Virology 518 (2018) 423–433

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/virology

A giant virus infecting green algae encodes key fermentation genes

Christopher R. Schvarcz, Grieg F. Steward*

Department of Oceanography, Daniel K. Inouye Center for Microbial Oceanography: Research and Education, University of Hawai'i at Mānoa, 1950 East-West Road, Honolulu, HI 96822, United States

ARTICLE INFO

Keywords:
NCLDV
Mimiviridae
Giant virus
Algal virus
Green algae
Pyruvate formate-lyase
Auxiliary metabolic genes

ABSTRACT

The family *Mimiviridae* contains uncommonly large viruses, many of which were isolated using a free-living amoeba as a host. Although the genomes of these and other mimivirids that infect marine heterokont and haptophyte protists have now been sequenced, there has yet to be a genomic investigation of a mimivirid that infects a member of the Viridiplantae lineage (green algae and land plants). Here we characterize the 668-kilobase complete genome of TetV-1, a mimivirid that infects the cosmopolitan green alga *Tetraselmis* (Chlorodendrophyceae). The analysis revealed genes not previously seen in viruses, such as the mannitol metabolism enzyme mannitol 1-phosphate dehydrogenase, the saccharide degradation enzyme alpha-galactosidase, and the key fermentation genes pyruvate formate-lyase and pyruvate formate-lyase activating enzyme. The TetV genome is the largest sequenced to date for a virus that infects a photosynthetic organism, and its genes reveal unprecedented mechanisms by which viruses manipulate their host's metabolism.

1. Introduction

Viruses display exceptional genetic and morphological diversity (Koonin and Dolja, 2014), and most remain undiscovered or are known only through sequence fragments (Brum et al., 2015; Paez-Espino et al., 2016). Analyses of novel viral genomes often uncover genes that were once exclusively reported in cellular organisms, continually expanding the known functional repertoire of viruses (Mann et al., 2003; Sharon et al., 2011; Rosenwasser and Ziv, 2016). Viruses having exceptionally large capsids and genomes, the so-called "giant" viruses, can reach a size (> 1 μm) and gene-coding capacity (> 2.5 Mb) rivaling that of free-living cells (Fischer, 2016). Although the difference in replication mechanism (assembly vs. fission) continues to clearly distinguish viruses from cells (Steward et al., 2013), the distinction between canonical cellular vs. viral metabolic functions is blurring with the analysis of new giant virus genomes. For example, the 1.2 megabase genome of Acanthamoeba polyphaga mimivirus (APMV) revealed the first examples of virus-encoded amino-acyl transfer RNA synthetases (Raoult et al., 2004; Abergel et al., 2005), and more recently, a viral metacaspase was identified in the single amplified genome of a marine giant virus (Wilson et al., 2017).

Despite the increasing genomic data available for giant viruses, the range of genetic potential represented among the largest viruses sequenced to date is likely biased by the limited number of hosts on which they have been isolated. Nine out of the ten largest viral genomes published in the National Center for Biotechnology Information (NCBI) Viral Genomes database (all > 600 kilobase pairs in length) are from viruses that were isolated on a single type of heterotrophic host namely, amoebae in the genus *Acanthamoeba* (Raoult et al., 2004; Yoosuf et al., 2012, 2014; Arslan et al., 2011; Philippe et al., 2013; Legendre et al., 2014, 2015; Antwerpen et al., 2015). The other is for a virus infecting the heterotrophic heterokont protist, *Cafeteria roenbergensis* (Fischer et al., 2010). This low host diversity equates to a low diversity of cellular metabolisms available for these viruses to exploit, resulting in fewer types of proteins that a virus might co-opt for successful infection. Investigating the genomes of giant viruses that infect a broader range of hosts may help to reveal new viral functions and infection strategies.

In this study, we present the physical characteristics and complete genome sequence for a novel virus in the family *Mimiviridae* that infects green algae in the genus *Tetraselmis* (Chlorodendrophyceae): Tetraselmis virus 1 (TetV-1; referred to hereafter as simply TetV). Only one other green alga-infecting mimivirid (Pyramimonas orientalis virus 01B; PoV) has previously been characterized (Sandaa et al., 2001), but its genome has yet to be sequenced. In the absence of a full genome analysis of any green alga-infecting mimivirid, the functional capacity of these viruses, and how they compare to other mimivirids, has been unclear.

TetV was isolated from the coastal waters of O'ahu, Hawai'i, and its host was isolated from seawater collected roughly 100 kilometers north of O'ahu, at the Hawai'i Ocean Time-series site Station ALOHA (Karl and Lukas, 1996). While this particular host strain was isolated from the oligotrophic, open ocean, members of the genus *Tetraselmis* have a

* Corresponding author.
E-mail address: grieg@hawaii.edu (G.F. Steward).

https://doi.org/10.1016/j.virol.2018.03.010
Received 22 November 2017; Received in revised form 8 March 2018; Accepted 11 March 2018

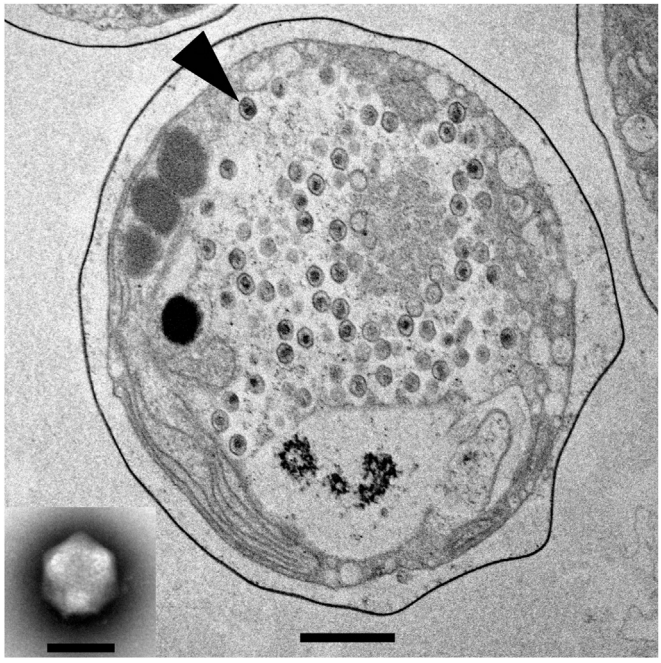

Fig. 1. Transmission electron micrograph of an infected *Tetraselmis* cell section. The arrow points to a TetV virion. Scale bar equals 1 µm. Inset: negatively stained TetV virion. Scale bar equals 200 nm.

cosmopolitan distribution and are commonly found in nutrient-rich marine and fresh waters. *Tetraselmis* spp. are widely used as a feed source in the aquaculture industry (Hemaiswarya et al., 2010) and have received attention from the biofuel industry as a model organism for the production of starch (Zheng et al., 2011; Yao et al., 2012; Ji et al., 2014). Investigations of *Tetraselmis*-infecting viruses might, therefore, ultimately find some practical applications.

2. Results and discussion

2.1. Virion properties

TetV virions have average minimum and maximum dimensions of 226 ± 9 nm and 257 ± 9 nm, respectively (Fig. 1), making them among the largest viruses characterized for algae. This is slightly larger than the closely related PoV, which has an average capsid size of 180–220 nm (Sandaa et al., 2001), but smaller than some other previously characterized or observed algal viruses (Johannessen et al., 2015; Dodds and Cole, 1980; Sicko-Goad and Walker, 1979).

The buoyant density of TetV particles in CsCl is approximately 1.386 g mL^{-1}, which is within the range typically observed for marine viruses (Lawrence and Steward, 2010).

2.2. Host range

The host range of TetV was tested using diverse phytoplankton strains collected from both Station ALOHA (origin of host) and Kāneʻohe Bay, Oʻahu, Hawaiʻi (origin of TetV). Included were other chlorophytes such as *Pyramimonas*, *Micromonas*, and prasinophyte clade VII, as well as other photosynthetic flagellates such as chlorarachniophytes, cryptophytes, dictyochophytes, haptophytes, and pelagophytes (Table S2). Permissiveness to TetV infection resulting in cell lysis was only observed in other *Tetraselmis* strains (all of the eight total *Tetraselmis* strains tested; 18S rDNA sequences deposited in GenBank under accession numbers MH055444-MH055457). A phylogenetic analysis of permissive strains showed that they included three distinct phylotypes in a clade encompassing *Tetraselmis cordiformis* and *Tetraselmis* symbionts of open ocean radiolaria (Fig. S1).

2.3. Phylogenetic analysis

Phylogenetic analyses of multiple marker genes support TetV's classification as a member of family *Mimiviridae* (Fig. 2). Based on the phylogenies of DNA-dependent DNA polymerase family B (PolB) and DNA mismatch repair protein MutS7, TetV appears to be most closely related to a previously isolated virus, PoV, infecting the prasinophyte *Pyramimonas orientalis*. The phylogeny for major capsid protein MCP1 shows a slightly different evolutionary history; however, these branches have lower support. Although an A32-like virion packaging ATPase from PoV was not available for comparison, the phylogeny for this gene, like the other three, suggests a clustering of TetV with other alga-infecting members of the family *Mimiviridae*. This is consistent with a comparative genomic analysis of members of mimivirids, and the suggestion that these alga-infecting viruses could constitute a subfamily within the *Mimiviridae* (Gallot-Lavallée et al., 2017). The analysis presented here hints at further substructure within this group and the possible emergence of a monophyletic clade of chlorophyte-infecting mimivirids founded by TetV and PoV.

2.4. Basic genome features

The complete TetV dsDNA genome (deposited in GenBank under accession number KY322437) was assembled as a single, circular sequence measuring 668,031 base pairs (bp), making it the largest genome sequenced to date for a virus infecting a photosynthetic organism. This is nearly 200 kb larger than the next largest for an algal virus, the recently published Chrysochromulina ericina virus (CeV) genome, sequenced at 473,558 bp (Gallot-Lavallée et al., 2015). The circular structure and size of TetV's genome was supported by pulsed-field gel electrophoresis (PFGE; Fig. S2 and Table S1). Undigested genomic DNA was retained in the well during PFGE, a phenomenon expected for large circular molecules under the conditions used (Levene and Zimm, 1987), and digestion with NotI, a restriction endonuclease predicted to have a single cut site in the genome, resulted in the expected single band. The estimated sizes of the bands resulting from digestion with NotI or SfiI agreed with the predicted sizes to within the error of the method (Table S1).

The TetV genome has a GC content of 41.2%, which is considerably higher than other mimivirids (23–32%; Abergel et al., 2015; Gallot-Lavallée et al., 2017; Schulz et al., 2017). Although the complete PoV genome is not available, the GC content of the genes for PolB (38%) and MCP1 (53%) suggest PoV might have a GC content more similar to TetV, consistent with the apparent phylogenetic affiliation of these two viruses.

