering the amoeba via a composite mechanism, although it cannot be excluded that other genes essential for virophage replication are located in the deleted regions.

A general defence mechanism against virophages

was recently proposed in mimivirus (Levasseur *et al.*, 2016). Of the three mimivirus lineages (A: mimiviruses, B: moulmouvirus, C: megaviruses), Zamilon can only replicate with lineage B and C strains, but not with lineage A mimiviruses. Genome analysis of lineage A viruses revealed an open reading frame (R349) with four 15-nt long repeats of a Zamilon DNA sequence. This so-called ‘mimivirus virophage resistance element’ (MIMIVIRE) was proposed to represent a CRISPR-Cas-like adaptive immune system against virophages. Silencing of the R349 gene allowed Zamilon to replicate in lineage mimiviruses. Two additional genes located close to the R349 gene, the helicase R350 and the nuclease R354, were also proposed to be part of this defence system, because silencing any of these two genes increased Zamilon replication in lineage A mimiviruses. Criticism has been voiced, however, over the alleged similarity to CRISPR-Cas systems, owing, amongst other things, to the lack of regularly spaced or flanking repeats that are typical for CRISPR-Cas (Claverie and Abergel, 2016; Mohanraju *et al.*, 2016).

Instead, Claverie and Abergel proposed a protein–protein interaction scenario, according to which the repeated Zamilon sequences in the R349 protein would compete for binding partners with the Zamilon ORF 4 protein, in which the repeated sequences are present (Claverie and Abergel, 2016). The R350 and R354 proteins would not be directly involved in virophage restriction, but be required for general replication or transcription processes and their silencing would thus indirectly benefit the replication of Zamilon. Clearly, the mechanism of the observed resistance to Zamilon in lineage A mimiviruses remains to be elucidated. Regardless of the specific workings of such a defence system, it becomes evident that host–parasite arms races are not restricted to cellular hosts and are also being waged between giant viruses and their viral parasites.

Virophage–host cell interactions

From the beginning, virophages were hypothesized to integrate into host genomes. This was based on the presence of integrases in Sputnik and mavirus, and the genetic relatedness between mavirus and the endogenous eukaryotic maverick/polinton elements (Fischer and Suttle, 2011; La Scola *et al.*, 2008). Mavirus/ALM and Sputnik/Zamilon

encode two different types of integrases. Whereas members of the genus *Sputnikvirus* have a bacteriophage lambda-like tyrosine recombinase (Gaia *et al.*, 2014; La Scola *et al.*, 2008), the mavirus-like virophages encode a retrovirus RVE-type integrase (Fischer and Suttle, 2011; Zhou *et al.*, 2013). The benefits of genome integration for virophages are apparent: this process physically ties the virophage to one of its two host components and increases its chances to encounter conditions favourable for replication. At the same time, integrated viral DNA is maintained and replicated by the host cell, whereas free virions are exposed to various environmental parameters that may destroy infectivity, such as ultraviolet irradiation, pH, salinity, etc.

A targeted computational search for virophages in published eukaryotic genomes revealed a group of closely related viral elements in the nuclear genome of the chlorarachniophyte alga *Bigelowiella natans* (Blanc *et al.*, 2015). The authors of this study identified 38 virophage-like elements with high similarity to each other (average 91% nucleotide similarity), ranging in size from 0.1 kb to 33.3 kb. Based on phylogenetic reconstruction of MCP and ATPase, the *B. natans* virophages are most closely related to YSLV 5 (Fig. 12.4). Interestingly, many of these virophage-like genes were found to be transcribed, including conserved genes such as MCP, mCP, and ATPase. The presence of integrated genomic fragments from putative viruses of the NCLDV clade suggests that *B. natans* is host to both giant DNA viruses as well as virophages (Blanc *et al.*, 2015). It remains unclear, however, how old these endogenous virophage-like elements are and how frequently they integrate into host cell genomes.

An experimental study in the *Cafeteria*–CroV–mavirus system was able to shed first light on these questions by showing that mavirus integrates very efficiently into the nuclear genome of *C. roenbergensis* during a coinfection with CroV (Fischer and Hackl, 2016). The authors sequenced and assembled the nuclear genome of a host strain before and after coinfection, which allowed them to determine the exact integration sites of mavirus. A least 11 distinct mavirus integrations were found within a single host genome and the integration sites displayed no obvious consensus motif, suggesting that integration is not sequence-specific. Host and virus population data from this study showed that nearly

every CroV-infected host cell is destined to lyse. The high percentage of *de novo* provirophages can be explained by the ability of mavirus to integrate into the host genome in the absence of CroV (M. Fischer, unpublished data). Unlike the *B. natans* virophages, the endogenous mavirus genes were found to be transcriptionally silent. Infection of a mavirus-bearing cell line with CroV, however, resulted in gene expression, DNA replication, and particle formation of the endogenous mavirus genomes.

