



Double-stranded DNA viroplankton dynamics and reproductive strategies in the oligotrophic open ocean water column

Elaine Luo ¹ · John M. Eppley¹ · Anna E. Romano¹ · Daniel R. Mende ¹ · Edward F. DeLong ¹

Received: 6 June 2019 / Revised: 17 December 2019 / Accepted: 30 January 2020 / Published online: 14 February 2020
© The Author(s) 2020. This article is published with open access

Abstract

Microbial communities are critical to ecosystem dynamics and biogeochemical cycling in the open oceans. Viruses are essential elements of these communities, influencing the productivity, diversity, and evolution of cellular hosts. To further explore the natural history and ecology of open-ocean viruses, we surveyed the spatiotemporal dynamics of double-stranded DNA (dsDNA) viruses in both viroplankton and bacterioplankton size fractions in the North Pacific Subtropical Gyre, one of the largest biomes on the planet. Assembly and clustering of viral genomes revealed a peak in viroplankton diversity at the base of the euphotic zone, where virus populations and host species richness both reached their maxima. Simultaneous characterization of both extracellular and intracellular viruses suggested depth-specific reproductive strategies. In particular, analyses indicated elevated lytic interactions in the mixed layer, more temporally variable temperate phage interactions at the base of the euphotic zone, and increased lysogeny in the mesopelagic ocean. Furthermore, the depth variability of auxiliary metabolic genes suggested habitat-specific strategies for viral influence on light-energy, nitrogen, and phosphorus acquisition during host infection. Most virus populations were temporally persistent over several years in this environment at the 95% nucleic acid identity level. In total, our analyses revealed variable distributional patterns and diverse reproductive and metabolic strategies of virus populations in the open-ocean water column.

Introduction

Viruses represent dynamic reservoirs of unexplored genetic diversity. On average an order of magnitude more abundant than cellular organisms [1], viruses occur at $\sim 10^7$ /mL in the surface layer of open oceans covering $\sim 40\%$ of Earth [2]. In this environment, dsDNA bacteriophages infect key microbial groups, including oxygenic photoautotrophs such as *Prochlorococcus* and *Synechococcus* (e.g., [3, 4]) and common bacterial heterotrophs such as *Pelagibacter* (SAR11), *Puniceispirillum* (SAR116), *Roseobacter*, and *Alteromonas* (e.g., [5–8]). Viruses can lyse their hosts at

estimated rates of 20–40% per day [9, 10], potentially contributing as much as 145 gigatonnes to annual global carbon flux [11, 12]. Viruses also influence the diversity and biogeochemistry of marine ecosystems by carrying auxiliary metabolic genes (AMGs) that manipulate host metabolism during infection (reviewed in [13]).

Recent developments in high-throughput DNA sequencing have allowed for exploration of viral diversity at unprecedented scales. The continuing description of viral genetic diversity highlights the importance of further cultivation-independent in situ characterization of environmental viral populations. Metagenomic virus surveys have focused on characterizing geographic variability across surface oceans [14–17], temporal variability at the surface ocean [18–21], and vertical variability across depth profiles [22–24]. Coupled studies of viral and host dynamics in both space and time at well-defined sample sites [25–27] further have the potential to provide additional perspective on the dynamics, patterns and consequences of viral diversity.

To further explore viral diversity, environmental distributions, and dynamics in the open ocean, we characterized virus genomes from seawater in virus-enriched (0.02–0.2 μm) and cell-enriched (>0.2 μm) size fractions

Supplementary information The online version of this article (<https://doi.org/10.1038/s41396-020-0604-8>) contains supplementary material, which is available to authorized users.

✉ Edward F. DeLong
edelong@hawaii.edu

¹ Daniel K. Inouye Center for Microbial Oceanography: Research and Education (C-MORE), University of Hawaii, Honolulu, HI 96822, USA

over time and depth in the North Pacific Subtropical Gyre (NPSG). The approach facilitated the exploration of both extracellular viroplankton particles as well as cell-associated phages, to better characterize reproductive and metabolic strategies of viruses in the open ocean. We use the term “reproductive strategies” here to refer life history differences between strictly lytic viruses with a singular strategy of host lysis, versus temperate phages that in addition to the lytic cycle, have the potential to either integrate into their host’s genome or reside as an extra-chromosomal element. In this study, we analyzed samples collected over 1.5 years over 12 depths (5–500 m) to characterize dsDNA viruses in both viroplankton and bacterioplankton size fractions. This virus genome dataset (referred to here as the ALOHA 2.0 viral database) provides new perspectives on the diversity, reproductive strategies, gene content, and ecology of viruses in the open ocean.

Methods

A schematic overview of our workflow is presented in Fig. S1.

Sample collection, extraction, and sequencing

Station ALOHA (22°45′ N, 158° W), a relatively seasonally stable environment located in the NPSG, is a well-characterized sampling site of the Hawaii Ocean Time-series (HOT) program. The ALOHA 2.0 dataset contains 374 metagenomic samples collected at 12 depths (5, 24, 45, 75, 100, 125, 150, 175, 200, 225, 250, and 500 m) at approximately monthly intervals of 16 timepoints spanning 1.5 years from 2014 to 2016 (Fig. S2). These collections correspond to HOT cruise numbers 267–283, for which physiochemical data are available in Table S1 and on Hawaii Ocean Time-Series HOT-DOGS application (<http://hahana.soest.hawaii.edu/hot/hot-dogs>).

All samples were collected using the following procedure. 2–4 L of seawater (2 L from 5 to 175 m, 4 L from 200–500 m) were collected using CTD-attached Niskin bottles and filtered, using peristaltic pumps at a flow rate of about 6 L/h, onto a 0.2 µm 25 mm Supor filter (VWR 28147-956) housed in polypropylene filter holder (Cole-Palmer EW-06623-32). These >0.2 µm cell-enriched samples were removed from filter holder and stored in 300 µL of RNALater (Ambion AM7021, Waltham MA) at –80 °C. 1–2 L (1 L from 5–175 m, 2 L from 200–500 m) of the < 0.2 µm filtrate were collected and filtered, using peristaltic pumps at a flow rate of about 1 L/h, onto a 0.02 µm Whatman Anotop filter (VWR 28138-017, Radnor PA). These corresponding 0.2–0.02 µm virus-enriched samples

were stored sealed in filter housing at –80 °C. DNA extraction, sequencing, and read quality-control are described in the Supplementary methods.

Quality-controlled reads, on average 9–10 million per sample (Table S1), were assembled within each sample using option “-k 21,33,55,77,99,127” on metaSPAdes v3.10.1 [28], generating a total of 83 million contigs amongst 416 metagenomes (Table S1). All sequences were used for viral reassembly and database curation. For downstream analyses, smaller metagenomes from samples with duplicate sequencing runs and samples with <1 million reads were removed, resulting in a total of 374 metagenomes. Sequences were submitted to NCBI SRA under project number PRJNA352737 and assemblies can be found under BioSample SAMN12604809.

Viral-specific reassembly

All >3 kb contigs were filtered using VIRSorter v1.03 [29] using the virome database, under regular mode for cell-enriched samples and decontamination mode for virus-enriched samples. Contigs from all identified viral categories were retained (Table S2). BWA-MEM v0.7.15 (Li 2013) and msamtools [30] was used to identify 809 million reads mapping to these putative viral contigs at >95% average nucleotide identity (ANI) across >45 bp (Table S2). Viral reads were reassembled using metaSPAdes v3.11.1, which was chosen due to improved genome recovery and low rate of generating false apparent circularity [31]. Reads from contigs with >10 coverage, a threshold representing 99% genome recovery [31], were pooled across each depth for 12 reassemblies. Reads from contigs with <10 coverage were pooled across all samples into a low-coverage reassembly to improve genome recovery.

