



OPEN ACCESS

EDITED BY

Sead Sabanadzovic,
Mississippi State University,
United States

REVIEWED BY

Jonatas Abrahao,
Federal University of Minas Gerais, Brazil
Felipe Hernandes Coutinho,
Spanish National Research Council (CSIC),
Spain

*CORRESPONDENCE

Frank O. Aylward
faylward@vt.edu

SPECIALTY SECTION

This article was submitted to
Virology,
a section of the journal
Frontiers in Microbiology

RECEIVED 17 August 2022

ACCEPTED 25 October 2022

PUBLISHED 25 November 2022

CITATION

Farzad R, Ha AD and Aylward FO (2022)
Diversity and genomics of giant viruses in
the North Pacific Subtropical Gyre.
Front. Microbiol. 13:1021923.
doi: 10.3389/fmicb.2022.1021923

COPYRIGHT

© 2022 Farzad, Ha and Aylward. This is an
open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that
the original publication in this journal is
cited, in accordance with accepted
academic practice. No use, distribution or
reproduction is permitted which does not
comply with these terms.

Diversity and genomics of giant viruses in the North Pacific Subtropical Gyre

Roxanna Farzad¹, Anh D. Ha¹ and Frank O. Aylward^{1,2*}

¹Department of Biological Sciences, Virginia Tech, Blacksburg, VA, United States, ²Center for Emerging, Zoonotic, and Arthropod-Borne Infectious Disease, Virginia Tech, Blacksburg, VA, United States

Large double-stranded DNA viruses of the phylum *Nucleocytoviricota*, often referred to as “giant viruses,” are ubiquitous members of marine ecosystems that are important agents of mortality for eukaryotic plankton. Although giant viruses are known to be prevalent in marine systems, their activities in oligotrophic ocean waters remain unclear. Oligotrophic gyres constitute the majority of the ocean and assessing viral activities in these regions is therefore critical for understanding overall marine microbial processes. In this study, we generated 11 metagenome-assembled genomes (MAGs) of giant viruses from samples previously collected from Station ALOHA in the North Pacific Subtropical Gyre. Phylogenetic analyses revealed that they belong to the orders *Imitervirales* ($n = 6$), *Algavirales* ($n = 4$), and *Pimascovirales* ($n = 1$). Genome sizes ranged from ~119–574 kbp, and several of the genomes encoded predicted TCA cycle components, cytoskeletal proteins, collagen, rhodopsins, and proteins potentially involved in other cellular processes. Comparison with other marine metagenomes revealed that several have broad distribution across ocean basins and represent abundant viral constituents of pelagic surface waters. Our work sheds light on the diversity of giant viruses present in oligotrophic ocean waters across the globe.

KEYWORDS

giant viruses, marine viruses, station ALOHA, *Prasinoviridae*, *Mesomimiviridae*, *Nucleocytoviricota*

Introduction

Nucleo-cytoplasmic large DNA viruses (NCLDVs, phylum *Nucleocytoviricota*), also known as “giant viruses,” are a lineage of eukaryotic viruses that include many animal and protist pathogens. In addition to several well-known families that infect vertebrates (e.g., *Poxviridae*, *Asfaviridae*, and *Iridoviridae*), several families in this phylum infect a variety of algae and other protists (e.g., *Phycodnaviridae*, *Marseilleviridae*, and *Mimiviridae*; Fischer, 2016; Wilhelm et al., 2016; Koonin and Yutin, 2019; Karki et al., 2021; Aylward and Moniruzzaman, 2022). Appreciation of the environmental prevalence of viruses within the *Nucleocytoviricota* somewhat lagged behind other viral groups because the large capsid sizes of many members of this phylum often precluded their recovery in diversity surveys

that focused on particles that could pass through a 0.2 μm filter. Nevertheless, pioneering studies focusing on marker genes provided early evidence that these large DNA viruses are widespread in the environment (Chen et al., 1996; Short and Suttle, 2002; Short, 2012), and later metagenomic studies revealed an enormous diversity in this group, particularly in marine environments (Yau et al., 2011; Hingamp et al., 2013; Bäckström et al., 2019; Endo et al., 2020; Moniruzzaman et al., 2020a; Schulz et al., 2020). Recent estimates suggest that there are at least 32 different families of giant viruses that reside in diverse ecosystems across the globe, and more will almost certainly be identified in the future (Aylward et al., 2021).

Giant viruses have complex and chimeric genomes that are the product of widespread gene exchange with various cellular lineages (Boyer et al., 2009; Yoshikawa et al., 2019; Moniruzzaman et al., 2020a,b). Besides the core machinery involved in virion structure and DNA replication, giant viruses also commonly encode genes involved in translation, glycolysis, TCA cycle, cytoskeletal dynamics, light-harvesting, nutrient transport, and other pathways involved in nutrient homeostasis (Schulz et al., 2017; Moniruzzaman et al., 2020a). Rhodopsins are also common in a wide range of marine giant viruses (Yutin and Koonin, 2012; Needham et al., 2019; Rozenberg et al., 2020). Rhodopsins are light-driven ion pumps that can be involved in energy production and signal transduction (Ernst et al., 2014; Govorunova et al., 2017). Viral rhodopsins may permit viruses to modify host phototaxis during infection, which may promote their proliferation (Gallot-Lavallée and Archibald, 2020). Proteins involved in cytoskeletal dynamics have also been found to be quite common in a variety of marine giant viruses; viral homologs to actin, myosin, and kinesin genes could potentially benefit viruses by manipulating the host's cytoskeleton by using host motor proteins to traffic virions or maintain the localization of viral machinery during infection (Ha et al., 2021; Kijima et al., 2021; Da Cunha et al., 2022). These recent findings collectively suggest that giant viruses use a broad assortment of functional genes to manipulate host physiology and alter the intracellular environment to promote virion propagation.

Although giant viruses are globally distributed in a variety of habitats, they appear to be particularly diverse and abundant in the ocean (Ghedini and Claverie, 2005; Monier et al., 2008; Endo et al., 2020; Moniruzzaman et al., 2020a). The majority of the ocean is made up of oligotrophic oceanic gyres, and it is therefore of particular interest to examine viral dynamics in these systems. One field site that has been particularly useful for examining microbial diversity in oceanic gyres is Station ALOHA (A Long-term Oligotrophic Habitat Assessment), located at 22°45'N, 158°W, nearly 100 km north of the Hawaiian island of Oahu (Karl and Church, 2014). Several recent studies have recently elucidated a rich diversity of viruses that are present at or near Station ALOHA (Aylward et al., 2017; Luo et al., 2017, 2020, 2022). In this study, we surveyed previously-sequenced metagenomes generated

from Station ALOHA to characterize the diversity of giant viruses in this habitat (Mende et al., 2017). Although metagenomes derived from <0.2 μm size fractions are typically used to evaluate viral diversity in marine systems, recent studies have found that many giant viruses are often found in larger size fractions along with bacteria and archaea (Endo et al., 2020; Moniruzzaman et al., 2020a). We also analyzed the encoded functions in the draft giant virus genomes that we recovered to gain insight into possible mechanisms they employ to manipulate their hosts during infection. Lastly, by examining publicly-available metagenomes from other marine environments we examined the distribution and biogeography of these viruses on a global scale. Because of Station ALOHA's location in the North Pacific Subtropical Gyre, examination of giant viruses found here provides a window into those lineages that are likely broadly distributed in oligotrophic ocean waters and may play important roles in marine ecological dynamics.

Materials and methods

Metagenomes used

We analyzed metagenomes that were generated in a previous study (Mende et al., 2017). This dataset consists of 107 samples from depths ranging from 25 m to 1,000 m that were collected over a 1.5 year sampling period at Station ALOHA on 11 cruises of the Hawaii Ocean Time-series (HOT). The methods used for sample collection and processing have been previously described (Mende et al., 2017). Briefly, water was filtered onto 0.2 μm filters, and after DNA extraction, libraries were created with Illumina TruSeq LT Nano kit, and metagenomes were sequenced using Illumina MiSeq and NextSeq 500 systems.

Generation of metagenome-assembled genomes

MAGs were generated using a workflow developed previously (Aylward and Moniruzzaman, 2021). Briefly, metagenomes were assembled with MegaHit v. 1.2.9 (with parameters `-min-contig-len 5000`), and contigs were subsequently binned using MetaBat2 v. 2.12.1 (with parameters `-s 100000, -m 10000, -minS 75, -maxEdges 75`; Kang et al., 2019). All bins were analyzed with ViralRecall v. 2.0 to identify those that corresponded to giant viruses, and only those that contained 4 of 5 NCLDV marker genes were retained. The marker genes used for this were the A32-like ATPase (A32), B-family DNA polymerase (PolB), virus late transcription factor 3 (VLTF3), major capsid protein (MCP), and superfamily II helicase (SFII). This resulted in 11 NCLDV bins ranging in size from 119,690 to 574,081 bp (Table 1). Genome statistics were compiled with SeqKit v. 2.2.0 (Shen et al., 2016), and we predicted proteins using Prodigal v. 2.6.3 (Hyatt et al., 2010) with default parameters (Table 1).

TABLE 1 General characteristics of 11 metagenome-assembled genomes (MAGs) of giant viruses generated from St. ALOHA.

