**Introduction**

Viruses are the most abundant biological entity in the ocean and are important players in biogeochemical cycling (Fuhrman, 1999). Viral lysis disrupts the flow of organic matter up the food chain by releasing labile dissolved organic matter (DOM) and intracellular resources such as ammonium (Fuhrman, 1999; Middelboe and Jørgensen, 2006; Shelford *et al.*, 2012; Lønborg *et al.*, 2013; Ankrah *et al.*, 2014). The composition of DOM in viral lysate is more complex than that of phytoplankton exudate, which affects the composition of the heterotrophic bacterial community (Zhao *et al.*, 2019). Though viruses are known for creation of DOM, viral infection can also cause microbes to produce sticky substances such as transparent exopolymers (TEP) which cause them to form particles (Peduzzi and Weinbauer, 1993; Vardi *et al.*, 2012; Lønborg *et al.*, 2013). Virus induced fluxes can be sporadically important in the environment (Brussaard *et al.*, 2005; Laber *et al.*, 2018; Sheyn *et al.*, 2018; Kranzler *et al.*, 2019). Viral infection has been implicated as a mechanism for cyanobacteria in oxygen deficient zones to become incorporated into particles (Fuchsman *et al.*, 2019). Reduced grazing in anoxic oxygen deficient zones may cause viral infection to be the main source of mortality in these regions (Zakem *et al.*, 2020).

Cyanophages, viruses that infect cyanobacteria, are abundant and potentially important causes of mortality for the marine picocyanobacteria *Prochlorococcus* and *Synechococcus* (Sullivan *et al.*, 2003; Parsons *et al.*, 2012; Baran *et al.*, 2018). *Prochlorococcus* and *Synechococcus,* are abundant under different ocean regimes, with *Prochlorococcus* dominating in the oligotrophic open ocean from 40°N to 40°S and the geographic extent of *Synechococcus* reaching to both polar regions and high nutrient regimes (Flombaum *et al.*, 2013). Picocyanobacteria contribute greatly to primary production, especially offshore (Rii *et al.*, 2016). They also contribute significantly to organic C inputs to oxygen deficient zones (<10 nM O2) when they are present (Fuchsman *et al.*, 2019). *Prochlorococcus* can be phylogenetically separated into ecotypes that are adapted to significantly different conditions including high light surface waters, very low light deep waters, warmer and colder temperature ranges and oxic or anoxic waters (Rocap *et al.*, 2003; Ahlgren *et al.*, 2006; Zinser *et al.*, 2007; Lavin *et al.*, 2010; Malmstrom *et al.*, 2010). *Synechococcus* also has ecotypes adapted to different temperatures, nutrient availability, and iron availability (Sohm *et al.*, 2016).

Viruses evolve in many important ways, including mutations that allow changes in host range or virulence (Marston *et al.*, 2012; Bull and Lauring, 2014; Schwartz and Lindell, 2017). Horizontal gene transfer, the exchange of DNA between contemporary organisms, is faster than mutation (Jain *et al.*, 2003). The encoding of host genes in viral genomes through horizontal gene transfer is a form of evolution that links viruses, hosts and the environment. Viruses interface with their host’s metabolism during infection and, therefore, are subject to environmental constraints imposed on their host (Kelly *et al.*, 2013; Motegi *et al.*, 2015). Most cellular transcription and translation stops under viral infection (Doron *et al.*, 2016; Waldbauer *et al.*, 2019). However, the host provides the energy and resources for virus production. To better enable this, the virus supplements key short-lived host proteins by having the host transcribe and translate viral encoded host-derived genes (Lindell *et al.*, 2005; Fridman *et al.*, 2017). Host genes are thought to increase viral fitness either by synthesizing nucleotides for virus replication (Lindell *et al.*, 2007; Weigele *et al.*, 2007), enabling their hosts to acquire limiting nutrients (Breitbart *et al.*, 2007; Kelly *et al.*, 2013), alleviating energy limitation (Thompson *et al.*, 2011; Mahmoudabadi *et al.*, 2017) or by otherwise dealing with the direct stress imposed by the environment (Lindell *et al.*, 2004). Examples of linkages between environmental stress and cyanophage host genes include host genes *hlii* and *psbA* that protect from or ameliorate the consequences from the destructive properties of high light in surface waters (Lindell *et al.*, 2004), or adaptations to an elongated latent period from energy limitation at depth by use of host genes for purine synthesis which are needed to create ATP and NAD (Fuchsman *et al.*, 2021).

There are several types of cyanophage: the T4-like *Myoviridae*, T7-like *Podoviridae*, *Siphoviridae* and unclassified BAC21E04 (Mann *et al.*, 2003; Sullivan *et al.*, 2003, 2009; Huang *et al.*, 2012; Labrie *et al.*, 2013; Mizuno *et al.*, 2013; Crummett *et al.*, 2016; Flores-Uribe *et al.*, 2019). Myo-cyanophage are larger and are generally more generalist while podo-cyanophage are smaller and more specific in their hosts (Zborowsky and Lindell, 2019). Myo-cyanophage often enter a cell before being neutralized by the host defense system and while host resistance to podo-cyanophage occurs outside the cell at the initial adsorption step (Zborowsky and Lindell, 2019). Each family of cyanophage encodes host derived genes with the number and organization of these genes varying (Labrie *et al.*, 2013; Crummett *et al.*, 2016; Marston and Martiny, 2016). The diverse of strategies used by different cyanophage families affect their acquisition of viral host genes. Though viral host genes can be important to viral success, they also increase the number of nucleotides required during replication and the N and P required for those nucleotides. Host-derived genes which are critical only in a specific environment (such as the surface ocean) will not provide fitness benefits in other environments (such as the deep euphotic zone) and therefore are not encoded (Bragg and Chisholm, 2008; Hellweger, 2009; Fuchsman *et al.*, 2021). Whether a cyanophage replaces a no longer favorable host gene with a different host gene or reduces their total host genes appears to depend on the packaging mechanism of the virus (Fuchsman *et al.*, 2021). Deletions or additions of host genes affect the number of nucleotides used in podophage genome replication, where genome concatemers are cleaved at specific recognition sites (Chung and Hinkle, 1990) but myoviruses, where genome concatemers are cleaved non-specifically, are only affected by capsid size (Streisinger *et al.*, 1967). Thus podo-cyanophage seem to reduce the number viral host genes when possible while myo-cyanophage may replace them (Fuchsman *et al.*, 2021). All these constraints affect viral (and host) evolution.

The euphotic zone has strong gradients of light, nutrients, and in some cases oxygen providing different environmental conditions for *Prochlorococcus* and cyanophages between surface waters and depth. *Prochlorococcus* ecotypes and abundances change dramatically over the span of meters in a depth profile (Ahlgren *et al.*, 2006; Johnson *et al.*, 2006; Zinser *et al.*, 2007). However, the scientific community is just beginning to investigate the corresponding changes in the cyanophage community or viral host-derived genes at the appropriate resolution in the euphotic zone (Luo *et al.*, 2020; Fuchsman *et al.*, 2021). Previously, we used phylogenetic read placement to quantitative compare 19 cyanophage host-derived genes to core cyanophage genes along a detailed depth profiles crossing from the oxic primary chlorophyll maximum to the secondary chlorophyll maximum within the ETNP Oxygen Deficient Zone, comparing to the fully oxic Geotraces GA03 St 10 (BATS) metagenomes in the oxic North Atlantic Subtropical Gyre (Biller *et al.*, 2018) (Fuchsman *et al.*, 2021). We found that the type and proportion of host-derived genes in cyanophages changed across the depth gradient in environmental conditions, corresponding to changes in *Prochlorococcus* ecotypes (Fuchsman *et al.*, 2021). Here we extend these analyses to cover more of the global ocean, including the entire GA03 North Atlantic Subtropical Gyre transect (Biller *et al.*, 2018), stations from the GP13 South Pacific transect (Biller *et al.*, 2018), Hawaii Ocean Timeseries (HOT) (Mende *et al.*, 2017; Luo *et al.*, 2020), the Mediterranean (Haro-Moreno *et al.*, 2018), and the Eastern Tropical South Pacific Oxygen Deficient Zone (ETSP ODZ) (Fuchsman *et al.*). With this increased dataset, we statistically link cyanophage host genes, cyanobacterial ecotypes, and environmental parameters. We find general trends in cyanophage host gene utilization that are consistent throughout the ocean basins.

