A Persistent Giant Algal Virus, with a Unique Morphology, Encodes an Unprecedented Number of Genes Involved in Energy Metabolism

Romain Blanc-Mathieu,^{a,b} Håkon Dahle,^{c,d,g} Antje Hofgaard,^e David Brandt,^f Hiroki Ban,^a Jörn Kalinowski,^f Hiroyuki Ogata,^a Ruth-Anne Sandaa^c

^aInstitute for Chemical Research, Kyoto University, Gokasho, Uji, Japan

^bLaboratoire de Physiologie Cellulaire & Végétale, CEA, Université Grenoble Alpes, CNRS, INRA, IRIG, Grenoble, France

^cDepartment of Biological Sciences, University of Bergen, Bergen, Norway

^dK.G. Jebsen Center for Deep Sea Research, University of Bergen, Bergen, Norway

^eDepartment of Biosciences, University of Oslo, Oslo, Norway

^fCenter for Biotechnology, Bielefeld University, Bielefeld, Germany

^gComputational Biology Unit, Department of Informatics, University of Bergen, Bergen, Norway

ABSTRACT Viruses have long been viewed as entities possessing extremely limited metabolic capacities. Over the last decade, however, this view has been challenged, as metabolic genes have been identified in viruses possessing large genomes and virions—the synthesis of which is energetically demanding. Here, we unveil peculiar phenotypic and genomic features of *Prymnesium kappa* virus RF01 (PkV RF01), a giant virus of the *Mimiviridae* family. We found that this virus encodes an unprecedented number of proteins involved in energy metabolism, including all four succinate dehydrogenase (SDH) subunits (A to D), as well as key enzymes in the β -oxidation pathway. The *SDHA* gene was transcribed upon infection, indicating that the viral SDH is actively used by the virus, potentially to modulate its host's energy metabolism. We detected orthologous *SDHA* and *SDHB* genes in numerous genome fragments from uncultivated marine *Mimiviridae* viruses, which suggests that the viral SDH is widespread in oceans. PkV RF01 was less virulent than other cultured prymnesioviruses, a phenomenon that may be linked to the metabolic capacity of this virus and is suggestive of relatively long coevolution with its hosts. It also has a unique morphology compared to those of other characterized viruses in the *Mimiviridae* family. Finally, we found that PkV RF01 is the only alga-infecting *Mimiviridae* virus encoding two aminoacyl-tRNA synthetases and enzymes corresponding to an entire base excision repair (BER) pathway, as seen in heterotroph-infecting *Mimiviridae* viruses. These *Mimiviridae* encoded-enzymes were found to be monophyletic and branching at the root of the eukaryotic tree of life. This placement suggests that the last common ancestor of *Mimiviridae* was endowed with a large, complex genome prior to the divergence of known extant eukaryotes.

IMPORTANCE Viruses on Earth are tremendously diverse in terms of morphology, functionality, and genomic composition. Over the last decade, the conceptual gap separating viruses and cellular life has tightened because of the detection of metabolic genes in viral genomes that express complex virus phenotypes upon infection. Here, we describe *Prymnesium kappa* virus RF01, an alga-infecting large virus with a unique morphology, an atypical infection profile, and an unprecedented number of genes involved in energy metabolism (such as the tricarboxylic acid [TCA] cycle and the β -oxidation pathway). Moreover, we show that the gene corresponding to one of these enzymes (the succinate dehydrogenase subunit A) is transcribed during infection and is widespread among marine viruses. This discovery provides evidence

Citation Blanc-Mathieu R, Dahle H, Hofgaard A, Brandt D, Ban H, Kalinowski J, Ogata H, Sandaa R-A. 2021. A persistent giant algal virus, with a unique morphology, encodes an unprecedented number of genes involved in energy metabolism. *J Virol* 95:e02446-20. <https://doi.org/10.1128/JVI.02446-20>.

Editor Joanna L. Shisler, University of Illinois at Urbana Champaign

Copyright © 2021 American Society for Microbiology. All Rights Reserved.

Address correspondence to Ruth-Anne Sandaa, ruth.sandaa@uib.no.

Received 13 January 2021

Accepted 15 January 2021

Accepted manuscript posted online 3 February 2021

Published 25 March 2021

that a virus has the potential to actively regulate energy metabolism with its own genes.

KEYWORDS β -oxidation, *Mimiviridae*, algal virus, aminoacyl-tRNA synthetases, coevolution, energy production, metabolism, persistent, succinate dehydrogenase

In their essay “Varieties of Living Things: Life at the Intersection of Lineage and Metabolism,” Dupré and O’Malley proposed to address Schrödinger’s question “What is Life?” by “describing a spectrum of biological entities that illustrates why no sharp dividing line between living and non-living things is likely to be useful” (1). Microbiologists have contributed considerably to this descriptive effort, both by reporting the existence of viruses endowed with genes coding for functions once thought to be exclusive to cellular life and by concomitantly proposing that actively infecting viruses are a “living form” (2–4). Genes encoding elements for photosynthesis (5, 6), carbon metabolism (7), and nitrogen (8) and sulfur cycling (9) have been found in bacterial viruses, where they are used to maintain or augment cellular processes during infection and to redirect energy and resources toward viral production (8, 10, 11). Genes for protein synthesis, including translation initiation, elongation, and termination, and genes encoding a range of aminoacyl-tRNA synthetases have been found in *Mimiviridae*, a group of giant viruses infecting single-celled eukaryotes (12–14). *Mimiviridae* and other large DNA viruses, including some bacterial viruses, also have tRNA genes (15, 16). Ribosomal proteins have recently been reported in viral genomes derived from metagenomes (17). Genes involved in other metabolic processes, such as fermentation (18), glycosylation (19), photosynthesis (20), and rhodopsin synthesis (21), are encoded in *Mimiviridae* and other related large eukaryotic DNA viruses. Metabolic genes are frequently observed within virus genomes (20, 22, 23); although they represent a tiny fraction of the viral gene pool, these genes have the potential to dramatically modify the phenotype of an actively infected cell and alter the ecological role of the host (7, 24, 25). The infected host in this state has been referred to as a virocell (2). One might expect that the interplay between viral genes and host genes in virocells would become increasingly fine-tuned and complex during prolonged virus-host coevolution, which also typically leads to lower virulence. Much of the complexity of virocells may still be undetected, as most *Mimiviridae* viruses isolated with their natural hosts (mostly algae) are highly virulent, with several being involved in rapid algal bloom termination events (26).

Viruses of the *Mimiviridae* family are known to infect heterotrophic and autotrophic microbial eukaryotes. This divide is also reflected in the phylogeny of these viruses, some of which are classified into two proposed subfamilies, “Megavirinae” and “Mesomimivirinae” (27). The former contains viruses with genomes larger than 1 Mbp, all isolated from Amoebozoa, while the latter includes viruses with smaller genomes isolated from haptophyte algae of class Prymnesiophyceae. Several *Mimiviridae* members outside these two groups have been characterized to some extent as well, namely viruses isolated from heterotrophs (*Cafeteria roenbergensis* virus, CroV; *Bodo saltans* virus, BsV; and ChoanoVirus, ChoanoV), autotrophs (*Aureococcus anophagefferens* virus, AaV; *Tetraselmis* virus 1, TetV; *Pyramimonas orientalis* virus, PoV; and *Prymnesium kappa* virus RF01, PkV RF01), a metazoan (Namao virus), and metagenomes (klosneuviruses). The Mesomimivirinae subfamily includes viruses infecting bloom-forming hosts, such as *Phaeocystis pouchetii*, *Phaeocystis globosa*, and *Prymnesium parvum* (*Phaeocystis pouchetii* virus [PpV], *Phaeocystis globosa* virus [PgV group I], and *Prymnesium parvum* DNA virus BW1 [PpDNAV], respectively) (28–30); it also includes several viruses infecting *Haptolina ericina* and *Prymnesium kappa*, which normally do not form massive blooms but are present at low densities in seawater year-round (31). In marine environments, viruses infecting low-density and non-bloom-forming algae may be the most common virus-host systems—that is, low-density hosts (nonblooming) and viruses that appear to have coevolved in response to this host growth strategy. Thus far, the only known

FIG 1 PkV RF01 morphology and reduced viral infectivity under chloroform treatment. (A) Screenshot of a cryo-electron tomogram of a PkV RF01 virion. (B) Composite image of 61 cryo-electron tomograms (-60° to 60° , imaged every 2°). Purple, capsid; green, inner membrane consisting of multiple irregular, convoluted membranes; blue, internal rod-shaped core filled with dense material. The full set of records is available on GitHub (see "Data availability"). Bar, 100 nm. (C) Reduction of PkV RF01 infectivity with chloroform. Experiments were set up in triplicate, and host cells were counted by flow cytometry. Chloroform-treated PkV RF01 was added to exponentially growing He UiO028 cells in a 1:10 volume ratio. Controls were He UiO028 cells incubated with chloroform-treated medium (control A), untreated PkV RF01 (control B), and untreated medium (control C). Standard deviations (SDs) are indicated with error bars.

representatives of such viruses are *Prymnesium kappa* viruses RF01 (PkV RF01) and RF02 (PkV RF02), *Haptolina ericina* virus RF02 (HeV RF02), and *Chrysochromulina ericina* virus (CeV 01B, infecting *Haptolina ericina*) (32, 33). Together with PgV, all of these viruses, except for PkV RF01, belong to the subfamily Mesomimivirinae, on the basis of their monophyletic relationship and, in the case of PgV and CeV, a shared genomic similarity (27). In contrast, phylogenetic analysis of two partially sequenced marker genes has placed PkV RF01 deep inside the *Mimiviridae* clade, and characterization of its life cycle has revealed an atypical infection profile (33). Here, we report new phenotypic features, as well as new viral functions inferred from analysis of the genome sequence of PkV RF01. We found that this virus has a unique morphology, is less virulent than most other alga-infecting viruses, and possesses an unprecedented number of energy-generating genes. We uncovered clues suggesting that members of *Mimiviridae* that potentially modulate the metabolism of their hosts are widespread in the ocean. Our findings of peculiar genomic features in a persistent virus provide new insights on virus-host coevolution and may stimulate further advances in modeling the history of their interaction.

RESULTS AND DISCUSSION

PkV RF01 has an atypical morphology. The icosahedral PkV RF01 particle is approximately 400 nm in diameter (Fig. 1A and B). Beneath the capsid, several convoluted inner membranes fill approximately 66% of the interior. Treatment with chloroform can be used to identify possible functions of lipid membranes, as it acts to

TABLE 1 Infection parameters of *Prymnesium kappa* viruses RF01 and RF02 and *Haptolina ericina* virus RF02

Viral species and hosts	Infectious progeny/ml (MPN) ^a	Host cells/ml (FCM) ^a	Total VLP/ml (FCM)	Burst size (no. of VLP) ^b	Infectivity (%) ^c	No. of infectious particles in a burst ^d
PkV RF01 (He UiO028)	$2.9 \times 10^6 (\pm 0.2 \times 10^6)$	4.9×10^5	$1.8 \times 10^8 (\pm 0.9 \times 10^8)$	363	2	6
PkV RF02 (Pk RCC3423)	$2.2 \times 10^8 (\pm 0.2 \times 10^8)$	4.6×10^5	$5.0 \times 10^8 (\pm 0.1 \times 10^8)$	1,093	44	483
HeV RF02 (He UiO028)	$5.8 \times 10^7 (\pm 0.2 \times 10^7)$	4.9×10^5	$4.4 \times 10^8 (\pm 0.0 \times 10^8)$	907	13	119

^aMeasurement performed in duplicates. FCM, flow cytometry.