Genome	Genome length	GC content %	Num of protein coding genes	Order	Family	Genus
HOT_MAG4	122,808	38.47	164	<i>Algavirales</i>	AG_04	g175
HOT_MAG14	147,238	37.49	200	<i>Algavirales</i>	AG_01	g177
HOT_MAG30	171,883	34.21	260	<i>Algavirales</i>	AG_01	g177
HOT_MAG20	119,690	33.82	174	<i>Algavirales</i>	AG_01	g177
HOT_MAG12	386,441	33.29	412	<i>Imitervirales</i>	IM_01	g336
HOT_MAG13	433,885	32.35	475	<i>Imitervirales</i>	IM_01	g342
HOT_MAG10	426,436	29.37	491	<i>Imitervirales</i>	IM_01	g342
HOT_MAG3	574,081	31.33	559	<i>Imitervirales</i>	IM_09	g279
HOT_MAG5	477,804	30.13	484	<i>Imitervirales</i>	IM_09	g279
HOT_MAG60	489,708	24.69	541	<i>Imitervirales</i>	IM_09	g274
HOT_MAG22	471,006	34.86	484	<i>Pimascovirales</i>	PM_01	NA

Phylogenetic construction

In order to explore the phylogenetic placement of the reconstructed MAGs, we used the protein predictions of the 11 MAGs generated in this study together with proteins from all reference giant viruses that have previously been compiled in the Giant Virus Database (Aylward et al., 2021). Subsequently, we used the program *ncldv_markersearch* to generate a concatenated alignment of 7 marker genes, as previously described¹ (Aylward et al., 2021). Briefly, this tool uses HMMER3 to identify 7 conserved marker genes (superfamily II helicase (SFII), virus-like transcription factor (VLTF3), B-family DNA polymerase (PolB), and A32-like ATPase (A32), a DNA-dependant RNA polymerase (RNAP) subunit, transcription elongation factor II-S (TFIIS), and a family II topoisomerase), and then uses Clustal Omega v1.2.4 to produce multi-sequence alignments, which are then concatenated. Proteins that are absent in a MAG are replaced with a series of X characters in the concatenated alignment. We then generated a maximum-likelihood phylogenetic tree using IQ-TREE v. 1.6.12 with 1,000 ultrafast bootstraps (parameters -m LG + F + I + G4 -bb 1,000 -wbt -nt AUTO -runs 3; Hoang et al., 2018; Minh et al., 2020). The tree was visualized in the interactive Tree of Life (iTOL; Letunic and Bork, 2019; Figure 1).

Average nucleotide and amino acid identity

In order to assess divergence from reference genomes, we calculated one-way average amino acid identity (AAI) and average nucleotide identity (ANI) of the 11 MAGs against all genomes in the Giant Virus Database. Both comparisons were done with LAST v. 959, and results were parsed with a custom Python script. Results are shown in Supplementary Table S1 (Kielbasa et al., 2011).

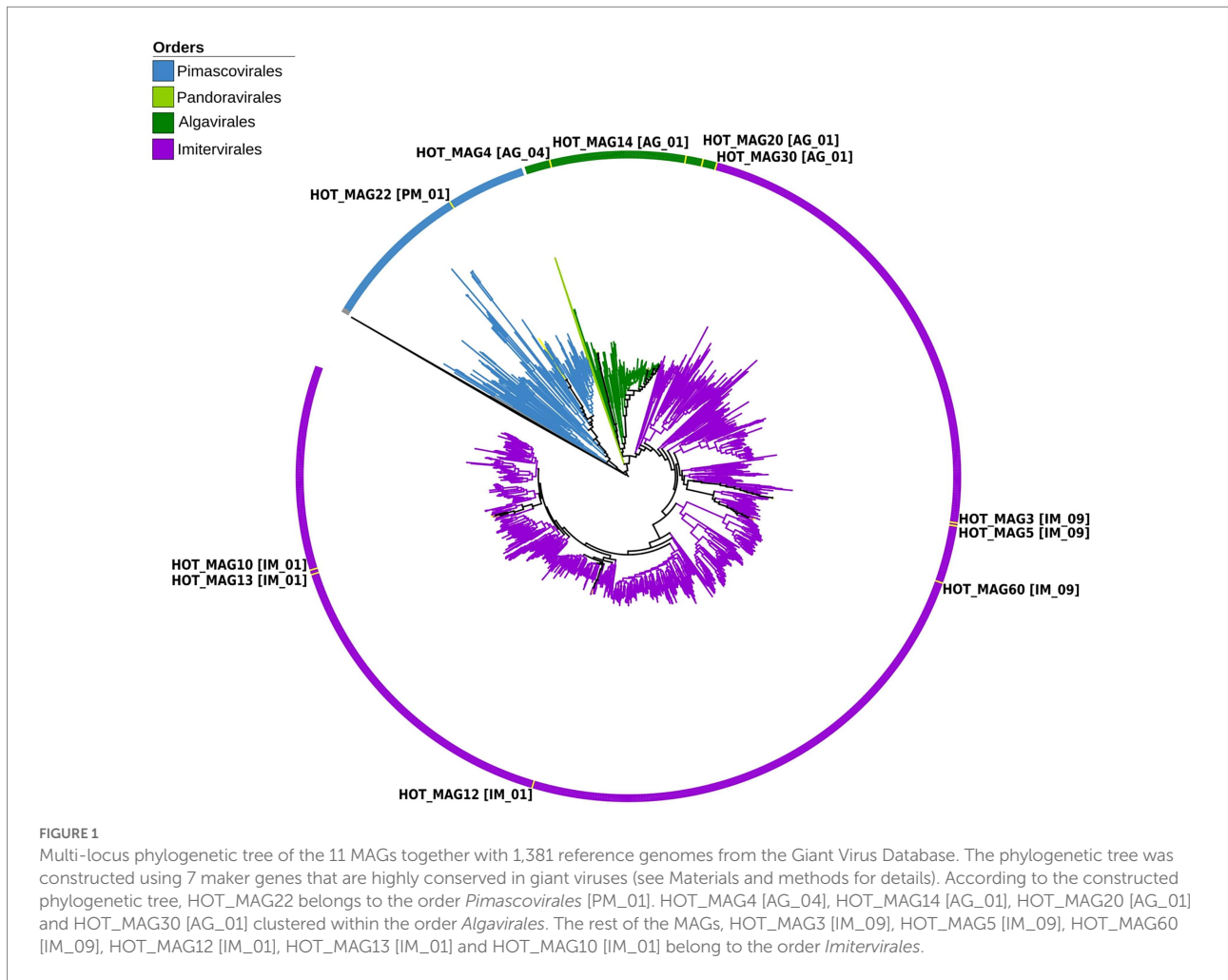
¹ https://github.com/faylward/ncldv_markersearch

Sequence similarity search and protein annotation

To perform sequence homology searches, we used LAST v. 959 (parameters -m 5000, -f BlastTab, -P 32, -u 2, -Q 0) to compare all protein predictions against a protein database that included RefSeq 207 as well as all protein predictions in the Giant Virus Database (O'Leary et al., 2016; Aylward et al., 2021). We indicated the sequence similarities between 11 MAGs and multiple organisms (bacteria, eukaryotes, viruses, and others) together with those which have no hits (Figure 2A). Moreover, we retained all the best hits to viruses and searched for homology between MAGs and different viral orders and families (Figure 2B). Subsequently, plots to visualize the results were made with ggplot2 v. 3.3.6 in R software (Wickham, 2011; Figure 2) and the final results of the sequence similarity search are accessible in Supplementary Table S2. For protein functional prediction, we annotated all predicted proteins in each genome by searching them against the Pfam database v. 34 (Nguyen et al., 2011) using HMMER3 v. 3.3 (parameter "-cut_nc") with all hits retained. These annotations are available in Supplementary Table S3. Protein annotations were manually inspected to detect the presence of genes involved in central carbon metabolism, DNA processing, light harvesting, amino acid metabolism, cytoskeleton dynamics, and other functions of interest.

Read-mapping analysis

We examined the distribution of our 11 MAGs by mapping the reads from other marine metagenomes onto them using coverM 0.6.1 (parameters -min-read-percent-identity 0.95, minimum 20% covered fraction; available from <https://github.com/wwood/CoverM>). We used a breadth cutoff of 20% (i.e., 20% covered fraction), consistent with a recent study that used read-mapping to determine the distribution of large bacteriophages (Weinheimer and Aylward, 2022). We used the metagenomic datasets corresponding to the GA02, GA03, GP13, and GA10 cruises from the bioGEOTRACES dataset (Biller et al., 2018).



These metagenomes were sequenced from 480 samples collected during 2010–2011 from 91 stations. The location and the sampling date of each of the transects are as follows: GA02 (from North to South Atlantic, May 2010–March 2011), GA03 (North Atlantic, October 2010–December 2011), GA10 (South Atlantic, October 2010–November 2010), GP13 (South Pacific Ocean, May 2011–June 2011). In addition, we evaluated the distribution of 11 MAGs in St. ALOHA across different depths (0–1,000 m) and across a 1.5-year time series at that location to evaluate their depth distribution and seasonal abundance in this oligotrophic ecosystem. The Hawaii Ocean Time-series (HOT) metagenomes were previously generated (Mende et al., 2017) during approximate monthly sampling periods from August 2010 to December 2011 at St. ALOHA. Reads from these metagenomes were mapped onto the 11 MAGs using the same coverM parameters as described above. For visualization, we used R and R-based tools (v. 3.3.2)² to draw world map plots and Ocean Data View software (v. 4.7.10)³ for contour and Hawaii Ocean Time-series (HOT) plots. Depth

distribution was interpolated in Ocean Data View in the DIVA gridding mode. The raw read mapping statistics of 11 MAGs on a global scale and in Station ALOHA are available in [Supplementary Tables S4](#) and [S5](#), respectively.