A total of 663 genes were predicted in the TetV genome, including 653 coding DNA sequences (CDSs) and 10 tRNAs (complete list of gene annotations in Supplementary Dataset S1). All of the tRNAs are encoded consecutively, on the same strand, in an 879-bp region of the genome, an arrangement that is also seen in Cafeteria roenbergensis virus BV-PW1 (CroV) and many chloroviruses (Fischer et al., 2010; Fitzgerald et al., 2007b, 2007a). The average CDS length is 960 bp, and the overall gene-coding density of the genome is 93.7%.

2.5. Protein BLAST top hit distributions

A blastp top hit analysis was performed using a combined database that included GenBank's non-redundant (nr) dataset and eukaryotic transcriptomes available through the Gordon and Betty Moore Foundation's Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP; Keeling et al., 2014). Sixty-seven percent (435) of TetV's predicted proteins share no similarity to known proteins (Fig. 3), a proportion considerably higher than what was found for other recently sequenced algal mimivirids (43–47% no hits; Santini et al., 2013;

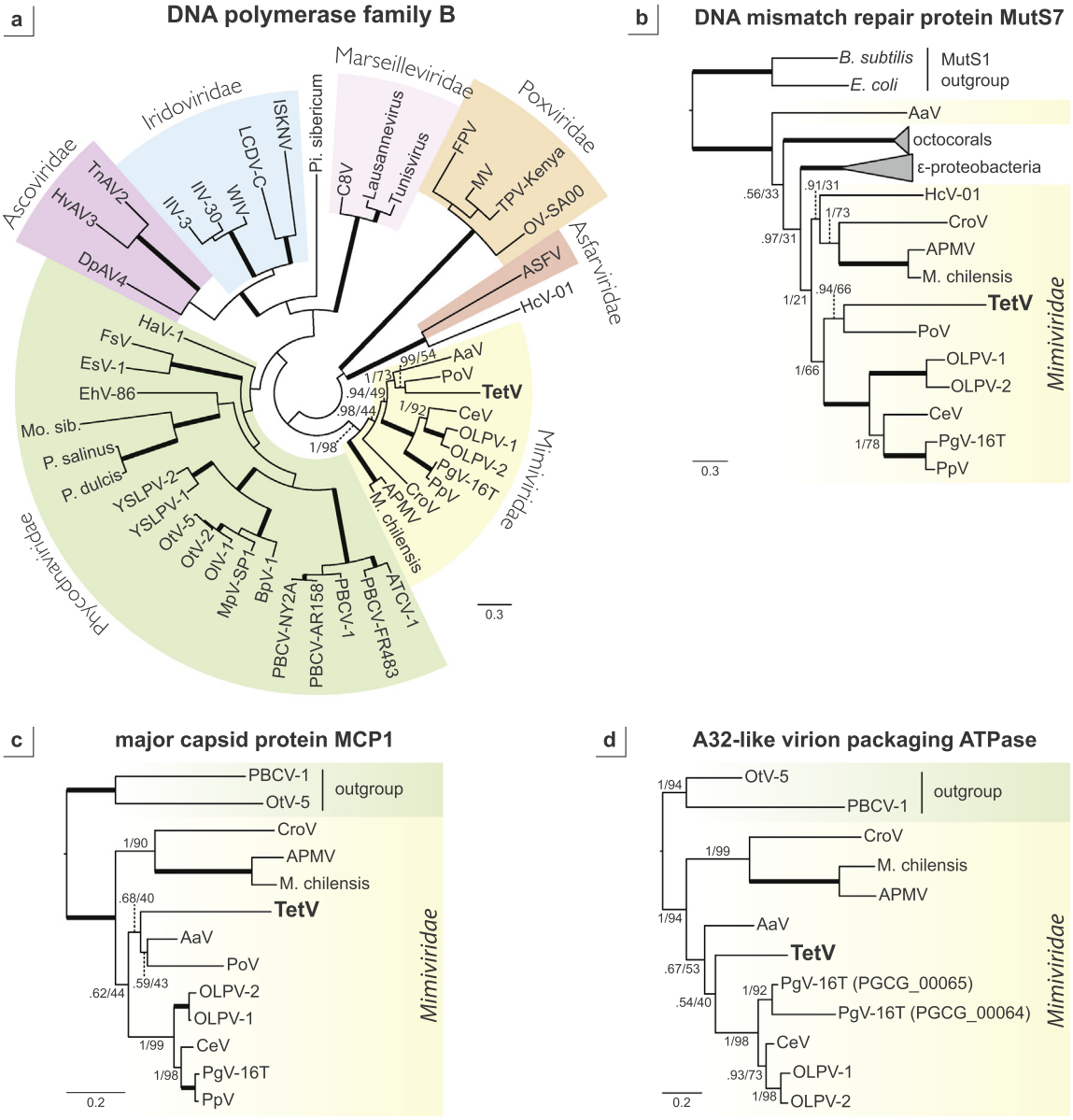

Fig. 2. Phylogenetic trees of select Nucleo-Cytoplasmic Large DNA Viruses (NCLDVs), based on protein sequence alignments of (a) DNA polymerase family B (PolB; 843 amino acid [aa] sites), (b) DNA mismatch repair protein MutS7 (674 aa sites; includes octocoral and Epsilonproteobacteria MutS7 sequences reported in Ogata et al., 2011), (c) major capsid protein MCP1 (359 aa sites), and (d) A32-like virion packaging ATPase (230 aa sites). All trees represent Bayesian majority consensus trees, with support values (Bayesian posterior probability/maximum-likelihood bootstrap percent, 1000 replicates) provided for branches within *Mimiviridae*. Bold branches indicate complete support (Bayesian posterior probability of 1 and maximum-likelihood bootstrap percent of 100). Scale bars are in units of amino acid substitutions per site. Abbreviations are as follows: AaV, Aureococcus anophagefferens virus; APMV, Acanthamoeba polyphaga mimivirus; ASFV, African swine fever virus; ATCV-1, Acanthocystis turfacea Chlorella virus 1; BpV-1, Bathycoccus sp. RCC1105 virus BpV1; C8V, Cannes 8 virus; CeV, Chrysochromulina ericina virus; CroV, Cafeteria roenbergensis virus BV-PW1; DpAV4, Diadromus pulchellus ascovirus 4a; EhV-86, Emiliania huxleyi virus 86; EsV-1, Ectocarpus siliculosus virus 1; FPV, Fowlpox virus; FsV, Feldmannia species virus; HaVDNA-1, Heterosigma akashiwo virus 01; HcV-01, Heterocapsa circularisquama DNA virus 01; HvAV3, Heliothis virescens ascovirus 3e; IIV-3, Invertebrate iridescent virus 3; IIV-30, Invertebrate iridescent virus 30; ISKNV, Infectious spleen and kidney necrosis virus; Lausannevirus, Lausannevirus; LCDV-C, Lymphocystis disease virus - isolate China; M. chilensis, Megavirus chilensis; Mo. sib., Mollivirus sibericum; MpV-SP1, Micromonas pusilla virus SP1; MV, Myxoma virus; OLPV-1, Organic Lake phycodnavirus 1; OLPV-2, Organic Lake phycodnavirus 2; OlV-1, Ostreococcus lucimarinus virus OlV1; OtV-2, Ostreococcus tauri virus 2; OtV-5, Ostreococcus tauri virus OtV5; OV-SA00, Orf virus; P. dulcis, Pandoravirus dulcis; P. salinus, Pandoravirus salinus; Pi. sibericum, Pithovirus sibericum; PBCV-1, Paramecium bursaria Chlorella virus 1; PBCV-AR158, Paramecium bursaria Chlorella virus AR158; PBCV-FR483, Paramecium bursaria Chlorella virus FR483; PBCV-NY2A, Paramecium bursaria Chlorella virus NY2A; PgV-16 T, Phaeocystis globosa virus 16 T; PoV, Pyramimonas orientalis virus 01B; PpV, Phaeocystis pouchetii virus; TetV, Tetraselmis virus 1; TnAV2, Trichoplusia ni ascovirus 2c; TPV-Kenya, Tanapox virus; Tunisvirus, Tunisvirus fontaine2; WIV, Wiseana iridescent virus; YSLPV-1, Yellowstone lake phycodnavirus 1; YSLPV-2, Yellowstone lake phycodnavirus 2.

Moniruzzaman et al., 2014; Gallot-Lavallée et al., 2015), indicating the novelty of TetV's genome relative to those currently sequenced. One hundred ten (17%) TetV proteins are most similar to eukaryotic proteins, followed by 63 (10%) top hits to viruses and 45 (7%) top hits to bacteria. This suggests that TetV contains a large number of eukaryote-derived proteins that are being observed for the first time in viruses. Forty-three (39%) of the 110 eukaryote hits are from organisms in the group Viridiplantae, which includes chlorophytes (such as *Tetraselmis*) and land plants. Of the Viridiplantae hits, 34 (31%) are from chlorophytes, and 19 (17%) of those are from *Tetraselmis* chlorophytes (2.9%

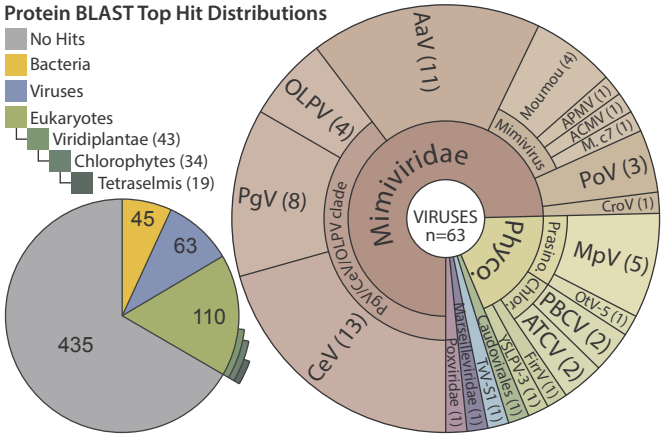

Fig. 3. Distribution of blastp top hits for TetV proteins searched against a combined database of nr plus peptide sequences from the MMETSP. For the domain-level pie chart, additional outer rings are drawn that indicate the number of hits belonging to the nested taxonomic groups Viridiplantae, chlorophytes, and *Tetraselmis*, where the arc length of the segment is proportional to the number of associated hits. The distribution of viral top hits is presented as a hierarchical pie chart, where inner rings represent higher level groupings (e.g. family and genus) for viral taxa in the outer rings. Abbreviations are as defined in Fig. 2, with the exceptions of: ACMV, Acanthamoeba castellanii mamavirus; ATCV, Acanthocystis turfacea Chlorella virus strains Canal-1 and NTS-1; Chlor., *Chlorovirus* genus; FirV, Feldmannia irregularis virus a; Moumou, Acanthamoeba polyphaga moumouvirus (n = 3) and Moumouvirus goulette (n = 1); M. c7, Megavirus courdo7; MpV, *Micromonas*-infecting strains MpV1 (n = 2), PL1 (n = 2), and SP1(n = 1); OLPV, Organic Lake phycodnavirus strains OLPV-1 (n = 3) and OLPV (n = 1); PBCV, Paramecium bursaria Chlorella virus strains KS1B and MT325; PgV, Phaeocystis globosa virus 16 T; Phyco., *Phycodnaviridae* family; Prasino., *Prasinovirus* genus; TvV-S1, Tetraselmis viridis virus S1; YSLPV-3, Yellowstone lake phycodnavirus 3.

of the total TetV proteome). For comparison, only 1.6% of AaV's proteins had top hits to the genome of its host *Aureococcus anophagefferens* (Moniruzzaman et al., 2014). Of the 63 top hits to other virus proteins, the majority (47; 75%) belong to other mimivirids.