The presence of the endogenous virophage did not benefit the CroV-infected host cell directly, nor was CroV replication inhibited as a result of mavirus reactivation. During subsequent rounds of infection, however, the reactivated mavirus particles inhibited CroV replication during regular coinfections, which led to a significantly reduced host cell mortality rate, depending on the MOI of mavirus and CroV (Fischer and Hackl, 2016). Thus, endogenous virophages can in principle protect host populations from giant virus infection, as proposed earlier (Fischer and Suttle, 2011). A crucial condition to ensure host survival, however, is that the spread of the lytic virus can be stopped before all cells are infected. The initially infected provirophage-bearing cells would be sacrificed in an altruistic manner to release virophage particles and eventually protect the population. Not all virophages may act in a way similar to mavirus though, and the sporadic distribution of integrases in virophage genomes, as well as the moderate effect of Zamilon on mimiviruses, suggests that not all virophages are beneficial to their cellular hosts. This should not, however, detract from the possible ecological importance of virophages on their viral and cellular hosts, which was proposed soon after the discovery of virophages (Fischer and Suttle, 2011; Yau *et al.*, 2011).

Several studies have explored host–virus–virophage dynamics *in silico* using mathematical models. Yau *et al.* (2011) presented a Lotka–Volterra simulation with OLV as a predator of a lytic giant virus which in turn preys on an alga. In this model, the presence of the virophage reduced the recovery time of the host population after giant virus lysis. Algal blooms were predicted to occur more frequently, with an overall increase in secondary production due to the action of the virophage. Wodarz explored the evolutionary dynamics of a

virophage-containing tripartite system by modeling cells infected with one, both, or none of the two viruses (Wodarz, 2013). This model suggested that the virophage evolves towards higher levels of pathogenicity on the giant virus, which may lead to instability and eventual extinction of giant virus and virophage. Thus, a refined model is needed to allow for persistence of the virophage, which is apparently the case for naturally occurring virophages and their hosts. Taylor *et al.* (2014) constructed a model in which they paid particular attention to different entry modes of virophages, i.e. entry either as a composite with the giant virus (Sputnik) or independently of the giant virus (mavirus). The authors found that both entry modes allow for stable coexistence of such a tripartite system, albeit with slight differences. Also, entry mode did not influence the beneficial effect of virophages on host populations as long as the virophage was pathogenic for the giant virus. Clearly more experimental data are needed to validate these theoretical predictions and explore the ecological role of virophages.

Conclusion

Viruses are often perceived as the ultimate parasites, encoding just the bare essentials for successfully infecting and manipulating a host cell. The discovery of giant DNA viruses with genome lengths upwards of 1 million base pairs contradicted this paradigm. Increasing complexity allows these viruses to become independent from the host cell in many biochemical pathways, such as DNA replication and repair, transcription, or glycosylation. On the other hand, building enzymatically rich replication compartments also bears the risk of becoming a target for other parasites. Virophages of the family *Lavidaviridae* have evolved to utilize the giant virus-encoded transcription machinery for their gene expression, instead of locating to the nucleus to use the DNA-dependent RNA polymerase of the host. Until recently, virophages have escaped human notice because they are primarily associated with unicellular eukaryotic hosts and cause no disease in plants or animals. Nevertheless, these viruses are probably very ancient and can be found all over the globe in diverse marine and freshwater environments, where they replicate within various heterotrophic and presumably also phototrophic protists, and are associated with different types of

giant DNA viruses. Many aspects of their infection biology, ecology and evolution remain to be studied, but from the very beginning of their discovery, these fascinating ‘viruses of viruses’ have captured the interest of scientists and the general public alike.