Results and discussion

Phylogenetic analysis of the giant virus metagenome-assembled genomes

Based on our multi-locus phylogenetic analysis, four of the MAGs can be placed within the order *Algavirales* (HOT_MAGs 4, 14, 20, and 30), six could be assigned to the *Imitervirales* (HOT_MAGs 3, 5, 60, 12, 13, and 10), and one could be placed in the *Pimascovirales* (HOT_MAG 22; [Figure 1](#)). Among the MAGs that fall within the *Algavirales*, MAGs 14, 30, and 20 belong to the family *Prasinoviridae* [AG_01] and the same genus-level group (g177) ([Table 1](#)). Viruses in this family are known to infect prasinophytes of the genera *Bathycoccus*, *Micromonas*, and *Ostreococcus* (Weynberg et al., 2017). Prasinophytes are picoeukaryotic algae that are broadly distributed in the ocean

² <https://www.R-project.org>

³ <http://odv.awi.de>

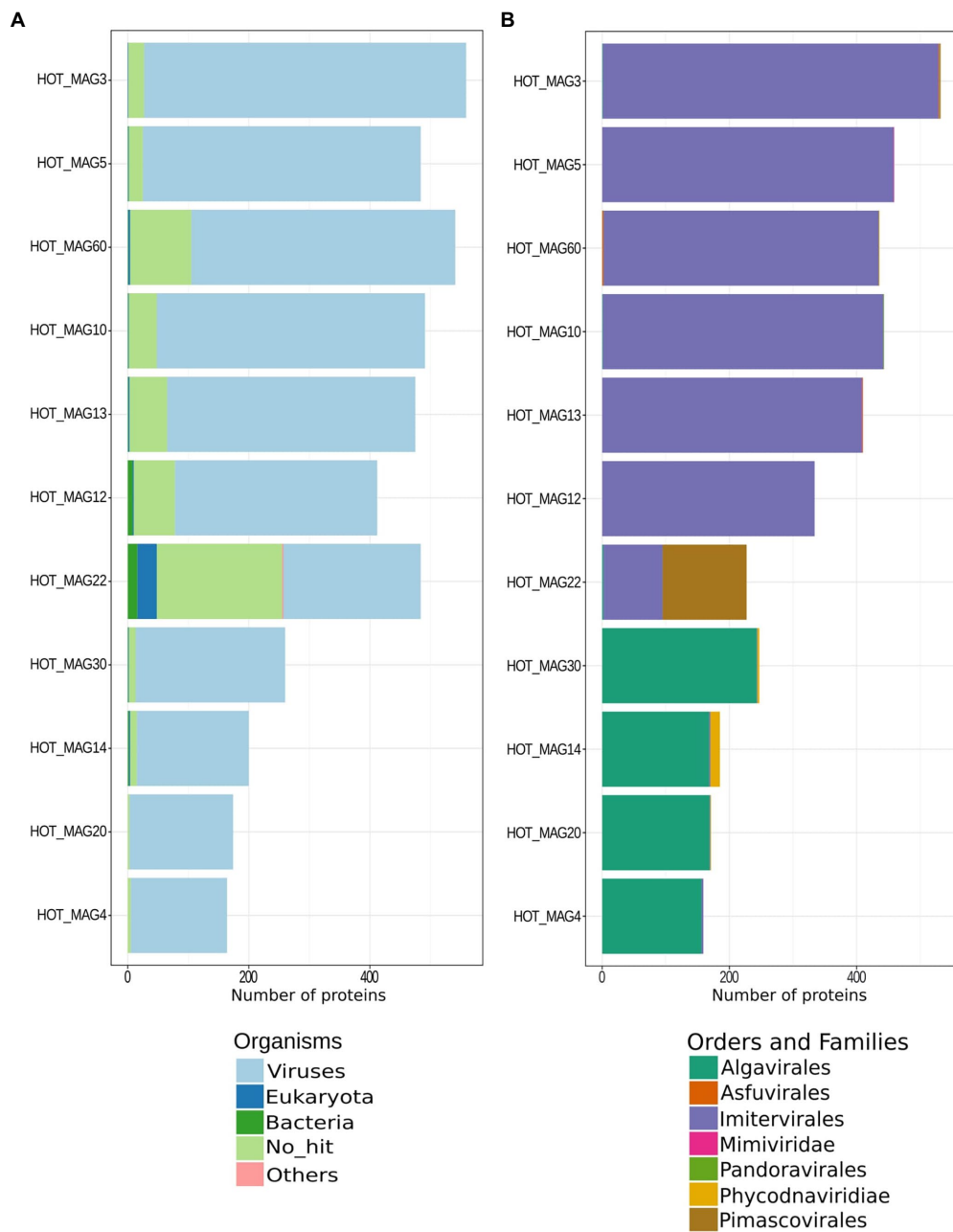


FIGURE 2
 Best matches of predicted proteins identified in each HOT MAG. The results are based on a LAST-based homology search, with best hits retained (see Materials and methods for details).

(Moreau et al., 2010; Lopes dos Santos et al., 2016). Prasinoviruses are correspondingly abundant in marine waters and play a crucial role in regulating the populations of their plankton hosts (Bellec et al., 2010). Cultivated representatives of these viruses have genome sizes ranging from 184 to 198 kbp and %GC contents from 37% to 45%. This is generally consistent with the prasinovirus MAGs that we recovered here, which range in size from ~120–172 kbp and in %GC content from 33.8%–37.5% (Table 1). The somewhat smaller size of HOT_MAGs 14 and 20 is potentially

an indication that these genomes are not complete, though it may also be a sign of genome reduction compared to their relatives. The last MAG that could be placed within the *Algavirales*, HOT_MAG 4, clustered within the family-level lineage AG_04 and the genus g175, both of which were recently demarcated (Aylward et al., 2021; Figure 1; Table 1). The sole cultivated representative of this lineage is *Heterosigma akashiwo virus* (HaV), which infects the eponymous raphidophyte that is responsible for causing harmful algal blooms (Nagasaki et al., 2005). Although

raphidophytes are commonly associated with algal blooms in coastal waters, they have also been reported in oligotrophic gyres (Leles et al., 2019), and it is, therefore, possible that the host of this MAG lies within this group.

Of the six *Imitervirales* MAGs, three fall within the proposed *Mesomimiviridae* family (IM_01: HOT_MAGs 10, 12, 13) (Figure 1; Table 1). HOT_MAGs 10 and 13 are clustered within the same genus g342 while HOT_MAG12 falls within the genus g336 (Table 1). Members of the *Mesomimiviridae* are particularly widespread in global marine and freshwater environments, and phylogenomic analysis suggests that roughly half of all currently available MAGs can be placed in this group (Aylward et al., 2021). The size of these MAGs ranges from 380 to 430 kbp (Table 1), which is consistent with those of cultivated viruses in this family that infect haptophyte hosts of the genera *Chrysochromulina* and *Phaeocystis* (Gallot-Lavallée et al., 2017). *Phaeocystis globosa* virus 16T (PgV-16T) is one of the most well-studied members of this family; it has a 150 nm diameter icosahedral virion with a 470 kbp genome size, which is comparable to the related viruses that infect *Chrysochromulina ericina* virus *C. parva* (Gallot-Lavallée et al., 2017; Stough et al., 2019).

The last three *Imitervirales* MAGs clustered within the recently demarcated family IM_09 (Figure 1; Table 1), which contains *Aureococcus anophagefferens* virus (AaV) and *Prymnesium kappa* virus (PkV). AaV is the smallest member of the *Imitervirales* isolated to date, with a genome size of 371 kbp and a virion diameter of 140 nm (Moniruzzaman et al., 2014; Gann et al., 2020). In contrast to this, PkV is rather large, with a 1.4 Mbp genome and a ~310 nm diameter virion (Johannessen et al., 2015; Blanc-Mathieu et al., 2021), underscoring the large range of genome and virion sizes within this group. HOT_MAGs 3 and 5 fall within the same genus g279 while HOT_MAG 60 clustered within genus g274. The three MAGs that fall within this group have genome sizes ranging from 470 to 575 kbp (Table 1), which is larger than AaV but quite a bit smaller than PkV, suggesting that they represent intermediate-sized members of this lineage.

The only MAG that clustered within the *Pimascovirales* is HOT_MAG 22, which falls within the recently demarcated family-level lineage PM_01 (Figure 1; Table 1). Interestingly, HOT_MAG 22 falls outside of previously demarcated genus-level groups and had low AAI to any references (highest match of 40% AAI to GVMAG-M-3300023184-117, Supplementary Table S1), suggesting that this MAG represents a new genus-level group. The order *Pimascovirales* includes several lineages that infect amoeba, such as Pithoviruses and Marseilleviruses, as well as other viruses that infect metazoan hosts ranging from insects to fish and frogs, such as the *Iridoviridae/Ascoviridae*. Pimascoviruses are not commonly viewed as widespread marine viruses, although, in the previous studies, other MAGs that fall within the PM_01 group have also been found in marine environments (Aylward et al., 2021), and in some cases have been found to be transcriptionally active (Ha et al., 2021). One study also reported several Pimascovirus MAGs from deep sea sediments, including one with phylogenetic placement near the Marseilleviruses and a notably

large genome (>700 kbp; Bäckström et al., 2019). The prevalence of these viruses in marine metagenomes suggests that they infect as-yet unknown protist hosts, and further research identifying their host range will be a necessary step toward clarifying the ecological impacts of marine pimascoviruses.

