Ubiquitous, B_{12}-dependent virioplankton utilizing ribonucleotide triphosphate reductase demonstrate interseasonal dynamics and associate with a diverse range of bacterial hosts in the pelagic ocean

Ling-Yi Wu[1], Gonçalo J. Piedade[2,3], Ryan M. Moore[4], Amelia O. Harrison[4], Ana M. Martins[3], Kay D. Bidle[5], Shawn W. Polson[4], Eric Sakowski[6], Jozef I. Nissimov[7], Jacob T. Dums[4,8], Barbra D. Ferrell[4], K. Eric Wommack[4,*]

[1] Theoretical Biology and Bioinformatics, Science4Life, Utrecht University, Padualaan 8, Utrecht, 3584 CH, the Netherlands

[2] Department of Marine Microbiology and Biogeochemistry, NIOZ Royal Netherlands Institute for Sea Research, 1797 SZ t'Horntje, The Netherlands

[3] Department of Oceanography and Fisheries and Ocean Sciences Institute-OKEANOS, University of the Azores, 9901-862 Horta, Faial, Azores, Portugal

[4] Delaware Biotechnology Institute, University of Delaware, 590 Avenue 1743, Newark, DE, 19713, United States

[5] Department of Marine and Coastal Sciences, Rutgers University, 71 Dudley Rd., New Brunswick, NJ, 08901, United States

[6] Johns Hopkins University, Department of Earth Sciences, Baltimore MD, United States

[7] Department of Biology, University of Waterloo, 200 University Ave. West, Waterloo, Ontario, N2L 3G1, Canada

[8] Biotechnology Program, North Carolina State University, 2800 Faucette Dr. Raleigh, NC, 27695, United States

Keywords: oceanic carbon cycle, compositional data, metagenomic amplicons, phage-host interactions, virus phylogeny/oligotrophic microbial community, long-read sequencing

* To whom correspondence should be addressed. Tel: +01 302-831-4362; Fax: +44 000 0000000; Email: wommack@udel.edu

ABSTRACT

Through infection and lysis of their coexisting bacterial hosts, viruses impact the biogeochemical cycles sustaining globally significant pelagic oceanic ecosystems. Currently, little is known of the ecological interactions between lytic viruses and their bacterial hosts underlying these biogeochemical impacts at ecosystem scales. This study focused on populations of lytic viruses carrying the B_{12}-dependent Class II monomeric ribonucleotide reductase (RNR) gene, ribonucleotide triphosphate reductase (RTPR), documenting seasonal changes in pelagic virioplankton and bacterioplankton using amplicon sequences of RTPR and the 16S rRNA gene, respectively. Amplicon sequence libraries were analyzed using compositional data analysis tools that account for the compositional nature of these data. Both virio- and bacterioplankton communities responded to environmental changes typically seen across seasonal cycles as well as shorter term upwelling–downwelling events. Defining RTPR-carrying viral populations according to major phylogenetic clades proved a more robust means of exploring virioplankton ecology than operational taxonomic units defined by percent sequence homology. Virioplankton RTPR populations showed positive associations with a broad phylogenetic diversity of bacterioplankton including dominant taxa within pelagic oceanic ecosystems such as *Prochlorococcus* and SAR11. Temporal changes in RTPR-virioplankton, occurring as both free viruses and within infected cells, indicated possible viral–host pairs undergoing sustained infection and lysis cycles throughout the seasonal study. Phylogenetic relationships inferred from RTPR sequences mirrored ecological patterns in virio- and bacterioplankton populations demonstrating possible genome to phenome associations for an essential viral replication gene.

INTRODUCTION

The subtropical pelagic ocean gyres are extensive, coherent regions that collectively represent the largest ecosystems on Earth, accounting for 40% of the earth's surface [1]. Circulation within the upper kilometer of subtropical gyres is primarily wind driven [2]. Due to their size, the horizontal and vertical motion of water within this upper layer of the gyres plays an outsized role in regulating global nutrient cycles and the size of the atmospheric carbon pool. Both of these issues are critical concerns for understanding anthropogenic impacts on fisheries, oceanic general circulation, and global climate. Ultimately, the activities of and the interactions between oceanic microbial communities—composed of protists, phytoplankton, bacterioplankton, and virioplankton—regulate the flow of carbon and nutrient elements within the pelagic ocean gyres [3,4]. Truly appreciating how oceans will respond to environmental changes brought on by anthropogenic activities requires detailed, mechanistic understanding of oceanic microbial communities [5–8]. This study focused on specific interactions between important and ubiquitous populations of lytic viruses [9,10] and bacterioplankton, the most abundant microbial host group within the low nutrient ecosystems that dominate much of the pelagic ocean.

Viral cell lysis is among the microbial community interactions most directly responsible for the

66 conversion of cellular biomass into dissolved and particulate organic carbon [11]. This process is one of the
67 major top-down controls on bacterioplankton productivity, responsible for between 10 and 40% of daily
68 bacterioplankton mortality in the ocean [12]. Upon lysis, cell contents not consumed by viral replication are
69 released, contributing to dissolved nutrient pools within marine ecosystems. Bacterioplankton cell lysis
70 comprises the largest known carbon flux process between biomass and the dissolved organic matter
71 (DOM) pool amounting to ~150 gigatons C/yr [6,13]. The conversion of biomass to DOM from viral lysis is
72 five times greater than that performed by other biological mechanisms such as grazing and programmed
73 cell death [6,13]. This flux from biomass to DOM sustains nutrient cycles, providing essential growth
74 substrates for bacterioplankton and phytoplankton productivity. The availability and concentrations of
75 different nutrients within the DOM pool comprise the bottom-up controls on bacterioplankton productivity
76 [6,13,14]. Besides influencing the composition, diversity, and productivity of marine bacterial communities
77 through top-down and bottom-up controls, viruses can influence bacterial physiology, improving fitness in
78 the face of environmental changes [15,16] by altering metabolic pathways [17,18] through horizontal gene
79 transfer; and through the expression of metabolic genes carried within viral genomes during viral infection
80 [19,20]. However, the behavior of individual virioplankton populations within ecosystems and their detailed
81 interactions with bacterioplankton host populations remain largely unknown. This study examined both
82 virioplankton and bacterioplankton dynamics within the North Atlantic Subtropical Gyre across time with
83 the objective of observing interactions at the community and population levels.
84 For some time, microbial ecologists have investigated bacterial populations and communities utilizing
85 16S rRNA gene amplicon sequences, a universal gene marker for bacterial diversity [21]. Due to their
86 polyphyletic origins, viruses do not have a universally conserved marker gene that could be similarly used
87 for tracing viral evolutionary history and community ecology investigations. Nevertheless, there are a few
88 widespread genes exhibiting robust evolutionary histories that can be used as markers for viral
89 community ecology studies within natural environments. Genes such as the DNA polymerase A gene
90 (encoding DNA polymerase I) [22,23], terminase gp20 gene [24–27], major capsid protein gp23 gene of T4-like
91 phages [28–30] , and ribonucleotide reductase (RNR) genes [9] have all informed studies assessing the
92 diversity of abundant viral populations. This study used amplicons of the ribonucleotide triphosphate
93 reductase (RTPR) gene for following the dynamics of virioplankton populations.
94 RNRs are the only known enzymes capable of reducing ribonucleotides to deoxyribonucleotides,
95 essential substrates for DNA synthesis [31,32]. RNRs are ancient enzymes that were essential to the
96 emergence of DNA [33]. As a consequence, RNR genes are under rigorous evolutionary selection
97 pressure, and among the most abundantly identified genes in marine virome libraries [34]. Importantly,
98 RNR genes are present in tailed phages within the class Caudoviricetes and have been identified within
99 the genomes of viruses infecting hosts within all three domains of life. Generally, lytic viruses carry RNR
100 genes, which ensures a steady supply of deoxyribonucleotides for DNA synthesis [35]. Therefore, RNRs
101 are most commonly observed within the genomes of lytic phages.
102 RNRs are biologically informative as this group of enzymes consists of three physiological classes

103 according to reactivity with oxygen. Class I RNRs are oxygen-dependent. Class II RNRs are oxygen-
104 independent and rely upon an adenosylcobalamin (vitamin B_{12}) cofactor for ribonucleotide reduction.
105 Class II RNRs are further divided according to structure and substrate utilization. Monomeric Class II
106 RTPR reduces ribonucleotide triphosphate to deoxyribonucleotide triphosphate (dNTPs), the substrate for
107 DNA synthesis. Dimeric Class II reduces ribonucleotide diphosphate to deoxyribonucleotide diphosphate
108 thus requiring a dinucleotide kinase for the final step in forming dNTPs. Class III RNRs are sensitive to
109 oxygen. As such, Class I and Class II RNRs, not Class III RNRs, are commonly present within
110 virioplankton communities found in oxygenated ocean water [10,32]. This study leveraged the predominance
111 of RTPR genes within viruses for examining the ecological dynamics of virioplankton populations. RTPR
112 and 16S rRNA amplicon sequence data were analyzed against a backdrop of microbial and
113 oceanographic observations using the latest compositional data analysis (CoDA) tools [36,37] [38,39] [40,41] for
114 assessing how lytic, B_{12}-dependent viral populations responded to seasonal dynamic changes within the
115 euphotic zone of a subtropical oceanic gyre, a region critically important within the global carbon cycle.

116 **MATERIALS AND METHODS**

117 **Study Site and Sample Collection**

118 Seawater samples were collected from a fixed station at South of Channel Faial-Pico islands (SCFP;
119 38°23'11.67"N, 28°35'6.14"W), located in the northeastern Atlantic, 6.4 nautical miles off São Mateus
120 (Pico Island) in the Azores archipelago (ICES subdivision Xa2). The station is representative of a larger
121 open ocean ecosystem, the Mid-Atlantic Ridge ecoregion, corresponding to the International Council for
122 the Exploration of the Seas (ICES) areas X and XII, which includes a high proportion of high seas [42]. The
123 Azores are located along the northern boundary of the North Atlantic Subtropical Gyre, which constitutes
124 a barrier between superficial cold waters of Nordic provenance and warmer waters of the gyre. Due to
125 the influence of large scale currents such as the Gulf Stream, the North Atlantic Current, and the Azores
126 Current, the islands are mostly exposed to the transport of warm waters of equatorial and tropical origin.
127 These currents result in high salinity, sharp horizontal temperature gradients, and high sea surface
128 temperatures (on average 15–16 °C in winter and 22–24 °C in summer). These physical conditions result
129 in low nutrient regimes, and correspond with low ecosystem productivity [43]. Nevertheless, frontal regions
130 and underwater features (seamounts, eddies, upwelling areas) can enhance local productivity [44,45].
131 Water samples were collected in acid-rinsed (0.1 M HCl) and seawater-flushed carboys between 9:30
132 and 11:00 local time from 0, 5, 25, 50, and 75 m depths during seven sampling events on May 12th, June
133 8th, June 22nd, July 11th, July 27th, August 5th, and September 8th of 2016. Surface seawater (0 m
134 depth) was collected by directly submerging a carboy. Seawater at other depths was collected using a
135 2.5 L Niskin bottle and subsequently transferred to a 5 L carboy. Seawater samples were transported
136 inside a light-tight cooler and laboratory-processed within 2 h of collection. Subsamples for flow
137 cytometric enumeration of phytoplankton were fixed immediately using glutaraldehyde (0.5% final
138 concentration). Flow cytometry samples were transported on ice (1–2 h) and stored at -80 °C until further

processing.

Water column conductivity, temperature, and depth (CTD) data was collected down to 150 m using a CTD instrument (MIDAS Valeport, Totnes, United Kingdom) (**Supplementary Table S1)**. Salinity (unitless, on the Practical Salinity Scale PSS-78) was expressed as a function of conductivity. Density (kg/m^3) was expressed as a function of pressure. CTD profiles were visualized using the cmocean colour palettes [46] on Ocean Data View 5.3.0 [47] (https://odv.awi.de). Temperature and salinity values between sampling days were interpolated using the weighted average grid method [47]. Temperature, salinity, and density were determined by averaging all values 0.5 m above and below the targeted sampling depths. Equipment malfunction prevented CTD data collection on September 8th, thus, temperature, salinity, and density metadata were estimated from the nearest time, pixel, and depth from the Marine Copernicus Global Ocean Physics model [48] (GLOBAL_ANALYSISFORECAST_PHY_CPL_001_015, http://marine.copernicus.eu/). Average values for chlorophyll a (mg/m^3), sea surface temperature (SST, °C), particulate organic carbon (POC, mg/m^3), and photosynthetically active radiation (PAR, $Einstein/m^2d$) data in 2016 were obtained from the NASA Ocean Color website (http://oceancolor.gsfc.nasa.gov/) at 1 km resolution. Satellite data was mapped using SeaDAS 7 [49] for the 9 km^2 area surrounding sampling station SCFP.

Phytoplankton Enumeration using Flow Cytometry

Phytoplankton cells were enumerated using a BD Influx Mariner 209s Flow Cytometer (BD Biosciences, San Jose, CA) equipped with a 488 nm argon laser and 525 ± 15, 542 ± 27, 585 ± 40, and 692 ± 40 nm emission filters with a nozzle size of 70 µm. After thawing and gentle vortexing, samples were run for 2 min [50]. A blank consisting of 0.22 µm filtered Milli-Q water was run before and after counting. The flow rate was determined by running Milli-Q water for 2 min and dividing the mass difference (determined with a scale) by time; mass was converted to volume using the density of 1 kg/L. Phytoplankton cells were identified based on chlorophyll fluorescence (692 ± 40 nm) from 488 nm excitation (**Supplementary Table S1**).

Virio- and Bacterioplankton Biomass Collection

The cellular (>0.22 µm) and subcellular (0.02–0.22 µm) fractions, biologically defined in this study as bacterioplankton and virioplankton, respectively, were separated through two consecutive syringe filtrations using 0.22 and 0.02 µm filters. All tubing and containers were acid-washed and deionized water-rinsed prior to sample filtration. The bacterioplankton fraction was collected by filtering 5 L of seawater through a 0.22 µm pore-size Sterivex filter (Merck Millipore, Darmstadt, Germany) under a vacuum pressure system linked to a filtration ramp (Merck Millipore, Burlington, MA). The sample collection flask was rinsed twice using the first 200 mL of 0.22 µm filtrate. Subsequently, the virioplankton fraction was collected by filtering 1 L of 0.22 µm filtrate through a 0.02 µm pore size Anotop filter (Whatman, Maidstone, United Kingdom) [51,52] connected to a 60 mL syringe, using a caulk gun which generated a steady pressure on the syringe. Both 0.22 µm and 0.02 µm filters with bacterio- or

175 virioplankton biomass were parafilm sealed and stored at -80 °C until DNA extraction.