References

- Bekliz, M., Verneau, J., Benamar, S., Raoult, D., La Scola, B., and Colson, P. (2015). A new Zamilon-like virophage partial genome assembled from a bioreactor metagenome. *Front. Microbiol.* 6, 1308. <https://doi.org/10.3389/fmicb.2015.01308>.
- Bekliz, M., Colson, P., and La Scola, B. (2016). The expanding family of virophages. *Viruses* 8, E317.
- Bellas, C.M., Anesio, A.M., and Barker, G. (2015). Analysis of virus genomes from glacial environments reveals novel virus groups with unusual host interactions. *Front. Microbiol.* 6, 656. <https://doi.org/10.3389/fmicb.2015.00656>.
- Blanc, G., Gallot-Lavallée, L., and Maumus, F. (2015). Provirophages in the *Bigelowiella* genome bear testimony to past encounters with giant viruses. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5318–26. <https://doi.org/10.1073/pnas.1506469112>.
- Boughalmi, M., Saadi, H., Pagnier, I., Colson, P., Fournous, G., Raoult, D., and La Scola, B. (2013). High-throughput isolation of giant viruses of the *Mimiviridae* and *Marseilleviridae* families in the Tunisian environment. *Environ. Microbiol.* 15, 2000–2007. <https://doi.org/10.1111/1462-2920.12068>.
- Boyer, M., Madoui, M.A., Gimenez, G., La Scola, B., and Raoult, D. (2010). Phylogenetic and phyletic studies of informational genes in genomes highlight existence of a 4 domain of life including giant viruses. *PLOS ONE* 5, e15530. <https://doi.org/10.1371/journal.pone.0015530>.
- Boyer, M., Azza, S., Barrassi, L., Klose, T., Campocasso, A., Pagnier, I., Fournous, G., Borg, A., Robert, C., Zhang, X., *et al.* (2011). Mimivirus shows dramatic genome reduction after intraamoebal culture. *Proc. Natl. Acad. Sci. U.S.A.* 108, 10296–10301. <https://doi.org/10.1073/pnas.1101118108>.
- Broyles, S.S. (2003). Vaccinia virus transcription. *J. Gen. Virol.* 84, 2293–2303.
- Burroughs, A.M., Iyer, L.M., and Aravind, L. (2007). Comparative genomics and evolutionary trajectories of viral ATP dependent DNA-packaging systems. *Genome Dyn* 3, 48–65. <https://doi.org/10.1159/000107603>.
- Byrne, D., Grzela, R., Lartigue, A., Audic, S., Chenivesse, S., Encinas, S., Claverie, J.M., and Abergel, C. (2009). The polyadenylation site of Mimivirus transcripts obeys a stringent ‘hairpin rule’. *Genome Res.* 19, 1233–1242. <https://doi.org/10.1101/gr.091561.109>.
- Campos, R.K., Boratto, P.V., Assis, F.L., Aguiar, E.R., Silva, L.C., Albarnaz, J.D., Dornas, F.P., Trindade, G.S., Ferreira, P.P., Marques, J.T., *et al.* (2014). Samba virus: a novel mimivirus from a giant rain forest, the Brazilian Amazon. *Virol. J.* 11, 95. <https://doi.org/10.1186/1743-422X-11-95>.