The phylogenetic placements of the 11 reconstructed MAGs are in agreement with the result from the homology search results of all proteins encoded by each MAG. Proteins in each MAG had best matches to viruses in the same order in which they were classified in the phylogeny (Figure 2). The only possible exception is the *Pimascovirales* MAG (HOT_MAG 22) which had a larger proportion of proteins with no known homologs or hits to the *Imitervirales*. This may be due to the relatively poor representation of marine *Pimascovirales*, together with the large number of *Imitervirales* in our reference genome databases.

Environmental distribution of the giant virus metagenome-assembled genomes

Previous studies have shown that the proposed *Mesomimiviridae* family is particularly prevalent in aquatic systems (Santini et al., 2013; Stough et al., 2019; Aylward et al., 2021). Consistent with this, we found that the three MAGs that fall within this clade were particularly widespread in the marine metagenomes that we surveyed here (HOT_MAGs 12, 13, and 10; Figure 3). HOT_MAGs 12, 13, and 10 were found in 12, 8, and 2 bioGEO TRACES sample locations, respectively. HOT_MAG 3, which can be classified into the family-level group IM_09, was detected in 15 distinct locations and is the most broadly-distributed of the MAGs that fall within the *Imitervirales* MAGs. HOT_MAGs 20 and 30, which fall within the *Prasinoviridae* family, were the most abundant MAGs in the bioGEO TRACES datasets (Figure 3). HOT_MAGs 20 and 30 were also quite widespread, occurring in 23 and 30 distinct locations, respectively. The third prasinovirus (HOT_MAG 14), and the other MAG that places within the *Algavirales* (HOT_MAG4, family AG_04) were each found in only 3 sample sites within the Atlantic Ocean (Figure 3). It is worth noting that HOT_MAGs 60 and 22 were not found in any reference bioGEO TRACES metagenomic dataset. HOT_MAG60 can be placed within the order *Imitervirales* (family IM_09) while HOT_MAG22 is the only member of the *Pimascovirales* we identified. This suggests that these viruses are either quite rare or typically found in low relative abundances and cannot be readily detected with metagenomic methods. Although these HOT_MAGs could only be detected at Station ALOHA, it seems unlikely that they are endemic to this region given that the prevailing conditions at this station are representative of widespread oligotrophic waters.

We examined the bioGEO TRACES GA02 transect in more detail because it traverses the north and south Atlantic and therefore allows for examination of trends in viral abundance across both latitude and depth (Figure 4). Many of the MAGs that fall within the *Imitervirales* were mostly concentrated in

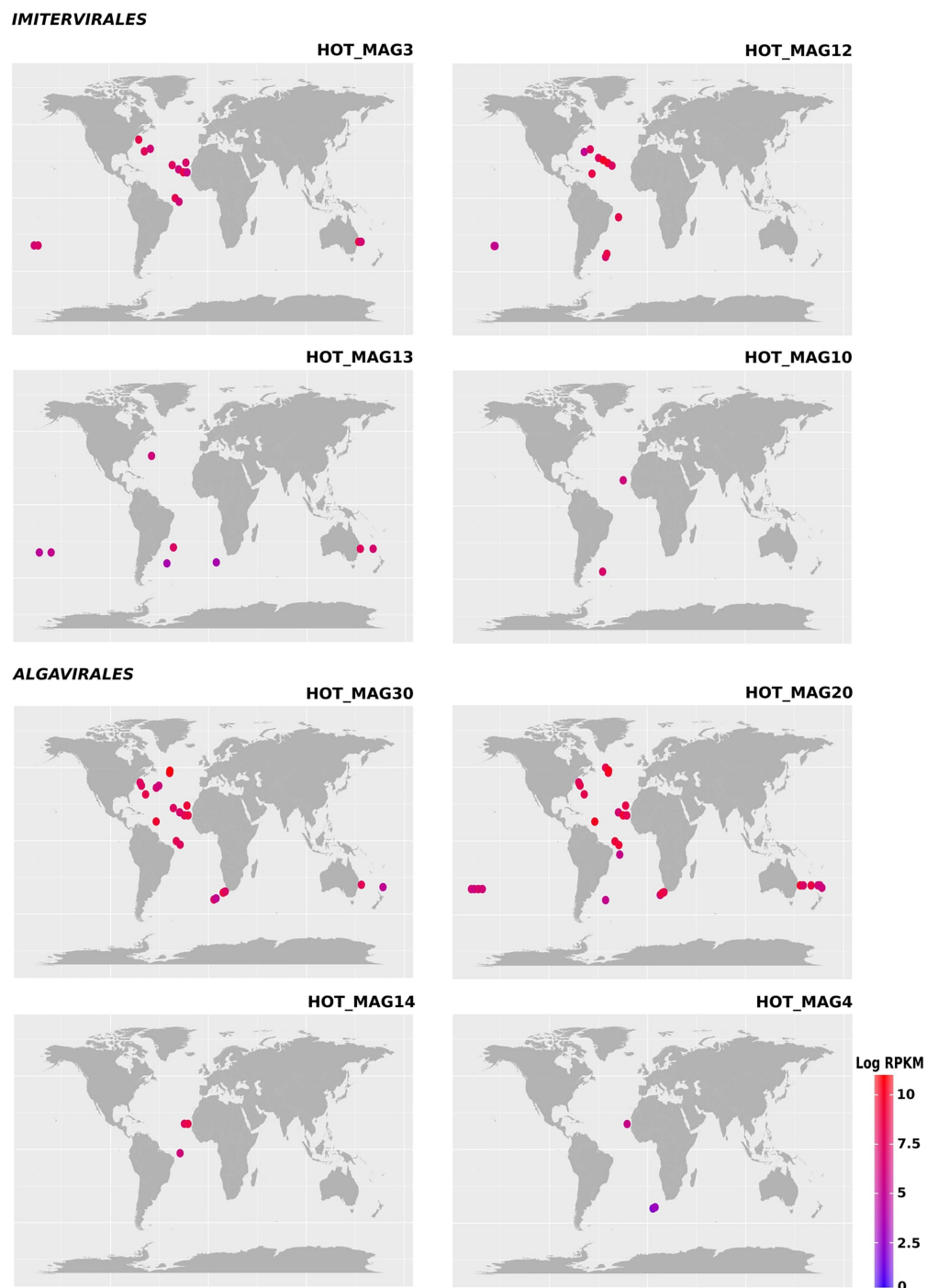


FIGURE 3

Environmental distribution of the HOT MAGs based on read-mapping of bioGEO TRACES metagenomes. HOT_MAGs (22,60) are not shown as they were not identified in bioGEO TRACES metagenomic dataset. HOT_MAG5 is not presented as it is closely related to HOT_MAG3 and its abundance is nearly the same as HOT_MAG3. Bubbles indicate the global distribution of the MAGs which was calculated based on RPKM. Moreover, the abundance of the MAGs in distinct sites with the same latitude and longitude are reported in parenthesis; HOT_MAG3 (15), HOT_MAG12 (12), HOT_MAG13 (8), HOT_MAG10 (2), HOT_MAG30 (23), HOT_MAG20 (30), HOT_MAG14 (3), and HOT_MAG4 (3).

oligotrophic surface waters (above 100 m), consistent with their initial identification at Station ALOHA. These MAGs were identified across a wide range of latitudes; while HOT_MAG3 showed the highest prevalence near the equator, HOT_MAGs

10, 12, and 13 could be detected at latitudes near 40 degrees. Among the MAGs within the *Algavirales*, we detected only the three *Prasinoviridae* MAGs in the GA02 transect metagenomes. Two of these HOT_MAGs (20 and 30) showed high abundance