176 **DNA Extraction from the Virioplankton Fraction and Tag-Encoded RTPR Amplicon Sequencing**

177 In brief, DNA was extracted from the 0.02 µm Anotop filters (virioplankton fraction) and quantified
178 using previously reported protocols [52]. RTPR PCR amplicons of ~750 bp were obtained using
179 virioplankton template DNA and degenerate primers designed using sequence alignments derived from
180 uncultivated virus populations observed within marine viromes. After magnetic bead purification,
181 amplicon molecules were tagged with a unique barcode sequence using ligation. Purified, barcoded
182 amplicons were then subjected to a limited cycle PCR with barcode sequences as template targets.
183 Enriched barcoded amplicons from each sample were then purified, pooled in equal proportions, and
184 sequenced using a PacBio RSII sequencer (Pacific Biosciences, Menlo Park, CA). Detailed methods
185 describing DNA extraction, amplification of RTPR target genes, and amplicon sequencing are provided in
186 **Supplementary Methods**.

187 **DNA Extraction from the Bacterioplankton Fraction, and 16S rRNA Amplicon and Tag-Encoded**
188 **RTPR Sequencing**

189 DNA was extracted from the 0.22 µm Sterivex filters (bacterioplankton fraction) using the
190 phenol/chloroform method [53]. The V3–V4 hypervariable region of the 16S rRNA gene was PCR-amplified
191 and sequenced from the bacterioplankton fraction on the Illumina MiSeq (Illumina, San Diego, CA) using
192 a Nextera XT DNA Library Preparation Kit (Illumina), which exploits a dual-indexing strategy for
193 multiplexed sequencing [54]. RTPR genes within the bacterioplankton fraction were PCR-amplified,
194 barcoded, and enriched using the same protocol as described for virioplankton. Detailed methods
195 describing DNA extraction, amplification of 16S rRNA and RTPR target genes, and amplicon sequencing
196 are provided in **Supplementary Methods**.

197 **RTPR amplicon quality control**

198 General and detailed bioinformatics analysis pipelines are summarized in **Fig. 1** and **Supplementary**
199 **Fig. S1**, respectively. Briefly, RTPR gene amplicon sequences from the virioplankton fraction (PacBio
200 RSII sequencer) and the bacterioplankton fraction (PacBio Sequel sequencer) were initially screened for
201 low quality bases and read length. Circular consensus sequence (CCS) reads were generated. Reads
202 with less than three full passes, less than 98% minimum predicted accuracy, and with a length shorter
203 than 250 bp or longer than 5,000 bp were excluded. Subsequently, sequence reads were demultiplexed,
204 and primer and barcode sequences were detected and removed. Reads were translated into predicted
205 amino acid sequences using a custom frameshift polishing pipeline
206 (https://github.com/dnasko/frameshift_polisher). Key catalytic residues within translated amplicons were
207 used for validating sequences as true RTPRs using the Protein Active Site Validation (PASV)
208 (https://github.com/mooreryan/pasv) pipeline [55]. Detailed methods describing bioinformatic processing of
209 RTPR amplicon sequences are provided in **Supplementary Methods**.

210 **Virioplankton RTPR 98% OTUs and phylogenetic clades**

211 Quality-controlled virioplankton RTPR amino acid sequences were run through 11 *de novo*
212 operational taxonomic unit (OTUs) clustering assessments varying the percent sequence identity (90–
213 100% amino acid sequence identity) using the *cluster_fast* command with default settings in USEARCH
214 (version 11.0) [56]. This heuristic process identified 98% OTU clusters as a balance between optimal
215 lumping and splitting of OTUs (**Supplementary Fig. S2**). Virioplankton RTPR 98% OTU amino acid
216 sequences were taxonomically divided into clades using a phylotyping method. OTUs were aligned using
217 MAFFT (version 7.450) with the ginsi method. A phylogenetic tree of virioplankton RTPR 98% OTUs was
218 built using FastTree (version 2.1.11) [57] and visualized using Iroki [58]. Virioplankton RTPR phylogenetic
219 clades were defined manually.

220 **16S quality control and classification**

221 General and detailed analysis pipelines are summarized in **Fig. 1** and **Supplementary Fig. S1.**
222 Demultiplexed, paired-end 16S rRNA reads from the bacterioplankton fraction were imported into QIIME2
223 (QIIME2 2019.1) [59] for quality filtering and generation of amplicon sequence variants (ASVs) using
224 DADA2 [60]. Forward and reverse reads were truncated at positions 35 and 300 and at positions 35 and
225 295, respectively. 16S ASVs were taxonomically identified (**Supplementary Table S2**) using the Silva
226 132 QIIME-compatible release (full length, seven-level taxonomy) [61] as a reference database in QIIME2.
227 Sequences taxonomically assigned to Archaea (22 ASVs) and chloroplasts (230 ASVs) were removed
228 prior to downstream analysis.
229 For proportionality tests with bacterioplankton 16S ASVs and virioplankton RTPR 98% OTUs, ASVs
230 were restricted to only those that occurred in the 29 samples with virioplankton RTPR amplicons. From
231 that subset of 29 samples, only ASVs with a total count greater than 100 and with at least 10 counts in at
232 least 10 samples were retained for downstream analysis. This avoided the risk of spurious associations
233 based on rare populations observed in few samples.

234 **RTPR and 16S community and population analyses**

235 Virioplankton RTPR 98% OTUs, virioplankton RTPR phylogenetic clades, and bacterioplankton 16S
236 ASVs were used in community and population analyses such as alpha and beta diversity, and population
237 proportionality. The influence of environmental conditions on virio- and bacterioplankton communities
238 were determined based on complete-linkage hierarchical clustering of RTPR 98% OTUs, RTPR
239 phylogenetic clades, or 16S ASV community profiles using the hclust method in R. Tested environmental
240 features include sampling day, depth, temperature, density, current velocity, salinity, nanoeukaryote
241 abundance, picoeukaryote abundance, *Prochlorococcus* abundance, *Synechococcus* abundance, total
242 autotroph abundance, viral abundance, and bacterial abundance. Clustering of virio- and
243 bacterioplankton communities was based on the Aitchison distance between the May 12th 5 m sample
244 and every other sample. Resulting dendrograms were visualized with Iroki. Correlations between virio-
245 or bacterioplankton communities and associated environmental features were tested using the *qiime*

diversity mantel [65] command in QIIME2 (permutation times: 999). Significant correlations were defined as $p < 0.05$. Associated environmental features that showed significant correlations with sample community composition were plotted next to the sample community composition dendrogram.

Alpha diversity of the virio- or bacterioplankton communities in each sample was estimated based on the Shannon index using the R package DivNet (version 0.3.2) [66] and displayed using ggplot2 (version 3.3.0) [67]. Significance of alpha diversity differences was tested using betta function in the R package DivNet breakaway (version 4.6.14) [68]. Estimated Shannon indices were converted to the effective number of species (ENS, Hill numbers) [69]. Correlations of ENS estimates based on 98% OTUs or RTPR phylogenetic clades were tested using the cor.test method in R. Beta diversity was visualized through principal component analysis (PCA) plots of sample community compositions using the R package compositions (version 1.40.3) [70]. Significance of beta diversity differences between sampling day or depth groups was tested using *beta-group-significance* command (permanova) method [71] in QIIME2.

Associations within or between virioplankton RTPR phylogenetic clades and bacterioplankton 16S ASV populations were explored using proportionality (*rho*, ρ) tests, with $-1 < \rho < 1$ [72]. Only positive ρ values reflecting a positive association between two populations were considered, as negative ρ values cannot be conclusively explained [73]. Only those associations with false discovery rates (FDRs) lower than 0.05 were retained for further analysis. Proportionality and FDR tests were performed using the R package propr (version 4.2.8). Associations with FDR <0.05 were extracted and plotted in a heatmap using ComplexHeatmap [74]. Columns and rows of the heatmap were clustered based on phylogenetic relations of virio- and bacterioplankton, respectively. Virioplankton RTPR clades were arranged by their topological order on the 98% OTU phylogenetic tree. Bacterioplankton 16S ASVs were arranged by their 16S phylogenetic placement or association patterns.

Virio- and bacterioplankton population analysis using common RTPR 98% OTUs

In 27 of 35 samples, RTPR amplicons were obtained from both virio- and bacterioplankton fractions yielding 47,997 quality-controlled RTPR amino acid sequences. Quality-controlled virio- and bacterioplankton RTPR amino acid sequences were run through 11 *de novo* operational taxonomic unit (OTU) clustering assessments varying the percent sequence identity (90–100% amino acid sequence identity) using the *cluster_fast* command with default settings in USEARCH. This heuristic process identified 5,411 98% OTU clusters as a balance between optimal lumping and splitting of OTUs (**Supplementary Fig. S3**). OTUs with less than ten counts were removed prior to downstream analysis. The COUNTIF function in Excel was used for binning the remaining 254 RTPR 98% OTUs into those containing sequences from both fractions, only the virioplankton fraction, or only the bacterioplankton fraction. RTPR 98% OTU counts were calculated using the SUM function in Excel prior to zero replacement and clr transformation. An arbitrary clr value of 5 or -5 was assigned to those common 98% OTUs found exclusively in the virio- or bacterioplankton fraction, respectively. The difference in clr abundance between fractions of each RTPR 98% OTU was calculated using the subtraction formula in Excel. Venn diagrams were plotted to show the total clr abundance of each OTU and the distribution of

283 each OTU (difference in clr abundance) in the virio- and/or bacterioplankton fraction.

284 Only the August 5th sampling date yielded RTPR amplicons from virio- and bacterioplankton
285 fractions at all depths. Thus, this date was chosen for assessing depth-resolved distribution patterns of
286 RTPR virioplankton populations within the virio- and bacterioplankton fractions. Calculation of clr
287 abundance values for the August 5th OTUs was performed by removing any OTU not present in at least
288 one August 5th sample, leaving 1,477 of the 5,411 OTUs observed in the combined virio- and
289 bacterioplankton datasets. Twenty-nine August 5th OTUs contained more than 50 sequences and had
290 two or fewer zero replacements, and were retained for further analysis. Similarity between the
291 phylogenetic distance of these 29 RTPR 98% OTUs and their depth-resolved distribution patterns within
292 the virioplankton fraction, the bacterioplankton fraction, and the combined virio- and bacterioplankton
293 fraction were tested by plotting the cophenetic distance versus Aithison distance for all combinations of
294 29 OTUs. Combinations rather than permutations of OTUs were used, as the distances used are
295 symmetric (e.g., d(A, B) = d(B, A), where d is the distance function, and A and B are OTUs). Self-
296 distances (e.g., d(A, A), d(B, B)) were not included. Linear models and significance tests were performed
297 with the lm function in R.

298 **Data transformation**

299 Compositional data analysis (CoDA) methods were applied to all analyses of community diversity and
300 population interactions. Zero values in the feature tables [virioplankton RTPR 98% OTUs
301 (**Supplementary Table S3**); RTPR phylogenetic clade (**Supplementary Table S4**); bacterioplankton 16S
302 ASVs (**Supplementary Table S5**); bacterioplankton 16S ASVs retained in the 29 samples with RTPR
303 amplicons (**Supplementary Table S6**); and common RTPR 98% OTUs found in both the bacterio- and
304 virioplankton fractions (**Supplementary Tables S7 & S8**)] were replaced by imputed values using the
305 count zero multiplicative method from the R package zCompositions (version 1.3.3.1) [62] in R (version
306 3.5.3) [63]. The centered log-ratio (clr) transform was applied to the zero-replaced data set to produce clr
307 abundance using the R package CoDaSeq (version 0.99.4) [64]. Within the August 5th subset (sampling
308 date with RTPR amplicons in both fractions at all depths) of common RTPR 98% OTUs (**Supplementary**
309 **Table S9**), zero count values were replaced using *cmultRepl* with the method count zero multiplicative
310 (CZM) in the zCompositions package (version 1.3.4) of R (version 4.0.2). Subsequently, the clr
311 transformation was performed using base 2 in the R package CoDaSeq.

312 **RTPR 98% OTUs within global RNR context**

313 RTPR 98% OTUs from both virio- and bacterioplankton fractions, RTPRs identified through stringent
314 homology search of the *Tara* Ocean dataset [75], and RNRdb [76] databases were aligned using Geneious
315 10.0.9 MAFFT alignment FFT-NS-2 and trimmed to the region of interest (H346 to S643 in *Lactobacillus*
316 *leichmannii* monomeric class 2 adenosylcobalamin-dependent ribonucleotide-triphosphate reductase). A
317 phylogenetic tree of trimmed sequences was built using FastTree and visualized with Iroki.

318 **Data Deposition**

319 RTPR and 16S rRNA gene sequences were deposited in NCBI BioProject under the accession
320 number PRJNA842570 and Zenodo under DOI 10.5281/zenodo.7313881
321 (https://doi.org/10.5281/zenodo.7313881).

322 **RESULTS**

323 **Water Column Environment**

324 Temperature of surface waters at station SCFP increased from May 12th (16.4 ºC) to August 5th
325 (22.0 ºC) (**Fig. 2A**), agreeing with broader sea surface temperature (SST) data derived from satellite
326 imagery (**Supplementary Fig. S4A**). Surface salinity (using the Practical Salinity Scale) increased from
327 May 12th (35.9) to July 11th (36.1), decreased by 0.1 on July 27th, and increased again to 36.1 on
328 August 5th (**Fig. 2B**). Surface waters were well mixed down to 50 m as indicated by temperature and
329 salinity contours prior to June (**Fig. 2**). Water column stratification intensified throughout the sampling
330 period, accompanying the mixed layer depth rising from 43 m (May 12th) to 5 m (August 5th). Salinity
331 showed distinct sub-surface increases during the study period, indicating substantial water movement in
332 the sampling region. Salinity changes in the sub-surface were centered at the 75 m depth, from June 8th
333 to July 27th, especially on July 11th. The observations suggested the presence of an anticyclonic
334 (downwelling) vortex at the sampling location around July 11th. Uplifts of isotherms and isohalines found
335 in temperature and salinity profiles between sampling dates suggested the presence of cyclonic
336 (upwelling) vortices, on June 22nd (more pronounced) and on July 27th (less pronounced) extending to
337 August 5th. Satellite-derived data in 2016 (**Supplementary Fig. S4A**) showed one phytoplankton bloom,
338 based on chlorophyll a and particulate organic carbon (POC), that occurred between February and April,
339 with its maximum in March.
340 The sampling period started after the seasonal spring phytoplankton bloom and spanned the
341 environmental and biological conditions observed from late spring to late summer. Observed changes in
342 phytoplankton abundance corresponded with seasonal environmental changes during the sampling
343 period (**Fig. 2A** and **Supplementary Fig. S4B**). High phytoplankton abundance (19–61 x 10^3 cells/mL)
344 was observed throughout the water column from 0 to 75 m depth on May 12th, indicating the end of the
345 annual spring bloom. Subsequently, phytoplankton abundance fell below 2 x 10^3 cells/mL with the
346 exception of a sub-surface peak of ~55 x 10^3 cells/mL at 50 m on June 22nd, coinciding with the
347 observed cyclonic upwelling event (**Supplementary Table S1**). The lowest phytoplankton abundance
348 was observed on July 11th, coinciding with the anticyclonic downwelling event. Another increase in
349 phytoplankton abundance coincided with the second, albeit less pronounced, cyclonic upwelling event
350 during the subsequent two sampling dates, July 27th and August 5th.