- Claverie, J.M., and Abergel, C. (2009a). Mimivirus and its viroplasm. *Annu. Rev. Genet.* 43, 49–66. <https://doi.org/10.1146/annurev-genet-102108-134255>.
- Claverie, J.M., and Abergel, C. (2016). CRISPR-Cas-like system in giant viruses: why MIMIVIRE is not likely to be an adaptive immune system. *Virol. Sin.* 31, 193–196. <https://doi.org/10.1007/s12250-016-3801-x>.
- Claverie, J.M., and Abergel, C. (2009b). Mimivirus and its viroplasm. *Annu. Rev. Genet.* 43, 49–66. <https://doi.org/10.1146/annurev-genet-102108-134255>.
- Colson, P., De Lamballerie, X., Yutin, N., Asgari, S., Bigot, Y., Bideshi, D.K., Cheng, X.W., Federici, B.A., Van Etten, J.L., Koonin, E.V., et al. (2013). 'Megavirales', a proposed new order for eukaryotic nucleocytoplasmic large DNA viruses. *Arch. Virol.* 158, 2517–2521. <https://doi.org/10.1007/s00705-013-1768-6>.
- Desnues, C., and Raoult, D. (2010). Inside the lifestyle of the viroplasm. *Intervirology* 53, 293–303. <https://doi.org/10.1159/000312914>.
- Desnues, C., and Raoult, D. (2012). Viroplasm question the existence of satellites. *Nat. Rev. Microbiol.* 10, 234. <https://doi.org/10.1038/nrmicro2676-c3>.
- Desnues, C., Boyer, M., and Raoult, D. (2012a). Sputnik, a viroplasm infecting the viral domain of life. *Adv. Virus Res.* 82, 63–89. <https://doi.org/10.1016/B978-0-12-394621-8.00013-3>.
- Desnues, C., La Scola, B., Yutin, N., Fournous, G., Robert, C., Azza, S., Jardot, P., Montell, S., Campocasso, A., Koonin, E.V., et al. (2012b). Proviruses and transposons as the diverse mobilome of giant viruses. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18078–18083. <https://doi.org/10.1073/pnas.1208835109>.
- Fischer, M.G. (2011). Sputnik and Mavirus: more than just satellite viruses. *Nat. Rev. Microbiol.* 10, 78. <https://doi.org/10.1038/nrmicro2676-c1>.
- Fischer, M.G. (2015). Viroplasm go nuclear in the marine alga *Bigelowiella natans*. *Proc. Natl. Acad. Sci. U.S.A.* 112, 11750–11751. <https://doi.org/10.1073/pnas.1515142112>.
- Fischer, M.G. (2016). Giant viruses come of age. *Curr. Opin. Microbiol.* 31, 50–57.
- Fischer, M.G., and Hackl, T. (2016). Host genome integration and giant virus-induced reactivation of the viroplasm mavirus. *Nature* 540, 288–291. <https://doi.org/10.1038/nature20593>.
- Fischer, M.G., and Suttle, C.A. (2011). A viroplasm at the origin of large DNA transposons. *Science* 332, 231–234. <https://doi.org/10.1126/science.1199412>.
- Fischer, M.G., Allen, M.J., Wilson, W.H., and Suttle, C.A. (2010). Giant virus with a remarkable complement of genes infects marine zooplankton. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19508–19513. <https://doi.org/10.1073/pnas.1007615107>.
- Fischer, M.G., Kelly, I., Foster, L.J., and Suttle, C.A. (2014). The virion of *Cafeteria roenbergensis* virus (CroV) contains a complex suite of proteins for transcription and DNA repair. *Virology* 466–467, 82–94. <https://doi.org/10.1016/j.virol.2014.05.029>.
- Gaia, M., Colson, P., Desnues, C., and La Scola, B. (2013a). Viroplasm Concept. *The InLS*, (Chichester, UK: John Wiley & Sons, Ltd), pp. 1–12.