^bThe number of viral particles released from each host cell, estimated from the total number of host cells preinfection and the total number of virus-like particles (VLPs) produced during the infection cycle.

^cEstimated as the percentage of infectious progeny of all VLPs produced during the infection cycle.

^dNumber of infectious particles released per host cell.

^eMPN, most probable number.

remove lipid molecules that might be essential for successful infection (34). Some algal viruses in the nucleocytoplasmic large DNA virus (NCLDV) group are sensitive to chloroform (30, 35, 36), suggesting that lipid-containing inner or outer membranes are involved in the infection process (35, 37). In our experiment, chloroform treatment of PkV RF01 drastically reduced the infectivity of the virus (Fig. 1C). As no outer membrane was detected by cryo-electron tomography, the sensitivity to chloroform might be linked to lipid components in either the capsid or the inner convoluted membranes. Internal lipid-containing membranes have been detected in several icosahedron-shaped double-stranded DNA viruses, including algal viruses belonging to the families *Phycodnaviridae* and *Mimiviridae*, mimiviruses, and various bacteriophages (38–43). In all of these viruses, the inner membranes are suggested to play a role in the release of the viral nucleoprotein core or genome by fusing with the host plasma membrane (40, 42, 43). Inner membranes in currently described NCLDVs more or less adopt the icosahedral morphology defined by the outer layer of capsomers (44, 45). We detected several convoluted inner membranes in PkV RF01 that do not follow the structure of the capsid. To our knowledge, this structural inconsistency has not been previously detected in any double-stranded DNA viruses, which calls for further investigation to understand the assembly process of PkV RF01 and how it enters its host. Another striking feature of the PkV RF01 virion is an internal rod-shaped core (ca. 55 nm in diameter), which is filled with dense material and positioned in the center of the virus particle. Similar features have been observed in transmission electron microscopy (TEM) images of large virus-like particles (VLPs) (300 to 700 nm) occurring in waste vacuoles of phaeodarian radiolarians collected from different oceans (46) and in zoospores of the green alga *Chlorococcus minutum* (47). To our knowledge, however, these features have not thus far been described in isolated viruses.

PkV RF01 has an atypical infection strategy. Only 2% of the total PkV RF01 viral particles produced during infection of *Haptolina ericina* UiO028 (He UiO028) were infectious (able to produce progeny) (Table 1). This infectivity was much lower than that of the other two prymnesioviruses, HeV RF02 and PkV RF02, which produced 13% and 44% infectious progeny, respectively (Table 1). The portion of infectious particles of PkV RF01 is also low compared to those of other algal viruses (48, 49). In addition, the latent period of PkV RF01 was previously reported to be longer (ca. 24 to 32 h [33]) in comparison with other prymnesioviruses (28, 29, 32, 33), and it has been demonstrated that PkV RF01 is also able to infect multiple species (33), which is another unusual trait among algal viruses (26).

The hosts of PkV RF01, PkV RF02, and HeV RF02 all belong to the order Prymnesiales, whose members are normally present in low abundance but cooccur year-round (*K* strategists) (50). PkV RF01, PkV RF02, and HeV RF02 are less virulent, as shown in the present study, and have longer latent periods than those of viruses infecting bloom-forming haptophytes (*r* strategists). Two of these viruses (PkV RF01 and HeV RF02) are also able to infect multiple species (generalists) (33). Longer replication time and reduced virulence, as hosts becomes scarce, increases the chances of vertical transmission rather than horizontal transmission of a virus. As vertical parent-to-offspring transmission depends on host reproduction, it has been argued that such transmission should select for reduced virulence because

the virus depends on host survival and reproduction for its transmission (51, 52). High virulence, on the other hand, may be supported by large, dense host populations, e.g., algal blooms, because high host densities ensure successful horizontal transmission of viral progeny to new hosts (51, 53). Viruses infecting the recurrent bloom-forming haptophytes *Phaeocystis pouchetii* virus (PpV) and *Phaeocystis globosa* virus (PgV) are indeed highly virulent, with between 60% and 100% of virus particles produced being infectious, resulting in rapid lysis of their hosts (48, 54). A broad host range might also increase the chance of transmission in an environment with low host abundances (*K* strategists). Such a strategy requires a trade-off whereby the virus decreases its opportunity of transmission by evolving longer replication times, higher decay rates, and reduced infectivity (discussed in references 55 and 56). This fits well with our two multispecies-infecting haptophyte viruses, PkV RF01 and HeV RF02, which have reduced proportions of infectious particles and longer replication times (33) relative to those of other haptophyte viruses with restricted host ranges (specialists), e.g., *Emiliania huxleyi* virus (EhV), PpV, and PgV.

The balance between fitness traits such as virulence, latent period and host range and trade-offs is the result of the adaptive evolution between viruses and their hosts and results in relationships ranging from acute to stable coexistence (persistence). In the ocean, persistent relationships—such as those between PkV RF01 and its hosts—seem to be most common among viruses infecting unicellular algae; this has been demonstrated by several metabarcoding studies revealing the persistence of dominance of viral operational taxonomic units (OTUs) over several months (57, 58). The atypical infection strategy of PkV RF01 evokes a persistent nature that is different from the vast majority of other algal viruses characterized thus far.

PkV RF01 has the largest genome among algal viruses. The genome of PkV RF01 was assembled as a linear DNA sequence of 1,421,182 bp. This size is more than twice that of the genome of TetV, which means that PkV RF01 has the largest reported genome of any virus infecting a photosynthetic organism (Fig. 2A). Evidence for the linear structure of this genome is the presence of ~5-kbp terminal inverted repeats. Despite being phylogenetically more closely related to alga-infecting *Mimiviridae* members, the genome size of PkV RF01 is in the range of those of heterotroph-infecting *Mimiviridae* members. The overall G+C content of PkV RF01 is 22.8%, which is low compared with that of other *Mimiviridae* (23% to 41%). Similarly to other *Mimiviridae*, the average G+C content of PkV RF01 in intergenic regions is relatively low, at 17.8%. This lower G+C content may reflect an ongoing loss of G and C nucleotides that is more prevalent in noncoding than coding regions because of weaker background selection in noncoding regions. The genome of PkV RF01 is predicted to contain 1,161 genes comprising 1,121 protein-coding DNA sequences (CDSs) and 40 tRNA genes corresponding to 13 amino acids. Most tRNA genes (30 out of 40) are clustered in three genomic regions that lack predicted CDSs, a feature also observed in other *Mimiviridae*. For example, all tRNAs of TetV ($n=10$) and CroV ($n=22$) are encoded consecutively on the same strand (18, 59). The average CDS length is 1,046 bp (minimum, 297; maximum, 1,493). Intergenic regions average 217 bp in length, with a cumulative sum of 244,005 bp, which corresponds to a gene density of 82.8%.

Of the 1,121 predicted CDSs, 641 (57%) exhibited sequence similarities (BLASTP *E* value conservative cutoff of 1×10^{-5}) to protein sequences in the UniRef90 database (Fig. 2A). Among these, 165 were most similar to *Mimiviridae*. Curiously, among the CDSs most similar to *Mimiviridae*, 60 were closest to ChoanoVirus, which was isolated from choanoflagellate cultures, followed by the subfamilies Mesomimivirinae ($n=49$) and “Klosneuvirinae” ($n=30$) (Fig. 2B). Among the 181 closest homologs found in eukaryotic organisms, 23 were haptophytes. A sequence-based homology search of corrected nanopore reads and scaffolds composing the initial assembly against *Lavidaviridae* proteomes (BLASTX; matrix: BLOSUM45; *E* value $< 1 \times 10^{-5}$) yielded no significant alignments against

FIG 2 Structure and gene taxonomic composition of the PkV RF01 genome sequence. (A) Rhizome and genomic features of the PkV RF01 genome. As illustrated by the rhizome (inner part of the figure), ORFans comprise the largest set of PkV RF01 genes, and a substantial portion (15%) have their best BLAST hits (in UniRef90) against “*Mimiviridae*.” Colors indicate taxonomic origin. Intergenic regions are white. Percentages of hits per taxonomic group greater than 5% of total genes are indicated. In the outermost ring, rectangles indicate the positions of glycosyltransferases (white), lipid-related enzymes (black), and succinate dehydrogenase genes (red), and the numbers correspond to *Mimiviridae* key enzymes (1 and 3, DNA-directed RNA polymerase II subunits 1 and 2, respectively; 2, DNA mismatch repair protein MutS7; 4, packaging ATPase; 5, VLTf3, 6, major capsid protein; 7: eukaryotic translation initiation factor 4E; 8, asparagine synthase; 9, DNA polymerase family B). The ring adjacent to the outermost ring shows G+C skew over a 10-KB window. (B) Taxonomic breakdown of 180 genes with best hits to virus genes. Mega, “Megavirinae”; AaV, *Aureococcus anophagefferens* virus; TetV, *Tetraselmis* virus 1; PoV, *Pyramimonas orientalis* virus.

any major or minor *Lavidaviridae* capsid proteins, which suggests that virophages were absent from the sample used for sequencing.

A previous analysis of PkV RF01 family B DNA polymerase (PolB) and the major capsid protein (MCP) placed this virus in the family *Mimiviridae* (33). We also reported recently that the PkV RF01 genome has additional NCLDV core genes, such as A32-like virion packing ATPase (nucleo-cytoplasmic virus orthologous group 0249 [NCVOG0249]) and RNApol (RNA pol subunit I [NCVOG0274] and subunit II [NCVOG0271]), and orthologous genes that are specific to *Mimiviridae*, namely, MutS7 (NCVOG2626) and asparagine synthase (AsnS; NCVOG0061) (60). Phylogenetic reconstruction using five NCLDV core genes confirmed the deep branching of PkV RF01 within the *Mimiviridae* family and suggested that PkV RF01, along with ChoanoV1, TetV, and AaV, is more closely related to Mesomimivirinae than to Megavirinae (Fig. 3A). In support of this evolutionary relationship, PkV RF01 has an additional copy of the second-largest RNA polymerase subunit gene (*rpb2*). This *rpb2* duplication is shared with all other *Mimiviridae* that infect algae, including Mesomimivirinae members, AaV (whose second copy is very short), and TetV, and was previously proposed as a useful feature to discriminate between the two main clades (autotroph- versus heterotroph-infecting viruses) within the *Mimiviridae* family (27). This additional *rpb2* copy is not found in other *Mimiviridae*, with the exception of ChoanoV1, whose genome was derived from a single cell metagenome in choanoflagellate cultures. Phylogenetic analysis indicates that these two *rpb2* copies were present in the ancestor of alga-infecting *Mimiviridae* and ChoanoV1 (Fig. 3B). In agreement with the phylogeny based on the five NCLDV core genes, this suggests that PkV RF01 and ChoanoV1, although evolutionarily distant, are more related to each other than to any other *Mimiviridae* member.