351 **Changing environmental conditions and virioplankton and bacterioplankton community**
352 **dynamics**

353 For unknown reasons, only 29 of 35 virioplankton samples yielded RTPR amplicons (**Supplementary**
354 **Tables S3 & S10**). From a total of 30,473 virioplankton RTPR amino acid sequences, 3,697 RTPR 98%
355 OTUs were identified (**Supplementary Fig. S1 and S2**). The RTPR 98% OTUs clustered into 17
356 phylogenetic clades as manually determined by phylogenetic tree topology (**Fig. 3**). The three clades
357 topologically located between clades 12 and 13 (gray color branches in **Fig. 3**) were excluded from
358 downstream analyses due to their low abundance (7 OTUs, 12 amino acid sequences) and distant
359 phylogenetic relations to other clades, leaving 3,690 98% OTUs (30,461 amino acid sequences,
360 **Supplementary Table S3**) in 17 clades (**Supplementary Table S4**). Phylogenetic clade 5 was the most
361 abundant based on both amino acid sequence counts (3,860) and clr abundance (1.07). The 17 RTPR
362 phylogenetic clades were used in subsequent compositional data analyses, including community diversity
363 and population interactions analyses, as clades of related OTUs provided clearer ecological signals than
364 the larger collection of more highly resolved RTPR 98% OTUs. However, as shown in supplemental
365 analyses, RTPR 98% OTU community diversity within and between samples was also determined for
366 comparison with the phylogenetic approach.

367 The V3–V4 region of the 16S rRNA gene was PCR-amplified from 35 bacterioplankton fraction
368 samples and sequenced (**Supplementary Tables S2 & S10**). After sequence quality filtering, 1,779
369 amplicon sequence variants (ASVs) were identified from a total of 918,792 sequences across all samples
370 (**Supplementary Table S5**). Complete-linkage hierarchical clustering based on the Aitchison distance of
371 clr-transformed sequence abundance between the first (May 12th 5 m) water sample and every other
372 sample was performed for each of three datasets (virioplankton RTPR 98% OTUs, RTPR phylogenetic
373 clades, or bacterioplankton 16S ASVs). Mantel tests assessed each Aitchison distance matrix with
374 distance matrices of all collected environmental metadata (**Supplementary Table S11**) to identify
375 environmental factors correlated with and potentially driving the clustering of virio- or bacterioplankton
376 communities. Sample depth, collection day (displayed as Julian day), temperature, and water density
377 were significantly (Spearman and Pearson p-value <0.02) correlated with virioplankton community
378 composition (RTPR phylogenetic clades and RTPR 98% OTUs) and bacterioplankton community
379 composition. For the significant environmental variables, correlation coefficients with virioplankton
380 populations were generally higher when using the RTPR phylogenetic clades.

381 Hierarchical clustering of virioplankton communities based on RTPR phylogenetic clades formed four
382 large groups (vA, vB, vC, and vD) (**Fig. 4A**) based on their distance from the May 12th 5 m sample.
383 Community groups vA/vB and groups vC/vD were organized into equally distant supergroups. Within
384 each group, communities had similar profiles for depth/density and for temperature/day. Groups vA/vB
385 contained communities generally occurring at deeper depths with lower water temperatures, whereas
386 groups vC/vD contained communities generally collected at shallower depths with higher water
387 temperatures. However, for all groups, there were outliers likely reflecting some of the low rank

correlation coefficients (0.19–0.51) between community structure and environmental metadata (**Supplementary Table S11**). Bacterioplankton communities formed three groups (bA, bB, and bC) (**Fig. 4B**) based on their distance to the May 12th 5 m sample. Communities in groups bA and bB were observed at lower water temperatures, whereas group bC communities were found in higher water temperatures at shallow depths (<25 m). The division between groups bA and bB was more subtle and may have been forced by depth as group bB contained communities in only 50 and 75 m samples.

The effects of sampling date (**Fig. 5A**) and depth (**Supplementary Fig. S5A**) on viral and bacterial ENS were significant (betta <0.05; data not shown); therefore, these environmental parameters were selected for analysis of their interactions with alpha and beta diversity of the virio- and bacterioplankton communities. Over the interseasonal sampling period, virioplankton alpha diversity according to RTPR phylogenetic clades (**Fig. 5A)** fluctuated slightly reaching a maximum around the July 11th anticyclonic downwelling event (**Fig. 2)**. This peak was shouldered by minimum diversity estimates on June 8th and July 27th corresponding with cyclonic upwelling events. Virioplankton alpha diversity fluctuated slightly with depth (**Supplementary Fig. S5A)**. Maximal and minimal diversity estimates occurred at 25 and 75 m, respectively. ENS estimates of alpha diversity based on 98% OTUs (**Supplementary Fig. S6A and B**) were three to six-fold higher than ENS estimates for virioplankton based on RTPR phylogenetic clades (**Fig. 5A and Supplementary Fig. S5A**) as there were thousands of 98% OTUs versus only 17 RTPR phylogenetic clades. The general trends of ENS estimates based on 98% OTUs and RTPR phylogenetic clades were consistent (Spearman's rank correlation: $\rho = 0.82$, p value = 0.034; Pearson's product-moment correlation: $r = 0.78$, p value = 0.037). Increases in RTPR 98% OTU ENS (**Supplementary Fig. S6A**) were observed on June 22nd and July 11th, corresponding with the observed cyclonic upwelling event. The maximum ENS was observed on September 8th; unfortunately its association with an upwelling event is unknown as equipment failure prevented temperature measurements on this date. The maximum and minimum virioplankton RTPR 98% OTU ENS in samples by depth were 5 and 50 m, respectively (**Supplementary Fig. S6B**). Bacterioplankton alpha diversity steadily declined from its May 12th peak to a minimum value on July 11th followed by a slight increase over the remaining sampling dates (**Fig. 5A**). In contrast to virioplankton, bacterioplankton alpha diversity did change dramatically with depth demonstrating steady increases from minimal diversity estimates in surface waters to the maximal diversity estimates observed at 75 m (**Supplementary Fig. S5A**).

Relationships between virioplankton RTPR phylogenetic clade communities were examined using principal components analysis (**Fig. 5B)**. Sixty-five percent of the variability between communities was explained by the first two principal components (PC1 48.9%, PC2 16.0%), and sampling date (**Fig. 5B)** seemed to explain this variability more than sample depth (**Supplementary Fig. S5B**). Using the 98% OTU populations, ~35% of the variability between communities was explained by the first two principal components (PC1 22.8%, PC2 12.1%), and was associated with both sampling date and depth (**Supplementary Fig. S6C and D**). The variability in virioplankton communities explained by the RTPR 98% OTUs was approximately half of that explained by phylogenetic clades. This trend was attributed to

425 the higher number and granularity of RTPR 98% OTU populations in describing virioplankton community
426 structure. For either means of defining virioplankton communities, PCA analyses showed that the May
427 12th virioplankton communities grouped closer to the origins of both the PC1 and PC2 axes, apart from
428 most other sampling dates, and the August 5th samples tended to place toward the ends of PC1 and PC2
429 axes. Virioplankton communities as defined by RTPR phylogenetic clades demonstrated significant
430 differences based on date and depth (PERMANOVA *p* value <0.05; **Supplementary Table S12A & D)**,
431 largely agreeing with observations of 98% OTUs **(Supplementary Table S12B & E**). May 12th
432 virioplankton communities differed from most other timepoints. June 8th communities differed from July
433 27th, and August 5th communities differed from those on June 8th, June 22nd, and July 11th. Depth-
434 based virioplankton community structure (**Supplementary Fig. S5B)** did not group as strongly on the
435 PCA plots as date-based communities (**Fig. 5B**). Nevertheless, surface communities (0 m) were different
436 from those at 50 and 75 m, and 5 m communities differed from 75 m communities.
437 Relationships between bacterioplankton 16S ASV communities were examined using principal
438 components analysis as well (**Fig. 5C)**. Approximately 38% of the variability between communities was
439 explained by the first two principal components (PC1 25.2%, PC2 12.4%), and was associated with both
440 sampling date (**Fig. 5C**) and depth (**Supplementary Fig. S5C**). As seen with viral communities defined
441 by either RTPR phylogenetic clades or 98% OTUs, the May 12th bacterioplankton samples grouped
442 toward the origin of both PC1 and PC2, apart from most other sampling dates. The August 5th
443 bacterioplankton communities showed a greater spread across PC1 than the virioplankton communities
444 for this sampling date. These observations were statistically confirmed. Bacterioplankton communities
445 on May 12th and June 8th were different from those observed on all other sampling dates, except the pair
446 of samples from June 8th and June 22nd (PERMANOVA *p* value and/or *q* value <0.05; **Supplementary**
447 **Table S12C**). Bacterial 16S ASV communities were also separated by depth, with 50 m and 75 m
448 communities occurring along the left side of PC1 (**Supplementary Fig. S5C**). Surface bacterioplankton
449 communities at 0 and 5 m significantly diverged from the deep communities at 50 and 75 m; 25 m
450 communities diverged from the 75 m communities; and 50 m communities diverged from the 75 m
451 communities (PERMANOVA *p* value and/or *q* value <0.05; **Supplementary Table S12F**).

452 **Associations between virioplankton RTPR phylogenetic clades and bacterioplankton 16S ASVs**

453 The proportionality test was used for comparing changes in the clr abundance of 17 virioplankton
454 RTPR populations (as defined by phylogenetic clades on **Fig. 3**) and the subset of 221 bacterioplankton
455 16S ASV populations in 27 samples with both virioplankton RTPR and bacterioplankton 16S amplicons
456 (**Supplementary Table S13**). Virio- and bacterioplankton population pairs with ρ values ≥0.48 had
457 positive associations as the false discovery rates were ≤0.05 (**Supplementary Fig. S7**). Nine of the 17
458 virioplankton RTPR phylogenetic clades had positive associations (ρ ≥.48) with 39 of 221
459 bacterioplankton ASV populations (**Fig. 6**). Associating virioplankton clades were phylogenetically
460 diverse and were observed at both high (clade 2 and 5; clr 1.03 and 1.07, respectively) and low
461 abundance (clade 3 and 11; clr -0.90 and -2.57, respectively) (**Fig. 3**). Virio- and bacterioplankton

462 abundance did not appear to drive these associations as low-abundance RTPR virioplankton associated
463 with high-abundance bacterioplankton and vice versa (**Supplementary Fig. S8)**. The phylogenetic
464 diversity of positively-associating bacterioplankton populations was broad, encompassing seven bacterial
465 phyla observed among the included 16S ASV populations. The bacterioplankton 16S ASV populations
466 ranged in clr abundance from the most abundant taxon 3.65 (Prochlorococcus_MIT9313) to one of the
467 least abundant taxon -1.99 (SAR116_clade).
468 Virioplankton RTPR phylogenetic clades 1 and 3 were found to be associated with the most abundant
469 bacterioplankton 16S ASV, Prochlorococcus_MIT9313, and several 16S ASVs from Proteobacteria, such
470 as SAR86, SAR11, and the Aegan169 marine group, well known to be abundant within oceanic
471 ecosystems. Virioplankton clade 2 showed associations with multiple bacterioplankton ASVs from
472 Proteobacteria and Bacteroidetes. Interestingly, RTPR phylogenetic clade 2 was not associated with any
473 of the same 16S ASVs as clades 1 and 3, despite its close phylogenetic relationship with these clades
474 (**Fig. 3**). Phylogenetic clades 3 and 4 were both associated with SAR11_clade_IV_1, and clade 4 was
475 also associated with SAR11_clade_IV_2. Virioplankton clades 11 and 15 were each only associated with
476 a single bacterioplankton 16S ASV, uncultured_Thiotrichaceae and SAR86_clade_4, respectively.
477 Virioplankton clade 5, the most abundant RTPR phylogenetic clade (clr 1.07, **Fig. 3**), showed the most
478 positive associations across a wide phylogenetic breadth of bacterioplankton populations, encompassing
479 four of the nine bacterial phyla and 15 of the 39 total positively-associated 16S ASVs, including
480 SAR11_clade_II.
481 Fourteen out of 136 non-self comparisons between virioplankton RTPR populations showed a
482 positive association (**Supplementary Fig. S9**, **Supplementary Table S13**). These associations between
483 virioplankton clades did not follow a pattern according to phylogenetic distance as only three of the 14
484 associations were between neighboring RTPR clades (clades 2 and 3; clades 5 and 6; and clades 8 and
485 9).
486 Of the 221 bacterioplankton 16S ASV populations, 36 showed positive, non-self associations with
487 other ASV populations (**Supplementary Fig. S10**, **Supplementary Table S13**). Overall, there were 390
488 positive non-self associations among the 36 ASVs, ~5% of all the non-self comparisons (1260).
489 Positively associating bacterioplankton 16S ASV populations were clearly divided by their association
490 with either SAR11_clade_II or Prochlorococcus_MIT9313 (red boxes in **Supplementary Fig. S10**). No
491 16S ASVs associated with both SAR11_clade_II and Prochlorococcus_MIT9313. The SAR_clade_II
492 population was associated with only other low abundance (clr <0) ASVs, whereas
493 Prochlorococcus_MIT9313 was associated with mostly high abundance (clr >0) ASVs. Beyond the
494 delineation of SAR11_clade_II from Prochlorococcus_MIT9313, positively associating bacterioplankton
495 ASV populations could be divided into three larger groups (**Supplementary Fig. S10** x-axis). Nearly all of
496 the ASV populations within group 1 all showed positive associations with SAR11_clade_II and nine other
497 taxa (clr values ranging from -1.38 to -0.35) across five phyla (Marinimicrobia:
498 Marinimicrobia_SAR406_clades 1 (clr -1.22) & 2 (clr -1.13); Actinobacteria: Candidatus_Actinomarina (clr

499 -1.38); Planctomycetes: uncultured_Rubinisphaeraceae (clr -1.11); Verrucomicrobia: Roseibacillus (clr -
500 0.35); Proteobacteria: UBA10353_marine_group (clr -1.12), Ga0077536 (clr -1.28), OM60_NOR5_clade
501 (clr -1.03), and Rhodobacteraceae (clr -1.02)). ASV populations within group 2 were defined by their
502 consistent association with the most dominant ASV taxa Cyanobacteria: Prochlorococcus_MIT9313 (clr
503 3.65). In addition to this singular representative of the Cyanobacteria, group 2 consisted of 13
504 *Proteobacteria* (clr values ranging from -0.45 to 2.15) and three *Bacteriodetes* (clr values ranging from -
505 1.45 to 0.19). Several of the ASV taxa in group 2, such as Proteobacteria:
506 uncultured_Ectothiorhodospiraceae_2 (clr 1.72), Proteobacteria: SAR11_clade_IV_1 (clr 2.15) & 2 (clr
507 1.95), and Proteobacteria: SAR86_clade 3 (clr 1.51) were the most abundant bacterioplankton
508 populations observed in the study. ASV populations within group 3, which consisted of only three
509 Proteobacteria taxa, showed no consistent associations within the group and few associations overall.