**Methods**

*Metagenomes*

Cellular metagenomes (>0.2 µm) were obtained from the entire GA03 North Atlantic Subtropical Gyre transect sampled in November 2011(Biller *et al.*, 2018), stations from the GP13 South Pacific transect (Biller *et al.*, 2018), Hawaii Ocean Timeseries (HOT) cruises from August (HOT 224) and September (HOT 225) of 2010 and May (HOT 272), August (HOT 275) and November (HOT278) of 2015 (Mende *et al.*, 2017; Luo *et al.*, 2020), the Mediterranean in October 2015 (Haro-Moreno *et al.*, 2018), the Eastern Tropical South Pacific Oxygen Deficient Zone (ETSP ODZ) Station 9 from July 2013 (Fuchsman *et al.*) and the Eastern Tropical North Pacific Oxygen Deficient Zone in April 2012 (Fuchsman *et al.*, 2017). Most of the results from the Eastern Tropical North Pacific Oxygen Deficient Zone have already been published (Fuchsman *et al.*, 2021). The Mediterranean samples were prefiltered through a 5 µm prefilter (Haro-Moreno *et al.*, 2018). The HOT 2010 samples were prefiltered through a 1.6 µm prefilter (Mende *et al.*, 2017). All other samples were not prefiltered.

BioGeotraces metagenomic sequences (Biller *et al.*, 2018) were downloaded from GenBank bioproject PRJNA385854. CTD and phosphate data were downloaded from the British Oceanographic Data Centre (<https://www.bodc.ac.uk/geotraces/>). HOT metagenomes (Mende *et al.*, 2017; Luo *et al.*, 2020) were downloaded from Bioproject PRJNA352737 (2015). Nutrient and CTD measurements for these cruises can be found at Hawaii Ocean Time Series Data Organization and Graphical System (<https://hahana.soest.hawaii.edu/hot/hot-dogs/>) and in the original papers (Mende *et al.*, 2017; Luo *et al.*, 2020). Metagenomes from the Eastern Tropical South Pacific St 9 can be downloaded from Bioproject PRJNA704804. Hydrographic and nutrient data from this ETSP cruise are deposited at NODC as accession 0128141 and are previously published (Fuchsman *et al.*; Peters *et al.*, 2018). Metagenomes from the Eastern Tropical North Pacific can be downloaded from Bioproject PRJN350692. Hydrographic and nutrient data from this ETNP cruise are deposited at <http://data.nodc.noaa.gov/accession/0109846> . Data for ETNP St 136 can be seen in (Fuchsman *et al.*, 2017). Mediterranean metagenomes were downloaded from BioProject PRJNA352*798*, and CTD and nutrient information is contained in the paper (Haro-Moreno *et al.*, 2018). Metadata including accessions numbers, latitude and longitude coordinates, oxygen and nutrient concentrations for all metagenomes can be found in Table S1. Mixed layer depths were determined by examining density profiles for each station.

*Phylogenetic trees and metagenomic read placement*

Reference trees were adapted from Fuchsman et al (2021). Assembled contigs from BioGeotraces transects GA03 and GP13 (Biller *et al.*, 2018) were downloaded and annotated with Prokka (Seemann, 2014) and created into a custom BLAST database (Altschul *et al.*, 1997). Appropriate genes were obtained by blasting (blastp) this database. These new assembled proteins were added to the sequences from the pre-existing trees. These combined full length sequences were aligned using MUSCLE (Edgar, 2004) in order to construct a protein maximum likelihood phylogenetic tree using RaxML-ng with bootstrap analysis (n=100) (Kozlov *et al.*, 2019). Groups within the phylogenetic trees were then labeled based on the references within that group.

The sequences making up the tree were then BLASTed (tblastn) very broadly (e-value=10^-5) against an ETNP, ETSP, HOT, GP13, GA03, and Mediterranean metagenomic read databases. The short reads were then aligned to the reference tree using PaPaRa Parsimony-based Phylogeny-Aware Read Alignment program 2.0 (Berger and Stamatakis, 2011). Non-overlapping paired end reads were then combined into one aligned sequence and placed on the tree by EPA-ng (Barbera *et al.*, 2019). Placed reads have a pendant length indicating the similarity between a query read and the location it places on the tree. Reads that placed with a pendant length greater than 2 were removed. The remaining reads were enumerated for each taxonomic group using the assign subcommand of gappa (Czech *et al.*, 2020). Taxonomic read counts were normalized using the method previously described (Fuchsman *et al.*, 2019) where normalization factors for each sample were determined by dividing the number of good quality reads in a sample by the 100 m ETNP sample. The read counts were multiplied by the sample normalization factor, divided by the length of gene, and then multiplied by 100 in order to make visualization easier.

Terminase large subunit (*terL*) was used as a core gene for cyanophage groups. Siphocyanophage included sequences from (Mizuno *et al.*, 2013) as well as (Sullivan *et al.*, 2009; Huang *et al.*, 2012). Uncategorized BAC21E04 (Flores-Uribe *et al.*, 2019) cyanophage were also examined. To determine the proportion of myo or podo cyanophage containing a particular viral host gene, a ratio of cyanophage host gene/ cyanophage *terL* was obtained. These ratios were specific to each family of cyanophage.

Some exceptions to the above pipeline apply. Cyanobacteria ecotypes were determined using the ITS region of rRNA. The nucleotide tree was previously used (Fuchsman *et al.*, 2021) and is originally from (Lavin *et al.*, 2010). As noted previously (Fuchsman *et al.*, 2021), the *psbA* tree was also in nucleotide space as the distinction between actual host genes and viral host genes is too small in amino acid space for this gene. The assimilatory nitrite reductase *nirA* phylogenetic tree was created using the cyanobacterial amino acid tree from (Widner, Fuchsman, *et al.*, 2018) and identified cyanophage contigs from (Gazitúa *et al.*, 2021) as well as assembled proteins from BioGeotraces (Biller *et al.*, 2018). The *speD* phylogenetic tree was created from scratch.

*Statistics*

To explore the statistical interactions between viruses and cyanobacterial groups, we generated statistical association networks using a variation on the graphical lasoo approach (Tibshirani, 1996). This approach used lassoo regression to identify which cyanobacteria and environmental parameters predicted the proportion of cyanophage with each viral host gene. However, our method differed from graphical lasso in that it did not explore associations between different viral genes nor between cyanobacterial groups.

To carry out this analysis, two matrices were generated. Matrix \*\*X\*\* contained the normalized abundances of all cyanobacteria ecotype reads, as well as the values of a suite of a suite of environmental and location parameters. Environmental parameters were: temperature, salinity, oxygen concentration, nitrate, nitrite, phosphate. Location parameters were binary dummy variables corresponding to the ocean region in which the sample was collected. They included the following geographic regions: East and West and Central North Atlantic, Eastern Tropical North and South Pacific Oceans. Location variables also indicated whether the sample was collected within the mixed layer, in a deep chlorophyll maximum, below the deep chlorophyll maximum, or in hypoxic (<60 µM O2), or oxygen deficient zone waters (<10 nM O2). Matrix \*\*Y\*\* contained the proportion of cyanophage with each viral host gene.

We used the glmnet package in R to perform lasso regressions using all variables in the matrix \*\*X\*\* to predict each variable in \*\*Y\*\*. Lasso requires specification of a tuning parameter, \*λ\*. We identified a single \*λ\* value that optimized the predictive capability of the approach. To identify predictive capability, we used a ten-fold cross validation approach to minimize residual mean squared error of our predictions. In this approach the data were split into ten even separate chunks. In an iterative process each chunk was held out as test data, the model was trained on the remaining nine chunks. The performance of the model at predicting the test data was calculated by averaging the residual mean squared error (RMSE) of model predictions for each variable in \*\*Y\*\*, given the matrix \*\*X\*\* a given value of \*λ\*. We then used the `optimize()` function in `R`'s `stats` package to identify the lambda value associated with the lowest, cross-validated, combined RMSE. We recorded which variables were kept by the lasso approach, and by their relative penalized coefficients. We visualized the results of the lasso regressions using the `igraph` package in R to show which viruses were statistically associated with each cyanobacterial host. We removed the environmental nodes from the visualization, but not the calculations, in order to allow the lasso to account for and factor out associations between viral genes and regions and environmental conditions.