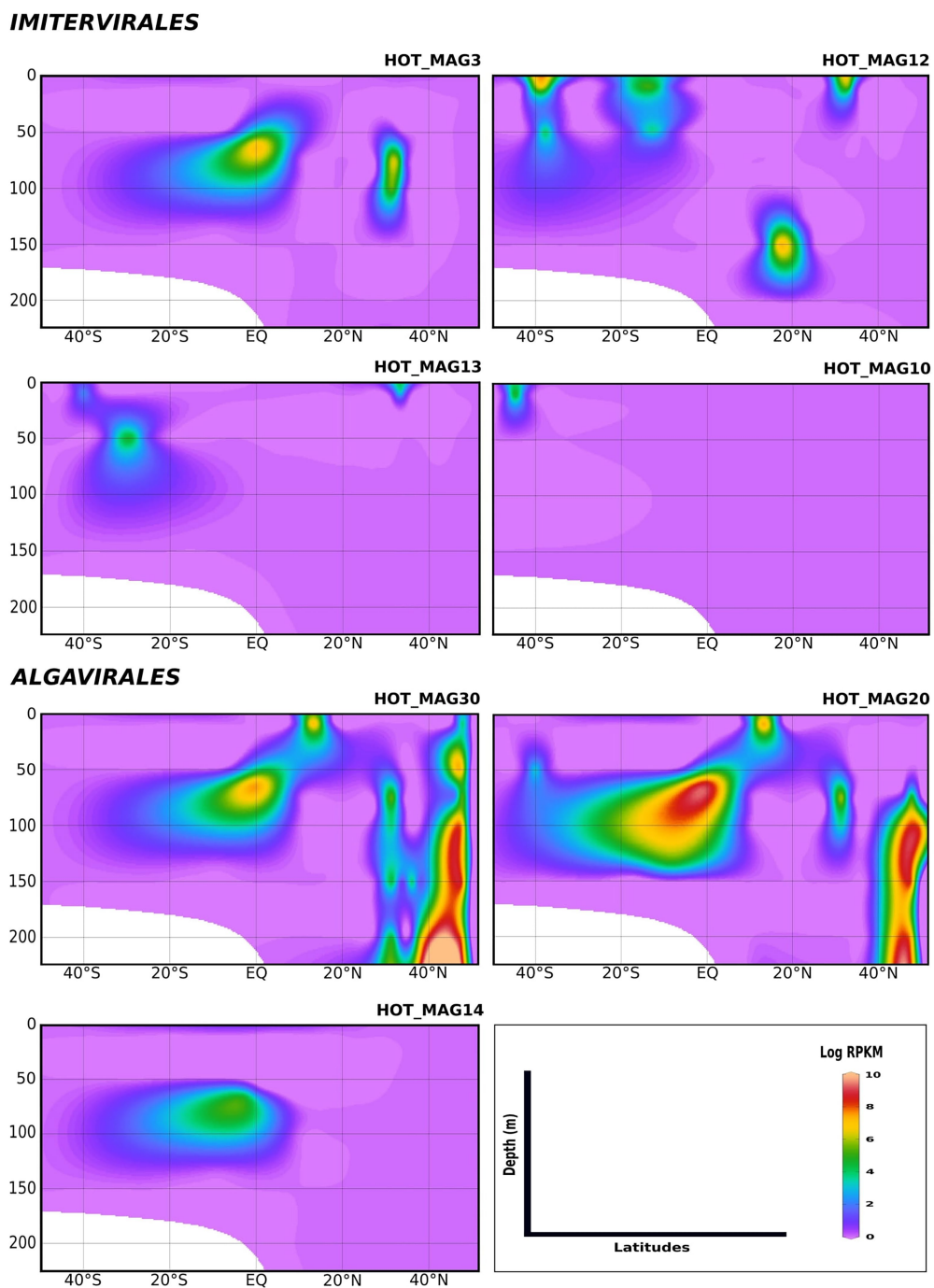


FIGURE 4
 Environmental distribution of the HOT MAGs across a depth profile of the GA02 transect (North Atlantic to South Atlantic). The contour plots were drawn based on the results from mapping GA02 metagenome dataset onto the 11 MAGs (see Materials and methods for details). The total abundance of the MAGs was calculated based on RPKM. Y-axis and x-axis represent depths(m) and latitudes, respectively and the colorful bar refers to the log of total abundance (RPKM) for each of the MAGs (high abundance=red and low abundance=purple). HOT_MAGs (22, 4, 60) are not shown as they were not found in the GA02 metagenomic dataset. HOT_MAG5 is also not presented as it is closely related to HOT_MAG3 and its abundance is almost the same as HOT_MAG3.

both in equatorial waters as well as high northern latitudes (Figure 4). In equatorial waters all prasinovirus MAGs were found in waters above 100 m, but in the north Atlantic HOT_MAGs 20 and 30 could be found in waters near 200 m, possibly

due to sinking water masses in this area (i.e., overturning circulation).

Lastly, in order to evaluate temporal trends in viral abundance, we also examined the presence of the MAGs across a 1.5 year

IMITERVIRALES

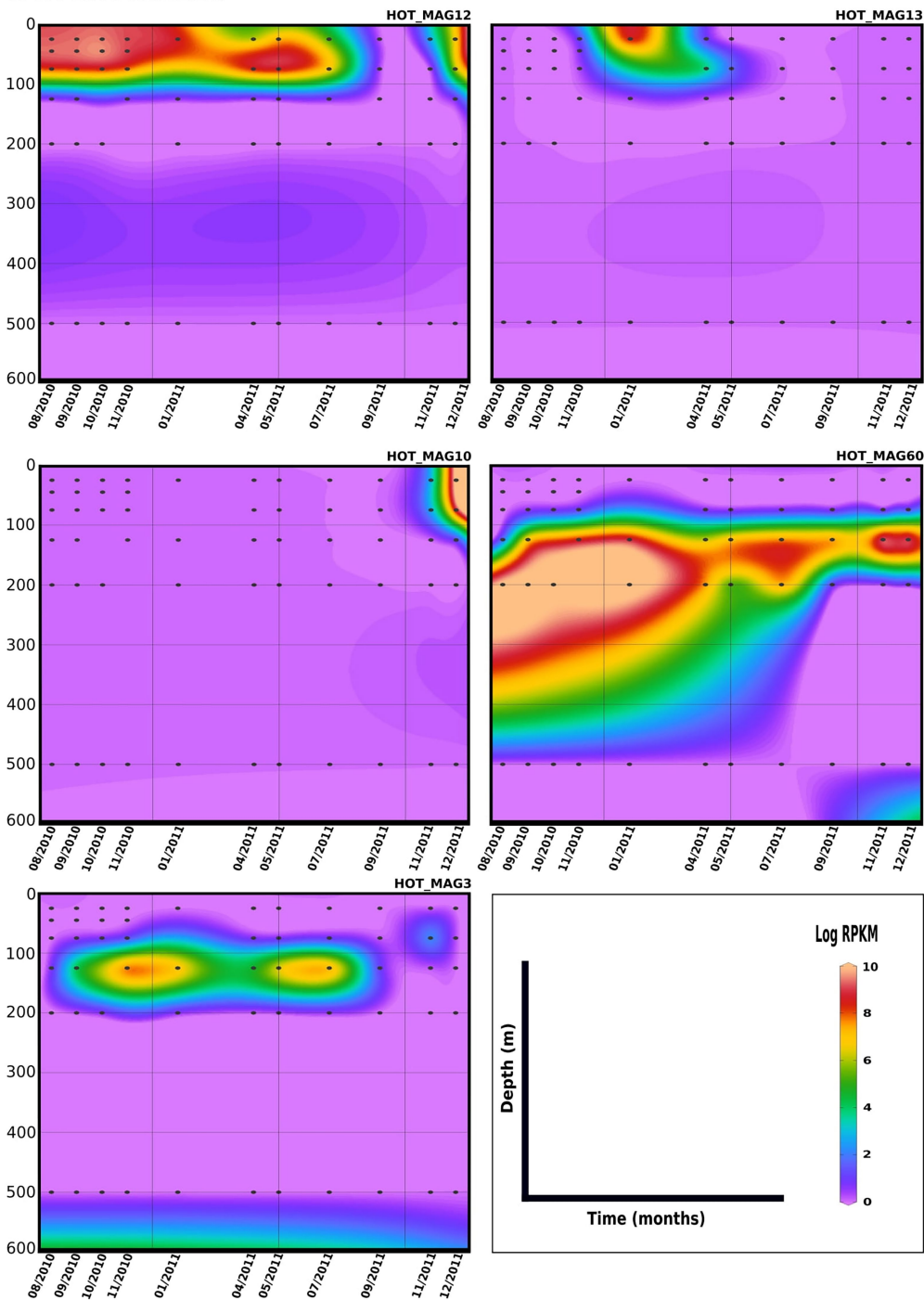


FIGURE 5

Distribution of 11 MAGs in the order *Imitervirales* in a 1.5-year sampling from St. ALOHA. Time series plots were drawn based on the results from mapping Hawaiian Ocean Time series metagenome dataset onto the 11 MAGs (see Materials and methods for details). The total abundance of the MAGs was calculated based on RPKM. Y-axis and x-axis represent depths(m) and latitudes, respectively and the colorful bar refers to the log of total abundance (RPKM) for each of the MAGs (high abundance=red and low abundance=purple). Only the distribution of the MAGs between 0 and 600 m were shown in the plots.

time-series at Station ALOHA from August 2010 to December 2011, which encompasses the same metagenomes used to generate these MAGs (Figures 5, 6). Here the *Algavirales* MAGs seem to

be more prevalent than the *Imitervirales* MAGs during the sampling period from St. ALOHA. Among those MAGs that fall within the family *Mesomimiviridae*, all are present in surface waters