510 **Depth- and time-resolved viral–host associations**

511 Spatiotemporal patterns in bacterioplankton associations with clade 5, the most abundant RTPR
512 virioplankton population, were explored by plotting changes in clr abundance at each sampling depth over
513 time (**Supplementary Fig. S11**). Clade 5 RTPR viral populations showed their highest abundances at
514 sampling depths of 25 m and deeper, with consistently high abundances occurring at 75 m. Deeper
515 samples also contained a greater number and diversity of bacterioplankton ASV taxa showing
516 associations with clade 5 RTPR virioplankton. Among the sampling dates, July 11th was noteworthy as
517 several of the ASV taxa associated with clade 5 RTPR virioplankton showed dramatic reductions in clr
518 abundance corresponding with the anticyclonic downwelling event. Correspondingly, the clr abundance
519 of clade 5 virioplankton held steady or declined on the two subsequent sampling dates (July 27th and
520 August 5th). A similar spatiotemporal analysis of associations between virio- and bacterioplankton
521 populations was conducted for abundant bacterioplankton populations, Prochlorococcus_MIT9313 and
522 three SAR11 clades, and four associating RTPR virioplankton populations (**Supplementary Fig. S12**).
523 Clades 1 and 3 RTPR virioplankton populations showed associations with Prochlorococcus_MIT9313
524 (**Fig. 6**). Within a sampled depth across time, the clr abundance of clade 1 and 3 viruses was relatively
525 steady, however, clr abundance of these two populations decreased in the 75 m samples as did the
526 abundance of their associated Prochlorococcus_MIT9313 population. RTPR clade 3 was also associated
527 with SAR11_clade_IV_1 which, like Prochlorococcus_MIT9313, was most abundant in surface waters.
528 RTPR clade 4 showed positive associations with SAR11_clade_IV_1 and _2 **(Fig. 6),** which were the
529 second and third most abundant bacterioplankton taxa showing associations with RTPR virioplankton.
530 While these SAR11 populations were most abundant at the surface (clr values of around 3.75), they were
531 also abundant throughout the water column and changed only modestly with time. The behavior of RTPR
532 clade 4 showed a similar pattern as it was most abundant in surface waters (clr values ≥ 1) and then
533 lower abundance at 50 and 75 m (clr values ~ 0). RTPR clade 5 associated with SAR11_clade_II which
534 was the rarest of the SAR11 taxa showing associations with an RTPR virioplankton population.
535 SAR11_clade_II populations increased with depth, showing their highest abundances (clr values ~ 0) at

536 75 m. Similarly, RTPR clade 5 was less abundant at shallow depths (clr values < 1) and then increased
537 to its highest abundances at 75 m (clr values > 1).

538 **Free and actively replicating viral populations: RTPR 98% OTUs occurring in virio- and**
539 **bacterioplankton fractions**

540 Twenty-nine virioplankton and 31 bacterioplankton samples provided RTPR amplicons with an
541 overlap of 27 samples providing amplicons from both. These 27 paired samples provided 47,997 RTPR
542 quality-filtered amino acid sequences, which were clustered *de novo* into 5,411 RTPR common 98%
543 OTUs (**Supplementary Table S7**). After removing 98% RTPR OTUs with less than ten sequences, 21
544 OTUs were observed only in the virioplankton fraction, 40 only in the bacterioplankton fraction, and 193 in
545 both fractions (**Fig. 7** and **Supplementary Table S8**). An encompassing phylogenetic analysis of these
546 RTPR 98% OTUs, sequences from the RNR database, and viromes from the *Tara* Oceans expedition
547 showed that the RTPR 98% OTUs from the virio- and bacterioplankton fractions occurred within a section
548 of the tree containing few sequences from known bacteria or viruses (RNRdb) but several sequences
549 from *Tara* Oceans viromes (**Supplementary Fig. S13**). The most abundant OTUs were primarily found in
550 the shared fraction (**Fig. 7A**). Less abundant OTUs were largely found in either virio- or bacterioplankton
551 fractions only. Examining the relative contribution of virio- or bacterioplankton amplicon sequences to
552 RTPR 98% OTUs required assignment of an arbitrary clr of 5 and -5 to those OTUs found only within the
553 virio- or bacterioplankton fractions, respectively. The difference in clr abundance within each RTPR 98%
554 OTU (clr(virioplankton)-clr(bacterioplankton), **Fig. 7B**) found that within most of the shared OTUs,
555 virioplankton contributed a greater number of amplicons (66% of all shared OTUs (128 of 193), red
556 colored OTUs above the dashed line in **Fig. 7B**).

557 All samples (fractions and depths) from the August 5th sampling date provided RTPR amplicon
558 sequences, enabling coordinate examination of virioplankton populations observed both as free viruses
559 (i.e., within the virioplankton fraction) and viruses replicating within cells (i.e., within the bacterioplankton
560 fraction). Twenty-nine RTPR 98% OTU populations were selected, each containing greater than 50 total
561 sequences with two or fewer zero-replaced samples (**Supplementary Table S9**). These heuristics
562 prevented the inclusion of rare and inconsistently observed virioplankton populations. Phylogenetic
563 analysis of these 29 RTPR 98% OTUs alongside observations of clr abundance of each OTU within the
564 virio- and bacterioplankton fractions addressed the hypothesis that the evolutionary distance between
565 virioplankton populations reflects similarity in their ecological behavior (**Fig. 8**). Statistical testing
566 comparing the cophenetic distance of the OTUs with the Aitchison distance of the OTUs according to
567 their clr abundances across the August 5th samples demonstrated a significant (p = 1.6 e^{-5}), weak (R^2 =
568 0.045), positive correlation (correlation r = 0.212) between evolutionary distance and clr abundance in the
569 bacterioplankton fraction (**Supplementary Fig. S14**). There was no significant relationship between
570 cophenetic distance and clr abundance in the virioplankton fraction or when combining the virio- and
571 bacterioplankton fractions (data not shown). Thus, more closely related RTPR virioplankton populations
572 demonstrated similar ecological behavior in terms of abundance as viruses replicating within cells (i.e.,

573 the bacterioplankton fraction), but not as free viruses.

574 The 29 August 5th RTPR 98% OTUs occurred across nearly all of the major RTPR clades identified
575 in **Fig. 3**, with the exception of clades 10 and 11 which were among the rarest clades observed in the
576 study (clr values -0.95 and -2.57, respectively). Clade 7 was the most represented among the August 5th
577 OTUs. These OTUs demonstrated a distinctive evolutionary split with a group of closely related OTUs (9,
578 15, 7, 16 and 26) clustering away from OTU 25 (**Fig. 8**). The clr abundance heat maps of the closely
579 related group were similar with the majority of sequences occurring in the virioplankton fraction above 25
580 meters. In contrast, OTU 25 sequences were more abundant below 25 m demonstrating high abundance
581 in the bacterioplankton fraction at 50 m. By and large, closely related August 5th OTUs showed similar
582 clr abundance patterns supporting the finding of a significant relationship between evolutionary distance
583 of RTPR virioplankton populations and Aithison distance.

584 **DISCUSSION**

585 **Ribonucleotide triphosphate reductase: a hallmark gene within viruses**

586 As the only enzyme capable of reducing ribonucleotides to deoxyribonucleotides, RNR effectively
587 controls the rate of DNA synthesis [33]. Thus, it makes intuitive sense that RNR genes would be common
588 within the genomes of lytic viruses as speed and efficiency in genome replication is under positive
589 selection pressure for viruses exhibiting a virulent lifecycle. Indeed, RNR genes are among the most
590 commonly observed genes within dsDNA viral genomes and viromes, lending support to this hypothesis
591 [34,77–79]. RNRs are ancient proteins as enzymatic ribonucleotide reduction was a critical step from the RNA
592 world to the emergent dominance of DNA as the principle information-carrying molecule for life on Earth
593 [80–82]. Thus, it is possible that RNRs existed within viral genomes not long after the emergence of DNA.
594 Phylogenetic analysis of known and environmental RTPRs (**Supplementary Fig. S13**) indicated that viral
595 RTPR proteins diverged from cellular RTPRs and are as deeply branching, reflecting both the unique
596 selective pressure on viral RTPRs and their long evolutionary history. However, the tree also shows the
597 emergence of cellular RTPRs from clades dominated by viral sequences and vice versa. These
598 instances may represent viral to cellular (and cellular to viral) gene transfer events. Throughout the long
599 evolutionary history of RNRs there is evidence of gene transfer between cellular kingdoms, most notably
600 the transfer of Class I RNRs from bacteria to archaea and ultimately eukaryotes [31] where Class I RNRs
601 predominate (**Supplementary Fig. S15A**). In the case of nucleotide metabolism proteins, such as RNRs,
602 it is possible that greater levels of biochemical innovation occur within lytic viruses than within cells, as
603 viruses typically carry one allele of a particular nucleotide metabolism gene which is under intense
604 selective pressure for fast and efficient DNA replication. Certainly, at the level of RNR class, it is clear
605 that selective processes may influence the distribution of RNR classes across the tree of life, where
606 viruses demonstrate a distinctly high frequency of Class II RTPRs in comparison with cellular kingdoms.
607 Among known viruses, Class II RTPRs demonstrate a distinctive skew in taxonomic distribution favoring
608 tailed phages having a myoviridae or siphoviridae morphology (**Supplementary Figs. S15A & B**).

609 Interestingly, a breadth of microbial diversity was associated with several of the RTPR viral
610 phylogenetic clades (**Fig. 6**). There are a few possible explanations for this. One possibility is that
611 physiochemical factors were stronger drivers of microbial dynamics than viral-induced mortality, resulting
612 in positive associations between RTPR viral clades and a sub-community of similarly adapted organisms,
613 both host and non-host alike. However, it is also possible that the diverse taxa positively associated with
614 several of the RTPR viral clades are a reflection of the hosts for viruses within these clades. Abundant
615 marine phages, like 37-F6, have been predicted to infect hosts spanning different phyla [83], while direct
616 linkages between viral RNR genes and 16S rRNA genes from diverse bacterial taxa have been observed
617 [84]. In the latter case, viruses associated with diverse taxa were less efficient at infecting these hosts than
618 viruses with narrow host ranges [84], but even infrequent cross-taxa infections could provide ample
619 opportunities for horizontal gene transfer (HGT) of RNR genes between viral populations, resulting in
620 similar RNR sequences across a diverse group of viruses infecting different hosts. Furthermore,
621 horizontal gene transfer of RNRs across all three domains of life [31], as well as between phage
622 populations [34], is well-established, suggesting that HGT has been critical to shaping viral ecology and
623 cobalamin cycling in marine environments. Answering questions surrounding cellular–viral gene
624 exchanges and their contribution towards the emergence of biochemical innovation will improve
625 understanding of the evolutionary mechanisms behind adaptation in microbial communities.

626 **Biochemistry constrains RTPR virioplankton population ecology**

627 RNR is an essential gene for cellular life as all cells contain dsDNA genomes. However, the
628 frequency of RNR types varies substantially among known cellular organisms and viruses
629 (**Supplementary Fig. S15A**). Meta-analysis of RNR protein sequences within the RNRdb [76] along with
630 sequences in the UniProt knowledgebase [85] showed the high frequency of RTPR (i.e., Class II
631 monomeric RNR) within known viruses (~22%) as compared to cellular kingdoms (2–3%). Within the
632 Caudovirales most RTPR genes occurred within siphophages (**Supplementary Fig. S15B**). In contrast,
633 Archaea and Bacteria favor Class II dimeric or Class III RNRs for anaerobic ribonucleotide reduction, a
634 biochemical strategy strikingly different from viruses. In the case of RTPR, this monomeric Class II RNR
635 requires a adensyolcobalamin (B_{12}) co-factor (like all Class II RNRs), can perform ribonucleotide
636 reduction under aerobic or anaerobic conditions, and, along with Class III RNRs, utilizes a ribonucleotide
637 triphosphate (NTP) substrate [86]. The B_{12} requirement and use of NTP substrates have intriguing
638 implications for the ecology of viruses carrying RTPR. Since cobamide nutrient (including B_{12}) synthesis
639 is complex, specialized, and energy intensive, few marine bacterial and archaeal phyla possess the entire
640 synthesis pathway [87,88]. Most bacteria and phytoplankton are auxotrophic for B_{12} and are thus dependent
641 on the few taxa capable of B_{12} production. RTPR virioplankton populations must infect a host that can
642 either synthesize B_{12} or assimilate and subsequently modify a B_{12} precursor containing the corrin ring.
643 Moreover, B_{12} has several chemical analogs depending on the combination of ligands at the α and β sites
644 of the corrin ring [89,90].
645 There is a growing recognition that selectivity in B_{12} analogs may drive interaction networks within

646 microbial communities as cobamide specificity is commonly observed in microorganisms [91]. For example,
647 pseudocobalamin, produced by cyanobacteria such as *Prochlorococcus* (the dominant bacterioplankton
648 taxa observed in this study), is poorly utilized by B_{12} auxotrophic microalgae [90,92]. Specificity in B_{12}
649 requirements among bacterioplankton taxa may explain the separation observed between
650 bacterioplankton populations associating with either SAR11_clade_II populations (**Supplementary Fig.**
651 **S10**, Group 1) or Prochlorococcus_MIT9313 (**Supplementary Fig. S10**, Group 2). Positive association
652 patterns between virioplankton RTPR phylotype populations and bacterioplankton 16S ASV populations
653 reflected the split seen between SAR11 and *Prochlorococcus* associations. RTPR virioplankton clades 5
654 and 6 associated with SAR11_clade_II and 14 other bacterioplankton populations (**Fig. 6**), three of which
655 (Magnetospiraceae, Rhodobacteraceae, Planktomarina) may be capable of B_{12} production [90]. Eleven of
656 these bacterioplankton populations are associated with SAR11_clade_II (**Supplementary Fig. S10**,
657 Group 1). Similarly, RTPR virioplankton clades 1 and 3 associated with Prochlorococcus_MIT9313, and
658 six other bacterioplankton populations that also demonstrated associations with *Prochlorococcus*
659 (**Supplementary Fig. S9**, Group 2). Three of these seven (AEGEAN_169_marine_group_1,
660 AEGEAN_169_marine_group_2, and Prochlorococcus_MIT9313) occur in an order, family, or genera
661 containing species capable of B_{12} production (**Fig. 6**).
662 Time- and depth-resolved observations of RTPR virioplankton clades 1, 3, and 5 (**Supplementary**
663 **Figs. S11** and **S12**) indicate that changes in the abundance of these virioplankton populations followed
664 those of their associated bacterioplankton populations throughout the euphotic zone. It seems unlikely
665 that the parallels in association patterns observed between bacterioplankton populations
666 (**Supplementary Fig. S10**) and those observed between RTPR virioplankton and bacterioplankton
667 populations (**Fig. 6**) are happenstance. It is likely that these positive virio- to bacterioplankton association
668 patterns reflect the B_{12}-driven association patterns among their bacterioplankton hosts [91]. It is likely that a
669 few heterotrophic bacterial taxa and *Prochlorococcus* were the principal B_{12}-producing populations in the
670 euphotic surface waters of the Azores. Although Thaumarchaeota are also important oceanic B_{12}
671 producers, metagenomic biogeographic surveys indicate that these archaeal populations predominate at
672 high latitudes or in mesopelagic waters [88].
673 It has been hypothesized that virioplankton influence oceanic primary and secondary productivity by
674 controlling the flux of limiting trace nutrients, such as iron [93], through cell lysis. The common presence
675 and dynamic nature of RTPR virioplankton populations now places viral–host interactions as another
676 important component within the vitamin traffic pathways of this trace nutrient in the oligotrophic ocean [94].
677 While technically B_{12} is not a nutrient as it is not incorporated into cellular biomass, it displays oceanic
678 distribution patterns reminiscent of nutrients such as nitrogen. Within surface waters of the oligotrophic
679 oceanic gyres, B_{12} occurs in vanishingly low concentrations of less than 1.75 picomolar [90,95]. However,
680 deeper waters typically demonstrate higher B_{12} concentrations [96]. Depth-resolved dissolved B_{12}
681 concentrations occurring at our study site were likely highest below 25 m, similar to those recorded for in
682 April and August at the western-most station of a recent study of eastern north Atlantic central waters off