ALOHA 2.0 virus database curation

Viral contigs across the 13 reassemblies, along with viral contigs from two previous smaller datasets near Station ALOHA [26, 32], were clustered with cd-hit-est v4.6 [33] at >95% ANI to form 1.5 million nonredundant viral populations. Populations were filtered through VIRSorter and 262 197 putative viruses from all categories were retained. Proteins were predicted using Prodigal v2.6.3 [34] and functionally annotated using HMMer v3.2 [35] against the PFAM-A v30 database [36]. Populations containing one or more known viral marker proteins were retained (bit score >30 to capsid, head, neck, tail, spike, portal, terminase, clamp loader, T4 proteins, T7 proteins, Mu proteins, excisionase, phage integrase, repressor protein CI, or Cro), resulting in 56,559 high-confidence virus populations ranging from 0.5 to 366 kbp in length. To focus on full

genomes or large genomic fragments, we retained only >10 kbp contigs, resulting in 17,369 populations that form the ALOHA 2.0 virus database. Functional annotations were inspected to ensure that no ribosomal proteins are present, with the exception of S21 found also in a cultivated *Pelagibacter* phage, S33 found enriched in aquatic viruses, and L7/12 found in assembled viral contigs [37]. One population containing L11 was retained due to the protein's proximity and interaction with L7/12 [38].

The high proportion of novel viral diversity in our samples precludes using reference genomes for the detection of chimeras, which are expected to occur at a frequency of ~0.5% using metaSPAdes [31]. As a result, we inspected clusters for chimeric signature through self-alignment using LAST v756 [39] to identify stretches of repeats at >95% ANI across >10 kbp. Fifteen populations displayed this signature and were noted as chimeras (Table S3).

Genomic completion

We used terminal repeats (apparent circularity) to identify a nonredundant set of complete genomes [40], using a combination of four different methods: i. 439 were identified using Virsorter ii. 411 from check_circularity.pl [41] iii. 790 from LAST to identify overlaps at 95% ANI across 100 bp–10 kbp within 200 bp of both ends; and iv. 1131 from NUCmer v3.1 [42] to identify direct terminal repeats 20bp–10kbp in length within 200 bp of both ends [43]. 1543 complete genomes were identified pooled amongst these four methods (Table S3). LAST was used to assess redundant circularly permuted contigs at 95% ANI across 150 bp, yielding 961 nonredundant complete genomes (16,787 total, Fig. S1).

Viral and prokaryotic contribution to total DNA

Viral contribution was calculated for each sample as the proportion of reads mapping to the ALOHA 2.0 viral database (>95% across >45 bp). Prokaryotic contribution was calculated for each sample as the proportion of reads classified as bacterial or archaeal with Kaiju v1.6.2 [44].

Spatiotemporal distribution and abundance

Reads from each sample were mapped using BWA-MEM to virus populations and filtered using msamtools at >95% across >45 bp. Anvi'o v3 [45] was used to calculate coverage profiles for every sample, using interquartile range (IQR) coverage, which diminishes the effect of conserved or hypervariable regions in respectively over- and underestimating coverage. For analyses including all virus populations, relative nucleotides mapped (nucleotides mapped to population divided by nucleotides mapped

across all populations) was used to calculate relative abundances (Tables S4, S5). For analyses including only complete viral genomes, relative coverage (genome coverage divided by total coverage summed across all complete genomes) was used to approximate relative abundances (Tables S6, S7).

To examine temporal persistence, reads from a previous 2010–2011 dataset from Station ALOHA [46] were mapped to the 2014–6 ALOHA 2.0 virus database using BWA-MEM and filtered using msamtools at >95% across >45 bp. Populations with non-zero IQR coverage were considered present in the 2010–2011 dataset (Table S3).

Characterizing cellular assemblages

COG0012, a universal single-copy marker protein, was used to generate 2568 mOTUs representing cells at the near-species level, at higher resolution than with rRNA-based OTUs [46–48]. One *Crocospaera* mOTU not previously included was curated, and relative abundances were calculated using relative coverages of reads mapping to these 2569 mOTUs (Table S8).

Viral population identification

Taxonomy was assigned using protein LAST alignments to known phages in RefSeq84 ([49]; Table S9), as well as five viral metagenomic databases available as of 2018: uvMED [50], uvDEEP [23], GOV [15], EV [16], and MED2017 [24]. To avoid inflating the number of novel populations, we performed broad taxonomic assignments at >60% average amino acid identity (AAI) across >50% of proteins to any reference genome or contig, with the best hit assigned based on the highest AAI (Table S3). A broader cut-off of >50% of proteins at any AAI was used to identify phages infecting heterotrophic bacterioplankton in RefSeq84, due to lower sequence representation than picocyanophages in these databases (Table S3).

Proteins were annotated using HMMsearch against PFAM (bit score >30). Proteins with functional domains not found in previously reported datasets [15, 26] were considered novel viral genes (Table S10).

Putative temperate phages were identified using i. functional annotations to identify phages with the genomic potential for lysogeny and ii. VIRSorter to identify integrated prophages. i. Functional annotations identified 922 populations with temperate phage markers (>30 bit score to integrase, excisionase, Cro, or CI repressor). ii. VIRSorter identified prophages from original assemblies, of which 413 temperate phage populations shared significant homology (a minimum of 150 bp at 95% ANI); VIRSorter identified 73 final viral populations as prophages.

Archaeal virus identification

Putative archaeal viruses were identified in unannotated populations using archaeal protein markers and sequence similarity to Archaea or archaeal viruses in RefSeq84 using archaeal markers (PFAM bit score >30). Populations carrying protein markers included one with AmoC [51], 16 with archaeal holiday junction resolvase, and 37 with MCM DNA helicase of archaeal origin (Table S9), yielding a total of 53 archaeal virus marker populations. 632 populations were identified having ≥ 1 proteins with top hits (protein–protein LAST) to Archaea or archaeal viruses in the RefSeq84 database [49]. We refined this set with the expectation that archaeal viruses should display high ratio of top protein hits to Archaea and archaeal viruses divided by top hits to Bacteria (AB ratio) and large proportion of proteins without RefSeq84 hits, given the scarcity of open ocean archaeal viruses in current databases. The 53 archaeal virus marker populations all displayed >0.06 AB ratio and >0.8 proportion of proteins without RefSeq84 hits (Fig. S3b). We used modified, conservative cut-offs (>0.5 AB ratio, >0.8 proportion of proteins without RefSeq84 hits) to retain 161 putative archaeal viruses (Fig. S3c). Phylum-level classifications were assigned based on number of protein top hits to Thaumarchaeota and Euryarchaeota. Populations with ties in number of top hits to these phyla were considered unclassified archaeal viruses.

Eukaryotic virus identification

Putative eukaryotic viruses were identified using the nucleocytoplasmic large DNA virus capsid marker (PFAM bit score >30) and ≥ 2 proteins with top hits to eukaryotic viruses in RefSeq84 (Table S9). All taxonomic identifications were cross-referenced to confirm that a population is assigned only to one group.