Out of 1,121 predicted protein-coding genes in the genome of PkV RF01, only about a third could be annotated with some functional description based on their sequence homology with characterized proteins. Such a small percentage is typical of divergent eukaryotic viruses detected for the first time. A total of 339 proteins

FIG 3 Phylogenetic evidence for PkV RF01 as a distant relative of "Mesomimivirinae." (A) Bayesian phylogenetic tree of NCLDVs reconstructed from a concatenated alignment of five core nucleocytoplasmic virus orthologous genes. Values at branches are posterior probabilities support. The tree was rooted using *Poxviridae* as an outgroup. The scale bar indicates substitutions per site. (B) Maximum-likelihood phylogenetic tree of cellular and NCLDV DNA-directed RNA polymerase subunit beta (RPB2). Values at branches are Shimodaira-Hasegawa-like local support.

(30%) showed significant sequence similarity with proteins in the Clusters of Orthologous Genes (COG) database (61) (Fig. 4). The distribution of COG functions associated with these hits was dominated by "Posttranslational modification, protein turnover, chaperones" (43 proteins) and "Cell wall/membrane/envelope biogenesis" (42 proteins), which comprise approximately two times more proteins than found in other *Mimiviridae* members, except for *Tupanvirus*. Among other well-represented categories, numbers of proteins in "Replication, recombination and repair" (36 proteins) and "Transcription" (23 proteins) were similar to those of other *Mimiviridae*, while the categories of "Translation, ribosomal structure and biogenesis" (25 proteins) and "Amino acid transport and metabolism" (20 proteins) were in the same range or higher than those of heterotroph-infecting *Mimiviridae* (mimiviruses, BsV, and CroV). Interestingly, 24, 17, and 9 PkV RF01 proteins were respectively assigned, respectively, to the categories of "Lipid transport and metabolism," "Carbohydrates transport and metabolism," and "Energy production and conservation," all much higher numbers compared with those found in other *Mimiviridae* viruses (Fig. 5).

Similarly to other *Mimiviridae* members, PkV RF01 encodes several genes involved in DNA repair, transcription, and translation. Notably, this virus has the full set of enzymes required for the base excision repair (BER) pathway, which is also the case for all *Mimiviridae* members except for those with smaller genomes (PgV, CeV, and AaV). PkV RF01 BER enzymes are closer (i.e., have a greater alignment score) to heterotroph-infecting *Mimiviridae* than to cellular homologs, thus suggesting that this pathway was present in the last common ancestor of *Mimiviridae*. According to a previous

FIG 4 Clusters of orthologous groups (COG) functional distribution of 339 proteins encoded by PkV RF01.

phylogenetic analysis, *Mimiviridae* BER enzymes are monophyletic with regard to *Mimiviridae* and have not recently been acquired from eukaryotes (62).

Unlike alga-infecting *Mimiviridae*, PkV RF01 encodes two amino-acyl tRNA synthetases (aaRS): an isoleucyl-tRNA synthetase (IleRS; open reading frame [ORF] 480) and an asparaginyl-tRNA synthetase (AsnRS; ORF 764). Both of these synthetases are found in most lineages of heterotroph-infecting *Mimiviridae* (AsnRS is missing from CroV and BsV, and IleRS is missing from *Mimivirus* lineage A). Phylogenetic analyses of these two proteins revealed a deep branching of viral homologs, which formed a monophyletic clade well separated from cellular homologs (Fig. 6).

Virus-encoded succinate dehydrogenase and energy production genes. We found six predicted protein-coding genes (ORFs 893 to 900) related to energy production in an 8,026-bp region (Fig. 7A). Four ORFs (ORFs 893 and 898 to 900) were predicted to code for all four subunits (SDHA, D, C, and B) of a functional succinate dehydrogenase (SDH, or electron transport chain complex II) of the oxidative phosphorylation pathway (Fig. 7B). In eukaryotes, all four subunits of this enzyme are encoded in the nuclear genome. This enzyme acts in the mitochondrial respiratory chain and participates in both the tricarboxylic acid (TCA) cycle and the respiratory electron transfer chain. In the TCA cycle, this succinate dehydrogenase oxidizes succinate to fumarate, while its activity in the inner mitochondrial membrane involves the reduction of a flavin adenine dinucleotide (FAD) cofactor, followed by electron transfer through three Fe-S centers to ubiquinone (Fig. 7C).

SDH genes have recently been reported in viral genomes assembled from environmental samples for which functional experiments cannot be done (63). In a reverse transcription-PCR (RT-PCR) experiment using primers specific for the PkV RF01 gene for SDHA (here *vSDHA*), we detected transcripts of this gene in samples collected 24, 72, and 96 h postinfection (Fig. 8). The *vSDHA* primers were tested on an uninfected culture to ensure that only the viral version of the *SDHA* gene was amplified (Fig. 9). The MCP gene of PkV RF01 was used both for protocol optimization and later as an internal

FIG 5 Comparative COG functional distribution among *Mimiviridae* members. COG sequences were automatically searched against the proteomes of each virus, using BLASTP with an E value of 1×10^{-5} as the significant similarity threshold.

FIG 6 Bayesian phylogenetic trees of two viral amino-acyl tRNA synthetases and their cellular homologs. (A) Isoleucine tRNA synthetases. (B) Aspartyl tRNA synthetases. Branches supported by posterior probability (PP) values of >70% are indicated by circles whose diameters are proportional to the PP value.

positive control (Fig. 10). Although the transcription of the viral *SDHA* suggests that the viral SDH is functional, we can only speculate on the possible role of this enzyme during infection. One possibility is that the viral SDH sustains the carbohydrate metabolism of infected cells (i.e., virocells) to supply building blocks of viral particles such as amino acids and to support proper replication of this large virus. Another possibility is that PkV RF01 uses its SDH as a part of an arms race with its host to turn on the TCA cycle after the host had turned it off to counter viral replication, or more simply to boost the energy metabolism of the virocells to augment the fitness of the host and/or to maximize virus production efficiency.

The discovery of the viral SDH prompted us to search for other potential virus-encoded SDHA and SDHB homologs in marine metagenomes. These two subunits (SDHA and SDHB) form the catalytic core containing the redox cofactors that participate in electron transfer to ubiquinone; they are thus more conserved than SDHC and SDHD subunits. To test for the presence of this viral SDH in other viruses, we searched for *vSDHA* and *vSDHB* in marine metagenomes of the Tara Oceans expedition. The 50 most similar and nonredundant SDHA and SDHB sequences predicted from 101 Tara Oceans genome fragments were most likely derived from *Mimiviridae* viruses (Fig. 11). Indeed, out of 1,113 genes predicted from these 101 genome fragments, 681 were annotated at some taxonomic level, of which 449 were predicted to be cellular and 157 viral. Of the 157 viral genes, 146 and 130 had their last common ancestor in *Mimiviridae* and *Mesomimivirinae*, respectively. A total of 32 of the 101 genome fragments contained at least one gene predicted to be of *Mimiviridae* origin, and the larger the genome fragment, the more *Mimiviridae* genes it was found to encode (Fig. 11A). Functional analysis indicated that 12 of the 1,113 predicted genes were NCLDV hallmark genes (encoding five VLTf3s, two capsid proteins, two PCNAs, two helicases, and one PolB). The high proportion of unknown genes and genes annotated as *Mimiviridae* in the 101 Tara Oceans genome fragments encoding SDHA or SDHB strongly suggests that these fragments belong to *Mimiviridae* viruses. This finding demonstrates that the presence of SDH is not restricted to PkV RF01 and is arguably widespread among marine *Mimiviridae*. According to phylogenetic analyses of cellular and viral SDHA and SDHB, the viral homologs form a monophyletic group that branches deeply within eukaryotic lineages (Fig. 11B and C). Long-branch attraction bias could generate such topologies, but, as explained above for the IleRS and AsnRS, it is more likely that the viral SDHA and SDHB were acquired at an early stage in the radiation of eukaryotic

FIG 7 Genes in PkV RF01 predicted to encode enzymes of oxidative phosphorylation and β -oxidation pathways. (A) Gene organization in the succinate dehydrogenase-containing region. (B) Schematic representation of the canonical enzymatic complex II in the mitochondrial membrane. (C) Location of succinate dehydrogenase in the TCA cycle and electron transport chain as known in plants, and a schematic reconstruction of the PkV RF01-encoded β -oxidation metabolic pathway.

lineages. The transcription of *vSDHA* and its occurrence in marine environments calls for further investigation to understand the biological role and coevolutionary significance of this viral SDH.

Other genes related to energy production were detected in the 8,026-bp-long region. ORF 894 and ORF 896, corresponding to cytochrome *c* (CytC) and cytochrome *b₆f* complex iron-sulfur (Cyt *b₆f*) subunits, respectively, showed high sequence conservation with *Chrysochromulina* sp. strain CCMP291 proteins (78% and 59% amino acid [aa] identities, respectively). CytC is a short protein (~100 aa) involved in the oxidative phosphorylation pathway, where it accommodates the transfer of electrons between the coenzyme Q-cytochrome *c* reductase (complex III) and coenzyme Q-cytochrome *c* oxidase (complex IV). The presence of Cyt *b₆f* between oxidative phosphorylation genes is puzzling because the cytochrome *b₆f* complex is involved in photosynthesis. The core of the chloroplast *b₆f* complex, however, is similar to the analogous respiratory cytochrome *bc₁* complex. The other two predicted ORFs in this region are similar to ubiquinone biosynthesis protein UbiB (ORF 895) or contain an NAD-binding domain and an Fe-S cluster (ORF 897) and may thus be associated with electron transport as well. ORF 897

FIG 8 The viral SDHA gene is transcribed during infection. Gels of PCR and reverse transcription-PCR (RT-PCR) in combination with a Turbo DNA-free kit. Samples were taken 24, 72, and 96 h after infection. (A) PCR with *vSDHA*-specific primers was used to check for the presence of genomic DNA after RNA isolation treated with 1× and 2× DNase (upper and lower panels, respectively). P, positive control (PKV RF01 genomic DNA); N, negative control (sterile distilled water [sdH_2O]). (B) RT-PCR of RNA samples using *vSDHA*-specific primers. M, DNA marker (MassRuler DNA ladder mix, 80 to 10,000 bp; Thermo Fisher).

has two distant (25% to 31% aa identity) homologs in the PkV RF01 genome (ORF 456 and ORF 625).

Some other genes were predicted to encode enzymes involved in pyruvate metabolism. ORF 79 has sequence homology with L-lactate dehydrogenases; it might thus catalyze the conversion of lactate to pyruvate, an intermediary compound serving as a starting point for several major metabolic pathways, such as glycolysis, gluconeogenesis, and the TCA cycle. ORF 727 was predicted to code for an isochorismate hydrolase that also produces pyruvate from isochorismate. ORF 24 and ORF 726 share sequence homology with phosphoenolpyruvate synthase and a partial pyruvate kinase, respectively. The former catalyzes the conversion of pyruvate to phosphoenolpyruvate (PEP), while the latter catalyzes the reverse reaction. Formation of PEP is an initial step in gluconeogenesis.