683 the Iberian peninsula [95]. While RTPR virioplankton populations did show differences in their depth and
684 time distribution (**Fig. 8, Supplementary Figs. S11** and **S12**), the abundance of these populations as a
685 fraction of overall virioplankton abundance could not be discerned using an amplicon approach.
686 Nevertheless, as RTPR viral populations depend on B_{12} for a steady supply of deoxyribonucleotides, it is
687 possible that these virioplankton regulate or enhance B_{12} uptake or production during infection.
688 Subsequently, lysis by RTPR-utilizing virioplankton could increase the flux of B_{12} pools disproportionately
689 over virioplankton populations that do not utilize B_{12}-dependent proteins. New quantitative approaches
690 for observations of viral–host interactions based on detecting single viral genes using polonies [97–99] could
691 provide data on the *in situ* frequency and production rate of RTPR virioplankton populations, as has been
692 recently demonstrated for pelagic cyanophage populations [100].
693 Unlike Class I and dimeric Class II RNRs that use ribonucleotide diphosphate (NDP) substrates,
694 RTPR uses ribonucleotide triphosphate (NTP) substrates. This difference in substrates may correlate
695 with significant differences in viral population ecology. Virioplankton populations utilizing a Class I or
696 dimeric Class II RNR would be dependent on NDP production from RNA digestion [101] as well as an NDP
697 kinase for producing dNTPs [102]. In contrast, RTPR virioplankton populations have direct and immediate
698 access to the cellular pool of NTPs (which is substantially larger than the NDP pool [103]) for DNA
699 synthesis. These biochemical features of RTPR would favor a rapid lytic cycle. It is possible that RTPR-
700 carrying virioplankton are highly lytic viruses specifically infecting actively growing bacterioplankton
701 populations containing large intercellular pools of rNTP substrates.
702 However, there are possible costs to an RTPR strategy in terms of DNA synthesis regulation.
703 Improved control of dNTP substrate levels for DNA synthesis and allosteric regulation of RNR enzyme
704 activity have been proposed as possible advantages of NDP versus NTP reduction [82] . Thus, RTPR-
705 carrying viruses may pay a cost for faster DNA synthesis with lower fidelity as imbalances or excesses in
706 dNTP levels from runaway RTPR activity can increase mutagenesis rates [102,104]. The fact that RTPRs are
707 common within the virioplankton [34,77,78] and are proportionally more represented within viruses than within
708 cellular genomes (**Supplementary Fig. S15**) indicates that the possible fitness costs of RTPR do not
709 outweigh the benefits of utilizing this gene for dNTP synthesis. Nevertheless, it is possible that the
710 frequency of RTPR-carrying viruses may depend on environmental context as the majority of known
711 viruses carry a Class I or Class II dimeric RNR (**Supplementary Fig. S15**).

712 **RTPR virioplankton phylotypes demonstrated dynamic responses to ecosystem change**

713 At the community and population scale, RTPR-carrying virioplankton demonstrated dynamic behavior
714 throughout the four month study, responding to seasonal and oceanographic changes influencing the
715 activity of co-occurring phyto- and bacterioplankton communities. In this sense, RTPR virioplankton
716 populations demonstrated community dynamic behavior reminiscent of that seen from studies utilizing
717 amplicons of structural genes such as T4 major capsid protein [28–30], portal protein [24,105], and terminase [25–
718 27]. A unique aspect of this study was that primer sets targeting viral RTPR genes were used for
719 amplifying RTPR from both the virioplankton (0.02–0.22 μm) and bacterioplankton (>0.22 μm) fractions.

720 While viral genes have been amplified and sequenced from bacterioplankton fractions in past studies [29],
721 to our knowledge none have examined coordinate occurrence of OTUs across extracellular (i.e.,
722 virioplankton fraction) and intracellular (i.e., bacterioplankton fraction) virioplankton populations. The
723 majority of RTPR OTUs observed in both fractions showed a greater abundance in the virioplankton
724 fraction (**Fig. 7**) indicating host populations undergoing active lysis. These OTUs had the highest clr
725 abundance of all RTPR virioplankton populations observed in the study. This observation makes sense
726 as each viral particle released contains an RTPR gene copy and dozens of viral particles are released for
727 each lysed cell. RTPR OTUs observed solely within either the virio- or bacterioplankton fraction were less
728 frequent (8% virioplankton only; 16% bacterioplankton only) and typically showed low clr abundance.
729 Virioplankton-only RTPR OTUs likely represented viral populations either no longer actively infecting
730 hosts or somehow displaced from their host populations. Bacterioplankton-only RTPR OTUs may have
731 represented lysogenic viruses existing as prophages, free viral genomic DNA within the >0.22 μm size
732 fraction, some lytic viral populations yet to lyse infected cells, or inadvertent amplification of a bacterial
733 RTPR gene. Distinguishing between these four possibilities would be difficult; nevertheless, it is clear
734 that the majority of observed RTPR OTUs (84%) occurred within the virioplankton.
735 Individual RTPR virioplankton populations were actively infecting, replicating, and subsequently
736 disappearing from the pelagic euphotic zone. Using an amplicon approach provided the advantage of
737 deeply sampling RTPR virioplankton; it was less clear how these data should be partitioned into
738 ecologically meaningful populations. Community and population ecology analyses conducted using either
739 a heuristic approach (binning amplicons into 98% amino acid identity OTUs) or an evolutionary approach
740 (defining major phylogenetic clades of RTPR amplicons) demonstrated the superiority of the evolutionary
741 approach. Community beta diversity analyses based on clr-transformed OTU or clade abundance
742 showed that the evolutionary approach (clades, **Fig. 5B** and **Supplementary Fig. S5B**) captured a
743 greater proportion of variability in virioplankton communities than the heuristic approach (98% OTUs
744 **Supplementary Fig. S6C and D**). The evolutionary approach lumped RTPR virioplankton populations
745 into 17 clades as opposed to thousands of 98% OTUs, likely providing a greater signal to noise ratio for
746 observing virioplankton ecological behavior. Indeed, virioplankton evolutionary relationships gleaned
747 from RTPR phylogeny demonstrated a significant connection with the ecological behavior of virioplankton
748 populations on the August 5th sampling date (**Fig. 8**). The RTPR genes carried within each clade
749 population likely occurred within a diversity of genomic backgrounds. Thus, it was surprising that this
750 single gene showed a statistically significant positive relationship between evolutionary distance (RTPR
751 phylogenetic tree) and patterns of clr abundance within the bacterioplankton fraction (Aitchison distance)
752 **(Supplementary Fig. S14)**. The significant correlation of cophenetic distance (RTPR phylogeny) and
753 ecological behavior (Aitchison distance) of RTPR OTUs identified within the bacterioplankton and not
754 within free viruses is interesting as these populations represented those RTPR viral populations
755 undergoing active replication.
756 Examinations of connections between gene-based viral phylogeny and viral ecology have been

757 limited, in large part due to the widely acknowledged mosaicism within viral genomes [106]. Nevertheless,
758 when considering the influence of enzyme biochemistry or the presence/absence of particular auxiliary
759 metabolic genes [107], it is clear that single genes have a dramatic influence on a virus' ecology. Single-
760 gene amplicon approaches provide not only an evolutionary framework for understanding viral ecology,
761 but also have practical advantages over metagenomic approaches in terms of cost, depth and breadth of
762 sampling, and computational complexity [108]. Prior work examining genetic connections to phage host
763 range phenotypes found that while whole genome networks provided better predictive capability, single
764 gene groups could also show strong connections [109]. Interestingly, DNA replication and phage replication
765 were among those Rapid Annotation Subsystem gene groups [110] showing a stronger predictive ability for
766 host range. This might not be surprising when considering that phage genome and replication processes
767 may require intimate interaction with host anabolic systems. Examining possible connections between
768 single genes and whole genomes and other important phage phenotypes, especially those describing
769 infection dynamics, should ultimately improve the predictive ability of metagenome data for ecosystem
770 modeling of phage impacts on ecosystem processes.

771 **AUTHOR CONTRIBUTIONS**

772 GP designed the study, performed processing of water samples and DNA extraction following water
773 samples collection and part of the RTPR library preparation. LW conducted the majority of the RTPR
774 library preparation and all of the 16S rDNA library preparation, performed all of the bioinformatics and
775 biostatistical analyses following sequence processing, and wrote and edited the manuscript. RM, CL, AH,
776 JD, BF, AM, KB, SP, JN and EW contributed to the study design, protocol development, data analysis
777 and interpretation, and manuscript preparation. All authors read, edited, and approved the final
778 manuscript.

779 **ACKNOWLEDGEMENTS**

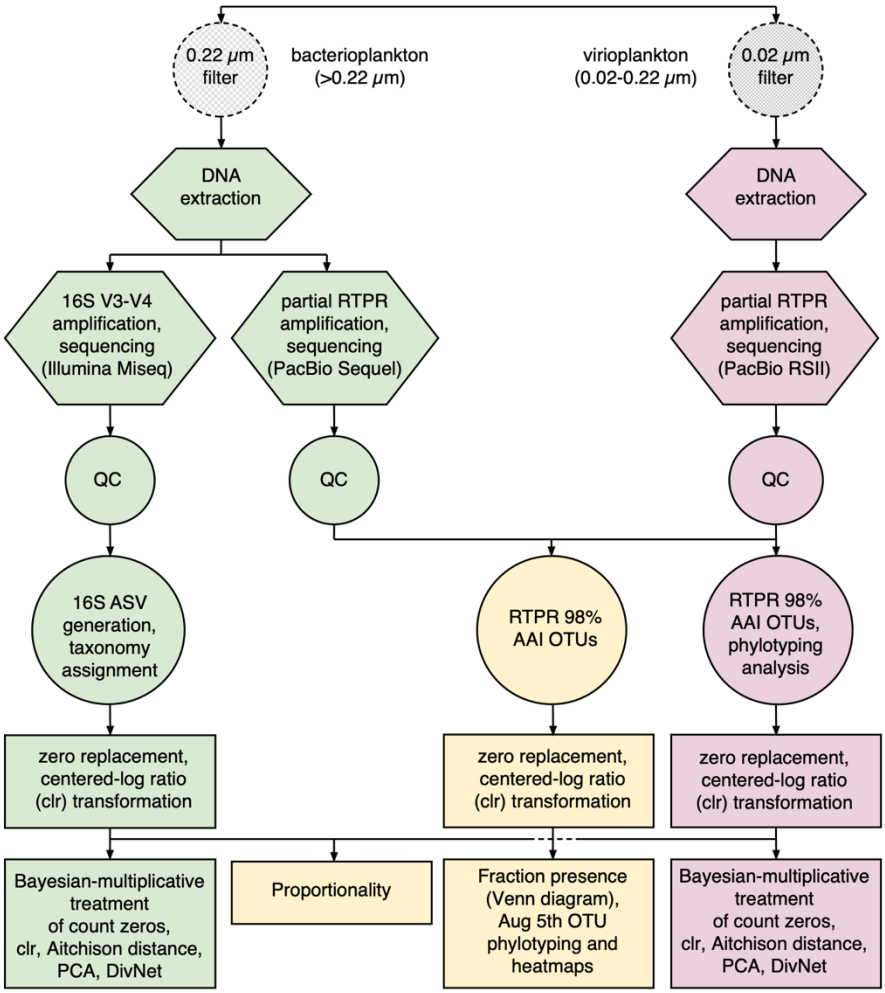
780 We would like to thank, by order of involvement, Catharina Pieper, Ricardo Fernandes, Christien Laber,
781 Prasanna Joglekar, Joana Botelho, Ana Pavon, Clara M. Loureiro, and Marilia Olio for the help collecting
782 and processing the samples. We would like to thank Dr. Daniel Nasko for input on library preparation and
783 statistical analysis. We would like to acknowledge the skipper Renato Bettencourt, and all laboratory and
784 facility management personnel from Departamento de Oceanografia e P, OKEANOS Institute of Marine
785 Research, and IMAR-Instituto do Mar from the University of the Azores, DMCS-RU and DBI-UD.

786 **FUNDING**

787 This research was supported by the Estagiar-L fellowship; EEA grant funded BIOMETOREzores
788 (PT02_Aviso2_0001) project; the mobility program FLAD-UA Crossing the Atlantic; Chinese Scholarship
789 Council (CSC), and National Science Foundation grant number 1736030. Support from the University of

790 Delaware Center for Bioinformatics and Computational Biology (CBCB) Core Facility, the University of
791 Delaware Sequencing and Genotyping Center, and use of the BIOMIX compute cluster was made
792 possible through funding from Delaware INBRE (NIGMS P20GM103446), the State of Delaware, and the
793 Delaware Biotechnology Institute.
794
795 *Conflict of interest statement.* None declared.

796 **FIGURES**



797
798 **Figure 1.** Overview of experimental methods and analysis pipeline. The 0.22 μm filter captured

bacterioplankton cells. The 0.02 µm filter captured virioplankton particles. Hexagons represent molecular genetic experimental methods. QC stands for a collection of quality control steps specific to the type of next-generation sequencing technology used in sequencing amplicon libraries. Circles represent conventional data analysis methods. Squares represent compositional data analysis (CoDA) methods.