Crocospaera putative phage identification

Putative *Crocospaera* phages were identified through a combination of methods. First, alignments to RefSeq84 revealed one putative *Crocospaera* phage (Table S3) with 11 of 15 proteins hitting to *Crocospaera*, and one hit to another Cyanobacteria (Table S9). For independent confirmation, this virus displayed abundance profiles expected from *Crocospaera* (i.e., summer bloom in the upper ocean). 170 other potential *Crocospaera* phages that displayed similar spatiotemporal distributions (Fig. S9) were also included as candidates for independent confirmation. Reads from samples capturing a *Crocospaera* bloom (and presumably its phages) near Station ALOHA around the same time [52] were mapped (>95% across >45 bp) to these 171 populations to confirm presence. 115 candidates that

recruited reads at >0 IQR coverage were retained. We then retained only candidates with higher number of top hits to non-*Prochlorococcus/Synechococcus* cyanobacterial proteins than top hits to *Prochlorococcus* or *Synechococcus* proteins, for a final set of one “putative” and six “potential” *Crocospaera* phages (Table S3).

Viral and prokaryotic diversity

To examine within-sample α -diversity, we calculated the Shannon diversity, evenness, and richness for each sample using the vegan package in R [53]. Respective assemblages were assessed using relative coverages of 16,787 viral populations and 2568 cellular mOTUs.

VC ratio

We define the Virus:Cell ratio (VC ratio) here as the log-ratio of a population’s relative abundance (nucleotides mapped) in the virus-enriched versus cell-enriched size fractions. To calculate VC ratios, samtools bedcov [54] was used to calculate nucleotides mapping to each population from all samples, normalized to the average library size of 1.3 million nucleotides per sample. Populations with zero nucleotides mapping in any sample were adjusted to one nucleotide for log-ratio calculation. Temporal variability of VC ratios was calculated for each population using the mean-normalized variance of its VC ratios within each depth.

Results and discussion

A total of 374 metagenomes were generated from seawater sampled at Station ALOHA across 16 timepoints at 12 depths (5–500 m). These samples were used to prepare DNA from cell-enriched >0.2 μm size fractions (encompassing cellular DNA, giant viruses, active infections, prophages, and absorbed phages), and virus-enriched 0.2–0.02 μm size fractions (capturing ultra-small cells, free virus particles, and ultra-small detritus). After metagenome assembly, curation, and classification (Fig. S1), the 4.2 TB of sequence data yielded 16,787 virus populations, with 8079 novel populations not represented in publicly available virus sequences. A total of 961 assembled populations were identified as complete via evidence of their terminal repeats [55, 56]. 1352 populations were identified as putative temperate phages, due to the presence of flanking cellular sequences or diagnostic temperate phage marker genes. Among novel viruses, 29 putative temperate *Pelagibacter* and *Puniceispirillum* phages were identified, along with 12 complete archaeal virus genomes, 25 genomic fragments of eukaryotic viruses, and a putative

Crocospaera phage-like element (Fig. S4, S5). The curated viral populations contained 236 novel viral proteins (Table S10) not previously reported from marine environments [15, 26]. These proteins included auxiliary metabolic genes in nitrogen (nitronate monooxygenase) and phosphorus cycling (transporter), phage holin, a toxin–antitoxin system (Fig. S4), and CRISPR-associated proteins.

Viral DNA contribution to total DNA

On average across all depths, the ALOHA 2.0 viral dataset recruited 42% of reads from the virus-enriched fraction and 8% of reads from the cell-enriched fraction (Fig. 1). Assuming that viruses account for 56% of the total DNA in the <0.22 μm fraction [57], our dataset recovered on average 75% of the total viral DNA from this habitat. In the virus-enriched fraction, viral contribution total DNA peaked near the base of the euphotic zone, 150–250 m. This increase in relative viral DNA contribution may reflect increased cellular turnover and viral production, increased temperate phage induction (described later), or decreased decay rates due to lower ambient UV light fluxes and temperature at these depths [58]. In the cell-enriched fraction, relative viral contribution to total DNA peaked near the deep chlorophyll maximum (DCM) around 100 m, extending previous observations of subsurface cyanophage maxima in cell-enriched fractions at Station ALOHA [59]. Since other sources of DNA contribution are mostly cellular in origin, the increase in ratio of viral-to-bacterial DNA likely indicates a combination of increased active infections and prophages in these cell populations.

Viral diversity hotspot at the base of the euphotic zone

Within-sample α -diversity and richness revealed hotspots for viroplankton diversity, and confirmed it for prokaryotic diversity [60], at the base of the euphotic zone between 150 and 250 m (Fig. 2). This peak in diversity could reflect both habitat variability and transitions in microbial metabolic diversity. For example, in this environment, photoautotrophic cyanobacteria dominate in the photic zone, while chemolithotrophic ammonium-oxidizing Thaumarchaeota and (presumptive) heterotrophic Euryarchaeota are both more abundant in deep waters. However, both cyanobacterial and archaeal groups do co-exist at transitional depths in and just below the deep chlorophyll maximum [60]. Similarly, while cyanophages were found predominantly in shallow waters and archaeal viruses mostly below the photic zone, both cyanophages and archaeal viruses were found at transitional depths between 150 and 250 m (annotations shown on Fig. 3, top bar). Previously uncharacterized viruses also increased in

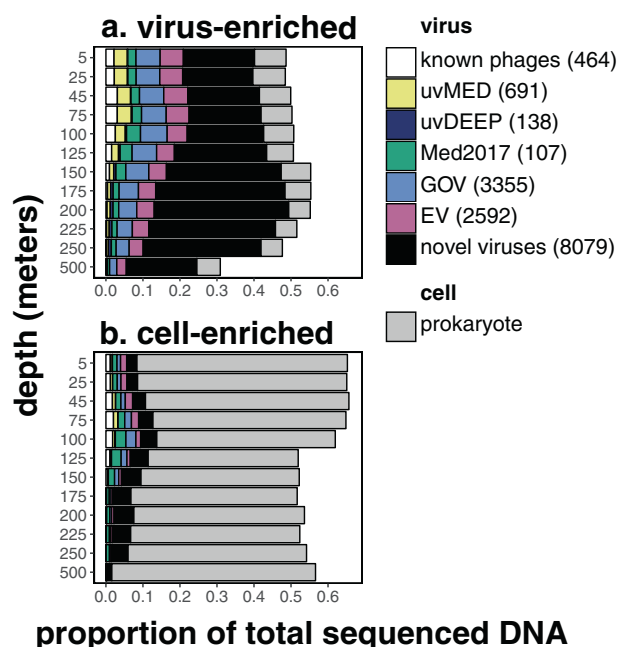


Fig. 1 Depth profiles of time-averaged viral and prokaryotic contributions to total sequenced DNA in **a** virus-enriched and **b** cell-enriched size fractions. Relative abundances of viral populations are colored by average amino acid identity (AAI) (>60% AAI across >50% genes) to known phages in RefSeq, and five other viral metagenomic datasets: uvMED [50], uvDEEP [23], Med2017 [24], GOV [15], and EV [16]. Legend shows number of populations identified in each category.

abundance below the upper surface waters, and represented >50% of viral assemblages below 125 m (Fig. 1). Elevated richness in oligotrophic picoplankton communities just below the photic zone was evident not only in Bacteria and Archaea, but also their viruses.

Depth-dependent patterns in temperate phage integration and induction

The community-wide abundances of 1352 putative temperate phages displayed depth-specific patterns (Fig. 4). In cell-enriched samples, we postulate that putative temperate phages represented integrated prophages, which increased at and below the DCM, consistent with previous reports of apparent increased lysogeny in deeper waters [26, 61, 62]. Depth-dependent changes in viral reproductive strategies were also evident within groups at the genus level, particularly in SAR11 phages with broad depth ranges (Fig. S7). Our results suggest that community-wide changes in reproductive strategies are driven not only by specific host groups, but also partly by environmental variability along the water column. Given that cellular host abundance generally decreases with depth (Fig. S2c), our results do not appear to support the piggyback-the-winner [63] hypothesis in this oligotrophic pelagic habitat. It seems probable that

Fig. 2 α -diversity depth profiles of viral and prokaryotic assemblages: a Shannon diversity, **b** richness (number of populations), and **c** evenness (Shannon diversity divided by log richness). Solid and open circles represent viral and prokaryotic assemblages, respectively, averaged through time (mean \pm SE).