Although TetV shows a greater phylogenetic affinity to AaV than the PgV/CeV/OLPV clade when analyzing selected genes (Fig. 2), a larger number of TetV proteins are more similar to those from viruses in the latter clade (n = 25 total) than AaV (n = 11). This may reflect AaV's relatively small genome and the loss of otherwise conserved genes through genome reduction; AaV has the smallest genome of characterized mimivirids, and the genome of its ancestor is hypothesized to have been even smaller (Moniruzzaman et al., 2014). A separate analysis of genes shared between TetV and other NCLDVs using Nucleo-Cytoplasmic Virus Orthologous Groups (NCVOGs; Yutin et al., 2009, 2014) produced similar results, in which TetV shared the highest number of genes with alga-infecting viruses of *Mimiviridae*, in particular PgV-16T, CeV, and OLPV (Figs. S3 and S4). At present, a complete genome sequence for TetV's presumed closest relative, PoV, is not available, precluding genome-wide comparisons between these viruses.

Although there are numerous alga-infecting mimivirid genomes now available, there are still a large number of TetV proteins (12; 19% of the hits to viruses) with highest BLAST similarity to viruses in the family *Phycodnaviridae*, many of which infect chlorophytes (e.g., viruses from the genera *Prasinovirus* and *Chlorovirus*). This may be indicative of horizontal gene transfer between phycodnavirids and TetV, facilitated by infections of the same group of hosts. Also notable was a single top hit to a hypothetical protein from a small (26,407-bp genome), unclassified virus that infects *Tetraselmis viridis* (Tetraselmis viridis virus S1; GenBank accession NC_020869), again suggesting horizontal gene transfer between TetV and divergent viruses infecting the same host.

Table 1
TetV genes representing novel viral homologs, seen for the first time in viruses.

		blastp Top Hit: nr + MMETSP	
TetV Locus	Annotation	Organism	E-value
TetV_320	mannitol 1-phosphate dehydrogenase*	*Heterosigma akashiwo* (Euk.)	4.28E-39
TetV_428	pyruvate formate-lyase[a]	*Chlamydomonas chlamydogama* (Euk.)	0.0
TetV_456	pyruvate formate-lyase activating enzyme[b]	*Tetraselmis* sp. (Euk.)	3.32E-51
TetV_601	alpha-galactosidase[c]	*Alkaliflexus imshenetskii* (Bact.)	9.42E-102

* Prasinoviruses encode a mannitol dehydrogenase belonging to a protein family similar to *Aspergillus fischeri* mannitol 2-dehydrogenase, which performs a different enzymatic reaction.

[a] Some phages encode a glycine radical protein that shares sequence similarity with the glycine radical domain at the C-terminal end of pyruvate formate-lyase, and many of these proteins have been misannotated as pyruvate formate-lyase.

[b] Some phages encode an anaerobic ribonucleoside-triphosphate reductase-activating protein that has been misannotated as a pyruvate formate-lyase activating enzyme.

[c] Some phages encode hypothetical proteins with very low-scoring (> 1E-3) partial domain hits (CDD) to alpha-galactosidase, which do not share BLAST similarity with any known alpha-galactosidases.

2.6. Gene function overview

Only 192 of the 653 predicted protein-coding genes in the TetV genome could be annotated with some level of functional description. At least four of these proteins were annotated with specific known functions that have not previously been observed in viruses (Table 1). One hundred thirty-three proteins had hits to the Clusters of Orthologous Groups (COG) database, and the distribution of COG functions associated with these hits is dominated by “Replication, recombination and repair” (23 proteins) and “Transcription” (18 proteins), as is seen in related mimivirids (Figs. S5 and S6). The next highest-represented categories were “Post-translational modification, protein turnover, chaperones” (14 proteins), owing to a large number of ubiquitination-related proteins, and “Signal transduction mechanisms” (14 proteins), including six serine/threonine protein kinases.

Despite TetV's larger genome, its complement of transcription- and translation-related genes is very similar to that found in other alga-infecting mimivirids, such as PgV-16 T and CeV (Fig. S7). TetV does, however, encode a eukaryotic initiation factor eIF-1A, which has not previously been observed in alga-infecting mimivirids.

Methylation appears to be another common function encoded in the TetV genome. TetV contains 12 putative methyltransferase proteins, including four DNA methyltransferases, one SET domain-containing protein methyltransferase, one FtsJ-like rRNA methyltransferase, and six other methyltransferases (three type 24, one type 11, one type 21 [FkbM-like], and one type 23).

TetV contains seven proteins related to DNA repair. These include two photolyases, DNA mismatch repair proteins MutS7 and MutS8, an ERCC4-type nuclease, a putative DNA repair ATPase SbcC, and another gene that appears to encode both the ATPase subunit (SbcC) and the exonuclease subunit (SbcD) of the DNA repair exonuclease SbcCD complex.

2.7. Ubiquitination-related genes

TetV contains 18 proteins with putative ubiquitination-related functions: 14 putative RING-finger-containing E3 ubiquitin ligases, one E2 ubiquitin-conjugating enzyme, two proteins related to de-ubiquitination, and one ubiquitin gene. Ubiquitination is the process of

covalently attaching ubiquitin(s) to a protein substrate, thereby altering the localization or activity of the protein or targeting it for degradation by the proteasome (Haglund and Dikic, 2005). Genes related to ubiquitin signaling can be found in all families of the NCLDVs, and it has been suggested that these genes may be used to counter host defenses (Iyer et al., 2006). Three of TetV's ubiquitination-related proteins appear to be most closely related to homologs in its *Tetraselmis* host. The genes for ubiquitin (TetV_289), an E2 ubiquitin-conjugating enzyme (TetV_621), and a MIEL1-like E3 ubiquitin ligase (TetV_138) all had top blastp hits to homologous proteins in the MMETSP *Tetraselmis* transcriptomes. These relationships are supported by phylogenetic analyses (Fig. S8), except in the case of ubiquitin, where the alignment region is too conserved for robust phylogenetics.

2.8. Fermentation genes and energy production

Of the four TetV genes representing novel viral homologs, the most interesting are perhaps pyruvate formate-lyase (PFL; TetV_428) and pyruvate formate-lyase activating enzyme (PFL-AE; TetV_456). PFL and PFL-AE are key enzymes in cellular fermentation pathways, allowing energy production in the absence of oxygen. Phylogenetic analyses for the TetV homologs suggest they were acquired by horizontal gene transfer from *Tetraselmis*, or a more ancestral chlorophyte host (Fig. 4). It is thought that green algae use fermentation to survive periods of low oxygen concentration (Catalanotti et al., 2013), which can occur in eutrophic environments with high rates of respiration, especially at night. In the absence of oxygen (and thus, in the absence of a functioning TCA cycle), fermentation provides a way to recycle oxidant that is needed for continued energy production through glycolysis. Similarly, TetV may use these genes to ensure that energy requirements are satisfied for successful infection during low oxygen conditions. There

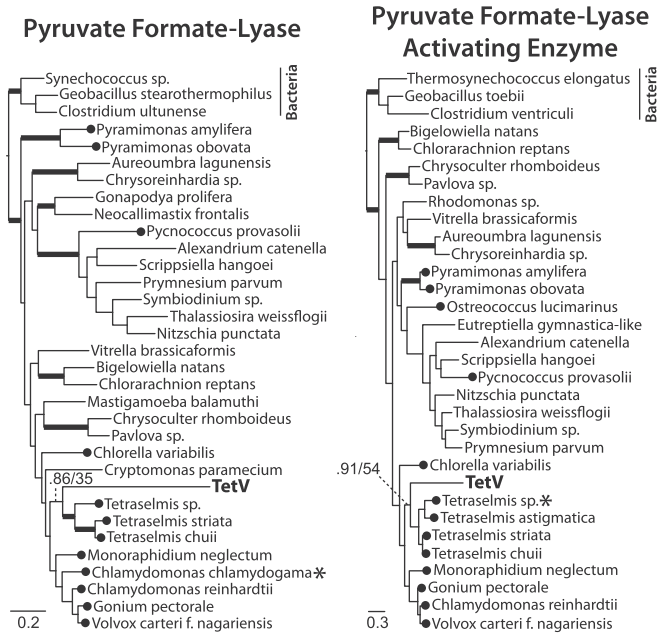

Fig. 4. Phylogenetic trees for the key fermentation genes pyruvate formate-lyase (PFL) and pyruvate formate-lyase activating enzyme (PFL-AE). Both trees represent Bayesian majority consensus trees inferred using alignments of 603 and 197 aa sites for PFL and PFL-AE, respectively, with support values (Bayesian posterior probability/maximum-likelihood bootstrap percent, 1000 replicates) provided for branches that group TetV's proteins with those of *Tetraselmis*. Bold branches indicate complete support (Bayesian posterior probability of 1 and maximum-likelihood bootstrap percent of 100). Taxa marked with filled circles are species of chlorophyte green algae, and an asterisk is used to mark the BLAST top hit for the respective TetV sequence. Scale bars are in units of amino acid substitutions per site.

are numerous reports of *Tetraselmis* blooms in coastal eutrophic waters, where concentrations can reach 10^4–10^5 cells mL^{-1}, turning surface waters green in color (Throndsen and Zingone, 1988; Jones and Rhodes, 1994; Collantes and Prado, 2006; Daoudi et al., 2013). High algal and bacterial respiration under such conditions could deplete dissolved oxygen in surface waters at nighttime. Furthermore, the spread of a virus through such a bloom and the ensuing mass lysis would dramatically increase bacterial respiration and the likelihood of anoxic conditions. Under such conditions, maintenance of the host's fermentation pathway could become highly advantageous.

The ability of TetV to manipulate its host's fermentation pathway may also hold interesting biochemical implications. In the case that TetV forces its host to utilize fermentation regardless of ambient oxygen concentrations, it could result in a greater release of fermentative products (such as ethanol) compared with uninfected cells, thus altering the composition of the dissolved organic carbon pool and the associated microbial response. Determining the extent to which TetV manipulates this pathway and the particular conditions under which this occurs is a priority for future experimental work with this model system.