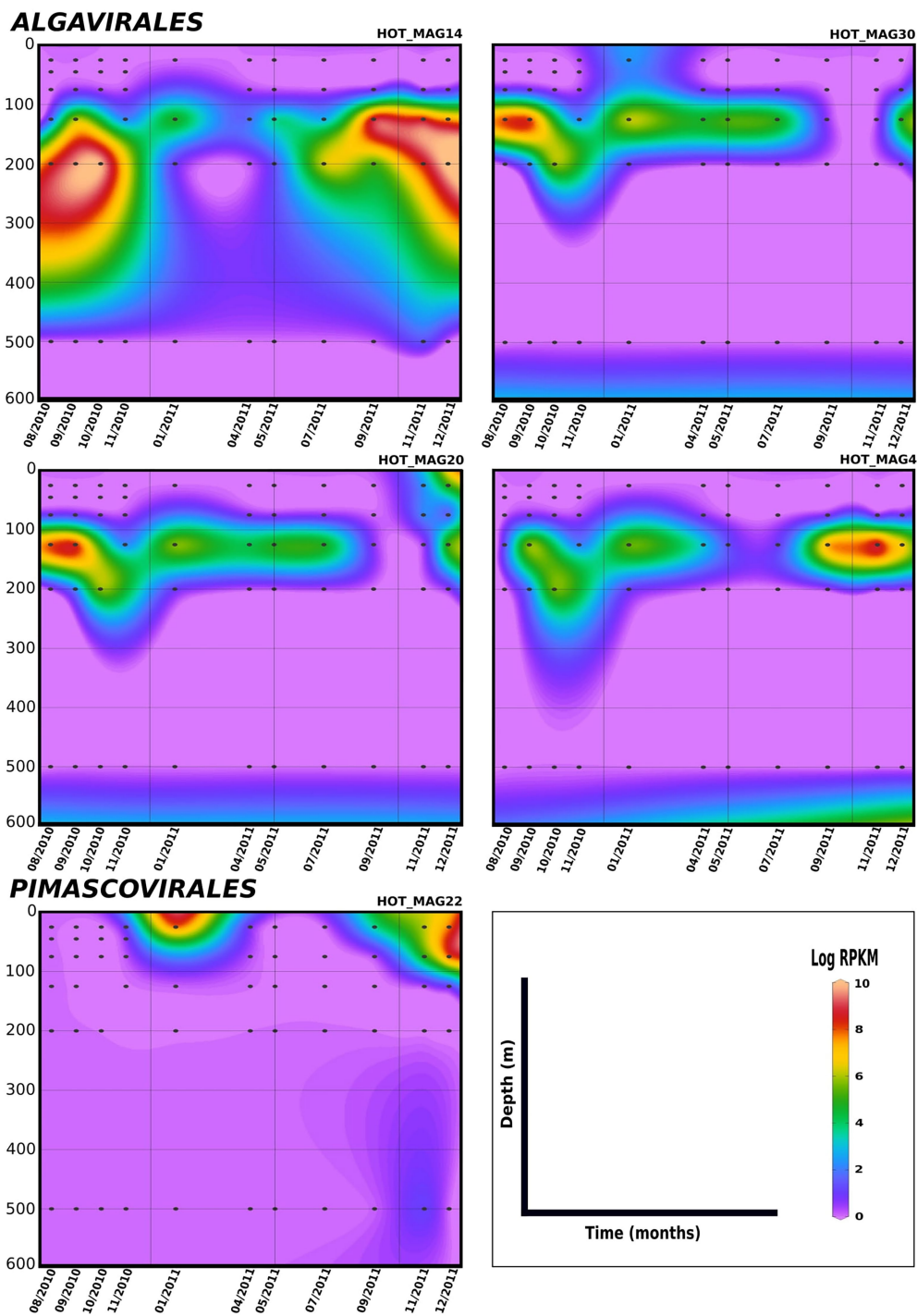


FIGURE 6
 Distribution of 11 MAGs in the orders *Algavirales* and *Pimascovirales* in a 1.5-year sampling from St. ALOHA. Time series plots were drawn based on the results from mapping Hawaiian Ocean Time series metagenome dataset onto the 11 MAGs (see Materials and methods for details). The total abundance of the MAGs was calculated based on RPKM. Y-axis and x-axis represent depths(m) and latitudes, respectively and the colorful bar refers to the log of total abundance (RPKM) for each of the MAGs (high abundance=red and low abundance=purple). Only the distribution of the MAGs between 0–600m were shown in the plots.

(above 100m), and HOT_MAG12 was prevalent throughout almost the entire 1.5 year sampling period (Figure 5). Within *Agavirales* MAGs, two members of the family *Prasinoviridae* (HOT_MAGs 20, 30) have almost the same pattern of their distribution and more likely to be present during August and

September, while HOT_MAG14 is abundant in greater depths (100–500 m) and prevalent in almost all seasons except for spring. HOT_MAG4, another *Agavirales* MAG, also has the same distribution trend as HOT_MAGs (20, 30), however, this MAG is highly abundant during November and December (Figure 6).

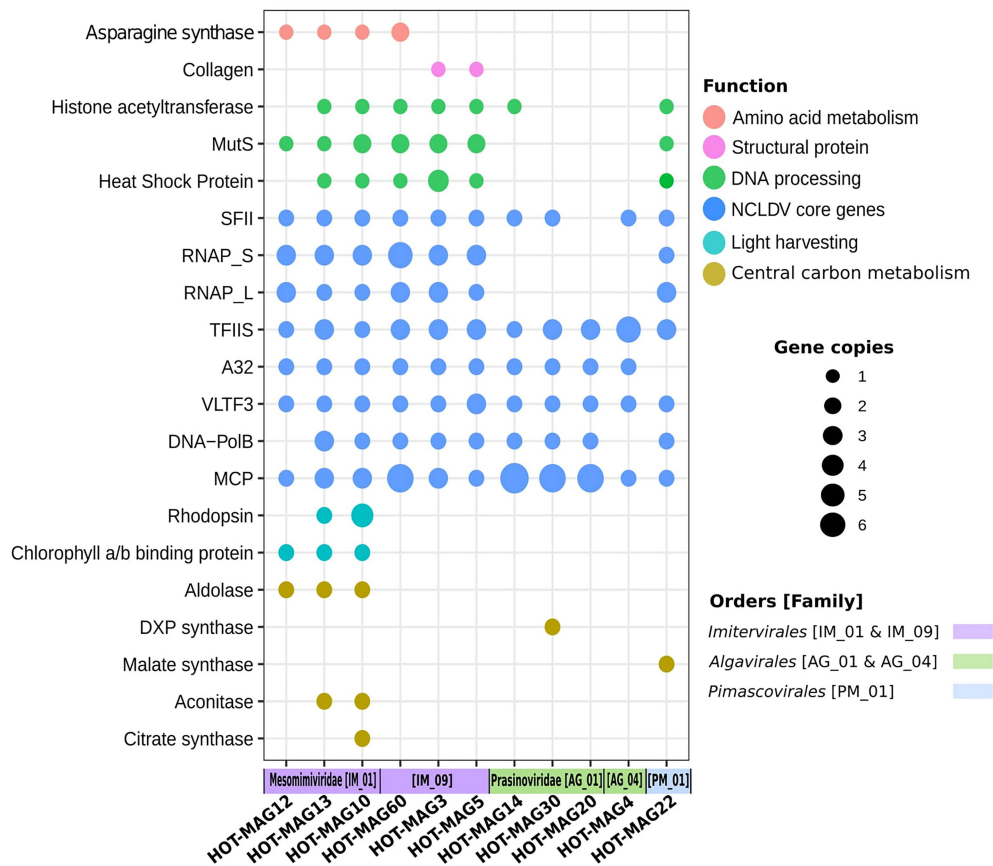


FIGURE 7
 Distribution of selected genes within 11 MAGs of giant viruses. The x-axis indicates the 11 MAGs of giant viruses together with the viral families and orders that they belong to. The size of the bubbles indicate the total abundance of the genes within the genomes of 11 MAGs. the abbreviations of the encoded genes are as follows: MutS: mismatch DNA repair, RNAP: DNA-dependent RNA polymerase Large and Small subunits, SFII: superfamily II helicase, TFIIIS: transcription factor IIB, A32: A32 ATPase, VLTF3: viral late transcription factor 3, MCP: major capsid protein, DXP synthase: 1-Deoxy-D-xylulose 5-phosphate synthase.

Overall, all of the MAGs within the *Algavirales* were prevalent in the 125 m samples, which was just below the DCM for the majority of cases. HOT_MAG30, HOT_MAG20 and HOT_MAG4 were found most frequently at 125 m, while HOT_MAG14 was also prevalent in several 200 m samples (Figure 6). This transition below the DCM is consistent with the large-scale microbial community turnover that occurs in this region at Station ALOHA (Mende et al., 2017). The only *Pimascovirales* MAG (HOT_MAG22), is distributed in shallow depths (0–100 m) and is highly concentrated between December and January (Figure 6).

The genomic repertoire encoded in the 11 metagenome-assembled genomes

Giant viruses have complex genomes that encode various genes that are not commonly found in viral lineages, including components of central carbon metabolism and translation-associated proteins. As expected, we found giant viruses core genes involved in DNA replication, transcription, and virion structure (Figure 7). We did not find RNAP subunits in the

prasinovirus MAGs, consistent with previous studies showing that this lineage lacks this enzyme (Moreau et al., 2010; Figure 7). The absence of RNAP subunits suggests that these viruses have a nuclear stage to their infection in which host RNAP is used for viral gene expression.

Asparagine synthase genes were identified among the *Imitervirales* MAGs, in particular among those members of the *Mesomimiviridae* (Figure 7). This enzyme was previously identified in *Phaeocystis globosa* virus, though it remains unclear what role it may play during infection (Santini et al., 2013). Collagen encoded genes were only found in *Imitervirales* MAGs belonging to the family IM_09, suggesting that these proteins may be a part of the structure of the virions of these viruses. Among the genes involved in DNA processing, chaperones from the heat shock protein families (HSP70, HSP90) which are presumably used for capsid protein folding (Legendre et al., 2010), DNA mismatch repair (MutS), and histone acetyltransferase known to be functional for packaging DNA within the capsid (Koonin and Yutin, 2010; Legendre et al., 2010), were mostly frequent among *Imitervirales* and *Pimascovirales* MAGs but were mostly absent from the *Algavirales* MAGs (Figure 7).