**Results**

*Cyanobacteria and their ecotypes*

We characterized the cyanobacteria and cyanophage across our dataset.For each station,the ITS region of rRNA was used to categorize picocyanobacteria into ecotypes. For *Prochlorococcus*, High Light I (HLI), High Light II (HLII), Low Light I (LLI), Low Light II (LLII), Not Cultured (NC1), Low Light IV (LLIV), and Low Light V (LLV) were examined. *Synechococcus* was only abundant in the North Atlantic GA03 transect and only in the Eastern part of this transect (Figure 1). We used the RNA polymerase *rpoB* gene to calculate the percent of bacteria and archaea that are marine picocyanobacteria (% community). The cyanophage single copy core gene terminase large subunit (*terL*) was used to identify families of cyanophage in cellular genomes. In general, myo-cyanophage were more abundant than podo-cyanophage. However, the differences were sometimes slight. Sipho-cyanophage were always much less abundant than myo and podo, as were unclassified BAC21E04 cyanophage (Figure 2-4, S1-S5). We used *terL* and *rpoB* data to create combined myo and podo cyanophage/cyanobacteria ratios in cellular metagenomes. Interestingly, the *terL/rpoB* cyanophage/cyanobacteria ratios consistently increased at the bottom of the euphotic zone in most datasets (Figure 2-4, S1-S5).

For the first section of the GA03 transect (St 1, 3, 4, 6, 8), the mixed layers were quite deep and there were no deep chlorophyll maxima (Figure 1). For stations 4, 6 and 8, the % community of picocyanobacteria were fairly constant throughout the mixed layer (10-15%) (Figure 1, S1). HLII *Prochlorococcus* was the dominant ecotype throughout the water column with *Synechococcus* as the second most abundant cyanobacteria (Figure 1). The maximum of Low Light I *Prochlorococcus* ecotype appears to be missing. However, when ecotypes are examined as % of picocyanobacteria, the low light ecotypes NC1, LLII, and LLI were more abundant HLII at the deeper depths, where the total percent community for picocyanobacteria was low. In GA03 transect section 1 there were a variety of *Synechococcus* ecotypes including Clade II, Clade XV, Clade III and Cyanobium (Figure S6). At station 3, there was a deep relic mixed layer at 84m, but a shallower actual mixed layer at 44m; picocyanobacteria % community had a strong maximum at 28m and the NC1 ecotype was found in abundance below the relic mixed layer. Thus station 3 seems to be between regimes. Cyanophage/cyanobacteria ratios in cellular metagenomes were low in the mixed layer, but increased steadily below the mixed layer, reaching up to 200 in the deep euphotic zone (Figure 2, S1), except for the most coastal station, GA03 St 1, which had very few picocyanobacteria (3% community) and had the highest cyanophage/cyanobacteria ratio in the mixed layer (Figure S1).

For the subtropical gyre section (section 2) of the North Atlantic GA03 transect (St 10, 14, 16, 18, 20, 22, 24), HLII dominated in surface waters (Figures 1-2). Below the mixed layer, LLI had its maxima, and then slightly deeper, LLII and NC1 had their maxima (Figures 1-2 and S2). Clade III was the dominant *Synechococcus* clade on part of the subtropical gyre section (Stations 10-18), but total *Synechococcus* were very low in abundance for Stations 20-24 (Figure S6). Many of the GA03 stations in the subtropical gyre (St 14, St 16, St 18, St 20) had maxima in picocyanobacteria % community (20-25%) below the mixed layer (Figure S2). We note that nutrients were undetectable in the mixed layer in the gyre, but measurable at the picocyanobacteria % community maxima (Figure 1). Like in section 1, in section 2 of the GA03 transect, cyanophage/cyanobacteria ratios in cellular metagenomes were low in the mixed layer, but increased to 200 in the deep euphotic zone (Figure 2, S2).

Similarly at Hawaii Ocean Timeseries (HOT) the *Prochlorococcus* ecotypes had a similar HLII to LLIV cascade with depth. In bulk water samples (2015) at HOT, picocyanobacteria had a maximum of 30-45% of the community in the mixed layer, dropping significantly in the lower euphotic zone (Figure 3, S4). Nutrients were also measurable in the mixed layer. *Synechococcus* was particularly not abundant in surface waters at HOT. Similarly to GA03, at HOT cyanophage/cyanobacteria ratios in cellular metagenomes were low in the mixed layer, but increased steadily below the mixed layer, reaching up to ~100 in the deep euphotic zone. The depth resolution was particularly fine in the HOT 2015 datasets and covered the entire euphotic zone. At the very bottom of the euphotic zone (for example 225m at HOT 278), the cyanobacteria types shift to more typical surface types. In HOT 272, HLI was found at 200m depth, but not in surface waters, so we hypothesize that it was not endemic to these deep depths.

At the South Pacific (GP13) stations, the top 5 samples were obtained in the mixed layer, and thus only 200m samples represent waters below the mixed layer. Nutrients were quite low but detectable in the mixed layer (Figures 3 and S3). Picocyanobacteria varied between 20-30% of community in the mixed layer (Figures 3 and S3). HLI was the most abundant ecotype in surface waters though HLII was still abundant (Figures 3 and S3). At 200m, NC1 was the most abundant ecotype (Figures 3 and S3). Cyanophage/cyanobacteria ratios in cellular metagenomes were low in the mixed layer but increased to 20-40 at 200m (Figures 3 and S3).

For the Mediterranean, HLI was the dominant ecotype in surface waters, but LLI had its maxima below the chlorophyll maximum at 60m and then LLII and NC1 had their maximum at 75m. However LLIV also had a large maximum at 75m. Once again picocyanobacteria had their maximum below the mixed layer, but picocyanobacteria only represented 10% of the prokaryotic community. Nitrate and ammonia were both measurable in the mixed layer (Haro-Moreno *et al.*, 2018).

The ODZ stations had two chlorophyll maxima, the primary chlorophyll maximum in oxic waters below the mixed layer and a thick secondary chlorophyll maximum in the ODZ. The ETSP had maximal chlorophyll at the surface and a thin secondary chlorophyll maximum in the ODZ. Nitrite concentrations were quite small at the top of the ETNP ODZ, but increased across the ETNP secondary chlorophyll maximum. In the ETSP, nitrite concentrations were high at the secondary chlorophyll maximum. Surface DNA samples weren’t obtained for the ETNP and ETSP, but the ecotype cascade was present at depth with a LLI maximum at the primary chlorophyll maximum, a LLII and NC1 maximum above the ODZ, and ODZ ecotype LLV in the ODZ. Picocyanobacteria were 3.5% of the prokaryotic community in the ETSP ODZ and 4% in the ETNP ODZ. The ETSP ODZ had much higher % of *Synechococcus* than the ETNP, but the ecotype of this *Synechococcus* could not be determined with our tree. In the ETNP and ETSP, the highest cyanophage/cyanobacteria ratios were in the oxyclines (hypoxic waters) above the ODZs (Figure 4). In fact these oxyclines were the only areas in the ocean where podo-cyanophage were more dominant in cellular metagenomes than myo-cyanophage (Figure 2-4, S1-S5). The cyanophage/cyanobacteria ratios then decreased in the ODZ and decreased further at the bottom of the secondary chlorophyll maximum (Figure 4). Thus the cyanophage/cyanobacteria ratio dynamics are completely different in the ODZs from in the oxic ocean.

*Cyanophage host gene/terL depth profiles*

Cyanophage host genes for purine nucleotide synthesis genes (*purN, purC, purM*), pyrimidine nucleotide synthesis genes orate phosphoribosyltransferase (*pyrE*) and thymidylate synthase (*thyX*), photosynthesis genes (*psbA, psbD*), polyamine biosynthesisgenepolyamine aminopropyltransferase (*speD*), and pentose phosphate pathway enzyme transaldolase (*talC*) were examined across ocean basins. We also examined *cobS* and *phoH,* which were suggested to be myo-cyanophage core genes rather than cyanophage host genes by their phylogenetic trees and consistent depth profiles (Fuchsman *et al.*, 2021). All these genes were compared to single copy core gene *terL* for myo-cyanophage or podo-cyanophage. *cobS* and *phoH* both had ratios with *terL* that were constant with depth and near 1 copy per genome (Figures S1-S5). Transaldolase (*talC*) also usually had a proportion near 1 copy per genome (Figures S1-S5).