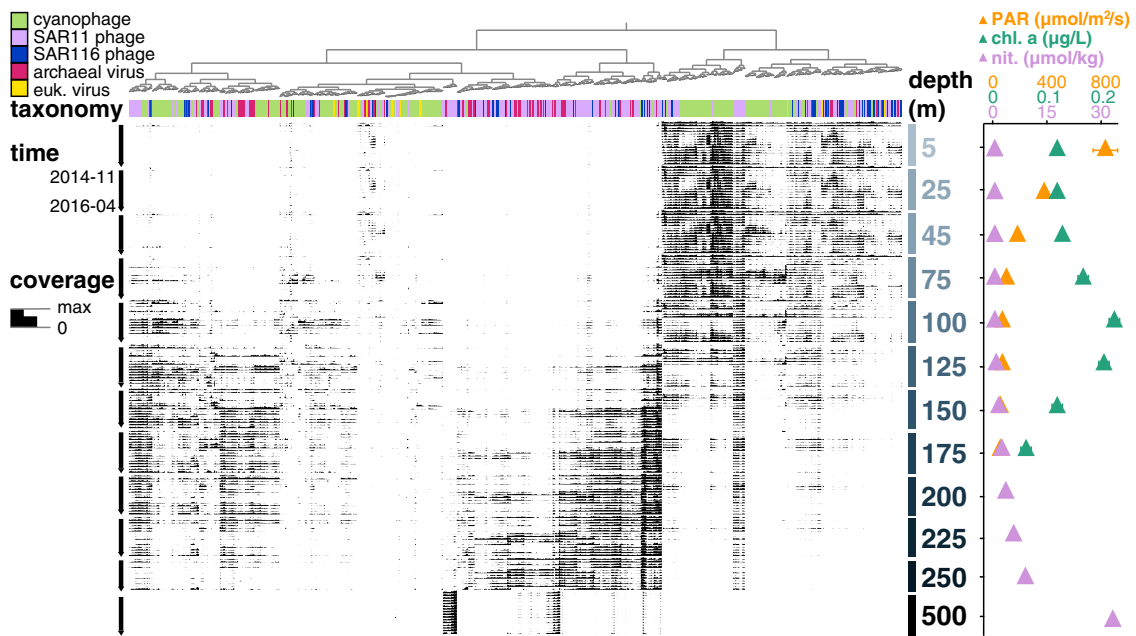
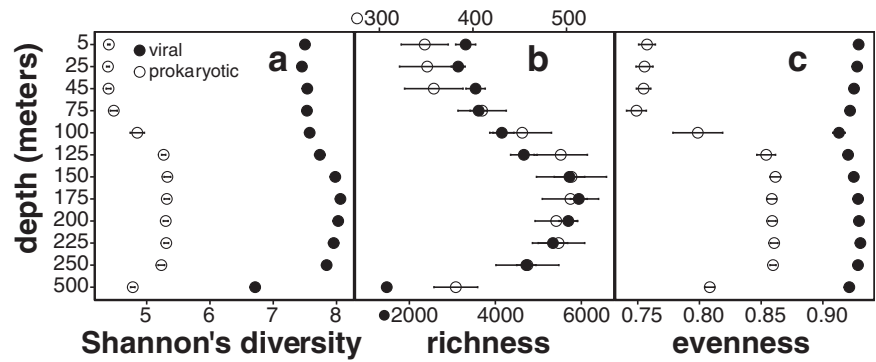


Fig. 3 Spatiotemporal distributions of annotated virus populations present in the virus-enriched size fraction. Each node on the top dendrogram and its associated column represents the coverage profile of one virus population, colored by its corresponding taxonomic annotation. Rows represent individual samples that are horizontally ordered by depth and time. The height of the black bar in every sample shows mean interquartile range (IQR) coverage of every population,

normalized to the maximum IQR coverage in that sample. Time-averaged depth profiles (mean \pm SE) of environmental variables of photosynthetically active radiation (PAR), fluorometric chlorophyll a, and inorganic nitrogen are shown in colored triangles on the right panel (data retrieved from Hawaii Ocean Time-Series HOT-DOGS application).

virus-host dynamics may have different trajectories in different environmental, biological, and ecological contexts, and may not be driven simply by bulk numerical host-cell and virus-particle ratios alone. In oligotrophic pelagic environments, high prokaryote cell densities in surface waters correspond to smaller average genome sizes [46, 64]. Within and below the DCM, low cell densities correspond to larger average genome sizes [46]. Host cell genome sizes (and therefore genomic “real estate” available for prophage, genomic islands and other mobile elements), rather than cell densities, might drive viral reproductive strategies [65].

In virus-enriched samples, we postulate that temperate phage signal represents temperate virus particles, which

peaked at 150–250 m (Fig. 4). This peak may reflect increased temperate phage productivity relative to deeper waters associated with more slowly growing hosts [66]. Temperate phage production at these intermediate depths appeared to be episodic, as indicated by temporal variability in populations’ extracellular to intracellular abundance (VC) ratios (Fig. 5). Low VC variability reflects phages with constant or no viral particle production. High VC variability indicates phages with more episodic virus particle production. The VC variability of temperate phages peaked at 150–250 m (Fig. 5), consistent with episodic induction and production. Episodic temperate phage induction may be driven by temporal resource variability or host abundance

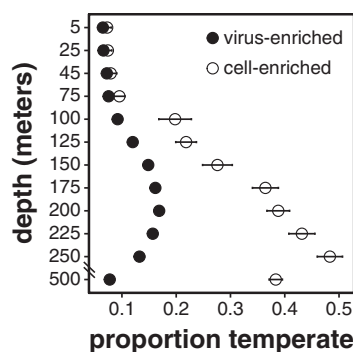


Fig. 4 Depth profiles of relative abundances of putative temperate phages. Circles represent time-averaged depth profiles (mean \pm SE) of the proportion of all viruses that were identified as temperate in the virus-enriched size fraction (closed circles) and cell-enriched size fraction (open circles).

[67–69], or increased cellular stress due to lower light for photoautotrophs at these depths (Fig. S2f).

Given the low abundance of temperate phages in the upper 75 m (Fig. 4), we infer that lytic phage–host interactions were prevalent in surface waters, potentially reflecting consistently high productivity and host abundance that can favor lytic strategies [66, 70–72]. Temperate phage induction and production appeared to peak at the base of the euphotic zone (Figs. 4, 5), potentially reflecting increased environmental variability at these depths that might favor a flexible reproductive strategy [69, 73]. Evidence for host-integrated prophages increased in mesopelagic waters (Fig. 4), possibly reflecting decreases in host productivity and abundance [72, 73], or an increase in genome size that might better accommodate prophages [46, 65]. Using size fractionation to separate intra- and extracellular viruses, our results revealed depth-specific viral reproductive strategies from the surface to mesopelagic ocean.

VC ratio to confirm genomic temperate phage identification

The VC ratio of each population represents its extracellular to intracellular abundance ratio. On average, temperate phages, or phages persisting intracellularly for long periods before initiating a lytic cycle, are expected to have lower VC ratios relative to strictly lytic phages. Indeed, temperate phages on average displayed lower VC ratios than that of other coexisting populations (Welsh's t test, $p = 0.02$). This trend provided further support for our temperate phage identifications using genomic markers. Nevertheless, intracellular temperate phage counts may reflect a variety of life history states. Intracellular temperate phage counts could represent prophage DNA still integrated in the host genome, replicating temperate phage that have excised from the host genome, or temperate phage that entered directly into the lytic cycle immediately post-infection (in eclipse phase).