Many of the MAGs encode multiple genes involved in central carbon metabolism, such as aldolase, 1-Deoxy-D-xylulose 5-phosphate synthase (DXP synthase), malate synthase, aconitase, and citrate synthase. Our findings indicate genes that belong to central carbon metabolism are mostly detected in *Mesomimiviridae* MAGs, notably those that are predicted to be involved in predicted to be involved in the TCA cycle (Figure 7). Interestingly, malate synthase and DXP synthase were only found in the *Pimascovirales* MAG and *Algavirales* MAG, respectively. Previous work has shown that enzymes involved in central carbon metabolism are quite common in many giant viruses (Moniruzzaman et al., 2020a), and our results here are consistent with those findings.

Previous studies have shown that rhodopsins and chlorophyll-binding proteins are quite common in a wide range of marine giant viruses (Yutin and Koonin, 2012; Needham et al., 2019; Moniruzzaman et al., 2020a; Rozenberg et al., 2020). Consistent with this, we found chlorophyll binding proteins in all three mesomimiviruses, and rhodopsin homologs in two (HOT_MAGs 10 and 13) (Figure 7). This suggests that these viruses likely infect phototrophic or mixotrophic hosts and manipulate light harvesting machinery during infection. All mesomimivirus MAGs were detected in shallow waters at Station ALOHA, consistent with the prevalence of their hosts in well-lit surface waters. Interestingly, we detected three rhodopsin homologs in HOT_MAG10; it is unclear what role these three enzymes would play during infection, but the presence of three distinct homologs suggests that they are an important component of the infection strategy of many marine giant viruses.

Conclusion

Our study sheds light on the phylogenetic diversity, genomics, and distribution of giant viruses in oligotrophic marine waters. We present 11 MAGs of giant viruses that we reconstructed from metagenomes generated from Station ALOHA in the North Pacific Subtropical Gyre. These MAGs fall within five families in the orders *Imitervirales*, *Algavirales*, and *Pimascovirales*. Those MAGs that fall within the *Prasinoviridae* and *Mesomimiviridae* families are the most widespread and abundant, and several of these MAGs were detected in diverse bioGEO TRACES metagenomes that were collected in different ocean basins. Several of the MAGs were found consistently at Station ALOHA over a 1.5 year period, suggesting they are persistent community members in oligotrophic waters. The MAGs encoded a diverse range of functions, including genes involved in central carbon metabolism and light harvesting, suggesting that they use a variety of strategies to manipulate the physiology of their hosts during infection. Our work contributes to a growing body of research that suggests that large DNA viruses are abundant and widespread components of marine systems that play key roles in ecological dynamics and biogeochemical cycling.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/Supplementary material.

Author contributions

FA and RF contributed in the conceptualization of the research and writing the manuscript. RF and AH performed bioinformatic analysis of the research. All authors contributed to the article and approved the submitted version.

Funding

This research was funded by a Simons Foundation Early Career Award in Marine Microbial Ecology and Evolution, NSF CAREER award 2141862, and NIGMS R35 award 1R35GM147290-01 to FA.

Acknowledgments

We thank members of the Aylward lab for helpful feedback. This work was performed using resources available at the Virginia Tech Advanced Research Computing Center.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2022.1021923/full#supplementary-material>

References

- Aylward, F. O., Boeuf, D., Mende, D. R., Wood-Charlson, E. M., Vislova, A., Eppley, J. M., et al. (2017). Diel cycling and long-term persistence of viruses in the ocean's euphotic zone. *Proc. Natl. Acad. Sci. U. S. A.* 114, 11446–11451. doi: 10.1073/pnas.1714821114
- Aylward, F. O., and Moniruzzaman, M. (2021). ViralRecall—A flexible command-line tool for the detection of Giant virus signatures in 'Omic data. *Viruses* 13:150. doi: 10.3390/v13020150
- Aylward, F. O., and Moniruzzaman, M. (2022). Viral complexity. *Biomolecules* 12:1061. doi: 10.3390/biom12081061
- Aylward, F. O., Moniruzzaman, M., Ha, A. D., and Koonin, E. V. (2021). A phylogenomic framework for charting the diversity and evolution of giant viruses. *PLoS Biol.* 19:e3001430. doi: 10.1371/journal.pbio.3001430
- Bäckström, D., Yutin, N., Jørgensen, S. L., Dharamshi, J., Homa, F., Zaremba-Niedwiedzka, K., et al. (2019). Virus genomes from Deep Sea sediments expand the ocean Megavirome and support independent origins of viral gigantism. *MBio* 10:e02497-18. doi: 10.1128/mBio.02497-18
- Bellec, L., Grimsley, N., Derelle, E., Moreau, H., and Desdèvises, Y. (2010). Abundance, spatial distribution and genetic diversity of *Ostreococcus tauri* viruses in two different environments. *Environ. Microbiol. Rep.* 2, 313–321. doi: 10.1111/j.1758-2229.2010.00138.x
- Billler, S. J., Berube, P. M., Dooley, K., Williams, M., Satinsky, B. M., Hackl, T., et al. (2018). Marine microbial metagenomes sampled across space and time. *Sci. Data* 5:180176. doi: 10.1038/sdata.2018.176
- Blanc-Mathieu, R., Dahle, H., Hofgaard, A., Brandt, D., Ban, H., Kalinowski, J., et al. (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. doi: 10.1128/JVI.02446-20
- Boyer, M., Yutin, N., Pagnier, I., Barrassi, L., Fournous, G., Espinosa, L., et al. (2009). Giant Marseillevirus highlights the role of amoebae as a melting pot in emergence of chimeric microorganisms. *Proc. Natl. Acad. Sci. U. S. A.* 106, 21848–21853. doi: 10.1073/pnas.0911354106
- Chen, F., Suttle, C. A., and Short, S. M. (1996). Genetic diversity in marine algal virus communities as revealed by sequence analysis of DNA polymerase genes. *Appl. Environ. Microbiol.* 62, 2869–2874. doi: 10.1128/aem.62.8.2869-2874.1996
- Da Cunha, V., Gaia, M., Ogata, H., Jaillon, O., Delmont, T. O., and Forterre, P. (2022). Giant viruses encode actin-related proteins. *Mol. Biol. Evol.* 39:2. doi: 10.1093/molbev/msac022
- Endo, H., Blanc-Mathieu, R., Li, Y., Salazar, G., Henry, N., Labadie, K., et al. (2020). Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. *Nat. Ecol. Evol.* 4, 1639–1649. doi: 10.1038/s41559-020-01288-w
- Ernst, O. P., Lodowski, D. T., Elstner, M., Hegemann, P., Brown, L. S., and Kandori, H. (2014). Microbial and animal rhodopsins: structures, functions, and molecular mechanisms. *Chem. Rev.* 114, 126–163. doi: 10.1021/cr4003769
- Fischer, M. G. (2016). Giant viruses come of age. *Curr. Opin. Microbiol.* 31, 50–57. doi: 10.1016/j.mib.2016.03.001
- Gallot-Lavallée, L., and Archibald, J. M. (2020). Evolutionary biology: viral Rhodopsins illuminate algal evolution. *Curr. Biol.* 30, R1469–R1471. doi: 10.1016/j.cub.2020.10.080
- Gallot-Lavallée, L., Blanc, G., and 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. doi: 10.1128/jvi.00230-17
- Gann, E. R., Jackson Gainer, P., Reynolds, T. B., and Wilhelm, S. W. (2020). Influence of light on the infection of *Aureococcus anophagefferens* CCMP 1984 by a giant virus. *PLoS One* 15:e0226758. doi: 10.1371/journal.pone.0226758
- Ghedini, E., and Claverie, J.-M. (2005). Mimivirus relatives in the Sargasso Sea. *Virol. J.* 2:62. doi: 10.1186/1743-422X-2-62
- Govorunova, E. G., Sineshchekov, O. A., Li, H., and Spudich, J. L. (2017). Microbial Rhodopsins: diversity, mechanisms, and Optogenetic applications. *Annu. Rev. Biochem.* 86, 845–872. doi: 10.1146/annurev-biochem-101910-144233
- Ha, A. D., Moniruzzaman, M., and Aylward, F. O. (2021). High transcriptional activity and diverse functional repertoires of hundreds of Giant viruses in a coastal marine. *Systems. mSystems* 6:e0029321. doi: 10.1128/mSystems.00293-21
- Hingamp, P., Grimsley, N., Acinas, S. G., Clerissi, C., Subirana, L., Poulain, J., et al. (2013). Exploring nucleocytoplasmic large DNA viruses in Tara oceans microbial metagenomes. *ISME J.* 7, 1678–1695. doi: 10.1101/gr.102582.109
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Hyatt, D., Chen, G.-L., Locascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Johannessen, T. V., Bratbak, G., Larsen, A., Ogata, H., Egge, E. S., Edvardsen, B., et al. (2015). Characterisation of three novel giant viruses reveals huge diversity among viruses infecting Prymnesiales (Haptophyta). *Virology* 476, 180–188. doi: 10.1016/j.virol.2014.12.014
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. doi: 10.7717/peerj.7359
- Karki, S., Moniruzzaman, M., and Aylward, F. O. (2021). Comparative genomics and environmental distribution of large dsDNA viruses in the family Asfarviridae. *Front. Microbiol.* 12:428683. doi: 10.1101/2021.01.29.428683
- Karl, D. M., and Church, M. J. (2014). Microbial oceanography and the Hawaii Ocean time-series programme. *Nat. Rev. Microbiol.* 12, 699–713. doi: 10.1038/nrmicro3333
- Kielbasa, S. M., Wan, R., Sato, K., Horton, P., and Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Res.* 21, 487–493. doi: 10.1101/gr.113985.110
- Kijima, S., Delmont, T. O., Miyazaki, U., Gaia, M., Endo, H., and Ogata, H. (2021). Discovery of viral myosin genes with complex evolutionary history within plankton. *Front. Microbiol.* 12:683294. doi: 10.3389/fmicb.2021.683294
- Koonin, E. V., and Yutin, N. (2010). Origin and evolution of eukaryotic large nucleocytoplasmic DNA viruses. *Intervirology* 53, 284–292. doi: 10.1186/1745-6150-7-34
- Koonin, E. V., and Yutin, N. (2019). Evolution of the large nucleocytoplasmic DNA viruses of eukaryotes and convergent origins of viral gigantism. *Adv. Virus Res.* 103, 167–202. doi: 10.1016/bs.aivir.2018.09.002
- Legendre, M., Audic, S., Poirot, O., Hingamp, P., Seltzer, V., Byrne, D., et al. (2010). mRNA deep sequencing reveals 75 new genes and a complex transcriptional landscape in Mimivirus. *Genome Res.* 20, 664–674. doi: 10.1101/gr.102582.109
- Leles, S. G., Mitra, A., Flynn, K. J., Tillmann, U., Stoecker, D., Jeong, H. J., et al. (2019). Sampling bias misrepresents the biogeographical significance of constitutive mixotrophs across global oceans. *Glob. Ecol. Biogeogr.* 28, 418–428. doi: 10.1111/geb.12853
- Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239
- Lopes dos Santos, A., Gourvil, P., Tragin, M., Noël, M.-H., Decelle, J., Romac, S., et al. (2016). Diversity and oceanic distribution of prasinophytes clade VII, the dominant group of green algae in oceanic waters. *ISME J.* 11, 512–528. doi: 10.1038/ismej.2016.120
- Luo, E., Aylward, F. O., Mende, D. R., and DeLong, E. F. (2017). Bacteriophage distributions and temporal variability in the Ocean's interior. *MBio* 8. doi: 10.1128/mBio.01903-17
- Luo, E., Eppley, J. M., Romano, A. E., Mende, D. R., and DeLong, E. F. (2020). Double-stranded DNA viroplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J.* 14, 1304–1315. doi: 10.1038/s41396-020-0604-8
- Luo, E., Leu, A. O., Eppley, J. M., Karl, D. M., and DeLong, E. F. (2022). Diversity and origins of bacterial and archaeal viruses on sinking particles reaching the abyssal ocean. *ISME J.* 16, 1627–1635. doi: 10.1038/s41396-022-01202-1
- Mende, D. R., Bryant, J. A., Aylward, F. O., Eppley, J. M., Nielsen, T., Karl, D. M., et al. (2017). Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat. Microbiol.* 2, 1367–1373. doi: 10.1038/s41564-017-0008-3
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Monier, A., Claverie, J.-M., and Ogata, H. (2008). Taxonomic distribution of large DNA viruses in the sea. *Genome Biol.* 9:R106. doi: 10.1186/gb-2008-9-7-r106
- Moniruzzaman, M., LeClerc, G. R., Brown, C. M., Gobler, C. J., Bidle, K. D., Wilson, W. H., et al. (2014). Genome of brown tide virus (AaV), the little giant of the Megaviridae, elucidates NCLDV genome expansion and host-virus coevolution. *Virology* 466–467. doi: 10.1016/j.virol.2014.06.031
- Moniruzzaman, M., Martínez-Gutiérrez, C. A., Weinheimer, A. R., and Aylward, F. O. (2020a). Dynamic genome evolution and complex virocell metabolism of globally-distributed giant viruses. *Nat. Commun.* 11:1710. doi: 10.1038/s41467-020-15507-2
- Moniruzzaman, M., Weinheimer, A. R., Martínez-Gutiérrez, C. A., and Aylward, F. O. (2020b). Widespread endogenization of giant viruses shapes genomes of green algae. *Nature* 588, 141–145. doi: 10.1038/s41586-020-2924-2