2.9. Other notable genes

In addition to the fermentation genes, TetV also contains the first viral homologs for the saccharide degradation protein alpha-galactosidase (TetV_601) and the mannitol metabolism protein mannitol 1-phosphate dehydrogenase (M1PHD; TetV_320). Interestingly, the rigid cell wall of *Tetraselmis* was shown to contain up to 7% galactose and 21% galacturonic acid (Becker et al., 1989), and it is possible that TetV's alpha-galactosidase assists in breaking down the cell wall upon lysis.

The M1PHD found in TetV is homologous to the eukaryotic M1PHD described in the brown macroalga *Ectocarpus siliculosus* (Rousvoal et al., 2010), where it plays a key role in converting the photosynthate fructose 6-phosphate to the carbon storage compound mannitol. Another mannitol cycle protein is found in prasinoviruses (Moreau et al., 2010) and appears to be similar to a bacterial mannitol 2-dehydrogenase (M2DH), which is used by some marine bacteria to metabolize algal-derived mannitol into fructose (Groisillier et al., 2015). Both M1PHD and M2DH are bidirectional enzymes and could potentially play similar roles in the mannitol metabolism of these virus-host systems, or they could simply be used for controlling redox balance in the host. It is also possible that mannitol metabolism is being manipulated by the virus to control intracellular osmolarity. *Tetraselmis* is known to produce mannitol for use as an osmolyte to acclimate to hypersaline conditions (Kirst, 1989). Forced production of mannitol could induce hypotonic shock, which would facilitate cell lysis.

TetV also contains the second-known viral example of a large-conductance mechanosensitive channel (MscL; TetV_044) and the second-known mimivirid example of an arabinose 5-phosphate isomerase (API; TetV_403; this protein is also found in some phages). Both of these proteins are also found in CroV. While the CroV API is most closely related to bacterial homologs, the TetV API is most similar to a *Tetraselmis* homolog (blastp top hit). Many chlorophytes, including *Tetraselmis*, contain KDO in their cell wall theca (Becker et al., 1991), and it is possible that TetV incorporates KDO into its virion to facilitate virion-cell recognition. However, API is the only KDO synthesis gene found in TetV, unlike in CroV, where the complete pathway is present.

2.10. Gene duplications and other repetitive genetic elements

Fifty-nine TetV proteins, representing 20.3% of the TetV genome (135,897 bp total), were most similar to other proteins in the TetV genome (blastp top hit), when searched against a combined database of nr, MMETSP proteins, and TetV proteins. This includes 17 pairs of proteins that were identified as reciprocal best hit (RBH) blastp

matches, as well as 7 additional protein pairs linked by a nonreciprocal best hit that were identified as being most closely related to each other based on maximum-likelihood phylogenetic analyses. These protein pairs likely represent recent gene duplication events. Some of the duplicate gene pairs share sequence similarity to larger sets of putative paralogs in the TetV genome (Table S3 and Fig. S9; 23 total putative paralog groups), the largest of which contains 8 gene copies. Most of these putative paralogs are annotated as hypothetical proteins, and 12 of the protein groups did not have blastp hits to proteins in either nr or MMETSP. Only one of the duplicated gene pairs (TetV_290 and TetV_368, which encode putative N-6 adenine-specific DNA methylases) appeared to be homologous to sequences from other NCLDV viruses.

An overview of nucleotide exact repeats in the TetV genome revealed two large clusters of repeats corresponding to fibronectin type III-like (FN3-like) domain repeats (Fig. S10). Further analysis identified 126 FN3-like domains across 15 TetV proteins (Table S4), including four of the six largest proteins in the TetV genome, the largest of which is 8212 amino acids long. The sequence space occupied by FN3-like domains represents approximately 6.4% (43,005 bp) of the TetV genome. While the FN3-like domains are encoded on both strands, their distribution is confined to one half of the circular genome (Fig. S11). Highly repetitive protein domains have been found in other mimivirids. For example, APMV and CroV contain numerous copies (> 400 copies in CroV) of a leucine-rich repeat (LRR) similar to the FNIP/IP22 domain (Fischer et al., 2010; O'Day et al., 2006), and approximately 11.3% of AaV's genome is composed of the DUF285 domain (Moniruzzaman et al., 2014), which is distantly related to LRRs. The roles of these repetitive viral domains are still unknown; however, studies of other LRRs and FN3 domains suggest that the main function of both these domains is to mediate protein-protein interactions (Kobe, 2001; Campbell and Spitzfaden, 1994).

3. Conclusions

The 668-kb genome of TetV is the largest sequenced genome for a virus that infects a photosynthetic organism. As such, it has revealed a large number of viral homologs for genes previously not seen in viruses, including the saccharide degradation enzyme alpha-galactosidase, the mannitol metabolism enzyme mannitol 1-phosphate dehydrogenase, and two key genes in algal fermentation pathways: pyruvate formate-lyase (PFL) and pyruvate formate-lyase activating enzyme (PFL-AE). The presence of PFL and PFL-AE suggests that TetV has the unprecedented capacity to manipulate its host's fermentation pathway, which holds intriguing implications for the ecology of the virus (and its potential to spread in hypoxic/anoxic environments) and ocean biochemistry (if TetV infections alter the production of fermentation products, such as ethanol). This highlights the growing importance of sequencing diverse reference genomes of cultivated viruses, especially in the case of giant viruses, which, because of their large size, are trapped on the same filters used to collect microbial samples. Such reference data is crucial for the accurate identification of giant viral sequences in microbial metagenomes, which are being generated at an unprecedented high rate, particularly for the oceans (Rusch et al., 2007; Wilkins et al., 2012; Konstantinidis et al., 2009; Bork et al., 2015).

The genomic analysis has also provided intriguing insights into TetV's evolutionary relationship with other viruses, as well as with its host. TetV is the second member of what appears to be an emerging clade of chlorophyte-infecting mimivirids. The relationship between TetV and its chlorophyte host seems to be a specific one, as indicated by the high proportion of proteins (17% of the TetV proteome) that are most similar to green algal and plant homologs. The long history of chlorophyte infections may have also facilitated horizontal gene transfer between TetV and chlorophyte-infecting phycodnavirids. However, additional chlorophyte-infecting mimivirid genomes will need to be sequenced before we can better understand the connections (if any) between the chlorophyte-infecting viruses of *Mimiviridae* and *Phycodnaviridae*.

4. Materials and methods

4.1. Host isolation

The *Tetraselmis* sp. host was isolated from whole seawater collected at Station ALOHA (22°45' N, 158°00' W) in the North Pacific Subtropical Gyre, on May 20, 2010, during Cruise #221 of the Hawaii Ocean Time-series program (Karl and Lukas, 1996). The seawater was collected from the upper euphotic zone (25–75 m depth) using Niskin bottles, after which the water was enriched with f/2 medium (Guillard, 1975) and incubated under fluorescent lights. The *Tetraselmis* was isolated into culture from the f/2-enriched seawater using three rounds of end-point dilution (10-fold serial dilutions). After isolation, batch cultures of the *Tetraselmis* were maintained in K medium (-Si; Keller et al., 2007) at 26 °C, under a 12:12 light/dark cycle with approx. 30 μmol m^{-2} s^{-1} photons of PAR.

4.2. Virus isolation

TetV was isolated from surface seawater collected from the O'ahu-side pier to Coconut Island (Moku o Lo'e), located in Kāne'ohe Bay, O'ahu, Hawai'i, on September 2, 2010. Approximately 175 L of surface seawater was collected and transported to the University of Hawai'i at Mānoa campus, where it was filtered through 142 mm diameter, 0.8 μm pore size membrane filters (Isopore, Millipore P/N: ATTP14250), followed by concentration using tangential flow filtration (TFF; Millipore Pellicon 2 Mini system, P/N: XX42PMINI), with three stacked 30 kDa nominal molecular weight limit (NMWL) filters (Millipore Ultracel, P/N: PLCTK-C, 0.1 m^2 ea.). The viral concentrate was then amended with K medium and used to challenge healthy *Tetraslemis* culture. After propagating the *Tetraselmis* lysate for multiple generations, the TetV virus was isolated from the lysate using three rounds of dilution-to-extinction (Nagasaki and Bratbak, 2010), performed in 96-well plates.

4.3. Host-range analysis

The host-range of TetV was investigated using in-house phytoplankton strains isolated from Station ALOHA (origin of the original TetV host) and Kāne'ohe Bay (origin of TetV). Strains included the chlorophytes *Tetraselmis* (n = 3, Sta. ALOHA; n = 5, Kāne'ohe Bay), *Pyramimonas* (n = 3, Kāne'ohe Bay), *Micromonas* (n = 1, Sta. ALOHA; n = 6, Kāne'ohe Bay) and prasinophyte clade VII (n = 1, Sta. ALOHA), the pelagophyte *Pelagomonas* (n = 2, Sta. ALOHA), the dictyochophytes *Florenciella* (n = 1, Sta. ALOHA) and *Rhizochromulina* (n = 1, Sta. ALOHA), the haptophyte *Chrysochromulina* (n = 1, Sta. ALOHA; n = 1, Kāne'ohe Bay), cryptophytes (n = 2, Kāne'ohe Bay), and chlorarachniophytes (n = 2, Sta. ALOHA).

TetV challenges were performed in 96-well plates. Dense, healthy culture of each strain was amended with an equal volume of media (K or f/2 media, depending on the strain), and 270 μL was added to each of six wells (three challenge wells and three control wells). Challenge wells were inoculated with 30 μL of fresh TetV lysate that had been filtered through a 0.45 μm Millipore Sterivex filter (P/N: SVHV010RS). Control wells were inoculated with 30 μL of the same 0.45 μm-filtered lysate that was additionally filtered through two 0.02 μm Whatman Anotop filters (P/N: 6809-2002). The challenges were monitored for up to two weeks by direct microscopy and in vivo chlorophyll fluorescence (measured using the Perkin Elmer 2030 Multilabel Reader, VICTOR X3). Strains exhibiting lysis were investigated further to see if the lytic effect could be propagated. This was done by performing new challenges using 30 μL of 1/100-diluted lysate from one of the lysed wells. All of such follow-up challenges resulted in complete lysis.