The proportion of myo-cyanophage with purine synthesis genes *purN*, *purM*, *purC* increased at the bottom of the euphotic zone throughout the ocean. Purine synthesis genes increase at the bottom of the euphotic zone even at the stations in the GA03 transect section 1 where HLII was the only abundant *Prochlorococcus* throughout the euphotic zone (Figure S1). However, we do see that at deeper depths where picocyanobacteria are present but not abundant, low light ecotypes were a higher proportion of cyanobacteria, and these were the depths where the proportion of cyanophage with purine synthesis genes increases (Figure 2, S1). Outside of the ETNP ODZ, we also see the proportion of myo-cyanophage with pyrimidine synthesis genes *pyrE* and *thyX* increase with depth (Figures 2-4, S1-S5).

The proportion of myo-cyanophage with photosynthesis gene *psbD* decreased with depth throughout the ocean, even at the stations in the GA03 transect section 1 (Figures 2-4, S1-S5). The proportion of podo-cyanophage with photosynthesis gene *psbA* generally decreased with depth except at the stations in the GA03 transect section 1 (Figures 2-4, S1-S5). The proportion of myo-cyanophage with *psbA* was constant with depth except in the ODZs, where it decreased significantly (Figures 2-4, S1-S5).

Some viral host genes were particular to the ODZs. The assimilatory nitrite reductase gene *nirA* was found on a cyanophage contig in the ETSP ODZ (Gazitúa *et al.*, 2021). In our dataset this cyanophage *nirA* was only abundant in the ODZs (Figures 2-4, S1-S5). However, cyanophage *nirA* was identified on contigs from the North Atlantic (GA03), so must be present there at low levels. In the ETSP, *nirA* reached 1 copy per myo-cyanophage, but in the ETNP, cyanophage *nirA* reached 2.5 copies per myo-cyanophage, probably indicating that *nirA* is also found in podo-cyanophage (Figure 4). The purine synthesis gene *purS* was only abundant in myo-cyanophage in ODZs, where itreached 1 copy per myo-cyanophage, but myo-cyanophage *purS* was present at the bottom of the euphotic zone at low levels at other stations reaching 0.1-0.2 copies per myo-cyanophage (Figures 2-4, S1-S5).

*Statistics*

Association networks indicated that some viral genes were associated with particular cyanobacterial ecotypes after we took into account environmental variables (temperature, salinity, nitrate, phosphate) (Figure 5). The proportion of myo-cyanophage with nitrite assimilation gene *nirA* strongly positively statistically associated with ODZ *Prochlorococcus* ecotype LLV. The proportion of myo-cyanophage with purine synthesis gene *purS* associated with both LLV and LLIV *Prochlorococcus*, purine synthesis gene *purC* associated with LLIV and LLII/NC1 *Prochlorococcus* and purine synthesis genes *purM* and *purN* and pyrimidine synthesis gene *pyrE* associated with LLII/NC1 *Prochlorococcus*. Polyamine aminopropyltransferase (*speD*) used in polyamine biosynthesis is never abundant in the myo-cyanophage (Figures 2-4, S1-S5), but the proportion of myo-cyanophage with *speD* positively statistically associated with *Synechococcus*. The proportion of myo-cyanophage with *psbD*, *phoH* and the proportion of podo-cyanophage with *psbA* positively associated with High Light *Prochlorococcus*. The proportion of myo-cyanophage with *psbA* was negatively associated with LLV *Prochlorococcus*.

*Assimilatory nitrite reductase nirA in Prochlorococcus*

Assimilatory nitrite reductase *nirA* is a variable gene in *Prochlorococcus*. Here we show a clear trend where the proportion of *Prochlorococcus* with *nirA* increases with depth (Figures 2, 3, S1-S5). ~20% of *Prochlorococcus* are able to use nitrite in the mixed layer of the GA03 transect (Figures S1-S2). ~40% of *Prochlorococcus* are able to use nitrite in the mixed layer at HOT (Figure S4-S5), and ~30% of *Prochlorococcus* are able to use nitrite in the mixed layer in the South Pacific (Figure S3). At all stations, ~100% of *Prochlorococcus* were able to use nitrite at the base of the euphotic zone. The ODZ stations did not have mixed layer samples, but all *Prochlorococcus* could use nitrite in the ODZs.

Previously, we examined *nirA* in LLV cyanobacteria in the ETNP ODZ and found ~6 copies per cyanobacterial genome (Widner, Fuchsman, *et al.*, 2018). In this reanalysis of the same dataset where the tree included viral *nirA*, the host copies of *nirA* were reduced to ~1 copy per genome in the ETNP ODZ.

*Cyanophage cobS phylotypes*

Viral host genes often form multiple distinct clusters on phylogenetic trees (Figure S7). In the ETNP, we saw previously that there was a sharp change in cyanophage host gene phylotypes between oxic and anoxic waters (Fuchsman *et al.*, 2021). Here we choose to examine phylotypes using the *cobS* gene because it is in the majority of myo-cyanophage and does not increase or decrease with depth. We find that while *cobS* is in the majority of myo-cyanophage (Figures S1-S5), there is a shift in phylotypes with depth in oxic waters. Upper euphotic zone phylotypes include phylotypes 4, 6 and 8 (Figure 6, S8). At some stations phylotype 2 dominates in the deep euphotic zone (HOT and GA03 St 3,4, 6, 8, 10) and at other stations cyanophage phylotype 1 dominates the deep euphotic zone (GA03 St 14, 16, 18, 20, 22, 24) (Figure 5, S8). Cyanophage phylotype 1 dominates the ODZ stations (Figure 6, S8).

**Discussion**

*Oceanic trends in cyanophage host genes*

The proportion of cyanophage with purine synthesis genes *purN*, *purM* and *purC* and pyrimidine synthesis gene *pyrE* increase with depth throughout the ocean (Figures 2-4, S1-S5). Contrastingly, we find that generally the proportion of podo-cyanophage with *psbA* and myo-cyanophage with *psbD* are ~1 in surface waters but are reduced with depth (Figures 2-4, S1-S5). These trends appear to be conserved between ocean regions (S1-S5) and over time (S4-S5). The proportion of cyanophage with these host genes depends on the ecotype of the host. Statistics indicate that *purN*, *purM* and *pyrE* genes are linked to Low Light *Prochlorococcus* ecotypes Low Light II and NCI (Figure 5). While also linked to Low Light II and NCI, *purC* is more strongly linked to LLIV *Prochlorococcus* (Figure 5). All these *Prochlorococcus* ecotypes live at extremely low light levels and are found deeper in the water column than the Low Light I ecotype. On the other hand, both *psbD* and podo *psbA* statistically associate with High Light *Prochlorococcus* ecotypes (Figure 5).

Some cyanophage host genes were specifically present or absent for the ODZ where the Low Light V ecotype of *Prochlorococcus* is specifically found. The proportion of myo cyanophage with *psbA* is ~1 but does not reduce with depth except in the ETNP and ETSP oxygen deficient zones. Myo *psbA* is negatively statistically associated with Low Light V. It was hypothesized that this reduction in myo *psbA* was due to the stable conditions in the ODZ, where high light levels never occur (Fuchsman *et al.*, 2021). Though it is unclear why this would be the case, the cyanomyophage purine synthesis gene *purS* was also specifically found in high abundance in the ODZs and is positively statistically associated with Low Light V *Prochlorococcus* (Figure 5). Myo-cyanophage with nitrite assimilation gene *nirA* also were only abundant in the ODZs and will be discussed more below.

Previously, we hypothesized that proteins for purine synthesis genes were needed by viruses because of elongated latent periods at depth (Fuchsman *et al.*, 2021). Purines are used in energy storage and transfer (ATP), signaling (cyclic-AMP), and in cofactors (NADH) (Zhao *et al.*, 2013). In fact, purine metabolism pathways were enriched in a global analysis of viral host genes in marine microbial metagenomes (Enav *et al.*, 2014). Latent period scales with host growth rates (Nabergoj *et al.*, 2018). Oxic *Prochlorococcus* living at 100-150m depths in the Pacific divide approximately once per week while *Prochlorococcus* in the mixed layer divide once per day (Vaulot *et al.*, 1995). Thus the cyanophage may need more energy and storage molecules created by the host due to their long latent periods at depth. However, outside of the ODZs, we also see the proportion of myo-cyanophage with pyrimidine synthesis gene *pyrE* increase with depth (Figure 2-4). This indicates that purines and pyrimidines may also be needed as nucleotides to form viral DNA.