- Moreau, H., Piganeau, G., Desdevises, Y., Cooke, R., Derelle, E., and Grimsley, N. (2010). Marine prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. *J. Virol.* 84, 12555–12563. doi: 10.1128/JVI.01123-10
- Nagasaki, K., Shirai, Y., Tomaru, Y., Nishida, K., and Pietrokovski, S. (2005). Algal viruses with distinct intraspecies host specificities include identical intein elements. *Appl. Environ. Microbiol.* 71, 3599–3607. doi: 10.1128/AEM.71.7.3599-3607.2005
- Needham, D. M., Yoshizawa, S., Hosaka, T., Poirier, C., Choi, C. J., Hehenberger, E., et al. (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. doi: 10.1073/pnas.1907517116
- Nguyen, C. D., Gardiner, K. J., and Cios, K. J. (2011). Protein annotation from protein interaction networks and gene ontology. *J. Biomed. Inform.* 44, 824–829. doi: 10.1016/j.jbi.2011.04.010
- O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufu, S., Haddad, D., McVeigh, R., et al. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 44, D733–D745. doi: 10.1093/nar/gkv1189
- Rozenberg, A., Oppermann, J., Wietek, J., Fernandez Lahore, R. G., Sandaa, R.-A., Bratbak, G., et al. (2020). Lateral gene transfer of anion-conducting Channelrhodopsins between green algae and Giant viruses. *Curr. Biol.* 30, 4910–4920.e5. doi: 10.1016/j.cub.2020.09.056
- Santini, S., Jeudy, S., Bartoli, J., Poirot, O., Lescot, M., Abergel, C., et al. (2013). Genome of Phaeocystis globosa virus PgV-16T highlights the common ancestry of the largest known DNA viruses infecting eukaryotes. *Proc. Natl. Acad. Sci. U. S. A.* 110, 10800–10805. doi: 10.1073/pnas.1303251110
- Schulz, F., Roux, S., Paez-Espino, D., Jungbluth, S., Walsh, D. A., Deneff, V. J., et al. (2020). Giant virus diversity and host interactions through global metagenomics. *Nature* 578, 432–436. doi: 10.1038/s41586-020-1957-x
- Schulz, F., Yutin, N., Ivanova, N. N., Ortega, D. R., Lee, T. K., Vierheilig, J., et al. (2017). Giant viruses with an expanded complement of translation system components. *Science* 356, 82–85. doi: 10.1126/science.aal4657
- Shen, W., Le, S., Li, Y., and Hu, F. (2016). SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11:e0163962. doi: 10.1371/journal.pone.0163962
- Short, S. M. (2012). The ecology of viruses that infect eukaryotic algae. *Environ. Microbiol.* 14, 2253–2271. doi: 10.1111/j.1462-2920.2012.02706.x
- Short, S. M., and Suttle, C. A. (2002). Sequence analysis of marine virus communities reveals that groups of related algal viruses are widely distributed in nature. *Appl. Environ. Microbiol.* 68, 1290–1296. doi: 10.1128/AEM.68.3.1290-1296.2002
- Stough, J. M. A., Yutin, N., Chaban, Y. V., Moniruzzaman, M., Gann, E. R., Pound, H. L., et al. (2019). Genome and environmental activity of a Chrysochromulina parva virus and its Virophages. *Front. Microbiol.* 10:703. doi: 10.3389/fmicb.2019.00703
- Weinheimer, A. R., and Aylward, F. O. (2022). Infection strategy and biogeography distinguish cosmopolitan groups of marine jumbo bacteriophages. *ISME J.* 16, 1657–1667. doi: 10.1038/s41396-022-01214-x
- Weynberg, K. D., Allen, M. J., and Wilson, W. H. (2017). Marine Prasinoviruses and their tiny plankton hosts: a review. *Viruses* 9:43. doi: 10.3390/v9030043
- Wickham, H. (2011). Ggplot2. Wiley Interdiscip. Rev. Comput. Stat 3, 180–185. doi: 10.1002/wics.147
- Wilhelm, S. W., Coy, S. R., Gann, E. R., Moniruzzaman, M., and Stough, J. M. A. (2016). Standing on the shoulders of Giant viruses: five lessons learned about large viruses infecting small eukaryotes and the opportunities they create. *PLoS Pathog.* 12:e1005752. doi: 10.1371/journal.ppat.1005752
- Yau, S., Lauro, F. M., DeMaere, M. Z., Brown, M. V., Thomas, T., Raftery, M. J., et al. (2011). Virophage control of antarctic algal host-virus dynamics. *Proc. Natl. Acad. Sci. U. S. A.* 108, 6163–6168. doi: 10.1073/pnas.1018221108
- Yoshikawa, G., Blanc-Mathieu, R., Song, C., Kayama, Y., Mochizuki, T., Murata, K., et al. (2019). Medusavirus, a novel large DNA virus discovered from hot spring water. *J. Virol.* 93:e02130-18. doi: 10.1128/JVI.02130-18
- Yutin, N., and Koonin, E. V. (2012). Proteorhodopsin genes in giant viruses. *Biol. Direct* 7:34. doi: 10.1186/1745-6150-7-34