4.4. Tetraselmis host DNA extraction and 18S sequencing

All of the TetV-permissive *Tetraselmis* strains identified in the host-range analysis were further characterized by small subunit rRNA gene (18S rDNA) sequencing. Cells for DNA extraction were harvested by centrifuging approximately 25 mL of culture at 3700 RCF for 10 min at 4 °C. DNA was extracted from the pellets using the ZymoBIOMICS DNA Mini Kit (P/N: D4304). The 18S rDNA was PCR-amplified from extracted DNA using universal primers Euk328f/Euk329r (der Staay et al., 2001) and the Roche Expand High Fidelity PCR System (P/N: 04738268001). PCR products were cloned using the Invitrogen TOPO TA Cloning Kit for Sequencing (P/N: K457501), and 2–3 colonies were picked for overnight culture in 3 mL Circlegrow media (MP Biomedicals, P/N: 3000-121) and extracted using the Qiagen QIAprep Spin Miniprep Kit (P/N: 27104). Full-length 18S rDNA was sequenced using primers M13f, M13r, 502f, and 1174r (Worden, 2006), and was performed using Sanger technology at the University of Hawai'i at Mānoa's facility for Advanced Studies in Genomics, Proteomics, and Bioinformatics (ASGPB).

4.5. Electron microscopy

Measurements of virion diameter were averaged from 23 particles imaged by electron microscopy, using negatively stained samples prepared from unfixed fresh lysate. Negatively stained samples were prepared using a variation of the direct application method (Doane and Anderson, 1987).

Thin sections of infected *Tetraselmis* cells were prepared using cells collected 24 h after inoculation of a healthy culture with 10% volume of TetV lysate. The infected cells were mixed 1:1 with fixative solution (2% glutaraldehyde, 0.2 M sodium cacodylate pH 7.2, 0.25 M sucrose, 10 mM CaCl) and stored for at least 2 h at 4 °C. Fixed cells were harvested by centrifugation at 1000 RCF for 10 min, washed with 0.1 M sodium cacodylate buffer (pH 7.4; 0.35 M sucrose), post-fixed for 1 h with 1% osmium tetroxide in 0.1 M sodium cacodylate buffer, dehydrated in a graded ethanol series, and embedded in LX 112 resin. Thin sections of the embedded cells were cut using an ultramicrotome and stained with 5% uranyl acetate and 0.3% lead citrate. Sections were examined with a Hitachi HT7700 electron microscope at the Biological Electron Microscope Facility, located at the University of Hawai'i at Mānoa.

4.6. Virus purification and buoyant density measurement

Viruses were produced by challenging 40 L of dense, healthy *Tetraselmis* culture with a 1.5% volume addition of previously produced lysate. After the majority of the cells had lysed (ca. 3 days), the remaining cells were removed by filtration through 142 mm diameter, 0.45 µm pore size Millipore Durapore filters (P/N: HVLP14250), overlaid with 125 mm diameter GF/C filters (Whatman P/N: 1822-125). The filtrate was concentrated to approximately 350 mL by TFF (system and filters described under *Virus Isolation*). Additional cell debris was removed by centrifuging the concentrate at 4000 RCF for 30 min and discarding the pellet. The sample was further concentrated by polyethylene glycol (PEG) precipitation (Lawrence and Steward, 2010). The PEG-concentrated viruses were purified on cesium chloride equilibrium buoyant density gradients, using an initial step gradient (SW 41 Ti rotor; 40,000 RPM for 2.5 h), followed by 3 consecutive continuous gradients (SW 41 Ti rotor; 30,000 RPM for ≥ 40 h; Lawrence and Steward, 2010). Gradient fractions were collected by either puncturing the tube with a needle or sipping from the top of the tube using a piston fractionator (Gradient Station, BioComp Instruments, Inc.). The virus-containing fractions were identified by a combination of epifluorescence microscopy (Suttle and Fuhrman, 2010) and fluorometric DNA measurements using the Quant-iT dsDNA high sensitivity assay kit (ThermoFisher Scientific). The buoyant density of TetV virions was determined by measuring the density of the peak viral fraction of the final continuous gradient using a positive displacement pipette (Lawrence and Steward, 2010).

4.7. Genome sequencing

TetV genomic DNA was extracted from the CsCl gradient-purified viral peak. The sample was first buffer exchanged three times with 500 µL TE buffer (100 mM Tris, 10 mM EDTA, pH 8.0) using an Amicon Ultra-0.5 30 kDa NMWL centrifugal ultrafilter (Millipore P/N: UFC503096), followed by DNA extraction with hot SDS and proteinase K and purification by selective precipitation (Masterpure DNA Purification Kit, Epicentre). Illumina sequencing was performed at the Georgia Genomics Facility, using Nextera XT library preparation and 250 bp paired-end sequencing on the MiSeq platform (2,305,807 paired reads). PacBio sequencing (P6-C4 chemistry) was performed by the University of Washington PacBio Sequencing Services. For the PacBio library preparation and sequencing, TetV genomic DNA was included in a pooled sample containing other distantly related viruses, and the TetV reads were subsequently identified and extracted using high-similarity BLAST searches against the TetV Illumina assembly. The TetV-specific PacBio dataset represented 7836 total reads, with a median length of 18,374 bp.

4.8. Genome assembly

De novo assembly of the TetV genome was attempted using both a hybrid strategy (SPAdes v3.6.2; Nurk et al., 2013) that utilized Illumina and PacBio data, as well as a PacBio-only approach (Canu v1.0; Berlin et al., 2015). Both assemblies produced a single scaffold representing the TetV genome. However, a closer comparison of the scaffolds revealed that the PacBio-only strategy preserved a higher number of tandem repeats in repetitive regions of the genome, suggesting that the shorter Illumina reads were effectively forcing these repeats to collapse in the hybrid assembly. We assumed the PacBio-only Canu assembly (95X mean coverage) better represented the true TetV genome sequence and proceeded with this.

The scaffold from the Canu assembly contained exact direct repeats corresponding to each end, suggesting that the genome is circular. The assembly was trimmed to represent a single, non-redundant copy, and the circularity of the genome was verified by mapping PacBio reads to the 100 kb partial genomic sequence that straddled both ends of the contig (50 kb from each end; reads mapped using BLASR v1.3.1; Chaisson and Tesler, 2012). The PacBio read coverage spanning the endpoints of the contig was similar to that of the rest of the sequence, thus validating the circularity of the genome sequence. The assembly was polished using a combination of pbalign v0.2.0.141024 and Quiver v2.0.0 (Koren et al., 2012, 2013; Berlin et al., 2015). The Quiver-polished assembly shared 100% sequence identity with the consensus of mapped Illumina reads (approx. 500X coverage), suggesting that the Quiver-polished *de novo* Canu assembly was high quality.

4.9. Gene prediction

Initial gene prediction was performed using Prodigal v2.6.3 (translation table 1; Hyatt et al., 2010), followed by additional steps to identify missed genes. Putative missed genes were identified by querying a larger set of potential open reading frames (ORFs; i.e., all potential genes found by Prodigal; option -s) against NCBI's Conserved Domain Database (CDD; default settings, E-value < 0.01; Marchler-Bauer et al., 2015). ORFs with high-scoring CDD hits that were also distinct from the default-predicted Prodigal ORFs were further analyzed with searches against pfam (http://pfam.xfam.org/; Finn et al., 2016), InterPro (https://www.ebi.ac.uk/interpro/; Jones et al., 2014), and the NCBI nr database (blastp, default settings; http://blast.ncbi.nlm.nih.gov/; Camacho et al., 2009). If an ORF was supported with high

similarity to known proteins, and if it did not overlap a default-predicted ORF having stronger evidence for protein similarity, then it was added to the annotations. Any default-predicted ORFs sharing significant overlap with the new ORF were removed. In this way, two new ORFs containing high-scoring protein domain hits were identified (TetV_107 and TetV_370), and one default-predicted ORF (TetV_630) was extended to include additional protein domains by moving its start position further upstream. tRNA genes were identified using the tRNAscan-SE v1.21 webserver (http://lowelab.ucsc.edu/tRNAscan-SE/; Schattner et al., 2005) with the default settings. Base pair numbering was initiated at the start of the first gene encountered on the original, linear genome assembly and, because of TetV's circular genome, does not hold any significance. The strand orientation was also arbitrarily chosen, based on the original genome assembly orientation.

4.10. Genome annotation

The annotation of each ORF was performed manually, after careful evaluation of protein similarity search results from the blastp (nr), CDD, pfam, and InterPro webservers, all using the default settings.

4.11. BLAST protein similarity analyses

Protein BLAST (blastp) searches were computed locally using a combined database that included the April 4, 2016 issue of the NCBI non-redundant (nr) database and peptide sequences from the Gordon and Betty Moore Foundation's Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP; Keeling et al., 2014), downloaded from iMicrobe (http://data.imicrobe.us/). All searches were performed using an E-value cutoff of 10^{-5}. Taxonomic information associated with the top blastp hits were used to create the standard pie chart and hierarchical pie chart in Fig. 3 (KronaTools v2.5; Ondov et al., 2011). In cases where the top hit belonged to an uncultivated organism, the next best hit to a cultivated organism was used instead. Only hits to cultivated organisms were included in the analysis.

4.12. Analysis of COG functional distributions

The functional distribution of TetV proteins, as well as the proteins of related mimivirids (AaV, PgV-16T, CeV, CroV, APMV, and Megavirus chilensis), were analyzed using the Clusters of Orthologous Groups of proteins (COGs) framework (Galperin et al., 2015). For consistency, new COG predictions were performed for all viruses included in the analysis, despite the data having been previously published for some of the viruses. The proteins for each genome were searched against the COG v1.0 database using the Batch Web CD-Search Tool (E-value < 0.01; http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi; Marchler-Bauer et al., 2015). For each protein, the highest scoring COG hit was matched to its assigned function(s), and distributions were created. Some COGs were associated with multiple functions, in which case each function was counted once.

4.13. Analysis of conserved NCLDV genes

The presence, absence, and duplication of conserved NCLDV genes in the TetV genome was investigated by assigning TetV proteins to their corresponding cluster of orthologous genes, using the Nucleo-Cytoplasmic Orthologous Groups (NCVOG) framework established by Yutin *et al.* (Yutin et al., 2009, 2014). This was done using a conservative approach relying on support from both the PSI-BLAST-based COGnitor program (Kristensen et al., 2010) and blastp protein similarity. For each protein, NCVOGs were predicted using COGnitor and compared with results from a blastp search against virus-classified proteins in the nr database (release 4/4/2016; E-value $\leq 10^{-5}$), where the db size was set equal to the original nr database size to allow an equivalent E-value cutoff. The highest-scoring blastp hit belonging to a protein with known NCVOG association (based on the curated 201409 NCVOG release; Yutin et al., 2014) was identified, and its NCVOG was compared with the COGnitor-assigned NCVOG for the same TetV protein. If the NCVOG assignment was the same for both blastp and COGnitor, it was retained. This procedure was also performed for AaV and CeV, since these genomes were not published at the time of the updated NCVOG release. Additionally, proteins from TetV, AaV and CeV that could not be assigned to any of the 201409 NCVOGs were compared against each other using COGtriangles (Kristensen et al., 2010) to identify new orthologous clusters among these three genomes.