- Gaia, M., Pagnier, I., Campocasso, A., Fournous, G., Raoult, D., and La Scola, B. (2013b). Broad spectrum of mimiviridae viroplasm allows its isolation using a mimivirus reporter. *PLOS ONE* 8, e61912. <https://doi.org/10.1371/journal.pone.0061912>.
- Gaia, M., Benamar, S., Boughalmi, M., Pagnier, I., Croce, O., Colson, P., Raoult, D., and La Scola, B. (2014). Zamilon, a novel viroplasm with Mimiviridae host specificity. *PLOS ONE* 9, e94923. <https://doi.org/10.1371/journal.pone.0094923>.
- Garza, D.R., and Suttle, C.A. (1995). Large double-stranded DNA viruses which cause the lysis of a marine heterotrophic nanoflagellate (*Bodo* sp) occur in natural marine viral communities. *Aquat. Microb. Ecol.* 9, 203–210.
- Gong, C., Zhang, W., Zhou, X., Wang, H., Sun, G., Xiao, J., Pan, Y., Yan, S., and Wang, Y. (2016). Novel viroplasm discovered in a freshwater lake in China. *Front. Microbiol.* 7, 5. <https://doi.org/10.3389/fmicb.2016.00005>.
- Gonzalez, J.M., and Suttle, C.A. (1993). Grazing by marine nanoflagellates on viruses and virus-sized particles – ingestion and digestion. *Mar. Ecol. Prog. Ser.* 94, 1–10.
- Greber, U.F., Willetts, M., Webster, P., and Helenius, A. (1993). Stepwise dismantling of adenovirus 2 during entry into cells. *Cell* 75, 477–486.
- Iyer, L.M., Aravind, L., and Koonin, E.V. (2001). Common origin of four diverse families of large eukaryotic DNA viruses. *J. Virol.* 75, 11720–11734. <https://doi.org/10.1128/JVI.75.23.11720-11734.2001>.
- Iyer, L.M., Balaji, S., Koonin, E.V., and Aravind, L. (2006). Evolutionary genomics of nucleocytoplasmic large DNA viruses. *Virus Res.* 117, 156–184.
- Iyer, L.M., Abhimanyu, S., and Aravind, L. (2008). A new family of polymerases related to superfamily A DNA polymerases and T7-like DNA-dependent RNA polymerases. *Biol. Direct* 3, 39. <https://doi.org/10.1186/1745-6150-3-39>.
- Jordan, I.K., Wolf, Y.I., and Koonin, E.V. (2004). Duplicated genes evolve slower than singletons despite the initial rate increase. *BMC Evol. Biol.* 4, 22. <https://doi.org/10.1186/1471-2148-4-22>.
- Kapitonov, V.V., and Jurka, J. (2006). Self-synthesizing DNA transposons in eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 103, 4540–4545.
- Kerepesi, C., and Grolmusz, V. (2017). The 'Giant Virus Finder' discovers an abundance of giant viruses in the Antarctic dry valleys. *Arch. Virol.* 162, 1671–1676. <https://doi.org/10.1007/s00705-017-3286-4>.
- King, A.M.Q., Adams, M.J., Carstens, E.B., and Lefkowitz, E.J. (2011). Virus Taxonomy. Ninth Report of the International Committee on Taxonomy of Viruses. (Elsevier Academic, London).
- Klose, T., Herbst, D.A., Zhu, H., Max, J.P., Kenttämä, H.I., and Rossmann, M.G. (2015). A mimivirus enzyme that participates in viral entry. *Structure* 23, 1058–1065. <https://doi.org/10.1016/j.str.2015.03.023>.
- Koonin, E.V., and Krupovic, M. (2017). Polintons, viroplasm and transposons: a tangled web linking viruses, transposons and immunity. *Curr. Opin. Virol.* 25, 7–15.
- Krupovic, M., and Cvikrkaite-Krupovic, V. (2011). Viroplasm or satellite viruses? *Nat. Rev. Microbiol.* 9, 762–763.
- Krupovic, M., and Koonin, E.V. (2016). Self-synthesizing transposons: unexpected key players in the evolution of