Despite the proportion of myo-cyanophage with *cobS* being constant with depth and near 1, the *cobS* phylotypes shift with depth in oxic waters (Figure 6). These phylotypes are from one transfer event, so this evolution should track the evolution of the virus. This implies that along with their hosts, and viral host genes, the cyanophage themselves are shifting with depth.

*Assimilatory nitrite reductase*

Most Low Light strains have the *nirA* gene for assimilatory nitrite reductase and some High Light strains have *nirA* (Berube *et al.*, 2019). More High Light strains have *nirA* in N limited regions such as the Indian Ocean and the Caribbean, so these traits are not strictly delineated at the ecotype level (Martiny *et al.*, 2009). In our dataset, the % of *Prochlorococcus* able to use nitrite clearly increased with depth with 20-50% of *Prochlorococcus* able to use nitrite in the mixed layer and ~100% able to use nitrite at the base of the euphotic zone (Figure 2-3, S1-S5).

Host nitrogen assimilation can be especially important to viruses because viruses are enriched in N and P compared to microorganisms (Jover *et al.*, 2014). Low environmental N concentrations can cause viral burst size to decrease significantly (Cheng *et al.*, 2015). In our dataset the cyanophage *nirA* was only detectable in the ODZs. However, cyanophage *nirA* was identified on contigs from the North Atlantic (GA03), so must be present there at low levels. In the ETSP, *nirA* reached 1 copy per myo-cyanophage, but in the ETNP, cyanophage *nirA* reached 2.5 copies per myo-cyanophage, probably indicating that *nirA* is also found in podo-cyanophage.

We note that though nitrite assimilation genes were found in 100% of *Prochlorococcus* cells at depth in oxic waters and 20-50% of *Prochlorococcus* in surface waters, viruses containing *nirA* were not detectable in oxic metagenomes (Figures 2-3, S1-S5). However, *nirA* genes were in 100% of myo-cyanomyophage in ODZ waters and were perhaps also in podo-cyanophage (Figure 4). This highlights the particular importance of nitrite as a N source to *Prochlorococcus* in ODZ waters. Gene analysis of Low Light V *Prochlorococcus* in ODZs indicated that they have the capability of utilizing ammonia, urea, nitrite and nitrate (Astorga-Elo *et al.*, 2015; Widner, Fuchsman, *et al.*, 2018; Ulloa *et al.*, 2021). However, Low Light V *Prochlorococcus* predominantly utilize nitrite in the ODZ as seen from in situ stable isotopes (Aldunate *et al.*, 2019) and *Prochlorococcus* *nirA* was found in proteomic analysis indicating activity (Fuchsman *et al.*, 2019). In the ETNP and ETSP ammonia is usually <10 nM levels in the ODZ (Widner, Fuchsman, *et al.*, 2018; Widner, Mordy, *et al.*, 2018) and *Prochlorococcus* must join the competition between anammox and nitrite oxidizers for this ammonia (Penn *et al.*, 2019). Additionally urea, also a form a reduced N, is usually below detection in the ODZ while nitrite concentration can reach several micromolar (Widner, Fuchsman, *et al.*, 2018).

*Physical zonation*

We find that mixed layer depths were important for cyanobacterial ecotype zonation in the water column. In the BioGeotraces GA03 transect, there were several stations in section 1 of the transect that had deep mixed layers (St 4, St 6. St 8) (Figure 1). At these stations, Hight Light II *Prochlorococcus* was the only *Prochlorococcus* ecotype abundant in the water column (Figure 1). Though previous work has found that LLI can survive mixing events (Malmstrom *et al.*, 2010), we see no evidence of LLI thriving at our stations with deep mixed layers (Figure 1). In fact NC1 appeared to be the first low light ecotype to recover from mixing, see station 3 (Figure 1). At section 1 stations with deep mixed layers, the proportion of podo-cyanophage with *psbA* was constant with depth, but the proportion of myo-cyanophage that had *psbD* was still reduced at depth and the proportion of myo-cyanophage with *purN*, *purM* and *purC* increased with depth. Despite never being abundant in the water column, Low Light ecotypes LLI, LLII, and NC1 were a greater proportion of the picocyanobacterial community at the bottom of the euphotic zone where cyanobacteria were present but not abundant (Fig 2, S1). At stations in section 2 of this transect where the mixed layer was shallower, the various Low Light ecotypes became important in the deeper water column, as is generally seen in the subtropical ocean (Braakman *et al.*, 2017).

The North Atlantic GA03 transect varied with some stations having maxima in picocyanobacteria % community of 15-30% at the surface (St 10, St 22, St 3, St 4, St 6, St 7, St 8), but others had maxima in picocyanobacteria of 20-25% below the mixed layer (St 14, St 16, St 18, St 20). We believe that the maxima below the mixed layer in the North Atlantic subtropical gyre is due to a lack of nutrients in the mixed layer (Fig 1). Nutrients were detectable below the mixed layer (Figure1) where the % community that is picocyanobacteria increased. This contrasts to HOT where the maximum in % community for picocyanobacteria was in the mixed layer, but nutrients were also measurable in the mixed layer (Figure 3, S4-S5). These variations in host abundance also affect cyanophage abundance profiles.

The HLI ecotype is known to be adapted to colder regions than the HLII ecotype (Johnson *et al.*, 2006), and we see an abundance of HLI in the colder South Pacific (Figure 3, S3). As seen previously (Garczarek *et al.*, 2007), HLI was also found in the Mediterranean, which also had colder surface waters. However, the switch between HLII and HLI ecotypes did not seem to affect the cyanophage host genes examined here.

The HOT 2015 datasets had very good coverage of the entire euphotic zone. We can see that at the deepest depths where the numbers of cyanobacteria are very small, the *speD/terL* ratios shift. When we look at the picocyanobacteria community at those depths, it is composed of *Synechococcus* and HLII (Figure 3), rather than the endemic Low Light ecotypes. These HL ecotypes could be transported from the surface by particles. *Synechococcus* has been previously found on particles and may be attracted to particles (Thiele *et al.*, 2015; Fuchsman *et al.*, 2019). We see that *speD/terL* ratios are statistically associated with *Synechococcus*. However, in HOT 272 at 200m, we see a shift in cyanophage host gene ratios back towards surface conditions and simultaneously see HLI ecotype of *Prochlorococcus* (Figure S4). HLI isn’t found in surface waters at this station, indicating potential horizontal advection. *Prochlorococcus* has the ability to survive extended periods of time under unideal light conditions (Roth-Rosenberg *et al.*, 2020), though such a situation could enhance viral infection.

*Cyanophage/Cyanobacteria ratios*

The ratio of cyanophage to cyanobacteria consistently increases with depth (Figure 2, 3, S1-S5). There are three reasons that these increasing profiles could occur. First, an increase in latent period with depth, It is extremely likely that viral latent period increases with depth, at least in myo cyanophage (Fuchsman *et al.*, 2021; Mruwat *et al.*, 2021). Latent period scales with host growth rates (Nabergoj *et al.*, 2018) and *Prochlorococcus* living in the deep euphotic zone divide approximately once per week (Vaulot *et al.*, 1995), which is consistent with *Prochlorococcus* cultures growing at very low light levels (Moore and Chisholm, 1999; Zinser *et al.*, 2007). A longer latent period would cause more viruses to be found in cells even with the same infection rate. However, larger cyanophage/cyanobacterial ratios could also be due to an increase in infection with depth. Stress can increase viral infection (Jiang and Paul, 1996; Thurber *et al.*, 2017). If low light causes stress, higher infection rates at depths where light is limiting would also cause the same profile. Thirdly, inputs of viruses from other sources such as sinking particles could increase cyanophage/cyanobacteria ratios. Infected cells form TEP and can stick together forming particles (Shibata *et al.*, 1997). Cyanophage to cyanobacteria ratios were found to be extremely high on particles in the ETNP (Fuchsman *et al.*, 2019). Thus shedding of viruses by particles at depths with few cyanobacteria could also increase the cyanophage/cyanobacteria ratios. However, importing viruses from surface waters would also import viral host gene signatures from surface waters. We only see this shift at HOT 272 200m, but cyanophage/cyanobacteria ratios are high at all the oxic stations.