4.14. Phylogenetics

All phylogenetic trees were constructed using a combination of Bayesian and maximum-likelihood methods. First, a Bayesian majority consensus tree was inferred using MrBayes v3.2.6 (Ronquist et al., 2012) with two runs of 4–8 chains, until the average standard deviation of split frequencies dropped below 0.01. Then, maximum-likelihood bootstrap values were generated for the Bayesian tree using RAxML v8.2.8 (Stamatakis, 2014) implemented with 1000 iterations of rapid bootstrapping. The specific procedures for preparing the sequence alignments varied by gene and are outlined below.

Sequence homologs used in the phylogenies were identified and retrieved based on BLAST similarity to the corresponding TetV virus/host sequence and/or published annotations. In the case of pyruvate formate-lyase (PFL), pyruvate formate-lyase activating enzyme (PFL-AE), MIEL1-like E3 ubiquitin ligase, and E2 ubiquitin conjugating enzyme, the blastp search database included proteins from both nr and the MMETSP. The DNA-dependent DNA polymerase family B (PolB) alignment was created using the MAFFT version 7 webserver (http://mafft.cbrc.jp/alignment/server/; E-INS-i strategy and inclusion of up to 200 homologs under the mafft-homologs option), and all other alignments were created using MAFFT v7.273 (Katoh and Standley, 2013; G-INS-i strategy for 18S rDNA, after trimming end sites that lacked data for any of the taxa; L-INS-i strategy for MutS7, MCP1, and A32-like virion packaging ATPase; E-INS-i strategy for PFL, PFL-AE, MIEL1-like E3 ubiquitin ligase, and E2 ubiquitin-conjugating enzyme). The PolB alignment was trimmed by removing sites with greater than 25% gaps, and all other protein alignments were trimmed using trimAL v1.4 (strict option; Capella-Gutierrez et al., 2009). Nucleotide and protein alignments were analyzed by jModelTest v2.1.10 (Darriba et al., 2012) and ProtTest v3.4 (Abascal et al., 2005), respectively, to select the best-fit model of nucleotide/protein substitution, prior to phylogenetic analysis.

See Supplementary Dataset S2 for a complete list of accession numbers for the sequences used in the phylogenetic reconstructions.

4.15. Identification of repetitive genetic elements and domains

Exact direct and inverted repeats greater than 20 bp in length were identified in the TetV genome using MUMmer v4.0 (Kurtz et al., 2004). For identification of fibronectin type III-like (FN3-like) domains, a search was performed using an HMM profile built from all FN3 domains (IPR003961) identified in TetV proteins by InterProScan v5.19–58. The InterProScan-identified domains were aligned in MAFFT v7.273 (L-INS-i strategy), and an HMM profile was created from the alignment using hmmbuild in HMMER v3.1b2 (http://www.hmmer.org). hmmsearch was used to identify the TetV FN3-like HMM in all TetV proteins, using a domain conditional E-value cutoff of 0.00001. The domain envelope coordinates were used in calculating the total nucleotides represented by the FN3-like domains.

4.16. Gene duplication identification

An investigation of potential gene duplications in the TetV genome was performed using a combination of blastp protein similarity, manual

inspections of protein alignments, and phylogenetics. First, TetV proteins were searched against a combined database of nr (Dec. 4, 2017), MMETSP transcriptomes, and all TetV proteins (blastp, e-value cutoff 0.00001). TetV queries with higher scoring hits to other TetV proteins than to non-TetV proteins in nr and MMETSP were identified. Some queries had high scoring hits to multiple TetV proteins, over non-TetV proteins. These groups of blastp-linked proteins were compared to each other, and instances of overlapping groups were merged together. Alignments of the groups were constructed (MAFFT v7.273; strategy automatically determined) and manually inspected. Where necessary, sequences lacking highly conserved residues were removed from groups, and in some cases these sequences were removed to create new groups. In cases where pairs of blastp-linked proteins were not reciprocal best hits (RBHs), a maximum-likelihood phylogenetic analysis (RAxML v8.2.6; automatic selection of protein substitution model; 1000 iterations of rapid bootstrapping) was performed on an alignment (MAFFT v7.273; strategy automatically determined) of the TetV pair that included up to 5 non-TetV top hits from the non-reciprocating protein, and pairs of non-RBHs that clustered together phylogenetically were retained as putative paralogs. Percent identity values for ungapped pairwise sequence alignments were calculated for the sequences of each group of putative paralogs, using alignments constructed with MAFFT v7.273 (strategy automatically determined).

4.17. Pulsed-field gel electrophoresis of restriction endonuclease-digested TetV genomic DNA

Pulsed-field gel electrophoresis of restriction endonuclease (RE)-digested TetV genomic DNA was used to provide an independent estimate of genome size, as well as to confirm the circularity of the genome (as predicted by the genome assembly). For this purpose, an additional sample of CsCl gradient-purified TetV was prepared from 10 L of lysate, using methods similar to those described above. Two RE enzymes were selected for the analysis: NotI (New England BioLabs [NEB] P/N: R0189S; one predicted digestion site) and SfiI (NEB P/N: R0123L; 10 predicted digestion sites). If the TetV genome exists in the virion as a circular molecule, then digestion with NotI and SfiI should yield one and ten linear segments, respectively, that correspond to *in silico*-predicted sizes. Alternatively, digestion of a linear molecule would result in two and eleven linear segments, respectively, and the size of one of the predicted segments would be inaccurately estimated.

TetV genomic DNA was extracted in agarose plugs (Steward and Culley, 2010; Sandaa et al., 2010), and RE digestions were performed according to protocols outlined in NEB product literature. Approximately 500 ng of genomic DNA was used for each NotI digestion and control sample and 1 μg for each SfiI digestion and control sample. Control plugs were prepared using the same buffer and temperature treatments as the digested samples, but excluding the addition of RE enzyme during the digestion step.

Pulsed-field gel electrophoresis was performed on the Bio-Rad CHEF-DR III System, using 1% agarose gels in 0.5X TBE running buffer. The NotI digest and control samples were run alongside the NEB Yeast Chromosome PFG Marker (NEB P/N: N0345S) using the following settings: 6.0 V/cm, 15 °C for 26 h; switch times of 70 s for 15 h and 120 s for 11 h; 120° included angle. The SfiI digest and control samples were run alongside the NEB MidRange I PFG Marker (NEB P/N: N3551S) using the following settings: 6.0 V/cm, 15 °C for 24 h; switch times ramped from 1 to 25 s; 120° included angle. Both gels were stained in 1X SYBR Gold in 0.5X TBE at room temperature for 40 min in the dark. Images were taken using the KODAK Gel Logic 200 System. Standard curves of relative distance migrated vs. DNA size were generated by measuring distances from the well to the leading edge of the bands on an enlargement of the gel image, correcting for uneven migration rate across the gel. For the first gel, the standard curve was linear and used six of the standards (three larger and three smaller than the TetV band). For the second gel, a square root transformation of band size was used to linearize the standard curve, and the standards spanning the visible TetV fragments were used (15–194 kb). The sizes of TetV genome (NotI digest) and genome fragments (SfiI digest) and the associated error (± 95% C. I.) were calculated by inverse prediction from the linear standard curves (Zar, 1996).

Acknowledgements

We would like to thank Tina M. Weatherby and Marilyn F. Dunlap at the University of Hawai'i (UH) at Mānoa Biological Electron Microscope Facility for their assistance and guidance with electron microscopy procedures, Jaclyn A. Miranda for assistance with culture maintenance, and Katharine A. Smith for assistance with figure creation. We also thank Chief Scientist Paul Lethaby, the Hawaii Ocean Time-Series team, and the crew of the R/V Ka'imikai-O-Kanaloa cruise KOK-1011. For the bioinformatics and phylogenetic analyses, the technical support and advanced computing resources from the UH Information Technology Services' Cyber infrastructure group are gratefully acknowledged.

Funding sources

This work was supported by the National Science Foundation (grant numbers EF 04-24599, OCE 09-26766, PLR 09-44851, DBI 10-40548, OCE 15-59356) and the Denise B. Evans Fellowship in Oceanographic Research awarded to C. R. Schvarcz by the Hawai'i Institute of Geophysics and Planetology at the University of Hawai'i at Mānoa.

Appendix A. Supplementary information

Supplementary information associated with this article can be found in the online version at doi:10.1016/j.virol.2018.03.010.

References

Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics 21 (April (9)), 2104–2105.
Abergel, C., Chenivesse, S., Byrne, D., Suhre, K., Arondel, V., Claverie, J.-M., 2005. Mimivirus TyrRS: preliminary structural and functional characterization of the first amino-acyl tRNA synthetasesynthetase found in a virus. Sect. F. Struct. Biol. Cryst. Commun. 61 (February (Pt 2)), 212–215.
Abergel, C., Legendre, M., Claverie, J.-M., 2015. The rapidly expanding universe of giant viruses: Mimivirus, Pandoravirus, Pithovirus and Mollivirus. FEMS Microbiol. Rev. 39 (October (6)), 779–796.
Antwerpen, M.H., Georgi, E., Zoeller, L., Woelfel, R., Stoecker, K., Scheid, P., 2015. Whole-genome sequencing of a pandoravirus isolated from keratitis-inducing acanthamoeba. Genome Announc 3 (January (2)) (e00136–15).
Arslan, D., Legendre, M., Seltzer, V., Abergel, C., Claverie, J.-M., 2011. Distant Mimivirus relative with a larger genome highlights the fundamental features of Megaviridae. Proc. Natl. Acad. Sci. USA 108 (October (42)), 17486–17491.
Becker, B., HARD, K., Melkonian, M., Kamerling, J.P., Vliegenthart, J.F.G., 1989. Identification of 3-deoxy-manno-2-octulosonic acid, 3-deoxy-5-O-methyl-manno-2-octulosonic acid and 3-deoxy-lyxo-2-heptulosaric acid in the cell wall (theca) of the green alga Tetraselmis striata Butcher (Prasinophyceae). Eur. J. Biochem. 182 (June (1)), 153–160.
Becker, B., Becker, D., Kamerling, J.P., Melkonian, M., 1991. 2-Keto-sugar acids in green flagellates: a chemical marker for prasinophycean scales. J. Phycol. 27 (August (4)), 498–504.
Berlin, K., Koren, S., Chin, C.-S., Drake, J.P., Landolin, J.M., Phillippy, A.M., 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat. Biotechnol. 33 (May (6)), 623–630.
Bork, P., Bowler, C., de Vargas, C., Gorsky, G., Karsenti, E., Wincker, P., 2015. Tara Oceans studies plankton at planetary scale. Science 348 (May (6237)), 873.
Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doulcier, G., Acinas, S.G., Alberti, A., Chaffron, S., Cruaud, C., de Vargas, C., Gasol, J.M., Gorsky, G., Gregory, A.C., Guidi, L., Hingamp, P., Iudicone, D., Not, F., Ogata, H., Pesant, S., Poulos, B.T., Schwenck, S.M., Speich, S., Dimier, C., Kandels-Lewis, S., Picheral, M., Searson, S., Tara Oceans Coordinators, Bork, P., Bowler, C., Sunagawa, S., Wincker, P., Karsenti, E., Sullivan, M.B., 2015. Patterns and ecological drivers of ocean viral communities. Science 348 (May (6237)), 1261498.
Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinforma 10 (1), 421.
Campbell, I.D., Spitzfaden, C., 1994. Building proteins with fibronectin type III modules. Structure 2 (5), 333–337.
Capella-Gutierrez, S., Silla-Martinez, J.M., Gabaldon, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25