A nearly complete virus-encoded β -oxidation pathway. In this study, 22 predicted genes were inferred to code for proteins involved in lipid synthesis or degradation, including key enzymes of the β -oxidation pathway (Table 2). Several genes were predicted to code for lipase-like proteins (ORFs 386, 481, 635, 653, and 690), including a triacylglycerol lipase (ORF 386) that can break down triacylglycerol into glycerol and fatty acids. Glycerol and fatty acids can be used as a starting point for ATP production by glycolysis and β -oxidation, respectively. In the β -oxidation pathway, fatty acids are fully oxidized to produce acetyl-coenzyme A (CoA), which can then enter the TCA cycle to yield NADH and FADH_2 ; these latter two products can funnel through to the electron transport chain to produce ATP (Fig. 7C). Each β -oxidation cycle itself also produces NADH and FADH_2 cofactors. We found that PkV RF01 encodes key β -oxidation

FIG 9 PCR optimization and confirmation of the SDHA gene in the PkV RF01 genome. (A and B) Results of PCR with SDHA primers using genomic PkV RF01 DNA (A) and genomic He UiO028 DNA (B) as the templates. Lanes 1 and 9, DNA ladder; lanes 2 to 7, optimization of the PCR annealing temperature from 55°C (lane 2) to 60°C (lane 7); lane 8, negative control (sdH₂O).

enzymes. First, two distantly related ORFs (ORF 142 and ORF 904, sharing 22% aa identity) have sequence homology with a long-chain fatty acyl-CoA synthetase. This enzyme catalyzes the formation of fatty acyl-CoA in the cytosol. Fatty acyl-CoA can be imported to mitochondria using a (carnitine) CoA-transferase that is also encoded in PkV RF01 (ORF 33). Once in the mitochondrial matrix, fatty acyl-CoA serves as a substrate on which an acyl-CoA dehydrogenase (ORF 1046) oxidizes the fatty acyl-CoA and reduces a FAD cofactor to produce a FADH₂ cofactor. We identified a 2,4-dienoyl-CoA reductase (ORF 30) that may facilitate the next oxidation step to produce a NADH cofactor. FADH₂ and NADH molecules produced by a β -oxidation cycle can both be

FIG 10 PCR and RT-PCR optimization using an internal control gene (*mcp*). PCR and RT-PCR were carried out after removal of genomic DNA using a Turbo DNA-free kit. Samples were taken 24, 72, and 96 h after infection. Two different protocols, both provided in the Turbo DNA-free kit manual, were used to optimize the reactions. (A) PCR check for the presence of genomic DNA after RNA isolation treated with 1× and 2× DNase (upper and lower panels, respectively). P, positive control (PkV RF01 genomic DNA); N, negative control (sdH₂O). (B) Result of RT-PCR of samples harvested 24, 72, and 96 h postinfection. M, DNA marker (MassRuler DNA ladder mix, 80 to 10,000 bp; Thermo Fisher).

FIG 11 Origin of PKV RF01 SDHA and SDHB and their most similar homologs in Tara Oceans metagenomes. (A) Taxonomy of genes predicted in Tara Oceans metagenome assembled-genome fragments encoding the 50 SDHAs and SDHBs most similar to PKV RF01 genes (for genome fragments having at least five predicted genes). (B and C) Phylogenetic trees of viral and cellular SDHAs (B) and SDHBs (C). Clades in green contain PKV RF01 SDHA or SDHB and their 50 most similar hits identified in Tara Oceans metagenomes (predicted to be *Mimiviridae* homologs from panel A). Red, eukaryotic phyla; black, unclassified eukaryotes. Trees are rooted with *Proteobacteria* and *Firmicutes* homologs (not shown). Circles indicate branches with posterior probability support of $\geq 50\%$.

TABLE 2 Genes related to lipid metabolism

ORF	Annotation ^a	KEGG orthology identifier	Pathway
30	2,4-Dienoyl-CoA reductase, mitochondrial (EC 1.3.1.34)	K13236	Beta-oxidation
33	Putative CoA-transferase	NS	Beta-oxidation
121	Glycerophosphoryl diester phosphodiesterase	K01126	Glycerophospholipids metabolisms
138	Fatty acid synthase (FASN)	K00665	Fatty acid biosynthesis
142	Long-chain fatty acid-CoA ligase ACSBG (EC 6.2.1.3)	K15013	Fatty acid degradation/biosynthesis/beta-oxidation
175	Acetyl-CoA carboxylase/biotin carboxylase 1 (EC 6.4.1.2, 6.3.4.14, and 2.1.3.15)	K11262	Fatty acid biosynthesis
236	Glutaryl-CoA dehydrogenase (EC 1.3.8.6)	K00252	Fatty acid degradation
293	Lysophospholipase like	NS	NS
357	Lysophospholipase like	NS	NS
386	Triacylglycerol lipase (EC 3.1.1.3)	K01046	Glycerolipid metabolism
481	Lipase-like	NS	NS
635	Lipase-like	NS	NS
653	Lipase-like	NS	NS
690	Lipase-like	NS	NS
774	Lysophospholipid acyltransferases (EC 2.3.1.22)	K14457	Glycerolipid metabolism
694	Lipase esterase (carbohydrate esterase CE10)	NS	NS
695	Lipase esterase (carbohydrate esterase CE10)	NS	NS
886	Stearoyl-CoA desaturase (delta-9 desaturase) (EC 1.14.19.1)	K00507	Biosynthesis of unsaturated fatty acids
902	Fatty acid synthase (FASN)	K00665	Fatty acid biosynthesis
904	Long-chain fatty acid-CoA ligase ACSBG (EC 6.2.1.3)	K15013	Fatty acid degradation/biosynthesis/beta-oxidation
1016	Cyclopropane-fatty-acyl-phospholipid synthase (EC 2.1.1.79)	K00574	NS
1046	Acyl-CoA dehydrogenase	K06445	Fatty acid degradation/beta-oxidation

^aCoA, coenzyme A; NS, non-significant.

oxidized in the electron transport chain to generate ATP. The enzymes involved in the two intermediate steps following each oxidation, either an enoyl-CoA hydratase or a β -ketothiolase, were not detected in our analysis.

Most of these genes have no homologs in reference viral genomes, and, to our knowledge, this is the first report of a virus possessing proteins that are directly involved in lipid-based energy production. By diverting host lipid machinery, interactions of viruses with lipids or lipid based-structures have long been known to have structural or signaling roles at different stages of the virus life cycle, such as entry, genome replication, morphogenesis, and exit (64–66). More recently, several studies on human viruses (two herpesviruses and one RNA virus) have shown that the metabolic state of an infected cell can be shifted toward energy generation to support viral replication (65). These studies have highlighted the increasing abundance—until up to 48 h after HCV infection—of enzymes involved in β -oxidation, amino acid catabolism, and the TCA cycle (67) and an increase in cellular β -oxidation following the release of free fatty acids caused by Dengue virus-induced autophagy (68). Among algal viruses, EhV remodels the transcription of host lipid genes for fatty acid synthesis to support viral assembly (69) and also to generate triacylglycerols stored in the virion and available as an energy pool in later infection phases (70). Besides diverting the host metabolism, EhV encodes seven proteins involved in the sphingolipid biosynthesis pathway (71). This pathway produces a viral sphingolipid that is a central component of EhV lipid membranes and that can also act as a signaling lipid and induce programmed cell death during the lytic infection phase (72). EhV also encodes a triglyceride lipase (with detectable homology to predicted PkV RF01 lipases ORF 635 and ORF653) that is highly expressed during late infection concomitantly with significant upregulation of host β -oxidation genes (69). These examples and our observations of several genes involved in β -oxidation clearly show that viruses can introduce new metabolism-related genes, sometimes representing entire pathways,

FIG 12 Comparative distribution of glycosyltransferase domains among viruses.

into the host, most likely to satisfy the high metabolic requirement of these giant viruses.

High representation of glycosyltransferases. Compared with other viruses, PKV RF01 was found to encode an unusually high number of glycosyltransferases (GTs) as well as other carbohydrate-active enzymes. Automated annotation of GTs (and other carbohydrate-active enzymes) in reference viral proteomes using dbCAN2 (73) revealed that the largest number of GT domains was encoded by PKV RF01 ($n = 48$), followed by CeV ($n = 13$), *Mimivirus* members, and CroV and AaV ($n = 8$ to 10) (Fig. 12). We uncovered 48 GT domains encoded in 40 ORFs, 8 of which were predicted to encode more than one GT domain. These domains correspond to 16 different GT families. Most domains were inferred to be functional, as 31 out of 48 covered at least 70% of the dbCAN2 reference domain, with coverage ranging

from 44% to 99%. GTs were found scattered across the genome of PkV RF01 but with some local clustering (Fig. 2A), the latter indicating possible involvement in the same pathway. GT32 was the most represented domain, with 11 proteins (as annotated by dbCAN2) and potentially three additional proteins (ORFs 40, 84, and 861). Eight proteins possessed a GT25 domain that can catalyze the transfer of various sugars onto a growing lipopolysaccharide chain during its biosynthesis. Among these eight predicted ORFs, four contained an additional nonoverlapping GT domain (two GT2s, one GT6, and one GT60). Functional analyses of GTs in mimiviruses (or in related *Paramecium bursaria* Chlorella viruses) have demonstrated that some of these enzymes are functional, being able to modify viral collagen-like proteins (74) and polymerize sugars (75). Conservation between PkV RF01 GTs and functionally characterized GTs in viruses and cells is absent or extremely low, which precludes any predictions as to the specific roles of these enzymes in the PkV RF01 life cycle. Nevertheless, this putative glycosylation-conducive autonomy possibly allows the virus to infect a variety of hosts, as the virus can modify its own glycans, which are used for host recognition, independently of the host system (76). In alphaviruses, flaviviruses, and herpesviruses, fusion is mediated by viral glycoproteins (40).

Other carbohydrate-active enzymes in the PkV RF01 genome include seven glycoside hydrolases (GHs), four carbohydrate esterases (CEs), one polysaccharide lyase (PL), one carbohydrate-binding module (CBM), and a putative sugar fermentation stimulation protein A (ORF 1003) that is possibly involved in maltose metabolism. These numbers are not excessively high compared with those in other viruses. Other detected ORFs were homologous to enzymes involved in carbohydrate transport and metabolism, notably a transketolase (ORF 528) involved in the pentose phosphate pathway in all organisms and in the Calvin cycle of photosynthetic organisms. Finally, we detected a 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase 2 (ORF 539) and a mannose-1-phosphate guanylyltransferase/mannose-6-phosphate isomerase (ORF 836) involved, respectively, in fructose and mannose metabolism.

Conclusions. The haptophyte virus PkV RF01 has been previously shown to have a longer replication cycle and a broader host range than those of other prymnesioviruses and most other algal viruses. Here, we revealed that PkV RF01 has atypical virion morphology and that infections yield several orders of magnitude fewer infectious particles than those of other tested prymnesioviruses. In-depth phylogenetic analysis using genes conserved in NCLDV confirmed that PkV RF01 belongs to *Mimiviridae* but is deeply diverged from existing members, although closer to alga-infecting *Mimiviridae* members than to heterotroph-infecting ones. Unlike other alga-infecting *Mimiviridae*, however, PkV RF01 has a large genome (1.4 Mb) and contains genes coding for two aminoacyl-tRNA synthetases and the complete BER pathway. All of these features are conserved in most heterotroph-infecting *Mimiviridae* and therefore must have been lost in other alga-infecting *Mimiviridae*. This outlier virus features an unprecedentedly high number of genes involved in energy metabolism and glycosylation machinery, which may enable its longer replication cycle and broader host range compared to those of other algal viruses. These genomic and phenotypic features are suggestive of a persistent infection behavior that probably evolved in response to the host growth strategy. Because of nutrient limitations, these persistent systems of slow-growing but ubiquitous hosts with less virulent viruses may represent the most common type of virocells in oceans.