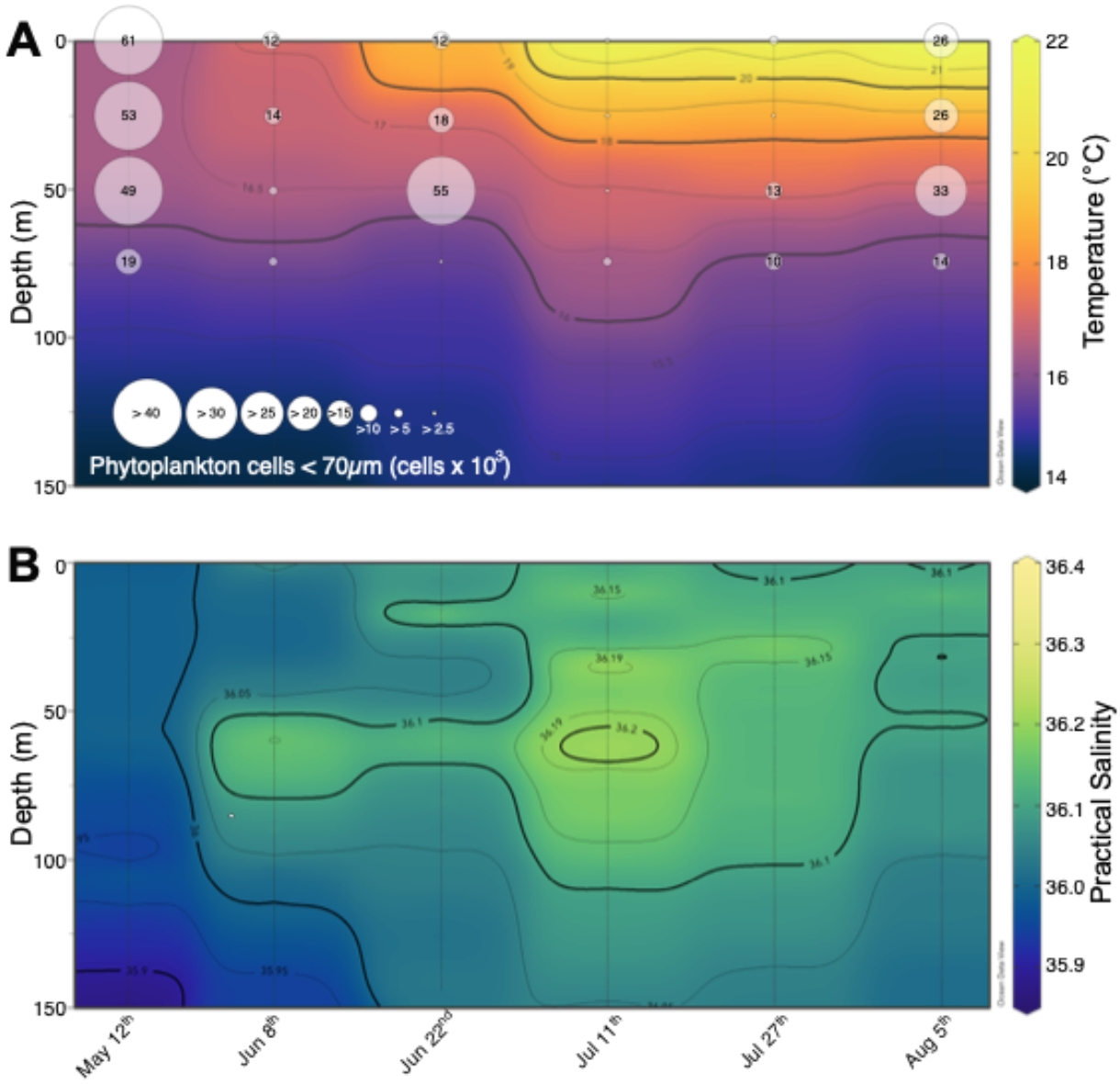


Figure 2. Vertical profile contours of **(A)** temperature and phytoplankton cell numbers, and **(B)** salinity at station South of Channel Faial-Pico islands (SCFP) from May 12th to August 5th, 2016. Data was not collected on September 8th due to equipment malfunction. Vertical gray lines indicate the sampling date and the maximum depth of data collection. Temperature and salinity data between sampling dates was

interpolated using the weighted average grid method. Circle sizes indicate phytoplankton abundances at each sampled depth (surface, 25, 50 and 75 m). Surface values were the average abundance at the 0 and 5 m depths.

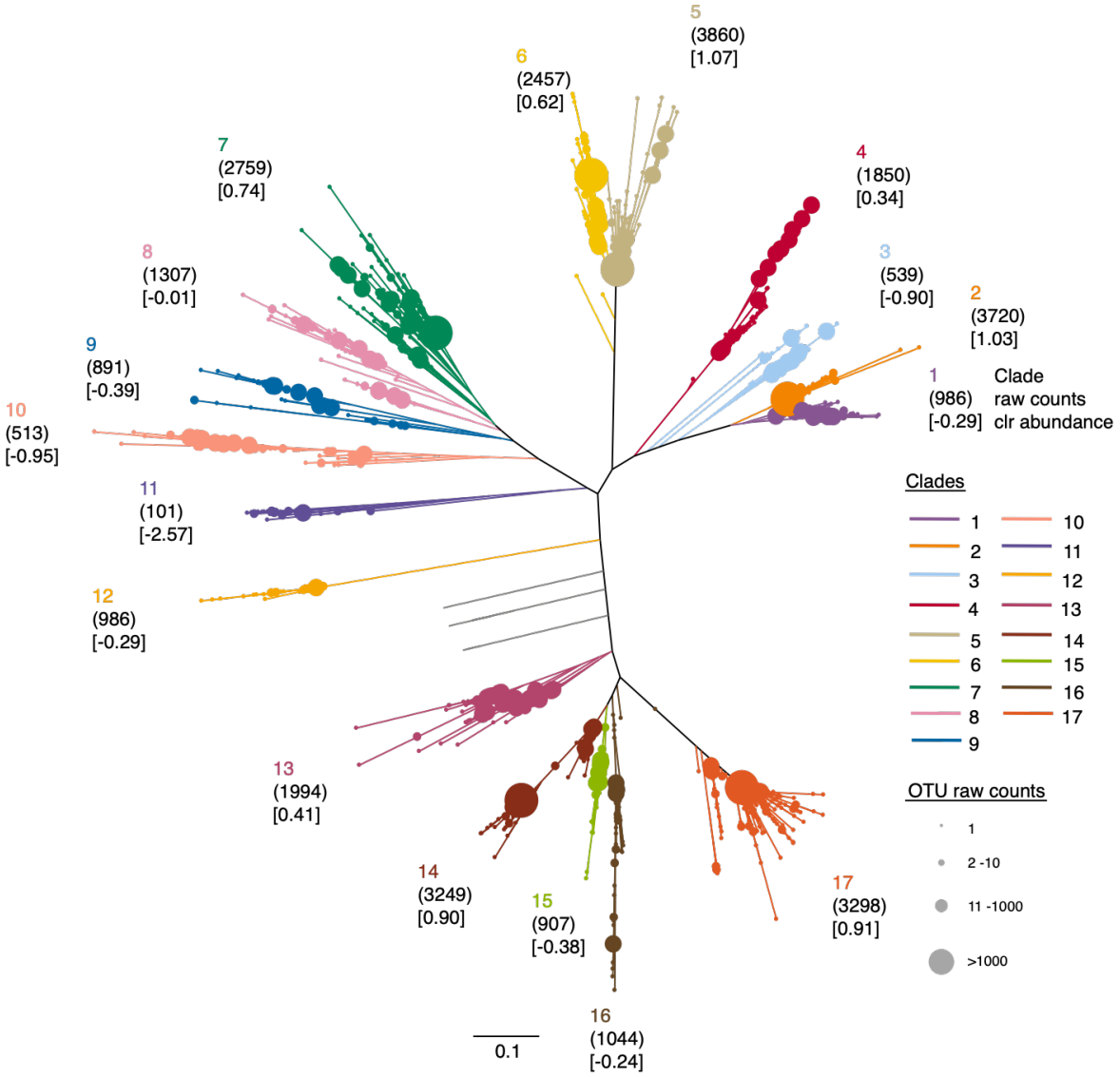


Figure 3. Approximate maximum likelihood tree of virioplankton RTPR 98% OTU representative sequences. Shared branch, node, and clade number colors indicate RTPR phylogenetic clade manual assignment. Seven OTUs (between clades 12 and 13) with low abundance and distant relationships with other phylogenetic clades are represented by gray branches and were excluded from downstream analysis. Node sizes indicate 98% OTU counts. Each phylogenetic clade's total amino acid sequence counts and clr abundance are indicated in rounded and squared parentheses, respectively. Scale bar indicates the average number of amino acid substitutions per site.

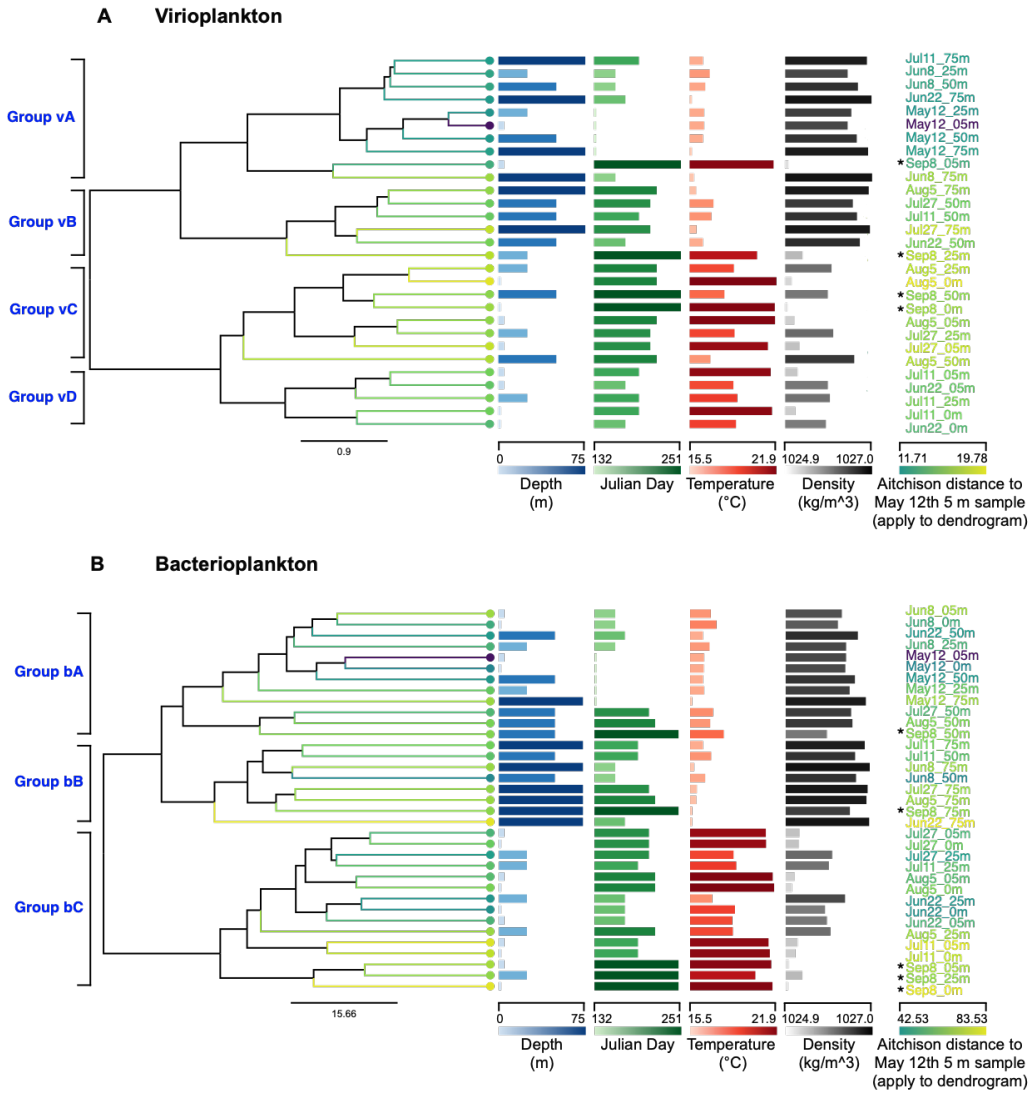


Figure 4. Phylogenetic clustering of virio- and bacterioplankton communities (dendrograms) based on **(A)** virioplankton RTPR phylogenetic clades and **(B)** bacterioplankton 16S ASVs, alongside environmental variables (columns of horizontal bars) demonstrating significant correlations (Spearman and Pearson tests, $p < .05$). Dendrograms are based on complete-linkage hierarchical clustering of each sample's Aitchison distance to the sample collected May 12th at 5 m (May12_05m) based on centered-log ratio (clr) abundance. Scale bar indicates Aitchison distance between samples. Major groups of virio- or bacterioplankton communities are labeled to the left of each dendrogram. Environmental variables are represented in columns, with horizontal bars indicating the measurement for each sample both by color gradient and by bar length compared to the numerical scale along each column's horizontal axis [Depth (blue), Day (green), Temperature (red), and Density (gray)]. Sample labels with asterisks (*) indicate samples where temperature and density were estimated using the Marine Copernicus Global Ocean Physics model.