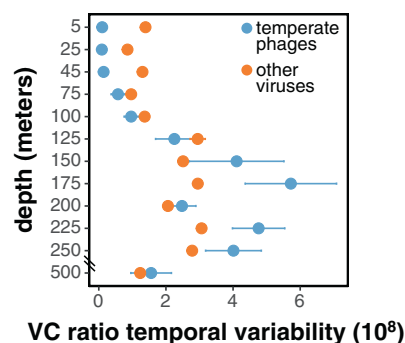


Fig. 5 Depth profiles of temporal variabilities of VC ratios (mean \pm SE) of 1352 inferred temperate phages (blue) and other viral populations (orange). Higher VC ratio temporal variability indicates episodic production of free viral particles, while lower variability indicates consistent production of free viral particles. Temporal variability of VC ratios is calculated for each population by pooling its VC ratios within each depth to determine the mean-normalized variance.

Presuming that the temperate phage signal in the virus-enriched fraction represents packaged phage particles, ~98% of putative temperate phages appeared to have the ability to produce viral particles in situ (Fig. S6b, Table S3). Hence, the lower VC ratios found in the temperate phages does appear to reflect their different lifestyles and reproductive strategies, relative to non-temperate phages.

Environmental distribution of viral AMGs

Viral copies of AMGs potentially encode for rate-limiting steps in host metabolism that are essential for viral propagation (reviewed in [13]). We show here that the abundance of three key AMGs in energy and nutrient acquisition vary with depth, potentially indicating cellular host variability or energy or nutrient limitation at specific depths.

Virus-encoded copies of photosystem reaction center genes were first observed in a *Synechococcus* phage genome, were thought to prevent photoinhibition by supplementing declining host photosynthesis during infection [74], and were found to be co-expressed with high-light gene in cyanophages infecting high-light *Prochlorococcus* hosts [4, 75]. Photosystem genes were subsequently observed in many cyanophage genomes [76], yet their environmental distribution in the water column remains relatively unexplored. Here, we observed a threefold increase in abundance of phage carrying photosystem genes from the surface waters to low-light environments around the DCM (Fig. 6a). Considering that cyanophages and cyanobacteria were most abundant in the surface ocean (Figs. 6a, S8a), a higher proportion of cyanophages carried photosystem genes at the DCM relative to surface waters. Our observations suggest that virus-encoded photosystem genes may be more advantageous in light-limited conditions, potentially to prevent light-energy limitation in hosts

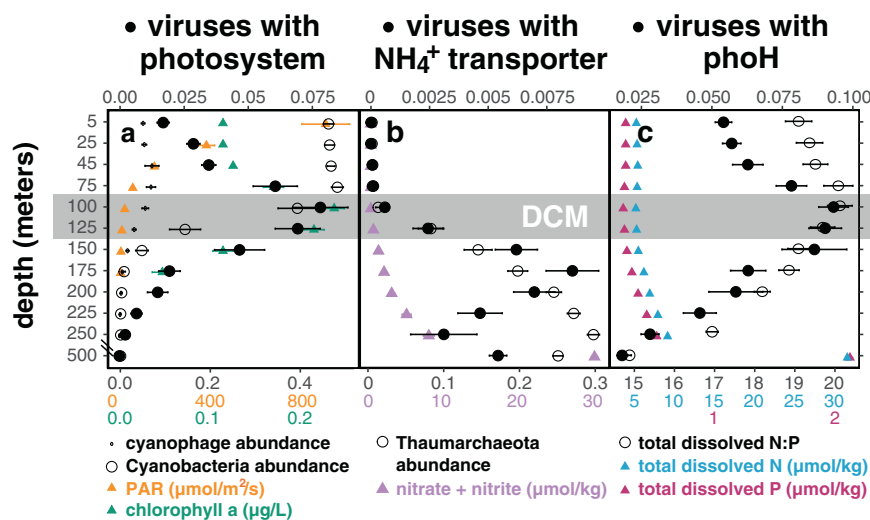


Fig. 6 Distribution of viruses containing specific auxiliary metabolic genes (AMGs) in the water column. Solid circles show depth profiles, averaged through time (mean \pm SE), of the abundance-normalized proportion of virus populations carrying auxiliary metabolic genes respectively for carbon, nitrogen, and phosphorus metabolism: **a** photosystem, **b** ammonium transporter, and **c** phoH. Time-averaged depth profiles of the following are included for environmental context: **a** relative abundances of cyanophages in the virus-enriched fraction (small open circles) and of cyanobacteria in the

cell-enriched fraction (large open circles), photosynthetically active radiation (PAR) (yellow triangles), and fluorometric chlorophyll a (green triangles); **b** Thaumarchaeal relative abundance in the cell-enriched fraction (open circles) and inorganic nitrogen concentrations (purple triangles). **c** total dissolved nitrogen (blue triangles), phosphorus (red triangles), and their ratio (open circles). Environmental metadata were retrieved from Hawaii Ocean Time-Series HOT-DOGS application. Gray box highlights depths around the deep chlorophyll maximum (DCM).

during infection. Alternatively, longer latent periods, consistent with our observed increase in viral DNA in the cell-enriched fraction at the DCM (Fig. 1a), might also favor phage-encoded photosystem genes at this depth [76].

Viral-encoded ammonium transporter genes, which shared highest similarity to bacterial homologs (Table S9), increased at and below the DCM in tandem with Thaumarchaeota abundance (Fig. 6b). Ammonium-oxidizing Thaumarchaeota in the ocean have demonstrated high affinity for ammonia [77] and therefore may compete with bacteria in ammonia-limited open-ocean environments [78]. As a result, bacteriophage-encoded copies of ammonium transporter genes might assist in ammonium acquisition during infection, particularly at depths with higher abundances of Thaumarchaeota competitors.

PhoH genes are used during phosphorus starvation [79], upregulated during cyanophage infection [80], present in diverse groups of marine viruses [5, 7, 81–83], and can be used as a marker gene for assessing viral assemblages [26, 84]. Viruses are richer in phosphorus relative to cells [85], potentially driving an enrichment of this auxiliary metabolic gene in marine viruses [84] that inhabit relatively nutrient-limited open oceans. At Station ALOHA, up to 10% of viruses encoded copies of phoH genes, peaking at the DCM in tandem with a maximum in the total dissolved nitrogen to phosphorus ratio (Fig. 6c). Our observations suggest that the phoH gene might assist in phosphorus acquisition during viral infection in relatively phosphorus-limited environments.

Spatiotemporal distributions: temporal persistence and depth distributions

Many viroplankton populations displayed temporal persistence through the 1.5-year time series (Figs. 3, S9), consistent with recent studies at Station ALOHA [26, 32]. However, some viroplankton populations occurred only at specific times of year (Fig. S10), potentially reflecting the seasonal variability of physiochemical variables and their cellular hosts. Of the 9579 viral populations that recruited reads from the ALOHA 2.0 cell-enriched samples from 2014–2016, 6959 (76%) also recruited reads from cell-enriched samples from 2010–2011 (Table S3). This strong temporal persistence indicates that viral genome evolution may be under stabilizing selection, constrained to very specific loci, or occur at finer resolutions than 95% ANI used here for read-mapping and defining populations. Our results show that many indigenous phage populations persist over multiannual cycles, indicative of continuous infection, replication, and lytic cycles that maintain this persistence.