MATERIALS AND METHODS

Culturing and infection. All algal host cultures were grown in liquid IMR/2 medium consisting of 70% aged seawater, 30% distilled water (25 practical salinity units [PSU]), and additional selenite (10 nM final concentration). The cultures were kept at 14°C and partially synchronized using a 14:10 h light:dark cycle with irradiance of $100 \mu\text{mol photons} \cdot \text{m}^{-2} \cdot \text{s}^{-2}$ supplied by white fluorescent tubes. Viruses were produced by adding freshly produced viral lysate (ca. 2×10^8 virus-like particles [VLP]/ml), propagated three times on the host before being added to exponentially growing host cultures (ca. 5×10^5 cells/ml) in a ratio of 1:10 (vol/vol). Infection was followed by flow cytometry (FCM) (77, 78) for 72 h by counting viral particles and host cells, as described previously (33). Burst size was calculated as the number of viral particles released from each host cell, estimated from the total number of host cells preinfection and the total number of VLPs produced during the infection cycle (33).

Infectious progeny. The percentage of viral infectious progeny was determined by comparing the most probable number (MPN) (endpoint dilution [78]) and flow cytometric total counts of viral particles produced during infection. The number of infectious particles released in a burst was determined based on the percentage of viral infectivity produced during the infection cycle and the burst size. Infectivity was tested using *Haptolina ericina* UiO028 as a host and was also compared with that of two other prymnesioviruses, HeV RF02 and PKV RF02 (33), propagated on He UiO028 and *Prymnesium kappa* RCC3423, respectively.

Briefly, 10 $\times$ dilutions were prepared from fresh viral lysate and added to exponentially growing host cells in 96-well microtiter plates (eight replicates for each dilution). The plates were incubated for 7 days under normal incubation conditions. Cell lysis was measured by monitoring *in situ* fluorescence on a plate reader (EnSpire 2300 multilabel reader; PerkinElmer) at 460/680 nm. Numbers of infectious particles were estimated from the proportion of lysed wells using the MPN_ver4.xls excel spreadsheet from Jarvis et al. (79).

Sensitivity to chloroform. The effect of chloroform on infectivity, used to infer the presence of a lipid membrane or lipid molecules in the capsid, was tested by adding 50% (vol/vol) chloroform to PKV RF01 lysate. After mixing, the chloroform phase was separated from the solution by centrifugation at 4,000 $\times g$ for 5 min. The tubes were incubated at 37°C for 2 h with the lids open to allow evaporation of any remaining chloroform.

Triplicates of exponentially growing He UiO028 cells (1.6×10^5 cells/ml) were incubated with 1:10 volumes of chloroform-treated viruses (ca. 2×10^8 VLP/ml). The incubation was followed for 7 days by counting host cells by FCM (78). Host cells in chloroform-treated or untreated medium at the same ratio used with the viral lysate were used as controls. Virus propagation in lysed cultures was confirmed by FCM.

Cryo-electron tomography. A small drop of concentrated PKV RF01 (8×10^9 viral particles/ml) was deposited on a glow discharged, 200-mesh copper grid with holey carbon film (R2/1 Cu 200; Quantifoil Micro Tools GmbH, Germany). The sample was blotted with filter paper and immediately plunge frozen in liquid ethane. Grids were transferred under liquid nitrogen to a cryo-transfer tomography holder (Fischione Instruments, USA) and inserted in a 200-kV transmission electron microscope (Talos F200C; Thermo Scientific) equipped with a Ceta 16M camera. Tilt series were recorded at $\times 45,000$ magnification and $-7 \mu\text{m}$ defocus between -60° and 60° in 2° increments. Finally, reconstruction, segmentation, and visualization of the tomograms was performed with IMOD v4.9 software (80).

Purification of viral particles and DNA isolation. Exponentially growing He UiO028 cultures (2 liters) were infected with 20 ml of PKV RF01 and inspected visually for lysis. An uninfected culture (100 ml) was used as a control. Lysed algal cultures were checked for viruses by FCM counting. Lysed cultures were first centrifuged to remove algal debris and some bacteria (5,500 rpm for 15 min). Viruses were then pelleted by ultracentrifugation at 25,000 rpm in an Optima L90K ultracentrifuge (Beckman Coulter) for 2 h. The pellets were resuspended in SM buffer (0.1 M NaCl, 8 mM $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$, 50 mM Tris-HCl, and 0.005% glycerol). Viral particles were further purified by OptiPrep gradient centrifugation (81). Fractions were checked for viruses by FCM and for infectivity by infection of He UiO028.

Isolation of high-quality DNA for sequencing was done following the protocol of Sandaa et al. (82) with some modifications. Viral particles were disrupted by one round of heating to 90°C for 2 min and then chilling on ice for 2 min. Disodium ethylenediaminetetraacetic acid and proteinase K at a final concentration of 20 mM and 100 $\mu\text{g} \cdot \text{ml}^{-1}$, respectively, were then added before incubation of the samples for 10 min at 55°C. Sodium dodecyl sulfate at a final concentration of 0.5% (wt/vol) was subsequently added, and samples were incubated for an additional 1 h at 55°C. Double-stranded DNA was then purified from the lysates using a Genomic DNA Clean & Concentrator-10 kit (Zymo Research, Irvine, CA) according to the manufacturer's protocols. To avoid shearing DNA, gentle pipetting and mixing (accomplished by turning the tubes instead of vortexing) were performed in all steps.

Genome assembly. Isolated DNA from PKV RF01 was subjected to Illumina TruSeq PCR-free library preparation (insert size, 350 bp). The generated library was sequenced on an Illumina MiSeq instrument in paired-end mode (2×300 bp) to yield approximately 1.9 million reads, which corresponds to about 400 $\times$ coverage. Reads were assembled into 2,498 contigs of 500 bp or more with a total assembly size of 4.75 Mb using Newbler (83). In addition, a ligation-based 1D² nanopore library (LSK-308) was constructed and sequenced using an Oxford Nanopore MinION Mk1b device and a Flo-Min107 flow cell, which resulted in 825 long reads with an N_{50} value of 13.6 kb and a total length of 9.89 Mb. To improve the assembly, short-read contigs were manually bridged with the long reads. Manual assembly using Consed (84) yielded a linear genome sequence of 1.4 Mb with inverted terminal repeats. After assembly, the consensus was polished using Nanopolish (85) and Pilon (86).

Phylogenetic analyses. (i) Five core genes, SDHA, and SDHB. The phylogenetic position of PKV RF01 was inferred from concatenated protein alignments of the following five core nucleocytoplasmic virus orthologous genes (NCVOGs) (87): D5-like helicase-primase (NCVOG0023), DNA polymerase elongation subunit family B (NCVOG0038), DNA or RNA helicases of superfamily II (NCVOG0076), packaging ATPase (NCVOG0249), and poxvirus late transcription factor VLTf3-like (NCVOG0262). Sequences were obtained from the NCVOG database (ftp.ncbi.nlm.nih.gov/pub/wolf/COGs/NCVOG/) (88). Additional sequences were obtained from genomes retrieved from GenBank and annotated with HMMER v3.12b using the *hmmsearch* (89) command with hidden Markov models (HMMs) available in Schulz et al. (13). Sequences from each NCVOG were aligned independently using MAFFT L-INS-i (90). The alignments were trimmed with trimAl v1.2 in *gapout* mode (91) prior to concatenation using a custom Python script. Bayesian phylogenetic trees were inferred with PhyloBayes 1.7 (92) using the CAT model and a general time-reversible (GTR) substitution matrix. Four chains were run for 34,500 to 35,500 generations.

TABLE 3 Forward and reverse PCR primers for amplification of *vSDHA* and MCP

Primer name	Sequence (5'–3')	PCR product (bp)
vSDHA-F1	ATGTGCCGAGAAGCTCCTAA	154
vSDHA-R1	CTGCACAGGCTGTTTCGATAA	
PkV-RF01-MCP-F	GATGAACCTTGCCCACT	256
PkV-RF01-MCP-F	GTGCATGGTACGTTTTCGTG	

The *bpcomp* command was used to check for convergence and stop when *maxdiff* was equal to 0.3. One chain was discarded, and a consensus tree was constructed using the remaining three chains.

For phylogenetic analyses of succinate dehydrogenase subunits, top hits of PkV RF01 SDHA and SDHB were retrieved from UniProt (<https://www.uniprot.org/>) using online PHMMR searches (<https://www.ebi.ac.uk/Tools/hmmer/search/phmmer>) and also from the Tara Oceans project using online BLASTP searches (<http://tara-oceans.mio.osupytheas.fr/ocean-gene-atlas/>) (93). Alignments generated with MAFFT L-INS-i were filtered with trimAl in *gappyout* mode. Maximum-likelihood phylogenies were inferred with RAxML 8.2.9 (94) using the PROTCATLG model and automatic bootstrapping with the following options: “-N autoMRE -f a -n autoresult.” Phylogenetic trees of PkV RF01, SDHA, and SDHB were visualized using interactive Tree Of Life (iTOL) (95).

(ii) Rpb2, IleRS, and AsnRS. To reconstruct a phylogenetic tree based on the second-largest RNA polymerase subunit, homologs were recruited by comparing *Mimivirus* Rpb2 against all proteins of viruses and selected organisms in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database using the GenomeNet BLASTP tool (<https://www.genome.jp/>). Organisms were manually selected from the KEGG list to ensure broad taxonomic coverage of the tree of life. The retrieved amino acid sequences were aligned using MAFFT-LINSI (90) and then trimmed using trimAl (91) with the following parameters: “-resoverlap 0.5 -seqoverlap 70 -gt 0.8 -st 0.001 -cons 50.” The tree was reconstructed using FastTree (96) as implemented in the GenomeNet TREE tool (<https://www.genome.jp/tools-bin/ete>). Isoleucine tRNA synthase and aspartyl tRNA synthetase viral and cellular homologs were retrieved and aligned in the same way. Trees were searched using PhyloBayes MPI (97) with the nonhomogeneous CAT+GTR model (98). For each protein, three chains were run until the *maxdiff* parameter reached <0.3 (0.27 for AsnRS and 0.16 for IleRS). One chain was discarded for IleRS, and a consensus tree was constructed using the remaining chains.

Gene prediction and functional and taxonomic annotation. GeneMarkS with the option “virus” (99) predicted 1,121 open reading frames (ORFs) in the fully assembled genome sequence of PkV RF01, while tRNAscan-SE (100) predicted 41 tRNAs. PkV RF01 CDS amino acid sequences were searched against Virus-Host DB (101), RefSeq (102), UniRef90 (103), and COG (61) databases using BLASTP with an *E* value of 1×10^{-5} as the significant similarity threshold and against the Conserved Domain Database (104) using RPS-BLAST with an *E* value threshold of 1×10^{-2} . The 10 best hits for each database were compiled in a single file and manually inspected to transfer annotations of subject sequences to our query. In ambiguous cases, such as distant homologs (often seen in viral genomes) or unclear or contradictory annotations of subject sequences, the query was searched against KEGG genes (105) to allow extensive manual checking using GenomeNet tools (<https://www.genome.jp/>; alignment quality, length comparison to canonical genes, and links with KEGG orthology). We automatically annotated glycosyltransferases (GTs) and other carbohydrate-active enzymes (glycoside hydrolases, GHs; polysaccharide lyases, PLs; carbohydrate esterases, CEs; and auxiliary activities, AAs) in PkV RF01 and all viral genomes in Virus-Host DB (as of June 2018) using the *hmm* option of the dbCAN2 pipeline and its profile database (73). We retained hits with *E* values of $<1 \times 10^{-5}$ and domain coverage of $>35\%$, which corresponded to default settings.