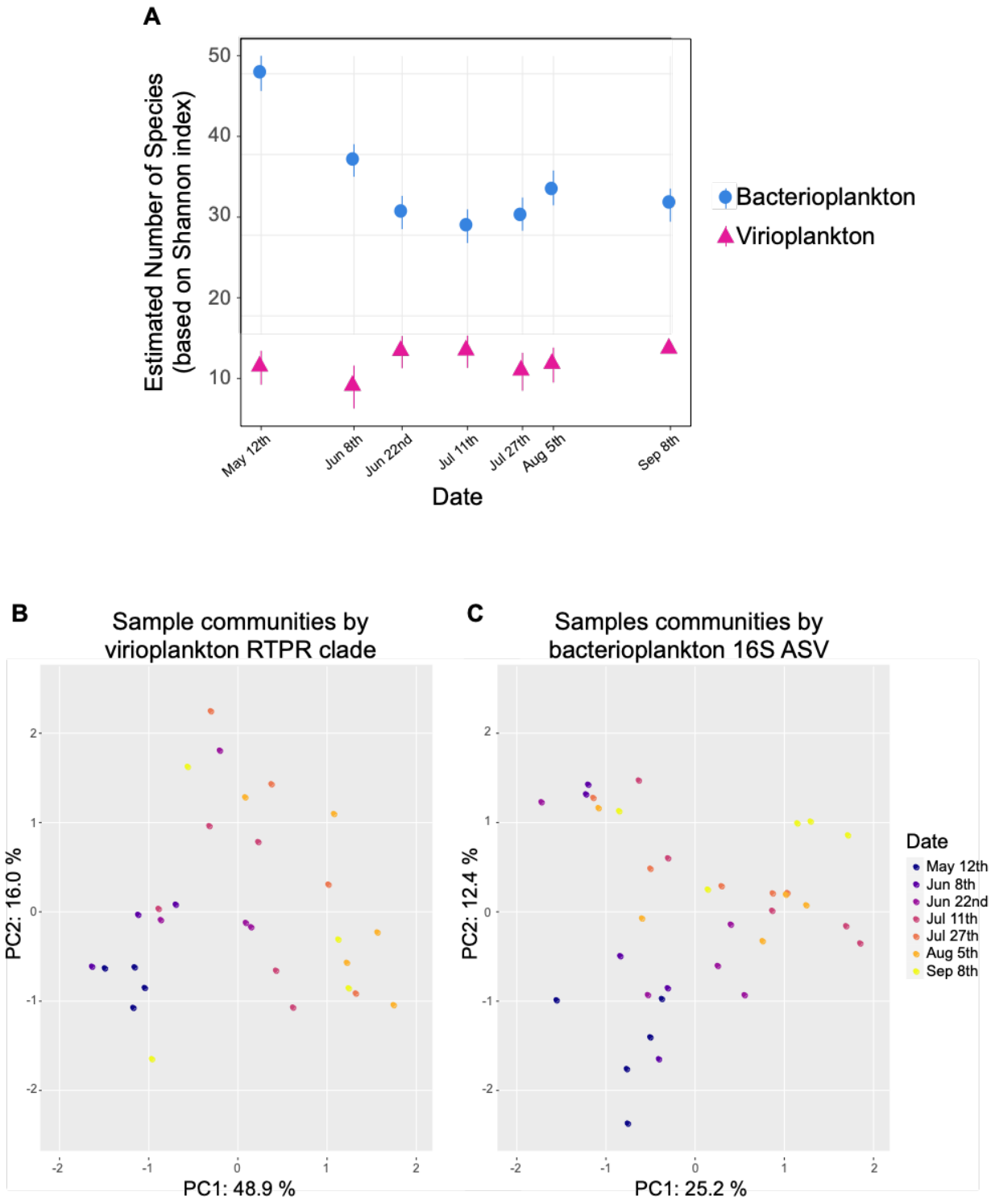
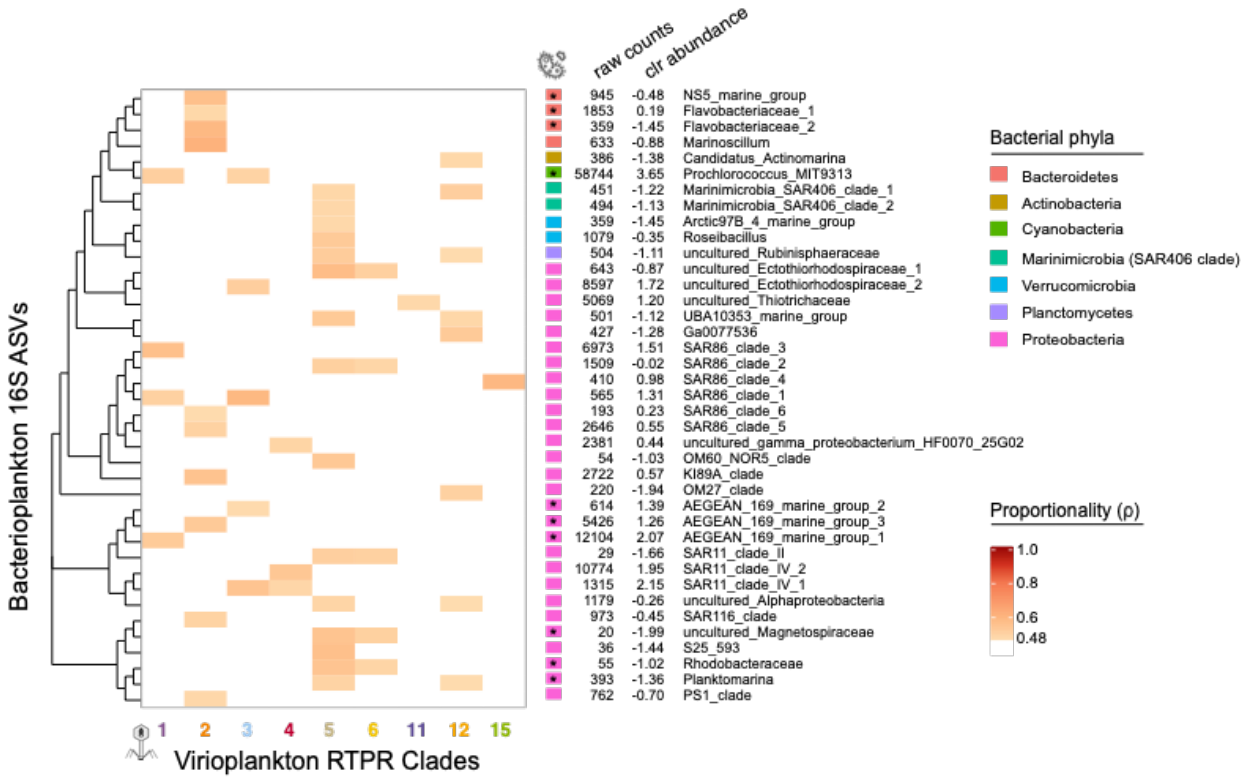


Figure 5. Sampling date influenced virioplankton (RTPR phylogenetic clade) and bacterioplankton (16S ASV) community alpha and beta diversity. **(A)** Effective number of species (ENS, based on Shannon

837 indices) of virioplankton RTPR phylogenetic clades and bacterioplankton 16S communities by sampling
838 date. Error bars represent two standard deviations of the estimates. Principal component analysis (PCA)
839 plots of beta diversity based on centered-log ratio (clr) transformed abundance of **(B)** virioplankton RPTR
840 phylogenetic clade communities or **(C)** bacterioplankton 16S ASV communities. Samples (circles) are
841 colored according to sampling date.



842

843 **Figure 6.** Matrix (ρ proportionality) of significantly associated (ρ ≥.48) virioplankton RTPR phylogenetic
844 clades (x-axis) and bacterioplankton 16S ASVs (y-axis) (**Supplementary Table S13**). Virioplankton
845 RTPR clades are colored and arranged by their topological order on the 98% OTU phylogenetic tree (**Fig.**
846 **3**). Bacterioplankton 16S ASVs are arranged according to phylogeny. Bacterioplankton 16S ASV phylum
847 (colored block), raw sequence count, clr abundance, and taxonomic assignment are shown on the right.
848 Asterisk indicates those ASVs within orders, families, or genera containing species capable of B_{12}
849 synthesis according to a pangenomic survey by Heal *et al*. 2017 [90].

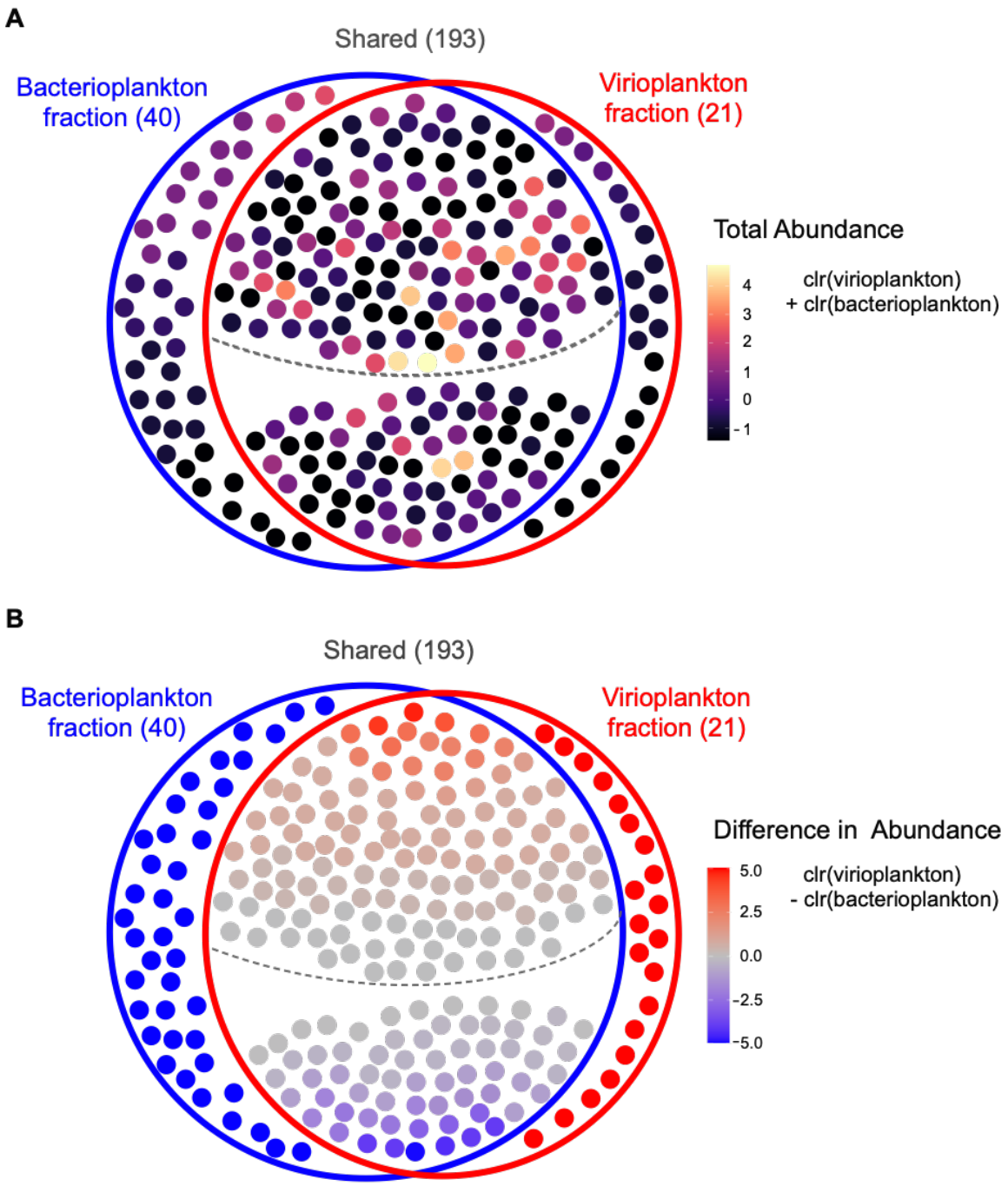


Figure 7. RTPR 98% OTUs (n=254) identified from amino acid sequences present in 27 paired samples with amplicons from the virio- and bacterioplankton fractions. RTPR 98% OTUs (filled circles) were observed in either only the bacterioplankton fraction (blue circle, n=40), only the virioplankton fraction (red circle, n=21), or both (Shared, n=193). The dashed line in the overlap region divides the shared OTUs more abundant in the virioplankton fraction (above the line, n=128) from those more abundant in the bacterioplankton fraction (below the line, n=65). **(A)** RTPR 98% OTUs are colored by their total centered

log ratio (clr) abundance in both fractions (clr[virioplankton]+clr[bacterioplankton]). **(B)** RTPR 98% OTUs are colored by the difference between clr abundance in the virioplankton and bacterioplankton fractions (clr[virioplankton] - clr[bacterioplankton]). Red colors indicate greater abundance of that OTU in the virioplankton fraction (values > 0, where 5 is an arbitrary maximum representing OTU presence in the virioplankton only). Blue colors indicate greater abundance of that OTU in the bacterioplankton fraction (values < 0, where -5 is an arbitrary minimum representing OTU presence in the bacterioplankton only).

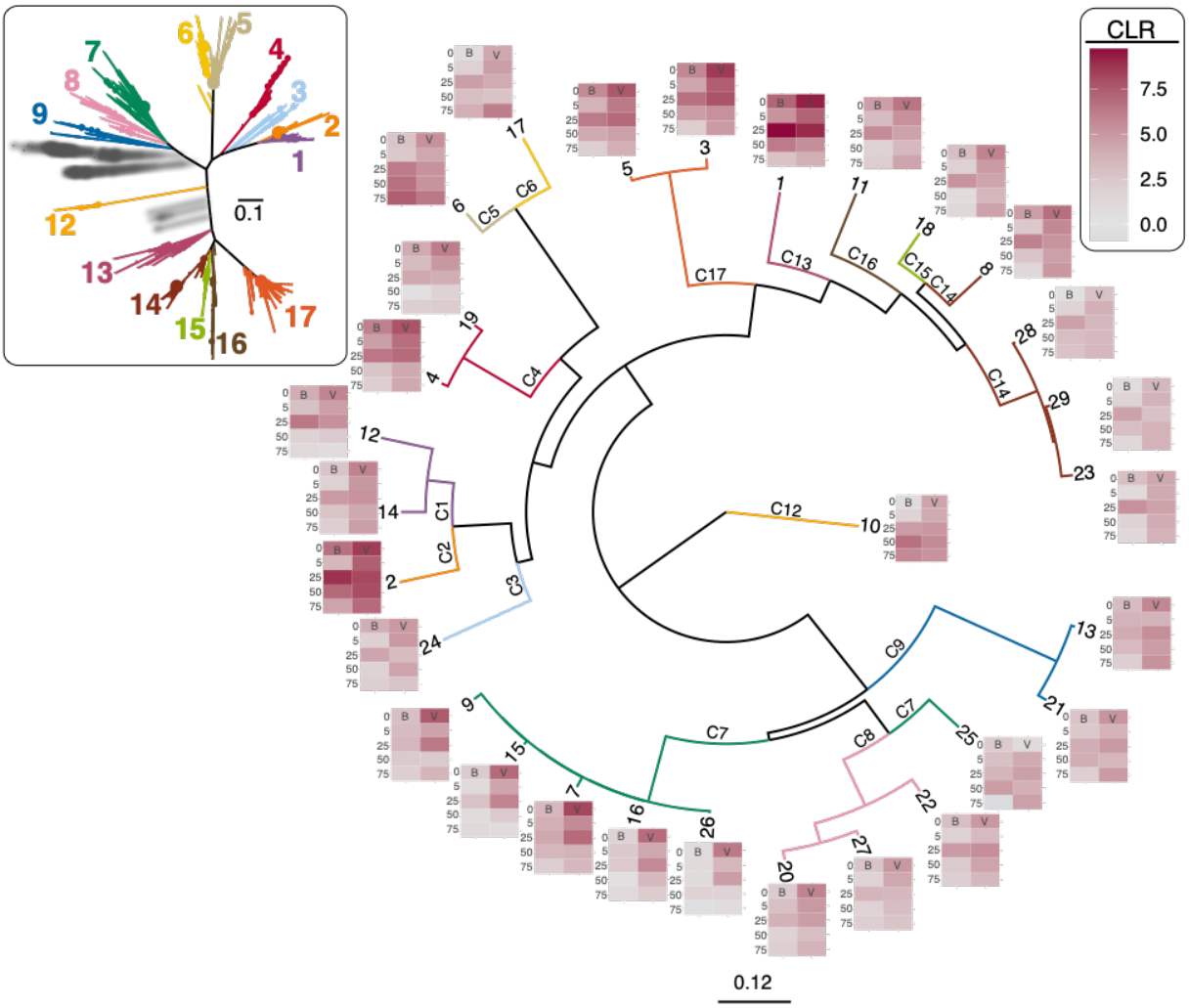


Figure 8. Phylogram of 29 RTPR 98% OTUs occurring within both the bacterioplankton (B) and virioplankton (V) fractions on the August 5th sampling date, with heat maps at nodes representing OTU centered-log ratio (clr) abundance in each sample (fraction and depth). Node labels indicate OTU number. Branches are colored and labeled according to membership of the OTUs in the RTPR clades (C1–C9; C12–C17) identified in the inset tree (**Fig. 3**). Scale bar indicates the average number of amino acid substitutions per site.