Viral assemblages displayed stratified depth structure, with shifting depth ranges along the transition zone between surface and mesopelagic waters (Figs. 3, S9). This depth structure likely reflects biogeochemical gradients [86] and depth partitioning of bacterial host communities [46]. Consistent with a previous study at Station ALOHA [26], we found no evidence of eurybathic viruses in either the cellular or viroplankton enriched size fractions.

Depth distributions of archaeal viruses and their hosts

Several different archaeal lineages, including Euryarchaeota and Thaumarchaeota (formerly Group I marine Crenarchaeota [87]) are commonly found along the water column at Station ALOHA [59, 88, 89]. Euryarchaeota apparently function as organic matter degrading heterotrophs, while Thaumarchaeota appear to function as chemolithoautotrophic ammonium oxidizers [78, 90]. We identified 161 putative archaeal virus populations, 12 of which represented complete genomes. We classified 67 of these populations as euryarchaeal and 60 as thaumarchaeal, with the latter encompassing crenarchaeal populations (Table S3). In both size fractions, total archaeal virus abundance peaked at and below the DCM (Fig. S8b), with putative thaumarchaeal viruses driving most of this change. Relative abundances of archaeal viruses were consistent, though somewhat lower in magnitude, with that of archaea (Fig. S8b), suggesting that our methods of archaeal virus identification were conservative but accurate.

Putative *Crocospaera* phages

The unicellular cyanobacterium *Crocospaera* can fix both nitrogen and carbon, and plays an important role in fueling primary production in nutrient-poor open ocean environments [91]. Although it is the third-most abundant cyanobacterium at Station ALOHA, after *Prochlorococcus* and *Synechococcus* [52], no *Crocospaera* phage sequences have yet been identified in current databases. We identified the genome of a putative *Crocospaera* phage or phage parasite based on multiple lines of evidence: sequence similarity to *Crocospaera* in 11 of 15 proteins, similar spatiotemporal abundance profiles to that expected from *Crocospaera* (summer in the upper ocean), and presence during a confirmed *Crocospaera* “bloom” near Station ALOHA around the same time [52]. This genome contains a higher GC content (41.7%) than other co-occurring viruses (37.3%), consistent with the higher GC content of *Crocospaera* at ~37.4% [92], compared with abundant surface *Prochlorococcus* at ~32% [46]. Accounting for up to 2% of viral DNA in the virus-enriched size fraction, this genome likely represents a phage, plasmid, or phage-inducible chromosomal island (PICI) packaged into phage particles. PICIs are phage parasitic DNA elements that can hijack another infecting phage’s packaging machinery and mobilize in infectious phage-like particles [93, 94]. The genome (11 kbp) appears complete and encodes a phage integrase (Fig. S6), characteristic of both temperate phages and PICIs. In addition to this genome, we identified six other potential *Crocospaera* phages (Table S3) based on similar Station ALOHA abundance profiles to

Crocospaera, their presence during a *Crocospaera* “bloom” in one set of samples, and their dissimilarity from *Prochlorococcus* or *Synechococcus* gene homologs.

Conclusion

Viruses impact the ecology and biogeochemistry of microbial communities across the oceans. Our study recovered a large fraction of dsDNA viruses found at Station ALOHA, revealing a hotspot for microbial diversity at the base of the euphotic zone, where the majority of viruses were distinct from those previously reported. The concurrent characterization of both intracellular and extracellular viruses provided independent support of temperate phage identification using marker genes, and revealed community-wide shifts in viral reproductive strategies. In this open-ocean environment, lytic interactions dominated the upper photic layer above the DCM, and temperate phages were most abundant in the mesopelagic ocean. Temporal variability in reproductive strategies appeared most prevalent in transitional depths at the base of the euphotic zone, marked by a peak in prophage integration and induction. Environmental distributions of viral AMGs also displayed depth-specific patterns in energy and nutrient acquisition. For example, photosystem genes were most abundant at light-limited depths around the DCM; bacteriophage-encoded ammonium transporter genes increased along with potential ammonia-oxidizing thaumarchaeal competitors; phosphorus starvation genes increased in tandem with N:P ratio. Although most viruses were temporally persistent over several years, some displayed temporal variability that appeared to reflect the seasonal distributions of specific hosts. These temporal patterns, in conjunction with other lines of evidence, led to the identification of putative *Crocospaera* phages or phage parasites that have not been previously identified. Taken together, these new data and analyses provide new insight on the spatiotemporal patterns of planktonic viral diversity, reproductive strategies and metabolic repertoires in the open ocean. Furthermore, simultaneous enumeration of extracellular and intracellular viruses and their hosts sets the stage for delineating more specific host-virus interactions.

Acknowledgements We thank the captain and crew of R/V Kilo Moana, R/V KOK, Hawaii Ocean Time-series, and SCOPE-ops team for cruise organization, sample collection, and oceanographic data acquisition. We thank Jackie Mueller and Grieg Stewart for their previous methods development using Anotop filters, and Paul Den Uyl for contributions to library preparation and sequencing. We thank Wei Qin, John Beaulaurier, Dominique Boeuf, Fuyan Li, Matthew Sullivan, and Murat Eren for discussions on data analysis. This project is funded by Simons Foundation (#329108) and the Gordon and Betty Moore Foundation (GBMF 3777) to EFD. Partial support for EL was