Taxonomic and functional analysis of vSDHA homologs in OM-RGCv1. We searched PkV RF01 SDHA and SDHB against OM-RGCv1 (106) using the Ocean Gene Atlas (93) BLAST-based tool and kept the top 50 hits with significant *E* values for further analysis. We then collected genome fragments (contigs) encoding these 50 SDHAs and 50 SDHBs by searching via BLASTN for identical hits over full *SDHA* or *SDHB* lengths against Tara Ocean assemblies (downloaded from EBI) used to construct OM-RGCv1. We predicted ORFs in these genome fragments using GeneMarkS. The resulting 1,113 amino acid sequences were functionally annotated by searching against Pfam protein families (107) using *hmmsearch* (108) and also taxonomically annotated using a last-common-ancestor strategy as in Carradec et al. (109); in brief, protein sequences were searched against a database composed of UniRef cell, MMETSP (110), and Virus-Host DB (101) data using DIAMOND (111). Selected hits were then used to derive the last common ancestor of the query using a NCBI taxonomic tree wired to reflect the taxonomy of NCLDV.

PCR and RT-PCR optimization. We designed specific primers (Table 3) targeting a 256-bp region of the *mcp* gene to use both as an internal control in the RT-PCR and to confirm that our protocols were optimized. For each PCR, a negative control (sterile distilled H₂O [sdH₂O]) was included. PCR amplifications were carried out in 50- μ l total volumes containing 1 μ l of template using a DNA HotStarTaq mastermix kit (Qiagen). The cycling protocol was as follows: 15 min at 95°C, followed by 35 cycles of 30 s at 94°C, 30 s at 59°C, and 30 s at 72°C, with a final extension of 12 min at 72°C.

RT-PCRs were performed using the SuperScript III one-step RT-PCR with Platinum *Taq* DNA polymerase system (Thermo Fisher). Cycling conditions were as follows: 16 min at 55°C and 2 min at 94°C, followed by 40 cycles of 15 s at 94°C, 30 s at 49°C, and 30 s at 68°C, and a final extension of 5 min at 68°C.

All PCR products were checked for the correct size on a 1.5% agarose gel stained with GelRed (Biotium). PCR products were further checked by sequencing using BigDye v3.1 (Thermo Fisher) for cycle sequencing (Seksverseringslaboriet, University of Bergen, Norway).

PCR amplification and RT-PCR analysis of *vSDHA*. To investigate whether the *vSDHA* gene is transcribed during infection, an infected culture of He_UiO028 plus PkV RF01 as well as an uninfected He_UiO028 culture (control) were set up as described above. Samples were collected from both cultures at 24, 72, and 96 h postinfection. RNA was extracted using an RNeasy Plus Universal minikit (Qiagen), with genomic DNA (gDNA) removed in an extra step using a Turbo DNA-free kit (Ambion).

Specific primers were designed to target a 150-bp region of the *vSDHA* gene (Table 3). For each PCR, two negative controls (sterile distilled H₂O and extracted DNA from He028) were included. As positive controls for the transcription, we used primers targeting the *mcp* gene (see above). As a positive PCR control, we used genomic PkV RF01 DNA. PCR amplifications were conducted in 50- μ l total volumes containing 1 μ l of template DNA using an *Ex Taq* kit (Takara). The cycling protocol was as follows: 5 min at 94°C, followed by 35 cycles of 30 s at 94°C, 30 s at 59°C, and 30 s at 72°C, with a final extension of 12 min at 72°C.

RT-PCRs were performed using a SuperScript III one-step RT-PCR with Platinum *Taq* DNA polymerase system (Thermo Fisher). Cycling conditions were as follows: 16 min at 55°C and 2 min at 94°C, followed by 40 cycles of 15 s at 94°C, 30 s at 49°C, and 30 s at 68°C, with a final extension of 5 min at 68°C. PCR products were checked as described above.

Data availability. Raw sequence reads and the PkV RF01 genome sequence were deposited at the European Bioinformatics Institute (EMBL-EBI) (<https://www.ebi.ac.uk>) under project identifier [PRJEB37450](https://www.ebi.ac.uk/PRJEB37450). The complete video records of a cryo-electron tomogram of a PkV RF01 virion and sequence data reported in this study, as well as the curated gene annotation table, are available at <https://github.com/RomainBlancMathieu/PkV-RF01>.

ACKNOWLEDGMENTS

The recording of tilt series was performed with the help of Sebastian Schultz at the Unit of Cellular Electron Microscopy, Norwegian Radium Hospital. Initial sequencing (MiSeq and PacBio) of PkV RF01 total DNA was performed at the Norwegian Sequencing Center (<https://www.sequencing.uio.no/>). We thank Hilde M. K. Stabell and Solveig Siqueland, Department of Biological Sciences, University of Bergen, Norway, for technical assistance with molecular biology experiments, as well as Christian Rückert, Bielefeld University, for support in manual finishing of genome assembly, and Minyue Fan, Kyoto University, for assistance in gene analysis. The Super Computer System, Institute for Chemical Research, Kyoto University, provided computational time. We thank Barbara Goodson from Edanz Group for editing the English text of a draft of the manuscript.

This work was supported by the Research Council of Norway project entitled “Uncovering the key players for regulation of phytoplankton function and structure: lesson to be learned from algal virus-haptophyte coexistence” (VirVar; project number 294364 to R.-A.S.). Additional funding was provided by the European Union Horizon 2020 research and innovation program, grant agreement no. 685778 (“Virus-X”) to R.-A.S. and D.B. This work was also supported by the Future Development Funding Program of the Kyoto University Research Coordination Alliance. H.O. was supported by JSPS/KAKENHI (grant 18H02279), and by Scientific Research on Innovative Areas from the Ministry of Education, Culture, Science, Sports and Technology (MEXT) of Japan (grants 16H06429, 16K21723, and 16H06437). We declare that we have no competing interests.

REFERENCES

- Dupré JO. 2009. Varieties of living things: life at the intersection of lineage and metabolism. *Philos Theory Biol* 1. <https://doi.org/10.3998/ptb.6959004.0001.003>.
- Forterre P. 2012. Theiroid concept. John Wiley & Sons, Ltd, Chichester, UK.
- Raoult D, Forterre P. 2008. Redefining viruses: lessons from *Mimivirus*. *Nat Rev Microbiol* 6:315–319. <https://doi.org/10.1038/nrmicro1858>.
- Claverie J-M. 2006. Viruses take center stage in cellular evolution. *Genome Biol* 7:110. <https://doi.org/10.1186/gb-2006-7-6-110>.
- Mann NH, Cook A, Millard A, Bailey S, Clokie M. 2003. Bacterial photosynthesis genes in a virus. *Nature* 424:741–741. <https://doi.org/10.1038/424741a>.
- Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. 2004. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci U S A* 101:11013–11018. <https://doi.org/10.1073/pnas.0401526101>.
- Hurwitz BL, Hallam SJ, Sullivan MB. 2013. Metabolic reprogramming by viruses in the sunlit and dark ocean. *Genome Biol* 14:R123. <https://doi.org/10.1186/gb-2013-14-11-r123>.
- Thompson LR, Zeng Q, Kelly L, Huang KH, Singer AU, Stubbe J, Chisholm SW. 2011. Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci U S A* 108: E757–764. <https://doi.org/10.1073/pnas.1102164108>.
- Anantharaman K, Duhaime MB, Breier JA, Wendt KA, Toner BM, Dick GJ. 2014. Sulfur oxidation genes in diverse deep-sea viruses. *Science* 344:757–760. <https://doi.org/10.1126/science.1252229>.
- Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. 2005. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* 438:86–89. <https://doi.org/10.1038/nature04111>.
- Fridman S, Flores-Urbe J, Larom S, Alalouf O, Liran O, Yacoby I, Salama F, Bailleul B, Rappaport F, Ziv T, Sharon I, Cornejo-Castillo FM, Philofof A, Dupont CL, Sánchez P, Acinas SG, Rohwer FL, Lindell D, Béjà O. 2017. A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. *Nat Microbiol* 2:1350–1357. <https://doi.org/10.1038/s41564-017-0002-9>.