- viruses and defense systems. *Curr. Opin. Microbiol.* 31, 25–33.
- Krupovic, M., and Koonin, E.V. (2015). Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat. Rev. Microbiol.* 13, 105–115. <https://doi.org/10.1038/nrmicro3389>.
- Krupovic, M., Bamford, D.H., and Koonin, E.V. (2014). Conservation of major and minor jelly-roll capsid proteins in Polinton (Maverick) transposons suggests that they are bona fide viruses. *Biol. Direct* 9, 6. <https://doi.org/10.1186/1745-6150-9-6>.
- Krupovic, M., Kuhn, J.H., and Fischer, M.G. (2016). A classification system for virophages and satellite viruses. *Arch. Virol.* 161, 233–247. <https://doi.org/10.1007/s00705-015-2622-9>.
- Legendre, M., Audic, S., Poirot, O., Hingamp, P., Seltzer, V., Byrne, D., Lartigue, A., Lescot, M., Bernadac, A., Poulain, J., et al. (2010). mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res.* 20, 664–674. <https://doi.org/10.1101/gr.102582.109>.
- Levasseur, A., Bekliz, M., Chabrière, E., Pontarotti, P., La Scola, B., and Raoult, D. (2016). MIMIVIRE is a defence system in mimivirus that confers resistance to virophage. *Nature* 531, 249–252. <https://doi.org/10.1038/nature17146>.
- Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* 353, aad5147. <https://doi.org/10.1126/science.aad5147>.
- Mutsaers, Y., Zauberman, N., Sabanay, I., and Minsky, A. (2010). Vaccinia-like cytoplasmic replication of the giant Mimivirus. *Proc. Natl. Acad. Sci. U.S.A.* 107, 5978–5982. <https://doi.org/10.1073/pnas.0912737107>.
- Nasir, A., Kim, K.M., and Caetano-Anolles, G. (2012). Giant viruses coexisted with the cellular ancestors and represent a distinct supergroup along with superkingdoms Archaea, Bacteria and Eukarya. *BMC Evol. Biol.* 12, 156. <https://doi.org/10.1186/1471-2148-12-156>.
- O'Day, D.H., Suhre, K., Myre, M.A., Chatterjee-Chakraborty, M., and Chavez, S.E. (2006). Isolation, characterization, and bioinformatic analysis of calmodulin-binding protein cmbB reveals a novel tandem IP22 repeat common to many Dictyostelium and Mimivirus proteins. *Biochem. Biophys. Res. Commun.* 346, 879–888.
- Oh, S., Yoo, D., and Liu, W.T. (2016). Metagenomics reveals a novel virophage population in a tibetan mountain lake. *Microbes Environ.* 31, 173–177. <https://doi.org/10.1264/jsme2.ME16003>.
- Parola, P., Renvoisé, A., Botelho-Nevers, E., La Scola, B., Desnues, C., and Raoult, D. (2012). *Acanthamoeba polyphaga mimivirus* virophage seroconversion in travelers returning from Laos. *Emerg. Infect. Dis.* 18, 1500–1502. <https://doi.org/10.3201/eid1809.120099>.
- Pei, J., Kim, B.H., and Grishin, N.V. (2008). PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nuc. Acids Res.* 36, 2295–2300. <https://doi.org/10.1093/nar/gkn072>.
- Pelletier, N., Raoult, D., and La Scola, B. (2009). Specific recognition of the major capsid protein of *Acanthamoeba polyphaga mimivirus* by sera of patients infected by *Francisella tularensis*. *FEMS Microbiol. Lett.* 297, 117–123. <https://doi.org/10.1111/j.1574-6968.2009.01675.x>.
- Piacente, F., Marin, M., Molinaro, A., De Castro, C., Seltzer, V., Salis, A., Damonte, G., Bernardi, C., Claverie, J.M., Abergel, C., et al. (2012). Giant DNA virus mimivirus encodes pathway for biosynthesis of unusual sugar 4-amino-4,6-dideoxy-D-glucose (Viosamine). *J. Biol. Chem.* 287, 3009–3018. <https://doi.org/10.1074/jbc.M111.314559>.
- Pritham, E.J., Putliwala, T., and Feschotte, C. (2007). Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. *Gene* 390, 3–17.
- Raoult, D., Audic, S., Robert, C., Abergel, C., Renesto, P., Ogata, H., La Scola, B., Suzan, M., and Claverie, J.M. (2004). The 1.2-megabase genome sequence of Mimivirus. *Science* 306, 1344–1350.
- Redrejo-Rodríguez, M., and Salas, M.L. (2014). Repair of base damage and genome maintenance in the nucleocytoplasmic large DNA viruses. *Virus Res.* 179, 12–25. <https://doi.org/10.1016/j.virusres.2013.10.017>.
- Reteno, D.G.I., Bajrai, L.H., Aherfi, S., Colson, P. and La Scola, B. (2018). Protozoal Giant Viruses. In: *Viruses of Microorganisms*, P. Hyman and S. T. Abedon, ed. (Norfolk, UK: Caister Academic Press), pp. 245–270. DOI: <https://doi.org/10.21775/9781910190852.11>
- Ronquist, F., and Huelsenbeck, J.P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574.
- Roux, S., Chan, L.-K., Egan, R., Malmstrom, R.R., McMahon, K.D., and Sullivan, M.B. (2017). Ecogenomics of virophages and their giant virus hosts assessed through time series metagenomics. *Nat. Commun.* 8, 858.
- Santini, S., Jeudy, S., Bartoli, J., Poirot, O., Lescot, M., Abergel, C., Barbe, V., Wommack, K.E., Noordeloos, A.A., Brussaard, C.P., et al. (2013). Genome of *Phaeocystis globosa* virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10800–10805. <https://doi.org/10.1073/pnas.1303251110>.
- La Scola, B., Audic, S., Robert, C., Jungang, L., de Lamballerie, X., Drancourt, M., Birtles, R., Claverie, J.-M., and Raoult, D. (2003). A giant virus in amoebae. *Science* 299, 2033.
- La Scola, B., Desnues, C., Pagnier, I., Robert, C., Barrassi, L., Fournous, G., Merchat, M., Suzan-Monti, M., Forterre, P., Koonin, E., et al. (2008). The virophage as a unique parasite of the giant mimivirus. *Nature* 455, 100–104. <https://doi.org/10.1038/nature07218>.
- La Scola, B., Campocasso, A., N'Dong, R., Fournous, G., Barrassi, L., Flaudrops, C., and Raoult, D. (2010). Tentative characterization of new environmental giant viruses by MALDI-TOF mass spectrometry. *Intervirology* 53, 344–353. <https://doi.org/10.1159/000312919>.
- Suhre, K. (2005). Gene and genome duplication in *Acanthamoeba polyphaga Mimivirus*. *J. Virol.* 79, 14095–14101.
- Sun, S., La Scola, B., Bowman, V.D., Ryan, C.M., Whitelegge, J.P., Raoult, D., and Rossmann, M.G. (2010). Structural studies of the Sputnik virophage. *J. Virol.* 84, 894–897. <https://doi.org/10.1128/JVI.01957-09>.