provided by the Natural Sciences and Engineering Research Council of Canada (PGSD3-487490-2016). This work is a contribution of the Simons Collaboration on Ocean Processes and Ecology and the Center for Microbial Oceanography: Research and Education.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Wigington CH, Sonderegger DL, Brussaard CPD, Buchan A, Finke JF, Fuhrman J, et al. Re-examining the relationship between virus and microbial cell abundances in the global oceans. *Nat Microbiol.* 2016;1:15024.
- Karl DM, Church MJ. Microbial oceanography and the Hawaii Ocean Time-series programme. *Nat Rev Microbiol.* 2014;12:1–15.
- Wilson WH, Joint IR, Carr NG, Mann NH. Isolation and molecular characterization of five marine cyanophages propagated on *Synechococcus* sp. strain WH7803. *Appl Environ Microbiol.* 1993;59:3736–43.
- Lindell D, Sullivan MB, Johnson ZI, Tolonen AC, Rohwer F, Chisholm SW. Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci.* 2004;101:11013–8.
- Rohwer F, Segall A, Steward G, Seguritan V, Breitbart M. The complete genomic sequence of the marine phage Roseophage. *Limnol Ocean.* 2000;45:408–18.
- Garcia-heredia I, Rodriguez-Valera F, Martin-Cuadrado A. Novel group of podovirus infecting the marine bacterium *Alteromonas macleodii*. *Bacteriophage.* 2013;3:e24766.
- Zhao Y, Temperton B, Thrash JC, Schwalbach MS, Vergin KL, Landry ZC, et al. Abundant SAR11 viruses in the ocean. *Nature.* 2013;494(Feb):357–60.
- Kang I, Oh H-M, Kang D, Cho J-C. Genome of a SAR116 bacteriophage shows the prevalence of this phage type in the oceans. *Proc Natl Acad Sci.* 2013;110:12343–8.
- Weinbauer MG. Ecology of prokaryotic viruses. *FEMS Microbiol Rev.* 2004;28:127–81.
- Weinbauer MG, Rassoulzadegan F. Are viruses driving microbial diversification and diversity? *Environ Microbiol.* 2004;6:1–11.
- Suttle CA. Viruses in the sea. *Nature.* 2005;437:356–61.
- Puxty RJ, Millard AD, Evans DJ, Scanlan DJ, Puxty RJ, Millard AD, et al. Viruses inhibit CO₂ fixation in the most abundant phototrophs on Earth. *Curr Biol.* 2016;26:1–5.
- Hurwitz BL, U'Ren JM. Viral metabolic reprogramming in marine ecosystems. *Curr Opin Microbiol.* 2016;31:161–8.
- Brum JR, Ignacio-espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, et al. Patterns and ecological drivers of ocean viral communities. *Science.* 2015;348:1261498.
- Roux S, Brum JR, Dutilh BE, Sunagawa S, Duhaime MB, Loy A, et al. Ecogenomics and biogeochemical impacts of uncultivated globally abundant ocean viruses. *Nature.* 2016;537:689–93.
- Paez-Espino D, Eloë-Fadrosh EA, Pavlopoulos GA, Thomas AD, Huntemann M, Mikhailova N, et al. Uncovering Earth's virome. *Nature.* 2016;536:425–30.
- Gregory AC, Zayed AA, Sunagawa S, Wincker P, Sullivan MB, Ferland J, et al. Marine DNA viral macro- and microdiversity from pole to pole. *Cell.* 2019;177:1–15.
- Chow CT, Fuhrman JA. Seasonality and monthly dynamics of marine myovirus communities. *Environ Microbiol.* 2012;4:2171–83.
- Needham DM, Chow C-ET, Cram JA, Sachdeva R, Parada A, Fuhrman JA. Short-term observations of marine bacterial and viral communities: patterns, connections and resilience. *ISME J.* 2013;7:1274–85.
- Chow CET, Winget DM, White RA, Hallam SJ, Suttle CA. Combining genomic sequencing methods to explore viral diversity and reveal potential virus-host interactions. *Front Microbiol.* 2015;6:1–15.
- Needham DM, Sachdeva R, Fuhrman JA. Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J.* 2017;11:1614–29.
- Hurwitz BL, Brum JR, Sullivan MB. Depth-stratified functional and taxonomic niche specialization in the 'core' and 'flexible' Pacific Ocean Virome. *ISME J.* 2015;9:472–84.
- Mizuno CM, Ghai R, Saghaï A, López-García P, Rodriguez-Valera F. Genomes of abundant and widespread viruses from the deep ocean. *MBio.* 2016;7:e00805–16.
- López-Pérez M, Haro-Moreno JM, Gonzalez-Serrano R, Parras-Moltó M, Rodriguez-Valera F. Genome diversity of marine phages recovered from Mediterranean metagenomes: Size matters. *PLoS Genet.* 2017;13:1–23.
- Goldsmith DB, Parsons RJ, Beyene D, Salamon P, Breitbart M. Deep sequencing of the viral *phoH* gene reveals temporal variation, depth-specific composition, and persistent dominance of the same viral *phoH* genes in the Sargasso Sea. *PeerJ.* 2015;3:e997.
- Luo E, Aylward FO, Mende DR, Delong EF. Bacteriophage Distributions and Temporal Variability in the Ocean's Interior. *MBio.* 2017;8:e01903–17.
- Sieradzki ET, Ignacio-Espinoza JC, Needham DM, Fichot EB, Fuhrman JA. Dynamic marine viral infections and major contribution to photosynthetic processes shown by spatiotemporal picoplankton metatranscriptomes. *Nat Commun.* 2019;10.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J Comput Biol.* 2012;19:455–77.
- Roux S, Enault F, Hurwitz BL, Sullivan MB. VirSorter: mining viral signal from microbial genomic data. *PeerJ.* 2015;3:e985.
- Arumugam M, Harrington ED, Raes J, Foerster KU, Arumugam M, Bork P. SmashCommunity: a metagenomic annotation and analysis tool. *Bioinformatics.* 2010;26:2977–8.
- Roux S, Emerson JB, Eloë-Fadrosh EA, Sullivan MB. Benchmarking viromics: an in silico evaluation of metagenome-enabled estimates of viral community composition and diversity. *PeerJ.* 2017;5:e3817.
- Aylward FO, Boeuf D, Mende DR, Wood-Charlson EM, Vislova A, Eppley JM, et al. Diel cycling and long-term persistence of