The ODZ regions show very different cyanophage/cyanobacteria ratios compared to oxic waters. The maximum of cyanophage/cyanobacteria ratios in the low oxygen waters above the ETNP ODZ is likely due to increased infection. Podo-cyanophage are the dominant cyanophage at these depths (Figure 4), which is unusual (Figure 2, 3, S1-S5). The community is shifting to an ecotype of *Prochlorococcus* that can withstand anoxia (LLV), and thus the other ecotypes must be feeling stress in these low oxygen waters. Interestingly, in the ODZ, despite low light, cyanophage/cyanobacteria ratios are low (Figure 4). There is an idea that as new *Prochlorococcus* ecotypes evolved, older ecotypes, such as LLII, were pushed into habitats that were less ideal for them (Braakman *et al.*, 2017). If this is the case, these low light ecotypes might be living under constant stress and may have higher viral infection rates. LLV, the ODZ ecotype, on the other hand, appears to be living in its ideal conditions, forming large maxima (Figure 4), experiencing reduced grazing (Zakem *et al.*, 2020) and low cyanophage/cyanobacteria ratios (Figure 4).

**Conclusions**

We find consistent patterns of key cyanophage host genes across the oxic ocean. These cyanophage host genes correlate with host ecotype rather than environmental parameters. By comparison to the oxic ocean, we can see what cyanophage host genes are especially important (*nirA* and *purS*) or unimportant (*psbA*) to Oxygen Deficient Zones. Cellular metagenomes from the oxic ocean are dominated by myo-cyanophage. However, in the low oxygen waters (oxycline) above the ODZ, podo-cyanophage dominate and cyanophage/cyanobacteria ratios are high, indicating active infection. Contrastingly, in ODZ waters cyanophage/cyanobacteria ratios are low, perhaps due to LLV *Prochlorococcus* being supremely well evolved for this niche. Thus high cyanophage/cyanobacteria ratios in the deep euphotic zone of the oxic ocean may support the idea (Braakman *et al.*, 2017) that oxic low light ecotypes are living in conditions that aren’t ideal for them.

References

Ahlgren, N.A., Rocap, G., and Chisholm, S.W. (2006) Measurement of *Prochlorococcus* ecotypes using real-time polymerase chain reaction reveals different abundances of genotypes with similar light physiologies. *Environ Microbiol* **8**: 441–454.

Aldunate, M., Henríquez-Castillo, C., Ji, Q., Lueders-Dumont, J., Mulholland, M.R., Ward, B.B., et al. (2019) Nitrogen assimilation in picocyanobacteria inhabiting the oxygen-deficient waters of the eastern tropical North and South Pacific. *Limnol Oceanogr* **65**: 437–453.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Ankrah, N.Y.D., May, A.L., Middleton, J.L., Jones, D.R., Hadden, M.K., Gooding, J.R., et al. (2014) Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *ISME J* **8**: 1089–1100.

Astorga-Elo, M., Ramirez-Flandes, S., DeLong, E.F., and Ulloa, O. (2015) Genomic potential for nitrogen assimilation in uncultivated members of Prochlorococcus from an anoxic marine zone. *ISME J* **9**: 1264–1267.

Baran, N., Goldin, S., Maidanik, I., and Lindell, D. (2018) Quantification of diverse virus populations in the environment using the polony method. *Nat Microbiol* **3**: 62–72.

Barbera, P., Kozlov, A.M., Czech, L., Morel, B., Darriba, D., Flouri, T., and Stamatakis, A. (2019) EPA-ng: Massively Parallel Evolutionary Placement of Genetic Sequences. *Syst Biol* **68**: 365–369.

Berger, S.A. and Stamatakis, A. (2011) Aligning short reads to reference alignments and trees. *Bioinformatics* **27**: 2068–2075.

Berube, P.M., Rasmussen, A., Braakman, R., Stepanauskas, R., and Chisholm, S.W. (2019) Emergence of trait variability through the lens of nitrogen assimilation in *Prochlorococcus*. *Elife* **8**: e41043.

Biller, S.J., Berube, P.M., Dooley, K., Williams, M., Satinsky, B.M., Hackl, T., et al. (2018) Data descriptor: Marine microbial metagenomes sampled across space and time. *Sci Data* **5**: 180176.

Braakman, R., Follows, M.J., and Chisholm, S.W. (2017) Metabolic evolution and the self-organization of ecosystems. *Proc Natl Acad Sci U S A* **114**: E3091–E3100.

Bragg, J.G. and Chisholm, S.W. (2008) Modeling the Fitness Consequences of a Cyanophage- Encoded Photosynthesis Gene. *PLoS One* **3**: e3550.

Breitbart, M., Thompson, L.R., Suttle, C.A., and Sullivan, M.B. (2007) Exploring the Vast Diversity of Marine Viruses. *Oceanography* **20**: 135–139.

Brussaard, C.P.D., Kuipers, B., and Veldhuis, M.J.W. (2005) A mesocosm study of *Phaeocystis globosa* population dynamics: I. Regulatory role of viruses in bloom control. *Harmful Algae* **4**: 859–874.

Bull, J.J. and Lauring, A.S. (2014) Theory and Empiricism in Virulence Evolution. *PLoS Pathog* **10**: e1004387.

Cheng, Y.S., Labavitch, J., and VanderGheynst, J.S. (2015) Organic and Inorganic Nitrogen Impact *Chlorella variabilis* Productivity and Host Quality for Viral Production and Cell Lysis. *Appl Biochem Biotechnol* **176**: 467–479.

Chung, Y.B. and Hinkle, D.C. (1990) Bacteriophage T7 DNA packaging. II. Analysis of the DNA sequences required for packaging using a plasmid transduction assay. *J Mol Biol* **216**: 927–938.

Crummett, L.T., Puxty, R.J., Weihe, C., Marston, M.F., and Martiny, J.B.H. (2016) The genomic content and context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* **499**: 219–229.

Czech, L., Barbera, P., and Stamatakis, A. (2020) Genesis and Gappa: processing, analyzing and visualizing phylogenetic (placement) data. *Bioinformatics* **36**: 3263–3265.

Doron, S., Fedida, A., Hernández-Prieto, M.A., Sabehi, G., Karunker, I., Stazic, D., et al. (2016) Transcriptome dynamics of a broad host-range cyanophage and its hosts. *ISME J* **10**: 1437–1455.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Enav, H., Mandel-Gutfreund, Y., and Béjà, O. (2014) Comparative metagenomic analyses reveal viral-induced shifts of host metabolism towards nucleotide biosynthesis. *Microbiome* **2**: 9.

Flombaum, P., Gallegos, J.L., Gordillo, R.A., Rincón, J., Zabala, L.L., Jiao, N., et al. (2013) Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus. *Proc Natl Acad Sci U S A* **110**: 9824–9829.

Flores-Uribe, J., Philosof, A., Sharon, I., Fridman, S., Larom, S., and Béjà, O. (2019) Brief Report A novel uncultured marine cyanophage lineage with lysogenic potential linked to a putative marine Synechococcus ‘ relic ’ prophage. *Environ Microbiol Rep* **11**: 598–604.

Fridman, S., Flores-Uribe, J., Larom, S., Alalouf, O., Liran, O., Yacoby, I., et al. (2017) A myovirus encoding both photosystem I and II proteins enhances cyclic electron flow in infected *Prochlorococcus* cells. *Nat Microbiol* **2**: 1350–1357.

Fuchsman, C.A., Carlson, M.C.G., Garcia Prieto, D., Hays, M.D., and Rocap, G. (2021) Cyanophage host-derived genes reflect contrasting selective pressures with depth in the oxic and anoxic water column of the Eastern Tropical North Pacific. *Environ Microbiol* **23**: 2782–2800.

Fuchsman, C.A., Cherubini, L., and Hays, M.D. An Analysis of Protists in Pacific Oxygen Deficient Zones: Implications for *Prochlorococcus* and N2 producing bacteria. *Environ Microbiol*.

Fuchsman, C.A., Devol, A.H., Saunders, J.K., McKay, C., and Rocap, G. (2017) Niche Partitioning of the N cycling microbial community of an offshore Oxygen Deficient Zone. *Front Microbiol* **8**: 2384.