(July (15)), 1972–1973.
Catalanotti, C.P., Yang, W., Posewitz, M.C., Grossman, A.R., 2013. Fermentation metabolism and its evolution in algae. Front. Plant Sci. 4 (May), 150.
Chaisson, M.J., Tesler, G., 2012. Mapping single molecule sequencing reads using basic local alignment with successive refinement (BLASR): application and theory. BMC Bioinforma. 13 (1), 238.
Collantes, G., Prado, R., 2006. Green bloom of Tetraselmis sp. in Valparaiso Bay. Harmful Algae News 30, 7.
Daoudi, M., Serve, L., Rharbi, N., El Madani, F., Vouve, F., 2013. Phytoplankton distribution in the Nador lagoon (Morocco) and possible risks for harmful algal blooms. Water Bull. 6 (1), 4–19.
Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9 (8), 772.
der Staay, S.Y.M.-v., De Wachter, R., Vaulot, D., 2001. Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. Nature 409 (6820), 607–610.
Doane, F.W., Anderson, N., 1987. Electron Microscopy in Diagnostic Virology: A Practical Guide and Atlas. Cambridge University Press, New York.
Dodds, J.A., Cole, A., 1980. Microscopy and biology of Uronema gigas, a filamentous eucaryotic green alga, and its associated tailed virus-like particle. Virology 100 (1), 156–165.
Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44 (January (D1)), D279–D285.
Fischer, M.G., Allen, M.J., Wilson, W.H., Suttle, C.A., 2010. Giant virus with a remarkable complement of genes infects marine zooplankton. Proc. Natl. Acad. Sci. USA 107 (November (45)), 19508–19513.
Fischer, M.G., 2016. Giant viruses come of age. Curr. Opin. Microbiol. 31 (June), 50–57.
Fitzgerald, L.A., Graves, M.V., Li, X., Feldblyum, T., Hartigan, J., Van Etten, J.L., 2007a. Sequence and annotation of the 314-kb MT325 and the 321-kb FR483 viruses that infect Chlorella Pbi. Virology 358 (February (2)), 459–471.
Fitzgerald, L.A., Graves, M.V., Li, X., Feldblyum, T., Nierman, W.C., Van Etten, J.L., 2007b. Sequence and annotation of the 369-kb NY-2A and the 345-kb AR158 viruses that infect Chlorella NC64A. Virology 358 (February (2)), 472–484.
Gallot-Lavallée, L., Pagarete, A., Legendre, M., Santini, S., Sandaa, R.-A., Himmelbauer, H., Ogata, H., Bratbak, G., Claverie, J.-M., 2015. The 474-Kilobase-Pair Complete Genome Sequence of CeV-01B, a Virus Infecting Haptolina (Chrysochromulina) ericina (Prymnesiophyceae). Genome Announc 3 (December (6)) (e01413–15).
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 (June (14)) (e00230–17).
Galperin, M.Y., Makarova, K.S., Wolf, Y.I., Koonin, E.V., 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43 (January (D1)), D261–D269.
Groisillier, A., Labourel, A., Michel, G., Tonon, T., 2015. The mannitol utilization system of the marine bacterium Zobellia galactanivorans. Appl. Environ. Microbiol. 81 (February (5)), 1799–1812.
Guillard, R.R.L., 1975. Culture of phytoplankton for feeding marine invertebrates. In: Culture of Marine Invertebrate Animals. Springer US, Boston, MA, pp. 29–60.
Haglund, K., Dikic, I., 2005. Ubiquitylation and cell signaling. EMBO J. 24 (19), 3353–3359.
Hemaiswarya, S., Raja, R., Kumar, R.R., Ganesan, V., Anbazhagan, C., 2010. Microalgae: a sustainable feed source for aquaculture. World J. Microbiol. Biotechnol. 27 (8), 1737–1746.
Hyatt, D., Chen, G.-L., LoCascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinforma 11 (March (1)), 1.
Iyer, L.M., Balaji, S., Koonin, E.V., Aravind, L., 2006. Evolutionary genomics of nucleocytoplasmic large DNA viruses. Virus Res. 117 (April (1)), 156–184.
Ji, C., Cao, X., Yao, C., Xue, S., Xiu, Z., 2014. Protein-protein interaction network of the marine microalga Tetraselmis subcordiformis: prediction and application for starch metabolism analysis. J. Ind. Microbiol. Biotechnol. 41 (8), 1287–1296.
Johannessen, T.V., Bratbak, G., Larsen, A., Ogata, H., Egge, E.S., Edvardsen, B., Eikrem, W., Sandaa, R.-A., 2015. Characterisation of three novel giant viruses reveals huge diversity among viruses infecting Prymnesiales (Haptophyta). Virology 476 (February), 180–188.
Jones, J.B., Rhodes, L.L., 1994. Suffocation of pilchards (Sardinops sagax) by a green microalgal bloom in Wellington Harbour, New Zealand. New Zeal. J. Mar. Fresh 28 (December (4)), 379–383.
Jones, P., Binns, D., Chang, H.Y., Fraser, M., Li, W., McAnulla, C., McWilliam, H., Maslen, J., Mitchell, A., Nuka, G., Pesseat, S., Quinn, A.F., Sangrador-Vegas, A., Scheremetjew, M., Yong, S.Y., Lopez, R., Hunter, S., 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30 (April (9)), 1236–1240.
Karl, D.M., Lukas, R., 1996. The Hawaii Ocean Time-series (HOT) program: background, rationale and field implementation. Deep Sea Res. Part 2 Top. Stud. Oceanogr. 43 (2–3), 129–156.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30 (March (4)), 772–780.
Keeling, P.J., Burki, F., Wilcox, H.M., Allam, B., Allen, E.E., Amaral-Zettler, L.A., Armbrust, E.V., Archibald, J.M., Bharti, A.K., Bell, C.J., Beszteri, B., Bidle, K.D., Cameron, C.T., Campbell, L., Caron, D.A., Cattolico, R.A., Collier, J.L., Coyne, K., Davy, S.K., Deschamps, P., Dyhrman, S.T., Edvardsen, B., Gates, R.D., Gobler, C.J., Greenwood, S.J., Guida, S.M., Jacobi, J.L., Jakobsen, K.S., James, E.R., Jenkins, B., John, U., Johnson, M.D., Juhl, A.R., Kamp, A., Katz, L.A., Kiene, R., Kudryavtsev, A., Leander, B.S., Lin, S., Lovejoy, C., Lynn, D., Marchetti, A., McManus, G., Nedelcu, A.M., Menden-Deuer, S., Miceli, C., Mock, T., Montresor, M., Moran, M.A., Murray, S., Nadathur, G., Nagai, S., Ngam, P.B., Palenik, B., Pawlowski, J., Petroni, G., Piganeau, G., Posewitz, M.C., Rengefors, K., Romano, G., Rumpho, M.E., Rynearson, T., Schilling, K.B., Schroeder, D.C., Simpson, A.G.B., Slamovits, C.H., Smith, D.R., Smith, G.J., Smith, S.R., Sosik, H.M., Stief, P., Theriot, E., Twary, S.N., Umale, P.E., Vaulot, D., Wawrik, B., Wheeler, G.L., Wilson, W.H., Xu, Y., Zingone, A., Worden, A.Z., 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 (June (6)), e1001889.
Keller, M.D., Selvin, R.C., Claus, W., Guillard, R.R.L., 2007. Media for the culture of oceanic ultraphytoplankton. J. Phycol. 23 (April (4)), 633–638.
Kirst, G.O., 1989. Salinity tolerance of eukaryotic marine algae. Annu. Rev. Plant. Phys. 40, 21–53.
Kobe, B., 2001. The leucine-rich repeat as a protein recognition motif. Curr. Opin. Struct. Biol. 11 (6), 725–732.
Konstantinidis, K.T., Braff, J., Karl, D.M., DeLong, E.F., 2009. Comparative metagenomic analysis of a microbial community residing at a depth of 4000 meters at Station ALOHA in the North Pacific Subtropical Gyre. Appl. Environ. Microbiol. 75 (August (16)), 5345–5355.
Koonin, E.V., Dolja, V.V., 2014. Virus world as an evolutionary network of viruses and capsidless selfish elements. Microbiol. Mol. Biol. Rev. 78 (June (2)), 278–303.
Koren, S., Schatz, M.C., Walenz, B.P., Martin, J., Howard, J.T., Ganapathy, G., Wang, Z., Rasko, D.A., McCombie, W.R., Jarvis, E.D., Phillippy, A.M., 2012. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat. Biotechnol. 30 (July (7)), 693–700.
Koren, S., Harhay, G.P., Smith, T.P., Bono, J.L., Harhay, D.M., Mcvey, S.D., Radune, D., Bergman, N.H., Phillippy, A.M., 2013. Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol. 14 (9), R101.
Kristensen, D.M., Kannan, L., Coleman, M.K., Wolf, Y.I., Sorokin, A., Koonin, E.V., Mushegian, A., 2010. A low-polynomial algorithm for assembling clusters of orthologous groups from intergenomic symmetric best matches. Bioinformatics 26 (June (12)), 1481–1487.
Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. Genome Biol. 5 (2), R12.
Lawrence, J.E., Steward, G.F., 2010. Purification of viruses by centrifugation. In: Wilhelm, S.W., Weinbauer, M.G., Suttle, C.A. (Eds.), Manual of Aquatic Viral Ecology. ASLO, May pp. 166–181.
Legendre, M., Bartoli, J., Shmakova, L., Jeudy, S., Labadie, K., Adrait, A., Lescot, M., Poirot, O., Bertaux, L., Bruley, C., Coute, Y., Rivkina, E., Abergel, C., Claverie, J.M., 2014. Thirty-thousand-year-old distant relative of giant icosahedral DNA viruses with a pandoravirus morphology. Proc. Natl. Acad. Sci. USA 111 (March (11)), 4274–4279.
Legendre, M., Lartigue, A., Bertaux, L., Jeudy, S., Bartoli, J., Lescot, M., Alempic, J.-M., Ramus, C., Bruley, C., Labadie, K., Shmakova, L., Rivkina, E., Couté, Y., Abergel, C., Claverie, J.-M., 2015. In-depth study of Mollivirus sibericum, a new 30,000-y-old giant virus infecting Acanthamoeba. Proc. Natl. Acad. Sci. USA 112 (September (38)), E5327–E5335.
Levene, S.D., Zimm, B.H., 1987. Separations of open-circular DNA using pulsed-field electrophoresis. Proc. Natl. Acad. Sci. USA 84 (12), 4054–4057.
Mann, N.H., Cook, A., Millard, A., Bailey, S., Clokie, M., 2003. Marine ecosystems: bacterial photosynthesis genes in a virus. Nature 424 (6950), 741.
Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., Lanczycki, C.J., Lu, F., Marchler, G.H., Song, J.S., Thanki, N., Wang, Z., Yamashita, R.A., Zhang, D., Zheng, C., Bryant, S.H., 2015. CDD: NCBI's conserved domain database. Nucleic Acids Res 43 (January (D1)), D222–D226.
Moniruzzaman, M., LeCleir, G.R., Brown, C.M., Gobler, C.J., Bidle, K.D., Wilson, W.H., Wilhelm, S.W., 2014. Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host-virus coevolution. Virology 466–467 (October), 60–70.
Moreau, H., Piganeau, G., Desdevises, Y., Cooke, R., Derelle, E., Grimsley, N., 2010. Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. J. Virol. 84 (December (24)), 12555–12563.
Nagasaki, K., Bratbak, G., 2010. Isolation of viruses infecting photosynthetic and non-photosynthetic protists. In: Wilhelm, S. W., Weinbauer, M. G., Suttle, C. A. (Eds.), Manual of Aquatic Viral Ecology. ASLO, May, pp. 92–101.
Nurk, S., Bankevich, A., Antipov, D., Gurevich, A.A., Korobeynikov, A., Lapidus, A., Prjibelski, A.D., Pyshkin, A., Sirotkin, A., Sirotkin, Y., Stepanauskas, R., Clingenpeel, S.R., Woyke, T., McLean, J.S., Lasken, R., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J. Comp. Biol. 20 (October (10)), 714–737.
O'Day, D.H., Suhre, K., Myre, M.A., Chatterjee-Chakraborty, M., Chavez, S.E., 2006. Isolation, characterization, and bioinformatic analysis of calmodulin-binding protein cmbB reveals a novel tandem IP22 repeat common to many Dictyostelium and Mimivirus proteins. Biochem. Bioph. Res. Co. 346 (August (3)), 879–888.
Ogata, H., Ray, J., Toyoda, K., Sandaa, R.-A., Nagasaki, K., Bratbak, G., Claverie, J.-M., 2011. Two new subfamilies of DNA mismatch repair proteins (MutS) specifically abundant in the marine environment. ISME J. 5 (January (7)), 1143–1151.
Ondov, B.D., Bergman, N.H., Phillippy, A.M., 2011. Interactive metagenomic visualization in a web browser. BMC Bioinforma 12 (1), 385.
Paez-Espino, D., Eloe-Fadrosh, E.A., Pavlopoulos, G.A., Thomas, A.D., Huntemann, M., Mikhailova, N., Rubin, E., Ivanova, N.N., Kyrpides, N.C., 2016. Uncovering Earth's virome. Nature 536 (August (7617)), 425–430.