- Taylor, B.P., Cortez, M.H., and Weitz, J.S. (2014). The virus of my virus is my friend: ecological effects of virophage with alternative modes of coinfection. *J. Theor. Biol.* 354, 124–136. <https://doi.org/10.1016/j.jtbi.2014.03.008>.
- Williams, T.A., Embley, T.M., and Heinz, E. (2011). Informational gene phylogenies do not support a fourth domain of life for nucleocytoplasmic large DNA viruses. *PLOS ONE* 6, e21080. <https://doi.org/10.1371/journal.pone.0021080>.
- Wodarz, D. (2013). Evolutionary dynamics of giant viruses and their virophages. *Evol. Ecol.* 3, 2103–2115. <https://doi.org/10.1002/ece.3.600>.
- Xiao, C., Kuznetsov, Y.G., Sun, S., Hafenstein, S.L., Kostyuchenko, V.A., Chipman, P.R., Suzan-Monti, M., Raoult, D., McPherson, A., and Rossmann, M.G. (2009). Structural studies of the giant mimivirus. *PLOS Biol.* 7, e92. <https://doi.org/10.1371/journal.pbio.1000092>.
- Yau, S., Lauro, F.M., DeMaere, M.Z., Brown, M.V., Thomas, T., Raftery, M.J., Andrews-Pfannkoch, C., Lewis, M., Hoffman, J.M., Gibson, J.A., *et al.* (2011). Virophage control of antarctic algal host-virus dynamics. *Proc. Natl. Acad. Sci. U.S.A.* 108, 6163–6168. <https://doi.org/10.1073/pnas.1018221108>.
- Yutin, N., Raoult, D., and Koonin, E.V. (2013). Virophages, polintons, and transpovirons: a complex evolutionary network of diverse selfish genetic elements with different reproduction strategies. *Virol. J.* 10, 158. <https://doi.org/10.1186/1743-422X-10-158>.
- Yutin, N., Wolf, Y.I., and Koonin, E.V. (2014). Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* 466–467, 38–52.
- Yutin, N., Kapitonov, V.V., and Koonin, E.V. (2015a). A new family of hybrid virophages from an animal gut metagenome. *Biol. Direct* 10, 19. <https://doi.org/10.1186/s13062-015-0054-9>.
- Yutin, N., Shevchenko, S., Kapitonov, V., Krupovic, M., and Koonin, E.V. (2015b). A novel group of diverse Polinton-like viruses discovered by metagenome analysis. *BMC Biol.* 13, 95. <https://doi.org/10.1186/s12915-015-0207-4>.
- Zablocki, O., van Zyl, L., Adriaenssens, E.M., Rubagotti, E., Tuffin, M., Cary, S.C., and Cowan, D. (2014). High-level diversity of tailed phages, eukaryote-associated viruses, and virophage-like elements in the metaviromes of antarctic soils. *Appl. Environ. Microbiol.* 80, 6888–6897. <https://doi.org/10.1128/AEM.01525-14>.
- Zhang, X., Sun, S., Xiang, Y., Wong, J., Klose, T., Raoult, D., and Rossmann, M.G. (2012). Structure of Sputnik, a virophage, at 3.5-Å resolution. *Proc. Natl. Acad. Sci. U.S.A.* 109, 18431–18436. <https://doi.org/10.1073/pnas.1211702109>.
- Zhou, J., Zhang, W., Yan, S., Xiao, J., Zhang, Y., Li, B., Pan, Y., and Wang, Y. (2013). Diversity of virophages in metagenomic data sets. *J. Virol.* 87, 4225–4236. <https://doi.org/10.1128/JVI.03398-12>.
- Zhou, J., Sun, D., Childers, A., McDermott, T.R., Wang, Y., and Liles, M.R. (2015). Three novel virophage genomes discovered from Yellowstone Lake metagenomes. *J. Virol.* 89, 1278–1285. <https://doi.org/10.1128/JVI.03039-14>.