Fuchsman, C.A., Palevsky, H.I., Widner, B., Duffy, M., Carlson, M.C.G., Neibauer, J.A., et al. (2019) Cyanobacteria and cyanophage contributions to carbon and nitrogen cycling in an oligotrophic oxygen-deficent zone. *ISME J* **13**: 2714–2726.

Fuhrman, J.A. (1999) Marine viruses and their biogeochemical and ecological effects. *Nature* **399**: 541–548.

Garczarek, L., Dufresne, A., Rousvoal, S., West, N.J., Mazard, S., Marie, D., et al. (2007) High vertical and low horizontal diversity of Prochlorococcus ecotypes in the Mediterranean Sea in summer. *FEMS Microbiol Ecol* **60**: 189–206.

Gazitúa, M.C., Vik, D.R., Roux, S., Gregory, A.C., Bolduc, B., Widner, B., et al. (2021) Potential virus-mediated nitrogen cycling in oxygen-depleted oceanic waters. *ISME J* **15**: 981–988.

Haro-Moreno, J.M., López-Pérez, M., de la Torre, J.R., Picazo, A., Camacho, A., and Rodriguez-Valera, F. (2018) Fine metagenomic profile of the Mediterranean stratified and mixed water columns revealed by assembly and recruitment. *Microbiome* **6**: 128.

Hellweger, F.L. (2009) Carrying photosynthesis genes increases ecological fitness of cyanophage in silico. *Environ Microbiol* **11**: 1386–1394.

Huang, S., Wang, K., Jiao, N., and Chen, F. (2012) Genome sequences of siphoviruses infecting marine *Synechococcus* unveil a diverse cyanophage group and extensive phage-host genetic exchanges. *Environ Microbiol* **14**: 540–558.

Jain, R., Rivera, M.C., Moore, J.E., and Lake, J.A. (2003) Horizontal gene transfer accelerates genome innovation and evolution. *Mol Biol Evol* **20**: 1598–1602.

Jiang, S.C. and Paul, J.H. (1996) Occurrence of lysogenic bacteria in marine microbial communities as determined by prophage induction. *Mar Ecol Prog Ser* **142**: 27–38.

Johnson, Z.I., Zinser, E.R., Coe, A., Mcnulty, N.P., Malcolm, E.S., Chisholm, S.W., et al. (2006) Partitioning Among *Prochlorococcus* Ecotypes Along Environmental Gradients. *Science (80- )* **311**: 1737–1740.

Jover, L.F., Effler, T.C., Buchan, A., Wilhelm, S.W., and Weitz, J.S. (2014) The elemental composition of virus particles: implications for marine biogeochemical cycles. *Nat Rev Microbiol* **12**: 519–528.

Kelly, L., Ding, H., Huang, K.H., Osburne, M.S., and Chisholm, S.W. (2013) Genetic diversity in cultured and wild marine cyanomyoviruses reveals phosphorus stress as a strong selective agent. *ISME J* **7**: 1827–1841.

Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stramatakis, A. (2019) RAxML-ng: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* **35**: 4453–4455.

Kranzler, C.F., Krause, J.W., Brzezinski, M.A., Edwards, B.R., Biggs, W.P., Maniscalco, M., et al. (2019) Silicon limitation facilitates virus infection and mortality of marine diatoms. *Nat Microbiol* **4**: 1790–1797.

Laber, C.P., Hunter, J.E., Carvalho, F., Collins, J.R., Hunter, E.J., Schieler, B.M., et al. (2018) Coccolithovirus facilitation of carbon export in the North Atlantic. *Nat Microbiol* **3**: 537–547.

Labrie, S.J., Frois-Moniz, K., Osburne, M.S., Kelly, L., Roggensack, S.E., Sullivan, M.B., et al. (2013) Genomes of marine cyanopodoviruses reveal multiple origins of diversity. *Environ Microbiol* **15**: 1356–1376.

Lavin, P., González, B., Santibáñez, J.F., Scanlan, D.J., and Ulloa, O. (2010) Novel lineages of *Prochlorococcus* thrive within. *Environ Microbiol Rep* **2**: 728–738.

Lindell, D., Jaffe, J.D., Coleman, M.L., Futschik, M.E., Axmann, I.M., Rector, T., et al. (2007) Genome-wide expression dynamics of a marine virus and host reveal features of co-evolution. *Nature* **449**: 83–86.

Lindell, D., Jaffe, J.D., Johnson, Z.I., Church, G.M., and Chisholm, S.W. (2005) Photosynthesis genes in marine viruses yield proteins during host infection. *Nature* **438**: 86–89.

Lindell, D., Sullivan, M.B., Johnson, Z.I., Tolonen, A.C., Rohwer, F., and Chisholm, S.W. (2004) Transfer of photosynthesis genes to and from *Prochlorococcus* viruses. *Proc Natl Acad Sci U S A* **101**: 11013–11018.

Lønborg, C., Middelboe, M., and Brussaard, C.P.D. (2013) Viral lysis of *Micromonas pusilla*: Impacts on dissolved organic matter production and composition. *Biogeochemistry* **116**: 231–240.

Luo, E., Eppley, J.M., Romano, A.E., Mende, D.R., and DeLong, E.F. (2020) Double-stranded DNA virioplankton dynamics and reproductive strategies in the oligotrophic open ocean water column. *ISME J* **14**: 1304–1315.

Mahmoudabadi, G., Milo, R., and Phillips, R. (2017) Energetic cost of building a virus. *Proc Natl Acad Sci U S A* **114**: E4324–E4333.

Malmstrom, R.R., Coe, A., Kettler, G.C., Martiny, A.C., Frias-Lopez, J., Zinser, E.R., and Chisholm, S.W. (2010) Temporal dynamics of *Prochlorococcus* ecotypes in the Atlantic and Pacific oceans. *ISME J* **4**: 1252–1264.

Mann, N.H., Cook, A., Millard, A., Bailey, S., and Clokie, M. (2003) Bacterial photosynthesis genes in a virus. *Nature* **424**: 741.

Marston, M.F. and Martiny, J.B.H. (2016) Genomic diversification of marine cyanophages into stable ecotypes. *Environ Microbiol* **18**: 4240–4253.

Marston, M.F., Pierciey, F.J., Shepard, A., Gearin, G., Qi, J., Yandava, C., and Schuster, S.C. (2012) Rapid diversification of coevolving marine *Synechococcus* and a virus. *Proc Natl Acad Sci* **109**: 4544–4549.

Martiny, A.C., Kathuria, S., and Berube, P.M. (2009) Widespread metabolic potential for nitrite and nitrate assimilation among Prochlorococcus ecotypes. *Proc Natl Acad Sci U S A* **106**: 10787–10792.

Mende, D.R., Bryant, J.A., Aylward, F.O., Eppley, J.M., Nielsen, T., Karl, D.M., and Delong, E.F. (2017) Environmental drivers of a microbial genomic transition zone in the ocean’s interior. *Nat Microbiol* **2**: 1367–1373.

Middelboe, M. and Jørgensen, N.O.G. (2006) Viral lysis of bacteria: An important source of dissolved amino acids and cell wall compounds. *J Mar Biol Assoc United Kingdom* **86**: 605–612.

Mizuno, C.M., Rodriguez-Valera, F., Garcia-Heredia, I., Martin-Cuadrado, A.B., and Ghai, R. (2013) Reconstruction of novel cyanobacterial siphovirus genomes from Mediterranean metagenomic fosmids. *Appl Environ Microbiol* **79**: 688–695.

Moore, L.R. and Chisholm, S.W. (1999) Photophysiology of the marine cyanobacterium *Prochlorococcus*: Ecotypic differences among cultured isolates. *Limnol Oceanogr* **44**: 628–638.

Motegi, C., Kaiser, K., Benner, R., and Weinbauer, M.G. (2015) Effect of P-limitation on prokaryotic and viral production in surface waters of the Northwestern Mediterranean Sea. *J Plankton Res* **37**: 16–20.

Mruwat, N., Carlson, M.C.G., Goldin, S., Ribalet, F., Kirzner, S., Hulata, Y., et al. (2021) A single-cell polony method reveals low levels of infected *Prochlorococcus* in oligotrophic waters despite high cyanophage abundances. *ISME J* **15**: 41–54.