- viruses in the ocean's euphotic zone. *Proc Natl Acad Sci*. 2017;201714821.
33. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006;22:1658–9.
 34. Hyatt D, Chen G, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics*. 2010;11.
 35. Eddy SR Accelerated Profile HMM Searches. *PLoS Comput Biol*. 2011;7.
 36. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz H, et al. The Pfam protein families database. *Nucleic Acids Res*. 2008;36 (Database issue):281–8.
 37. Mizuno CM, Guyomar C, Roux S, Lavigne R, Rodriguez-Valera F, Sullivan M, et al. Numerous cultivated and uncultivated viruses encode ribosomal proteins. *Nat Commun*. 2019;10:752.
 38. Gudkov AT. The L7 / L12 ribosomal domain of the ribosome: structural and functional studies. *FEBS Lett*. 1997;407:253–6.
 39. Kielbasa SM, Wan R, Sato K, Kiebas SM, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. *Genome Res*. 2011;21:487–93.
 40. Nishimura Y, Watai H, Honda T, Mihara T, Omae K, Roux S, et al. Environmental viral genomes shed new light on virus-host interactions in the ocean. *mSphere*. 2017;2:e00359–16.
 41. Imai T sprai = single pass read accuracy improver [Internet]. 2013. Available from: <http://zombie.cb.k.u-tokyo.ac.jp/sprai/>
 42. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol*. 2004;5:R12.
 43. Beaulaurier J, Luo E, Eppley J, Uyl P Den, Dai X, Turner DJ, et al. Assembly-free single-molecule nanopore sequencing recovers complete virus genomes from natural microbial communities. *bioRxiv*. 2019 Apr;619684.
 44. Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun*. 2016;7:1–9.
 45. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ*. 2015;3:e1319.
 46. Mende DR, Bryant JA, Aylward FO, Eppley JM, Nielsen T, Karl DM, et al. Environmental drivers of a microbial genomic transition zone in the ocean's interior. *Nat Microbiol*. 2017;2:1367–73.
 47. Mende DR, Sunagawa S, Zeller G, Bork P. Accurate and universal delineation of prokaryotic species. *Nat Methods*. 2013;10:881–4.
 48. Sunagawa S, Mende DR, Zeller G, Izquierdo-Carrasco F, Berger Sa, Kultima JR, et al. Metagenomic species profiling using universal phylogenetic marker genes. *Nat Methods*. 2013;10:1196–9.
 49. O'Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res*. 2016;44(Database issue):733–45.
 50. Mizuno CM, Rodriguez-Valera F, Kimes NE, Ghai R Expanding the Marine Virosphere Using Metagenomics. *PLoS Genet*. 2013;9.
 51. Ahlgren NA, Fuchsman CA, Rocap G, Fuhrman JA. Discovery of several novel, widespread, and ecologically distinct marine Thaumarchaeota viruses that encode amoC nitrification genes. *ISME J*. 2018;13:618–31.
 52. Wilson ST, Aylward FO, Ribalet F, Barone B, Casey JR, Connell PE, et al. Coordinated regulation of growth, activity and transcription in natural populations of the unicellular nitrogen-fixing cyanobacterium *Crocospaera*. *Nat Microbiol*. 2017;2:1–27.
 53. Oksanen J, Blanchet FG, Friendly M, Kindt R, Legendre P, McGlenn D, et al. vegan: Community ecology package [Internet]. 2019. Available from: <https://cran.r-project.org/package=vegan>
 54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment / Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
 55. Casjens SR, Gilcrease EB. Determining DNA Packaging Strategy by Analysis of the Termini of the Chromosomes in Tailed-Bacteriophage Virions. *Methods Mol Biol*. 2009;502:91–111.
 56. Grose JH, Casjens SR. Understanding the enormous diversity of bacteriophages: The tailed phages that infect the bacterial family Enterobacteriaceae. *Virology*. 2014;468:421–43.
 57. Brum JR, Steward GF, Karl DM. A novel method for the measurement of dissolved deoxyribonucleic acid in seawater. *Limnol Oceanogr Methods*. 2004;2:248–55.
 58. Suttle CA, Chen F. Mechanisms and rates of decay of marine viruses in seawater. *Appl Environ Microbiol*. 1992;58:3721–9.
 59. Delong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard N, et al. Community genomics among stratified microbial assemblages in the ocean's interior. 2006;311:496–503.
 60. Mende DR, Boeuf D, DeLong EF. Persistent core populations shape the microbiome throughout the water column in the North Pacific Subtropical Gyre. *Front Microbiol*. 2019;10(October):1–12.
 61. Weinbauer MG, Brettar I, Höfle MG. Lysogeny and virus-induced mortality of bacterioplankton in surface, deep, and anoxic marine waters. *Limnol Oceanogr*. 2003;48:1457–65.
 62. Williamson SJ, Cary SC, Williamson KE, Helton RR, Bench SR, Winget D, et al. Lysogenic virus-host interactions predominate at deep-sea diffuse-flow hydrothermal vents. *ISME J*. 2008;2: 1112–21.
 63. Knowles B, Silveira CB, Bailey BA, Barott K, Cantu VA, Cobián-Güemes AG, et al. Lytic to temperate switching of viral communities. *Nature*. 2016;531:533–7.
 64. Swan BK, Tupper B, Sczyrba A, Lauro FM, Martinez-Garcia M, González JM, et al. Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proc Natl Acad Sci U S A*. 2013;110:11463–8.
 65. Casjens S. Prophages and bacterial genomics: What have we learned so far? *Mol Microbiol*. 2003;49:277–300.
 66. Middelboe M. Bacterial growth rate and marine virus–host dynamics. *Microb Ecol*. 2000;40:114–24.
 67. McDaniel LD, Houchin LA, Williamson SJ, Paul JH. Lysogeny in marine *Synechococcus*. *Nature*. 2002;415:496.
 68. McDaniel L, Paul JH. Effect of nutrient addition and environmental factors on prophage induction in natural populations of marine *synechococcus* species. *Appl Environ Microbiol*. 2005;71:842–50.
 69. Brum JR, Hurwitz BL, Schofield O, Ducklow HW, Sullivan MB. Seasonal time bombs: dominant temperate viruses affect Southern Ocean microbial dynamics. *ISME J*. 2015;10:1–13.
 70. Stewart FM, Levin BR. The population biology of bacterial viruses: why be temperate. *Theor Popul Biol*. 1984;26:93–117.
 71. Moebus K. Marine bacteriophage reproduction under nutrient-limited growth of host bacteria. I. Investigations with six phage-host systems. *Mar Ecol Prog Ser*. 1987;144:1–12.
 72. Thingstad TF. Elements of a theory for the mechanisms controlling abundance, diversity, and biogeochemical role of lytic bacterial viruses in aquatic systems. *Limnol Oceanogr*. 2000;45:1320–8.
 73. Williamson SJ, Houchin LA, McDaniel L, Paul JH. Seasonal variation in lysogeny as depicted by prophage induction in Tampa Bay, Florida. *Appl Environ Microbiol*. 2002;68:4307–14.
 74. Mann NH, Cook A, Bailey S, Clokie M, Amanullah A, Azam N, et al. Bacterial photosynthesis genes in a virus. 2003;424:741–2.
 75. Lindell D, Jaffe JD, Johnson ZI, Church GM, Chisholm SW. Photosynthesis genes in marine viruses yield proteins during host infection. *Nature*. 2005;438:86–9.
 76. Sullivan MB, Lindell D, Lee JA, Thompson LR, Bielawski JP, Chisholm SW. Prevalence and evolution of core photosystem II

- genes in marine cyanobacterial viruses and their hosts. *PLoS Biol.* 2006;4:1344–57.
77. Santoro AE, Casciotti KL, Francis CA. Activity, abundance and diversity of nitrifying archaea and bacteria in the central California Current. *Environ Microbiol.* 2010;12:1989–2006.
78. Martens-Habbena W, Berube PM, Urakawa H, de la Torre JR, Stahl DA, Torre J, et al. Ammonia oxidation kinetics determine niche separation of nitrifying Archaea and Bacteria. *Nature.* 2009;461:976–9.
79. Kim S, Makino K, Amemura M, Shinagawa H. Molecular analysis of the *phoH* gene, belonging to the phosphate regulon in *Escherichia coli*. *J Bacteriol.* 1993;175:1316–24.
80. Lindell D, Jaffe JD, Coleman ML, Futschik ME, Axmann IM, Rector T, et al. Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature.* 2007;449:83–6.
81. Weigele PR, Pope WH, Pedulla ML, Houtz JM, Smith AL, Conway JF, et al. Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol.* 2007;9:1675–95.
82. Weynberg KD, Allen MJ, Ashelford K, Scanlan DJ, Wilson WH. From small hosts come big viruses: the complete genome of a second *Ostreococcus tauri* virus, OtV-1. *Environ Microbiol.* 2009;11:2821–39.
83. Moreau H, Piganeau G, Desdevises Y, Cooke R, Derelle E, Grimsley N. Marine Prasinovirus genomes show low evolutionary divergence and acquisition of protein metabolism genes by horizontal gene transfer. *J Virol.* 2010;84:12555–63.
84. Goldsmith DB, Crosti G, Dwivedi B, McDaniel LD, Varsani A, Suttle CA, et al. Development of *phoH* as a novel signature gene for assessing marine phage diversity. *Appl Environ Microbiol.* 2011;77:7730–9.
85. Jover LF, Effler TC, Buchan A, Wilhelm SW, Weitz JS. The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat Rev Microbiol.* 2014;12:519–28.
86. Karl DM, Lukas R. The Hawaii Ocean Time-series (HOT) program: Background, rationale and field implementation. *Deep Sea Res Part II Top Stud Oceanogr.* 1996;43:129–56.
87. DeLong EF. Archaea in coastal marine environments. *Proc Natl Acad Sci.* 1992;89:5685–9.
88. Karner MB, DeLong EF, Karl DM. Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature.* 2001;409:507–10.
89. Church MJ, Wai B, Karl DM, DeLong EF. Abundances of crenarchaeal *amoA* genes and transcripts in the Pacific Ocean. *Environ Microbiol.* 2010;12:679–88.
90. Iverson V, Morris RM, Frazar CD, Berthiaume CT, Morales RL, Armbrust EV, et al. Untangling genomes from metagenomes: Revealing an uncultured class of marine Euryarchaeota. *Science.* 2012;335:587–90.
91. Mohr W, Intermaggio MP, LaRoche J. Diel rhythm of nitrogen and carbon metabolism in the unicellular, diazotrophic cyanobacterium *Crocospaera watsonii* WH8501. *Environ Microbiol.* 2010;12(Feb):412–21.
92. Bench SR, Ilikchyan IN, James Tripp H, Zehr JP. Two strains of *Crocospaera watsonii* with highly conserved genomes are distinguished by strain-specific features. *Front Microbiol.* 2011;2:1–13.
93. Fillol-Salom A, Martínez-Rubio R, Abdulrahman RF, Chen J, Davies R, Penadés JR. Phage-inducible chromosomal islands are ubiquitous within the bacterial universe. *ISME J.* 2018;12:2114–28.
94. Penadés JR, Christie GE. The Phage-Inducible Chromosomal Islands: A Family of Highly Evolved Molecular Parasites. *Annu Rev Virol.* 2015;2:181–201.