Philippe, N., Legendre, M., Doutre, G., Couté, Y., Poirot, O., Lescot, M., Arslan, D., Seltzer, V., Bertaux, L., Bruley, C., Garin, J., Claverie, J.-M., Abergel, C., 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341 (July (6143)), 281–286.

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 (November (5700)), 1344–1350.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Hohna, S., Larget, B., Liu, L., Suchard, M.A., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61 (April (3)), 539–542.

Rosenwasser, S., Ziv, C., Creveld, S.G.v., Vardi, A., 2016. Virocell Metabolism: Metabolic Innovations During Host–Virus Interactions in the Ocean. Trends. Microbiol. 24 (10), Oct., 821–832.

Rousvoal, S., Groisillier, A., Dittami, S.M., Michel, G., Boyen, C., Tonon, T., 2010. Mannitol-1-phosphate dehydrogenase activity in Ectocarpus siliculosus, a key role for mannitol synthesis in brown algae. Planta 233 (October (2)), 261–273.

Rusch, D.B., Halpern, A.L., Sutton, G., Heidelberg, K.B., Williamson, S., Yooseph, S., Wu, D., Eisen, J.A., Hoffman, J.M., Remington, K., Beeson, K., Tran, B., Smith, H., Baden-Tillson, H., Stewart, C., Thorpe, J., Freeman, J., Andrews-Pfannkoch, C., Venter, J.E., Li, K., Kravitz, S., Heidelberg, J.F., Utterback, T., Rogers, Y.-H., Falcón, L.I., Souza, V., Bonilla-Rosso, G., Eguiarte, L.E., Karl, D.M., Sathyendranath, S., Platt, T., Bermingham, E., Gallardo, V., Tamayo-Castillo, G., Ferrari, M.R., Strausberg, R.L., Nealson, K., Friedman, R., Frazier, M., Venter, J.C., 2007. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. Plos Biol. 5 (March (3)), e77.

Sandaa, R.-A., Heldal, M., Castberg, T., Thyrhaug, R., Bratbak, G., 2001. Isolation and characterization of two viruses with large genome size infecting Chrysochromulina ericina (Prymnesiophyceae) and Pyramimonas orientalis (Prasinophyceae). Virology 290 (November (2)), 272–280.

Sandaa, R.-A., Short, S.M., Schroeder, D.C., 2010. Fingerprinting aquatic virus communities. In: Wilhelm, S. W., Weinbauer, M. G., Suttle, C. A. (Eds.), Manual of Aquatic Viral Ecology. ASLO, May, pp. 9–18.

Santini, S., Jeudy, S., Bartoli, J., Poirot, O., Lescot, M., Abergel, C., Barbe, V., Wommack, K.E., Noordeloos, A.A.M., Brussaard, C.P.D., Claverie, J.-M., 2013. Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. Proc. Natl. Acad. Sci. USA 110 (June (26)), 10800–10805.

Schattner, P., Brooks, A.N., Lowe, T.M., 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33 (Web Server), Jul., W686–W689.

Schulz, F., Yutin, N., Ivanova, N.N., Ortega, D.R., Lee, T.K., Vierheilig, J., Daims, H., Horn, M., Wagner, M., Jensen, G.J., Kyrpides, N.C., Koonin, E.V., Woyke, T., 2017. Giant viruses with an expanded complement of translation system components. Science 356 (April (6333)), 82–85.

Sharon, I., Battchikova, N., Aro, E.-M., Giglione, C., Meinnel, T., Glaser, F., Pinter, R.Y., Breitbart, M., Rohwer, F., Béjà, O., 2011. Comparative metagenomics of microbial traits within oceanic viral communities. ISME J. 5 (February (7)), 1178–1190.

Sicko-Goad, L., Walker, G., 1979. Viroplasm and large virus-like particles in the dinoflagellate Gymnodinium uberrimum. Protoplasma 99 (3), 203–210.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30 (May (9)), 1312–1313.

Steward, G.F., Culley, A.I., 2010. Extraction and purification of nucleic acids from viruses. In: Wilhelm, S.W., Weinbauer, M.G., Suttle, C.A. (Eds.), Manual of Aquatic Viral Ecology. ASLO, pp. 154–165.

Steward, G.F., Culley, A.I., Wood-Charlson, E.M., 2013. Marine viruses. In: Levin, S.A. (Ed.), Encyclopedia of Biodiversity. Elsevier, Waltham, MA, pp. 127–144.

Suttle, C.A., Fuhrman, J.A., 2010. Enumeration of virus particles in aquatic or sediment samples by epifluorescence microscopy. In: Wilhelm, S.W., Weinbauer, M.G., Suttle, C.A. (Eds.), Manual of Aquatic Viral Ecology. ASLO, pp. 145–153 May.

Throndsen, J., Zingone, A., 1988. Tetraselmis wettsteinii (Schiller) Throndsen comb. nov. and its occurrence in golfo di Napoli. G. Bot. Ital. 122 (January (3–4)), 227–235.

Wilkins, D., Lauro, F.M., Williams, T.J., DeMaere, M.Z., Brown, M.V., Hoffman, J.M., Andrews-Pfannkoch, C., Mcquaid, J.B., Riddle, M.J., Rintoul, S.R., Cavicchioli, R., 2012. Biogeographic partitioning of Southern Ocean microorganisms revealed by metagenomics. Environ. Microbiol. 15 (November (5)), 1318–1333.

Wilson, W.H., Gilg, I.C., Moniruzzaman, M., Field, E.K., Koren, S., LeCleir, G.R., Martínez Martínez, J., Poulton, N.J., Swan, B.K., Stepanauskas, R., Wilhelm, S.W., 2017. Genomic exploration of individual giant ocean viruses. ISME J. 11 (May (8)), 1736–1745.

Worden, A.Z., 2006. Picoeukaryote diversity in coastal waters of the Pacific Ocean. Aquat. Microb. Ecol. 43 (2), 165–175.

Yao, C., Ai, J., Cao, X., Xue, S., Zhang, W., 2012. Enhancing starch production of a marine green microalga Tetraselmis subcordiformis through nutrient limitation. Bioresour. Technol. 118, 438–444.

Yoosuf, N., Yutin, N., Colson, P., Shabalina, S.A., Pagnier, I., Robert, C., Azza, S., Klose, T., Wong, J., Rossmann, M.G., La Scola, B., Raoult, D., Koonin, E.V., 2012. Related giant viruses in distant locations and different habitats: Acanthamoeba polyphaga moumouvirus represents a third lineage of the Mimiviridae that is close to the megavirus lineage. Genome Biol. Evol. 4 (12), 1324–1330.

Yoosuf, N., Pagnier, I., Fournous, G., Robert, C., Raoult, D., La Scola, B., Colson, P., 2014. Draft genome sequences of Terra1 and Terra2 viruses, new members of the family Mimiviridae isolated from soil. Virology 452–453 (March), 125–132.

Yutin, N., Wolf, Y.I., Raoult, D., Koonin, E.V., 2009. Eukaryotic large nucleo-cytoplasmic DNA viruses: clusters of orthologous genes and reconstruction of viral genome evolution. Virol. J. 6 (1), 223.

Yutin, N., Wolf, Y.I., Koonin, E.V., 2014. Origin of giant viruses from smaller DNA viruses not from a fourth domain of cellular life. Virology 466–467 (October), 38–52.

Zar, J., 1996. Biostatistical Analysis, 3rd edition. Prentice Hall, Upper Saddle River, NJ.

Zheng, Y., Chen, Z., Lu, H., Zhang, W., 2011. Optimization of carbon dioxide fixation and starch accumulation by Tetraselmis subcordiformis in a rectangular airlift photobioreactor. Afr. J. Biotechnol. 10 (March (10)), 1888–1901.