12. Raoult D, Audic S, Robert C, Abergel C, Renesto P, Ogata H, La Scola B, Suzan M, Claverie J-M. 2004. The 1.2-megabase genome sequence of *Mimivirus*. *Science* 306:1344–1350. <https://doi.org/10.1126/science.1101485>.
13. Schulz F, Yutin N, Ivanova NN, Ortega DR, Lee TK, Vierheilig J, Daims H, Horn M, Wagner M, Jensen GJ, Kyrpides NC, Koonin EV, Woyke T. 2017. Giant viruses with an expanded complement of translation system components. *Science* 356:82–85. <https://doi.org/10.1126/science.aal4657>.
14. Abrahão J, Silva L, Silva LS, Khalil JYB, Rodrigues R, Arantes T, Assis F, Boratto P, Andrade M, Kroon EG, Ribeiro B, Bergier I, Seligmann H, Ghigo E, Colson P, Levasseur A, Kroemer G, Raoult D, Scola BL. 2018. Tailed giant *Tupanvirus* possesses the most complete translational apparatus of the known virosphere. *Nat Commun* 9:–12. <https://doi.org/10.1038/s41467-018-03168-1>.
15. Miller ES, Heidelberg JF, Eisen JA, Nelson WC, Durkin AS, Ciecko A, Feldblyum TV, White O, Paulsen IT, Nierman WC, Lee J, Szczypinski B, Fraser CM. 2003. Complete genome sequence of the broad-host-range vibriophage KVP40: comparative genomics of a T4-related bacteriophage. *J Bacteriol* 185:5220–5233. <https://doi.org/10.1128/jb.185.17.5220-5233.2003>.
16. Yoshikawa G, Askora A, Blanc-Mathieu R, Kawasaki T, Li Y, Nakano M, Ogata H, Yamada T. 2018. *Xanthomonas citri* jumbo phage XacN1 exhibits a wide host range and high complement of tRNA genes. *Sci Rep* 8:4486. <https://doi.org/10.1038/s41598-018-22239-3>.
17. Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan MB, Gillet R, Forterre P, Krupovic M. 2019. Numerous cultivated and uncultivated viruses encode ribosomal proteins. *Nat Commun* 10:752. <https://doi.org/10.1038/s41467-019-08672-6>.
18. Schvarcz CR, Steward GF. 2018. A giant virus infecting green algae encodes key fermentation genes. *Virology* 518:423–433. <https://doi.org/10.1016/j.virol.2018.03.010>.
19. Piacente F, Gaglianone M, Laugier ME, Tonetti MG. 2015. The autonomous glycosylation of large DNA viruses. *Int J Mol Sci* 16:29315–29328. <https://doi.org/10.3390/ijms161226169>.
20. Schulz F, Roux S, Paez-Espino D, Jungbluth S, Walsh D, Denef VJ, McMahon KD, Constantinidis KT, Elie-Fadrosh EA, Kyrpides N, Woyke T. 2020. Giant virus diversity and host interactions through global metagenomics. *Nature* 578:432–437. <https://doi.org/10.1038/s41586-020-1957-x>.
21. Needham DM, Yoshizawa S, Hosaka T, Poirier C, Choi CJ, Hehenberger E, Irwin NAT, Wilken S, Yung C-M, Bachy C, Kurihara R, Nakajima Y, Kojima K, Kimura-Someya T, Leonard G, Malmstrom RR, Mende DR, Olson DK, Sudo Y, Sudek S, Richards TA, DeLong EF, Keeling PJ, Santoro AE, Shirouzu M, Iwasaki W, Worden AZ. 2019. A distinct lineage of giant viruses brings a rhodopsin photosystem to unicellular marine predators. *Proc Natl Acad Sci U S A* 116:20574–20583. <https://doi.org/10.1073/pnas.1907517116>.
22. Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, Poulos BT, Solonenko N, Lara E, Poulain J, Pesant S, Kandels-Lewis S, Dimier C, Picheral M, Searson S, Cruaud C, Alberti A, Duarte CM, Gasol JM, Vaqué D, Bork P, Acinas SG, Wincker P, Sullivan MB, Tara Oceans Coordinators. 2016. Ecogenomics and potential biogeochemical impacts of globally abundant ocean viruses. *Nature* 537:689–693. <https://doi.org/10.1038/nature19366>.
23. Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, Blanc-Mathieu R, Yamamoto K, Hingamp P, Sako Y, Sullivan MB, Goto S, Ogata H, Yoshida T. 2017. Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere* 2:e00359-16. <https://doi.org/10.1128/mSphere.00359-16>.
24. Forterre P. 2013. The virocell concept and environmental microbiology. *ISME J* 7:233–236. <https://doi.org/10.1038/ismej.2012.110>.
25. Rosenwasser S, Ziv C, Creveld S. v, Vardi A. 2016. Virocell metabolism: metabolic innovations during host-virus interactions in the ocean. *Trends Microbiol* 24:821–832. <https://doi.org/10.1016/j.tim.2016.06.006>.
26. Coy SR, Gann ER, Pound HL, Short SM, Wilhelm SW. 2018. Viruses of eukaryotic algae: diversity, methods for detection, and future directions. *Viruses* 10:487. <https://doi.org/10.3390/v10090487>.
27. Gallot-Lavallée L, Blanc G, Claverie J-M. 2017. Comparative genomics of *Chrysochromulina ericina* virus and other microalga-infecting large DNA viruses highlights their intricate evolutionary relationship with the established *Mimiviridae* family. *J Virol* 91:e00230-17. <https://doi.org/10.1128/JVI.00230-17>.
28. Jacobsen A, Bratbak G, Haldal M. 1996. Isolation and characterization of a virus infecting *Phaeocystis pouchetii* (Prymnesiophyceae). *J Phycol* 32:923–927. <https://doi.org/10.1111/j.0022-3646.1996.00923.x>.
29. Baudoux A-C, Brussaard CPD. 2005. Characterization of different viruses infecting the marine harmful algal bloom species *Phaeocystis globosa*. *Virology* 341:80–90. <https://doi.org/10.1016/j.virol.2005.07.002>.
30. Wagstaff BA, Vladu IC, Barclay JE, Schroeder DC, Malin G, Field RA. 2017. Isolation and characterization of a double stranded DNA megavirus infecting the toxin-producing haptophyte *Prymnesium parvum*. *Viruses* 9:40. <https://doi.org/10.3390/v9030040>.
31. Egge ES, Johannessen TV, Andersen T, Eikrem W, Bittner L, Larsen A, Sandaa R-A, Edvardsen B. 2015. Seasonal diversity and dynamics of haptophytes in the Skagerrak, Norway, explored by high-throughput sequencing. *Mol Ecol* 24:3026–3042. <https://doi.org/10.1111/mec.13160>.
32. Sandaa RA, Haldal M, Castberg T, Thyraug R, Bratbak G. 2001. Isolation and characterization of two viruses with large genome size infecting *Chrysochromulina ericina* (Prymnesiophyceae) and *Pyramimonas orientalis* (Prymnesiophyceae). *Virology* 290:272–280. <https://doi.org/10.1006/viro.2001.1161>.
33. Johannessen TV, Bratbak G, Larsen A, Ogata H, Egge ES, Edvardsen B, Eikrem W, Sandaa R-A. 2015. Characterisation of three novel giant viruses reveals huge diversity among viruses infecting Prymnesiales (Haptophyta). *Virology* 476:180–188. <https://doi.org/10.1016/j.virol.2014.12.014>.
34. Feldman HA, Wang SS. 1961. Sensitivity of various viruses to chloroform. *Proc Soc Exp Biol Med* 106:736–738. <https://doi.org/10.3181/00379727-106-26459>.
35. Martínez M, Boere A, Gilg I, van Lent J, Witte H, van Bleijswijk J, Brussaard C. 2015. New lipid envelope-containing dsDNA virus isolates infecting *Micromonas pusilla* reveal a separate phylogenetic group. *Aquat Microb Ecol* 74:17–28. <https://doi.org/10.3354/ame01723>.
36. Mirza SF, Staniewski MA, Short CM, Long AM, Chaban YV, Short SM. 2015. Isolation and characterization of a virus infecting the freshwater algae *Chrysochromulina parva*. *Virology* 486:105–115. <https://doi.org/10.1016/j.virol.2015.09.005>.
37. Mackinder LCM, Worthy CA, Biggi G, Hall M, Ryan KP, Varsani A, Harper GM, Wilson WH, Brownlee C, Schroeder DC. 2009. A unicellular algal virus, *Emiliania huxleyi* virus 86, exploits an animal-like infection strategy. *J Gen Virol* 90:2306–2316. <https://doi.org/10.1099/vir.0.011635-0>.
38. Huiskonen JT, Kivelä HM, Bamford DH, Butcher SJ. 2004. The PM2 virion has a novel organization with an internal membrane and pentameric receptor binding spikes. *Nat Struct Mol Biol* 11:850–856. <https://doi.org/10.1038/nsmb807>.
39. Yan X, Chipman PR, Castberg T, Bratbak G, Baker TS. 2005. The marine algal virus PpV01 has an icosahedral capsid with T=219 quasiasymmetry. *J Virol* 79:9236–9243. <https://doi.org/10.1128/JVI.79.14.9236-9243.2005>.
40. Huiskonen JT, Butcher SJ. 2007. Membrane-containing viruses with icosahedrally symmetric capsids. *Curr Opin Struct Biol* 17:229–236. <https://doi.org/10.1016/j.sbi.2007.03.005>.
41. King AM, Lefkowitz E, Adams MJ, Carstens EB. 2011. Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses. Elsevier, London, UK.
42. Peralta B, Gil-Carmon D, Castaño-Díez D, Bertin A, Boulogne C, Oksanen HM, Bamford DH, Abrescia NGA. 2013. Mechanism of membranous tunnelling nanotube formation in viral genome delivery. *PLoS Biol* 11:e1001667. <https://doi.org/10.1371/journal.pbio.1001667>.
43. Philippe C, Krupovic M, Jaomanjaka F, Claisse O, Petrel M, Le Marrec C. 2018. Bacteriophage GC1, a novel tectivirus infecting *Gluconobacter cerinus*, an acetic acid bacterium associated with wine-making. *Viruses* 10:39. <https://doi.org/10.3390/v10010039>.
44. Yan X, Yu Z, Zhang P, Battisti AJ, Holdaway HA, Chipman PR, Bajaj C, Bergoin M, Rossmann MG, Baker TS. 2009. The capsid proteins of a large, icosahedral dsDNA virus. *J Mol Biol* 385:1287–1299. <https://doi.org/10.1016/j.jmb.2008.11.002>.
45. Klose T, Rossmann MG. 2014. Structure of large dsDNA viruses. *Biol Chem* 395:711–719. <https://doi.org/10.1515/hsz-2014-0145>.
46. Gowing MM. 1993. Large virus-like particles from vacuoles of phaeodarian radiolarians and from other marine samples. *Mar Ecol Prog Ser* 101:33–43. <https://doi.org/10.3354/meps101033>.
47. Gromov BV, Mamkaeva KA. 1981. A virus infection in the synchronized population of the *Chlorococcum minutum* zoospores. *Algological Studies/Archiv Für Hydrobiologie, Supplement Volumes* :252–259.
48. Bratbak G, Jacobsen A, Haldal M, Nagasaki K, Thingstad F. 1998. Virus production in *Phaeocystis pouchetii* and its relation to host cell growth and nutrition. *Aquat Microb Ecol* 16:1–9. <https://doi.org/10.3354/ame016001>.
49. Zimmerman AE, Bachy C, Ma X, Roux S, Jang HB, Sullivan MB, Waldbauer JR, Worden AZ. 2019. Closely related viruses of the marine picoeukaryotic