Nabergoj, D., Modic, P., and Podgornik, A. (2018) Effect of bacterial growth rate on bacteriophage population growth rate. *Microbiologyopen* **7**: e558.

Parsons, R.J., Breitbart, M., Lomas, M.W., and Carlson, C.A. (2012) Ocean time-series reveals recurring seasonal patterns of virioplankton dynamics in the northwestern Sargasso Sea. *ISME J* **6**: 273–284.

Peduzzi, P. and Weinbauer, M.G. (1993) Effect of concentrating the virus‐rich 2‐2nm size fraction of seawater on the formation of algal flocs (marine snow). *Limnol Oceanogr* **38**: 1562–1565.

Penn, J.L., Weber, T., Chang, B.X., and Deutsch, C. (2019) Microbial ecosystem dynamics drive fluctuating nitrogen loss in marine anoxic zones. *Proc Natl Acad Sci* **116**: 7220–7225.

Peters, B., Horak, R., Devol, A.H., Fuchsman, C.A., Forbes, M., Mordy, C.W., and Casciotti, K.L. (2018) Estimating fixed nitrogen loss and associated isotope effects using concentration and isotope measurements of NO3–, NO2–, and N2 from the Eastern Tropical South Pacific oxygen deficient zone. *Deep Res Part II Top Stud Oceanogr* **156**: 121–136.

Rii, Y.M., Karl, D.M., and Church, M.J. (2016) Temporal and vertical variability in picophytoplankton primary productivity in the North Pacific Subtropical Gyre. *Mar Ecol Prog Ser* **562**: 1–18.

Rocap, G., Larimer, F.W., Lamerdin, J., Malfatti, S., Chain, P., Ahlgren, N.A., et al. (2003) Genome divergence in two Prochlorococcus ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.

Roth-Rosenberg, D., Aharonovich, D., Vogts, T.L.-K.A., Nago, N., Grossart, H.-P., Voss, M., and Sher, D. (2020) *Prochlorococcus* Cells Rely on Microbial Interactions Rather than on Chlorotic Resting Stages To Survive Long-Term Nutrient Starvation. *MBio* **11**: e01846-20.

Schwartz, D.A. and Lindell, D. (2017) Genetic hurdles limit the arms race between *Prochlorococcus* and the T7-like podoviruses infecting them. *ISME J* **11**: 1836–1851.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**: 2068–2069.

Shelford, E.J., Middelboe, M., Møller, E.F., and Suttle, C.A. (2012) Virus-driven nitrogen cycling enhances phytoplankton growth. *Aquat Microb Ecol* **66**: 41–46.

Sheyn, U., Rosenwasser, S., Lehahn, Y., Barak-Gavish, N., Rotkopf, R., Bidle, K.D., et al. (2018) Expression profiling of host and virus during a coccolithophore bloom provides insights into the role of viral infection in promoting carbon export. *ISME J* **12**: 704–713.

Shibata, A., Kogure, K., Koike, I., and Ohwada, K. (1997) Formation of submicron colloidal particles from marine bacteria by viral infection. *Mar Ecol Prog Ser* **155**: 303–307.

Sohm, J.A., Ahlgren, N.A., Thomson, Z.J., Williams, C., Moffett, J.W., Saito, M.A., et al. (2016) Co-occurring Synechococcus ecotypes occupy four major oceanic regimes defined by temperature, macronutrients and iron. *ISME J* **10**: 333–345.

Streisinger, G., Emrich, J., and Stahl, M. (1967) Chromosome structure in phage T4, III. Terminal Redundancy and length determination. *Proc Natl Acad Sci* **57**: 292–295.

Sullivan, M.B., Krastins, B., Hughes, J.L., Kelly, L., Chase, M., Sarracino, D., and Chisholm, S.W. (2009) The genome and structural proteome of an ocean siphovirus: A new window into the cyanobacterial “mobilome.” *Environ Microbiol* **11**: 2935–2951.

Sullivan, M.B., Waterbury, J.B., and Chisholm, S.W. (2003) Cyanophages infecting the oceanic cyanobacterium *Prochlorococcus*. *Nature* **424**: 1047–1051.

Thiele, S., Fuchs, B.M., Amann, R., and Iversen, H. (2015) Colonization in the Photic Zone and Subsequent Changes during Sinking Determine Bacterial Community Composition in Marine Snow. *Appl Environ Microbiol* **81**: 1463–1471.

Thompson, L.R., Zeng, Q., Kelly, L., Huang, K.H., Singer, A.U., Stubbe, J., and Chisholm, S.W. (2011) Phage auxiliary metabolic genes and the redirection of cyanobacterial host carbon metabolism. *Proc Natl Acad Sci* **108**: E757–E764.

Thurber, R.V., Payet, J.P., Thurber, A.R., and Correa, A.M.S. (2017) Virus-host interactions and their roles in coral reef health and disease. *Nat Rev Microbiol* **15**: 205–216.

Tibshirani, R. (1996) Regression Shrinkage and Selection Via the Lasso. *J R Stat Soc Ser B* **58**: 267–288.

Ulloa, O., Henríquez-Castillo, C., Ramírez-Flandes, S., Plominsky, A.M., Murillo, A.A., Morgan-Lang, C., et al. (2021) The cyanobacterium Prochlorococcus has divergent light-harvesting antennae and may have evolved in a low-oxygen ocean. *Proc Natl Acad Sci U S A* **118**: e2025638118.

Vardi, A., Haramaty, L., Van Mooy, B.A.S., Fredricks, H.F., Kimmance, S.A., Larsen, A., and Bidle, K.D. (2012) Host-virus dynamics and subcellular controls of cell fate in a natural coccolithophore population. *Proc Natl Acad Sci U S A* **109**: 19327–19332.

Vaulot, D., Marie, D., Olson, R.J., and Chisholm, S.W. (1995) Growth of *Prochlorococcus*, a Photosynthetic Prokaryote, in the Equatorial Pacific Ocean. *Science (80- )* **268**: 1480–1482.

Waldbauer, J.R., Coleman, M.L., Rizzo, A.I., Campbell, K.L., Lotus, J., and Zhang, L. (2019) Nitrogen sourcing during viral infection of marine cyanobacteria. *Proc Natl Acad Sci U S A* **116**: 15590–15595.

Weigele, P.R., Pope, W.H., Pedulla, M.L., Houtz, J.M., Smith, A.L., Conway, J.F., et al. (2007) Genomic and structural analysis of Syn9, a cyanophage infecting marine *Prochlorococcus* and *Synechococcus*. *Environ Microbiol* **9**: 1675–1695.

Widner, B., Fuchsman, C.A., Chang, B.X., Rocap, G., and Mulholland, M.R. (2018) Utilization of urea and cyanate in waters overlying and within the eastern tropical north Pacific oxygen deficient zone. *FEMS Microbiol Ecol* **94**: fiy138.

Widner, B., Mordy, C.W., and Mulholland, M.R. (2018) Cyanate distribution and uptake above and within the Eastern Tropical South Pacific oxygen deficient zone. *Limnol Oceanogr* **63**: S177–S192.

Zakem, E.J., Mahadevan, A., Lauderdale, J.M., and Follows, M.J. (2020) Stable aerobic and anaerobic coexistence in anoxic marine zones. *ISME J* **14**: 288–301.

Zborowsky, S. and Lindell, D. (2019) Resistance in marine cyanobacteria differs against specialist and generalist cyanophages. *Proc Natl Acad Sci USA* **116**: 16899–16908.

Zhao, H., French, J.B., Fang, Y., and Benkovic, S.J. (2013) The purinosome, a multi-protein complex involved in the de novo biosynthesis of purines in humans. *Chem Commun* **49**: 4444–4452.

Zhao, Z., Gonsior, M., Schmitt-Kopplin, P., Zhan, Y., Zhang, R., Jiao, N., and Chen, F. (2019) Microbial transformation of virus-induced dissolved organic matter from picocyanobacteria: coupling of bacterial diversity and DOM chemodiversity. *ISME J* **13**: 2551–2565.

Zinser, E.R., Johnson, Z.I., Coe, A., Karaca, E., Veneziano, D., and Chisholm, S.W. (2007) Influence of light and temperature on *Prochlorococcus* ecotype distributions in the Atlantic Ocean. *Limnol Oceanogr* **52**: 2205–2220.