- alga *Ostreococcus lucimarinus* exhibit different ecological strategies. *Environ Microbiol* 21:2148–2170. <https://doi.org/10.1111/1462-2920.14608>.
50. Thomsen HA, Buck KR, Chavez FP. 1994. Haptophytes as components of marine phytoplankton. *DTU Res Database* :187–208.
 51. Berngruber TW, Froissart R, Choisy M, Gandon S. 2013. Evolution of virulence in emerging epidemics. *PLoS Pathog* 9:e1003209. <https://doi.org/10.1371/journal.ppat.1003209>.
 52. Day T, Proulx SR. 2004. A general theory for the evolutionary dynamics of virulence. *Am Nat* 163:E40–E63. <https://doi.org/10.1086/382548>.
 53. King AA, Shrestha S, Harvill ET, Bjørnstad ON. 2009. Evolution of acute infections and the invasion-persistence trade-off. *Am Nat* 173:446–455. <https://doi.org/10.1086/597217>.
 54. Brussaard CPD, Bratbak G, Baudoux A-C, Ruardij P. 2007. Phaeocystis and its interaction with viruses. *Biogeochemistry* 83:201–215. <https://doi.org/10.1007/s10533-007-9096-0>.
 55. Leggett HC, Buckling A, Long GH, Boots M. 2013. Generalism and the evolution of parasite virulence. *Trends Ecol Evol* 28:592–596. <https://doi.org/10.1016/j.tree.2013.07.002>.
 56. Woolhouse MEJ, Taylor LH, Haydon DT. 2001. Population biology of multihost pathogens. *Science* 292:1109–1112. <https://doi.org/10.1126/science.1059026>.
 57. Johannessen TV, Larsen A, Bratbak G, Pagarete A, Edvardsen B, Egge ED, Sandaa R-A. 2017. Seasonal dynamics of haptophytes and dsDNA algal viruses suggest complex virus-host relationship. *Viruses* 9:84. <https://doi.org/10.3390/v904084>.
 58. Gran-Stadniczenko S, Krabberød AK, Sandaa R-A, Yau S, Egge E, Edvardsen B. 2019. Seasonal dynamics of algae-infecting viruses and their inferred interactions with protists. *Viruses* 11:1043. <https://doi.org/10.3390/v11111043>.
 59. Fischer MG, Allen MJ, Wilson WH, Suttle CA. 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>.
 60. Sandaa R-A, Dahle H, Brussaard CPD, Ogata H, Blanc-Mathieu R. 2020. Algal viruses belonging to a subgroup within the *Mimiviridae* family. In Bamford D, Zuckerman M (ed), *Encyclopedia of virology*, 4th ed, Elsevier, London, UK.
 61. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33–36. <https://doi.org/10.1093/nar/28.1.33>.
 62. Blanc-Mathieu R, Ogata H. 2016. DNA repair genes in the Megavirales pangenome. *Curr Opin Microbiol* 31:94–100. <https://doi.org/10.1016/j.mib.2016.03.011>.
 63. Moniruzzaman M, Martinez-Gutierrez CA, Weinheimer AR, Aylward FO. 2020. Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. 1. *Nat Commun* 11:1710. <https://doi.org/10.1038/s41467-020-15507-2>.
 64. Ono A. 2010. Viruses and lipids. *Viruses* 2:1236–1238. <https://doi.org/10.3390/v2051236>.
 65. Heaton NS, Randall G. 2011. Multifaceted roles for lipids in viral infection. *Trends Microbiol* 19:368–375. <https://doi.org/10.1016/j.tim.2011.03.007>.
 66. Lange PT, Lagunoff M, Tarakanova VL. 2019. Chewing the fat: the conserved ability of DNA viruses to hijack cellular lipid metabolism. *Viruses* 11:119. <https://doi.org/10.3390/v11020119>.
 67. Diamond DL, Syder AJ, Jacobs JM, Sorensen CM, Walters K-A, Proll SC, McDermott JE, Gritsenko MA, Zhang Q, Zhao R, Metz TO, Camp DG, Waters KM, Smith RD, Rice CM, Katze MG. 2010. Temporal proteome and lipidome profiles reveal hepatitis C virus-associated reprogramming of hepatocellular metabolism and bioenergetics. *PLoS Pathog* 6:e1000719. <https://doi.org/10.1371/journal.ppat.1000719>.
 68. Heaton NS, Randall G. 2010. Dengue virus induced autophagy regulates lipid metabolism. *Cell Host Microbe* 8:422–432. <https://doi.org/10.1016/j.chom.2010.10.006>.
 69. Rosenwasser S, Mausz MA, Schatz D, Sheyn U, Malitsky S, Aharoni A, Weinstock E, Tzfadia O, Ben-Dor S, Feldmesser E, Pohnert G, Vardi A. 2014. Rewiring host lipid metabolism by large viruses determines the fate of *Emiliania huxleyi*, a bloom-forming alga in the ocean. *Plant Cell* 26:2689–2707. <https://doi.org/10.1105/tpc.114.125641>.
 70. Malitsky S, Ziv C, Rosenwasser S, Zheng S, Schatz D, Porat Z, Ben-Dor S, Aharoni A, Vardi A. 2016. Viral infection of the marine alga *Emiliania huxleyi* triggers lipidome remodeling and induces the production of highly saturated triacylglycerol. *New Phytol* 210:88–96. <https://doi.org/10.1111/nph.13852>.
 71. Wilson WH, Schroeder DC, Allen MJ, Holden MTG, Parkhill J, Barrell BG, Churcher C, Hamlin N, Mungall K, Norbertczak H, Quail MA, Price C, Rabinowitz E, Walker D, Craigon M, Roy D, Ghazal P. 2005. Complete genome sequence and lytic phase transcription profile of a coccolithovirus. *Science* 309:1090–1092. <https://doi.org/10.1126/science.1113109>.
 72. Vardi A, Van Mooy BAS, Fredricks HF, Popenodorf KJ, Ossolinski JE, Haramaty L, Bidle KD. 2009. Viral glycosphingolipids induce lytic infection and cell death in marine phytoplankton. *Science* 326:861–865. <https://doi.org/10.1126/science.1177322>.
 73. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res* 46:W95–W101. <https://doi.org/10.1093/nar/gky418>.
 74. Luther KB, Hülsmeier AJ, Schegg B, Deuber SA, Raoult D, Hennen T. 2011. Mimivirus collagen is modified by bifunctional lysyl hydroxylase and glycosyltransferase enzyme. *J Biol Chem* 286:43701–43709. <https://doi.org/10.1074/jbc.M111.309096>.
 75. Rommel AJ, Hülsmeier AJ, Jurt S, Hennen T. 2016. Giant mimivirus R707 encodes a glycogenin paralogue polymerizing glucose through α - and β -glycosidic linkages. *Biochem J* 473:3451–3462. <https://doi.org/10.1042/BCJ20160280>.
 76. Parakkottil Chothi M, Duncan GA, Armirotti A, Abergel C, Gurnon JR, Van Etten JL, Bernardi C, Damonte G, Tonetti M. 2010. Identification of an l-rhamnose synthetic pathway in two nucleocytoplasmic large DNA viruses. *J Virol* 84:8829–8838. <https://doi.org/10.1128/JVI.00770-10>.
 77. Marie D, Brussaard CPD, Thyrhaug R, Bratbak G, Vault D. 1999. Enumeration of marine viruses in culture and natural samples by flow cytometry. *Appl Environ Microbiol* 65:45–52. <https://doi.org/10.1128/AEM.65.1.45-52.1999>.
 78. Brussaard CPD. 2004. Optimization of procedures for counting viruses by flow cytometry. *Appl Environ Microbiol* 70:1506–1513. <https://doi.org/10.1128/aem.70.3.1506-1513.2004>.
 79. Jarvis B, Wilrich C, Wilrich P-T. 2010. Reconsideration of the derivation of most probable numbers, their standard deviations, confidence bounds and rarity values. *J Appl Microbiol* 109:1660–1667. <https://doi.org/10.1111/j.1365-2672.2010.04792.x>.
 80. Kremer JR, Mastrorade DN, McIntosh JR. 1996. Computer visualization of three-dimensional image data using IMOD. *J Struct Biol* 116:71–76. <https://doi.org/10.1006/jsbi.1996.0013>.
 81. Lawrence JE, Steward GF. 2010. Purification of viruses by centrifugation, p 166–181. In Wilhelm S, Weinbauer M, Suttle C (ed), *Manual of aquatic viral ecology*. American Society of Limnology and Oceanography, Waco, TX.
 82. Sandaa R-A, Storesund JE, Olesin E, Paulsen ML, Larsen A, Bratbak G, Ray JL. 2018. Seasonality drives microbial community structure, shaping both eukaryotic and prokaryotic host–viral relationships in an arctic marine ecosystem. *Viruses* 10:715. <https://doi.org/10.3390/v10120715>.
 83. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380. <https://doi.org/10.1038/nature03959>.
 84. Gordon D, Green P. 2013. *Consed*: a graphical editor for next-generation sequencing. *Bioinformatics* 29:2936–2937. <https://doi.org/10.1093/bioinformatics/btt515>.
 85. Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat Methods* 12:733–735. <https://doi.org/10.1038/nmeth.3444>.
 86. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
 87. Yutin N, Wolf YI, Raoult D, Koonin EV. 2009. Eukaryotic large nucleocytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. *Virol J* 6:223. <https://doi.org/10.1186/1743-422X-6-223>.
 88. Yutin N, Wolf YI, Koonin EV. 2014. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. *Virology* 466–467:38–52. <https://doi.org/10.1016/j.virol.2014.06.032>.
 89. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.

90. Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
91. Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
92. Lartillot N, Lepage T, Blanquart S. 2009. PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288. <https://doi.org/10.1093/bioinformatics/btp368>.
93. Villar E, Vannier T, Vernet C, Lescot M, Cuenca M, Alexandre A, Bachelier P, Rosnet T, Pelletier E, Sunagawa S, Hingamp P. 2018. The Ocean Gene Atlas: exploring the biogeography of plankton genes online. *Nucleic Acids Res* 46:W289–W295. <https://doi.org/10.1093/nar/gky376>.
94. Stamatakis A. 2014. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
95. Letunic I, Bork P. 2016. Interactive Tree Of Life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 44:W242–W245. <https://doi.org/10.1093/nar/gkw290>.
96. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* 26:1641–1650. <https://doi.org/10.1093/molbev/msp077>.
97. Lartillot N, Rodrigue N, Stubbs D, Richer J. 2013. PhyloBayes MPI: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Syst Biol* 62:611–615. <https://doi.org/10.1093/sysbio/syt022>.
98. Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-site heterogeneities in the amino-acid replacement process. *Mol Biol Evol* 21:1095–1109. <https://doi.org/10.1093/molbev/msh112>.
99. Besemer J, Lomsadze A, Borodovsky M. 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res* 29:2607–2618. <https://doi.org/10.1093/nar/29.12.2607>.
100. Lowe TM, Chan PP. 2016. tRNAscan-SE on-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res* 44:W54–57. <https://doi.org/10.1093/nar/gkw413>.
101. Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P, Goto S, Ogata H. 2016. Linking virus genomes with host taxonomy. *Viruses* 8:66. <https://doi.org/10.3390/v8030066>.
102. Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35:D61–D65. <https://doi.org/10.1093/nar/gkl842>.
103. Suzek BE, Wang Y, Huang H, McGarvey PB, Wu CH, the UniProt Consortium. 2015. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics* 31:926–932. <https://doi.org/10.1093/bioinformatics/btu739>.
104. Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He J, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, Bryant SH. 2015. CDD: NCBI's conserved domain database. *Nucleic Acids Res* 43:D222–D226. <https://doi.org/10.1093/nar/gku1221>.
105. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/nar/gkv1070>.
106. Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmiento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P, Tara Oceans Coordinators. 2015. Ocean plankton: structure and function of the global ocean microbiome. *Science* 348:1261359. <https://doi.org/10.1126/science.1261359>.
107. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. *Nucleic Acids Res* 47:D427–D432. <https://doi.org/10.1093/nar/gky995>.
108. Eddy SR. 1998. Profile hidden Markov models. *Bioinformatics* 14:755–763. <https://doi.org/10.1093/bioinformatics/14.9.755>.
109. Carradec Q, Pelletier E, Silva CD, Alberti A, Seeleuthner Y, Blanc-Mathieu R, Lima-Mendez G, Rocha F, Tirichine L, Labadie K, Kirilovsky A, Bertrand A, Engelen S, Madoui M-A, Méheust R, Poulain J, Romac S, Richter DJ, Yoshikawa G, Dimier C, Kandels-Lewis S, Picheral M, Searson S, Jaillon O, Aury J-M, Karsenti E, Sullivan MB, Sunagawa S, Bork P, Not F, Hingamp P, Raes J, Guidi L, Ogata H, Vargas C, d, Iudicone D, Bowler C, Wincker P, Tara Oceans Coordinators. 2018. A global ocean atlas of eukaryotic genes. *Nat Commun* 9:373. <https://doi.org/10.1038/s41467-017-02342-1>.
110. Keeling PJ, Burki F, Wilcox HM, Allam B, Allen EE, Amaral-Zettler LA, Armbrust EV, Archibald JM, Bharti AK, Bell CJ, Beszteri B, Bidle KD, Cameron CT, Campbell L, Caron DA, Cattolico RA, Collier JL, Coyne K, Davy SK, Deschamps P, Dyhrman ST, Edvardsen B, Gates RD, Gobler CJ, Greenwood SJ, Guida SM, Jacobi JL, Jakobsen KS, James ER, Jenkins B, John U, Johnson MD, Juhl AR, Kamp A, Katz LA, Kiene R, Kudryavtsev A, Leander BS, Lin S, Lovejoy C, Lynn D, Marchetti A, McManus G, Nedelcu AM, Menden-Deuer S, Miceli C, Mock T, Montresor M, Moran MA, Murray S, Nadathur G, Nagai S, Ngam PB, Palenik B, Pawlowski J, et al. 2014. The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLoS Biol* 12:e1001889. <https://doi.org/10.1371/journal.pbio.1001889>.
111. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>.