REFERENCES

1. McClain, C. R., Signorini, S. R. & Christian, J. R. Subtropical gyre variability observed by ocean-color

872 satellites. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* **51**, 281–301 (2004).
873 2. Huang, R. X. & Russell, S. Ventilation of the Subtropical North Pacific. *J. Phys. Oceanogr.* **24**, 2589–
874 2605 (1994).
875 3. Sherr, E. & Sherr, B. Understanding Roles of Microbes in Marine Pelagic Food Webs: A Brief History.
876 in *Microbial Ecology of the Oceans* (ed. Kirchman, D. L.) 27–44 (John Wiley & Sons, Inc., 2008).
877 doi:10.1002/9780470281840.ch2.
878 4. Goldman, J. C. Conceptual Role for Microaggregates in Pelagic Waters. *Bull. Mar. Sci.* **35**, 462–476
879 (1984).
880 5. Díez, B., Pedrós-Alió, C. & Massana, R. Study of Genetic Diversity of Eukaryotic Picoplankton in
881 Different Oceanic Regions by Small-Subunit rRNA Gene Cloning and Sequencing. *Appl. Environ.*
882 *Microbiol.* **67**, 2932 (2001).
883 6. Wilhelm, S. W. & Suttle, C. A. Viruses and Nutrient Cycles in the Sea: Viruses play critical roles in the
884 structure and function of aquatic food webs. *Bioscience* **49**, 781–788 (1999).
885 7. Ban, N. & Alder, J. How wild is the ocean? Assessing the intensity of anthropogenic marine activities
886 in British Columbia, Canada. *Aquat. Conserv.* **18**, 55–85 (2008).
887 8. Li, D. & Daler, D. Ocean pollution from land-based sources: East China Sea, China. *Ambio* **33**, 107–
888 113 (2004).
889 9. Sakowski, E. G. *et al.* Ribonucleotide reductases reveal novel viral diversity and predict biological and
890 ecological features of unknown marine viruses. *Proceedings of the National Academy of Sciences*
891 **111**, 15786–15791 (2014).
892 10. Harrison, A. O., Moore, R. M., Polson, S. W. & Wommack, K. E. Reannotation of the Ribonucleotide
893 Reductase in a Cyanophage Reveals Life History Strategies Within the Virioplankton. *Front.*
894 *Microbiol.* **10**, 134 (2019).
895 11. Lara, E. *et al.* Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Sci*
896 *Adv* **3**, e1602565 (2017).
897 12. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548
898 (1999).
899 13. Suttle, C. A. Viruses in the sea. *Nature* **437**, 356–361 (2005).
900 14. Wommack, K. E. & Colwell, R. R. Virioplankton: viruses in aquatic ecosystems. *Microbiol. Mol. Biol.*
901 *Rev.* **64**, 69–114 (2000).
902 15. Jiang, S. C. & Paul, J. H. Gene transfer by transduction in the marine environment. *Appl. Environ.*
903 *Microbiol.* **64**, 2780–2787 (1998).
904 16. A Schwartz, D. & Lindell, D. Genetic hurdles limit the arms race between Prochlorococcus and the
905 T7-like podoviruses infecting them. *ISME J.* **11**, (2017).
906 17. Ochman, H., Lawrence, J. G. & Groisman, E. A. Lateral gene transfer and the nature of bacterial
907 innovation. *Nature* **405**, 299–304 (2000).
908 18. Rosenwasser, S., Ziv, C., van Creveld, S. G. & Vardi, A. Virocell Metabolism: Metabolic Innovations

909 During Host–Virus Interactions in the Ocean. *Trends Microbiol.* **24**, 821–832 (2016).
910 19. Crummett, L. T., Puxty, R. J., Weihe, C., Marston, M. F. & Martiny, J. B. H. The genomic content and
911 context of auxiliary metabolic genes in marine cyanomyoviruses. *Virology* **499**, 219–229 (2016).
912 20. Thompson, L. R. *et al.* Phage auxiliary metabolic genes and the redirection of cyanobacterial host
913 carbon metabolism. *Proceedings of the National Academy of Sciences* **108**, E757–E764 (2011).
914 21. Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. 16S ribosomal DNA amplification for
915 phylogenetic study. *J. Bacteriol.* **173**, 697–703 (1991).
916 22. Schmidt, H. F., Sakowski, E. G., Williamson, S. J., Polson, S. W. & Wommack, K. E. Shotgun
917 metagenomics indicates novel family A DNA polymerases predominate within marine virioplankton.
918 *ISME J.* **8**, 103–114 (2014).
919 23. Brussaard, C. P. D., Short, S. M., Frederickson, C. M. & Suttle, C. A. Isolation and phylogenetic
920 analysis of novel viruses infecting the phytoplankton Phaeocystis globosa (Prymnesiophyceae). *Appl.*
921 *Environ. Microbiol.* **70**, 3700–3705 (2004).
922 24. Zhong, Y., Chen, F., Wilhelm, S. W., Poorvin, L. & Hodson, R. E. Phylogenetic diversity of marine
923 cyanophage isolates and natural virus communities as revealed by sequences of viral capsid
924 assembly protein gene g20. *Appl. Environ. Microbiol.* **68**, 1576–1584 (2002).
925 25. Short, C. M. & Suttle, C. A. Nearly identical bacteriophage structural gene sequences are widely
926 distributed in both marine and freshwater environments. *Appl. Environ. Microbiol.* **71**, 480–486
927 (2005).
928 26. Marston, M. F. & Sallee, J. L. Genetic diversity and temporal variation in the cyanophage community
929 infecting marine Synechococcus species in Rhode Island's coastal waters. *Appl. Environ. Microbiol.*
930 **69**, 4639–4647 (2003).
931 27. Goldin, S., Hulata, Y., Baran, N. & Lindell, D. Quantification of T4-Like and T7-Like Cyanophages
932 Using the Polony Method Show They Are Significant Members of the Virioplankton in the North
933 Pacific Subtropical Gyre. *Front. Microbiol.* **11**, 1210 (2020).
934 28. Comeau, A. M. & Krisch, H. M. The Capsid of the T4 Phage Superfamily: The Evolution, Diversity,
935 and Structure of Some of the Most Prevalent Proteins in the Biosphere. *Mol. Biol. Evol.* **25**, 1321–
936 1332 (2008).
937 29. Needham, D. M., Sachdeva, R. & Fuhrman, J. A. Ecological dynamics and co-occurrence among
938 marine phytoplankton, bacteria and myoviruses shows microdiversity matters. *ISME J.* **11**, 1614–
939 1629 (2017).
940 30. Liu, L., Cai, L. & Zhang, R. Co-existence of freshwater and marine T4-like myoviruses in a typical
941 subtropical estuary. *FEMS Microbiol. Ecol.* **93**, (2017).
942 31. Lundin, D., Gribaldo, S., Torrents, E., Sjöberg, B.-M. & Poole, A. M. Ribonucleotide reduction -
943 horizontal transfer of a required function spans all three domains. *BMC Evol. Biol.* **10**, 383 (2010).
944 32. Nordlund, P. & Reichard, P. Ribonucleotide Reductases. *Annu. Rev. Biochem.* **75**, 681–706 (2006).
945 33. Herrick, J. & Sclavi, B. Ribonucleotide reductase and the regulation of DNA replication: an old story

and an ancient heritage. *Mol. Microbiol.* **63**, 22–34 (2007).

34. Dwivedi, B., Xue, B., Lundin, D., Edwards, R. A. & Breitbart, M. A bioinformatic analysis of ribonucleotide reductase genes in phage genomes and metagenomes. *BMC Evol. Biol.* **13**, 33 (2013).
35. Sullivan, M. B., Coleman, M. L., Weigele, P., Rohwer, F. & Chisholm, S. W. Three Prochlorococcus cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* **3**, e144–e144 (2005).
36. McMurdie, P. J. & Holmes, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **10**, 1–12 (2014).
37. Weiss, S. *et al.* Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 27 (2017).
38. Gloor, G. B., Macklaim, J. M., Pawlowsky-Glahn, V. & Egozcue, J. J. Microbiome Datasets Are Compositional: And This Is Not Optional. *Front. Microbiol.* **8**, 2224 (2017).
39. McLaren, M. R. *et al.* Consistent and correctable bias in metagenomic sequencing experiments. *Elife* **8**, e46923 (2019).
40. Bian, G. *et al.* The Gut Microbiota of Healthy Aged Chinese Is Similar to That of the Healthy Young. *mSphere* **2**, (2017).
41. Fernandes, A. D. *et al.* Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis. *Microbiome* **2**, 15 (2014).
42. Publication Reports - EcosystemOverview_Azores_2019.pdf. http://www.ices.dk/sites/pub/Publication%20Reports/Forms/DispForm.aspx?ID=36442 doi:10.17895/ices.advice.5753.
43. Bashmachnikov, I., Belonenko, T. V. & Koldunov, A. V. Intra-annual and interannual non-stationary cycles of chlorophyll concentration in the Northeast Atlantic. *Remote Sens. Environ.* **137**, 55–68 (2013).
44. Arístegui, J. *et al.* Plankton metabolic balance at two North Atlantic seamounts. *Deep Sea Res. Part 2 Top. Stud. Oceanogr.* **56**, 2646–2655 (2009).
45. Mendonça, A. *et al.* Is there a seamount effect on microbial community structure and biomass? The case study of Seine and Sedlo seamounts (northeast Atlantic). *PLoS One* **7**, e29526 (2012).
46. True Colors of Oceanography: Guidelines for Effective and Accurate Colormap Selection | Oceanography. https://tos.org/oceanography/article/true-colors-of-oceanography-guidelines-for-effective-and-accurate-colormap doi:10.5670/oceanog.2016.66.
47. Schlitzer, R. Interactive analysis and visualization of geoscience data with Ocean Data View. *Comput. Geosci.* **28**, 1211–1218 (2002).
48. Sotillo, M. G. *et al.* The MyOcean IBI Ocean Forecast and Reanalysis Systems: operational products and roadmap to the future Copernicus Service. *Journal of Operational Oceanography* **8**, 63–79 (2015).

983 49. Baith, K., Lindsay, R., Fu, G. & McClain, C. R. Data analysis system developed for ocean color
984 satellite sensors. *Eos Trans. AGU* **82**, 202–202 (2001).
985 50. Marie, D., Partensky, F., Vaulot, D. & Brussaard, C. Enumeration of phytoplankton, bacteria, and
986 viruses in marine samples. *Curr. Protoc. Cytom.* **Chapter 11**, Unit 11.11 (2001).
987 51. Iker, B. C., Kitajima, M. & Gerba, C. P. Extraction and Purification of Viral Nucleic Acids from
988 Environmental Samples. in *Sample Preparation Techniques for Soil, Plant, and Animal Samples* (ed.
989 Micic, M.) 315–324 (Springer New York, 2016). doi:10.1007/978-1-4939-3185-9_22.
990 52. Mueller, J. A., Culley, A. I. & Steward, G. F. Variables influencing extraction of nucleic acids from
991 microbial plankton (viruses, bacteria, and protists) collected on nanoporous aluminum oxide filters.
992 *Appl. Environ. Microbiol.* **80**, 3930–3942 (2014).
993 53. Fuhrman, J. A., Comeau, D. E., Hagström, A. & Chan, A. M. Extraction from natural planktonic
994 microorganisms of DNA suitable for molecular biological studies. *Appl. Environ. Microbiol.* **54**, 1426–
995 1429 (1988).
996 54. Fadrosh, D. W. *et al.* An improved dual-indexing approach for multiplexed 16S rRNA gene
997 sequencing on the Illumina MiSeq platform. *Microbiome* **2**, 6 (2014).
998 55. Moore, R. M. *et al.* PASV: Automatic protein partitioning and validation using conserved residues.
999 *bioRxiv* 2021.01.20.427478 (2021) doi:10.1101/2021.01.20.427478.
1000 56. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–
1001 2461 (2010).
1002 57. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree 2 -- Approximately Maximum-Likelihood Trees for
1003 Large Alignments. *PLoS One* **5**, 1–10 (2010).
1004 58. Moore, R. M., Harrison, A. O., McAllister, S. M., Polson, S. W. & Wommack, K. E. Iroki: automatic
1005 customization and visualization of phylogenetic trees. *PeerJ* **8**, e8584 (2020).
1006 59. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using
1007 QIIME 2. *Nat. Biotechnol.* **37**, 852–857 (2019).
1008 60. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat.*
1009 *Methods* **13**, 581 EP– (2016).
1010 61. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and
1011 web-based tools. *Nucleic Acids Res.* **41**, D590–D596 (2012).
1012 62. Palarea-Albaladejo, J. & Martin-Fernandez, J. A. zCompositions -- R package for multivariate
1013 imputation of left-censored data under a compositional approach. *Chemometrics Intellig. Lab. Syst.*
1014 **143**, 85–96 (2015).
1015 63. Team, R. C. *R: A Language and Environment for Statistical Computing*. (2013).
1016 64. Gloor, G. B. & Reid, G. Compositional analysis: a valid approach to analyze microbiome high-
1017 throughput sequencing data. *Can. J. Microbiol.* **62**, 692–703 (2016).
1018 65. Mantel, N. The Detection of Disease Clustering and a Generalized Regression Approach. *Cancer*
1019 *Res.* **27**, 209–220 (1967).

1020 66. Willis, A. D. & Martin, B. D. DivNet: Estimating diversity in networked communities. *bioRxiv* (2018)
1021 doi:10.1101/305045.
1022 67. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer, 2016).
1023 68. Willis, A. & Bunge, J. Estimating diversity via frequency ratios. *Biometrics* **71**, 1042–1049 (2015).
1024 69. Kang, S., Rodrigues, J. L. M., Ng, J. P. & Gentry, T. J. Hill number as a bacterial diversity measure
1025 framework with high-throughput sequence data. *Sci. Rep.* **6**, 38263 (2016).
1026 70. van den Boogaart, K. G., Tolosana-Delgado, R. & Bren, M. compositions: Compositional Data
1027 Analysis. *Austral Ecol.* **26**, 32–46 (2018).
1028 71. Anderson, M. J. A new method for non-parametric multivariate analysis of variance. *Austral Ecol.* **26**,
1029 32–46 (2001).
1030 72. Quinn, T. P., Richardson, M. F., Lovell, D. & Crowley, T. M. propr: An R-package for Identifying
1031 Proportionally Abundant Features Using Compositional Data Analysis. *Sci. Rep.* **7**, 16252 (2017).
1032 73. Erb, I. & Notredame, C. How should we measure proportionality on relative gene expression data?
1033 *Theory Biosci.* **135**, 21–36 (2016).
1034 74. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in
1035 multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
1036 75. Alberti, A. *et al.* Viral to metazoan marine plankton nucleotide sequences from the Tara Oceans
1037 expedition. *Sci Data* **4**, 170093 (2017).
1038 76. Lundin, D., Torrents, E., Poole, A. M. & Sjöberg, B.-M. RNRdb, a curated database of the universal
1039 enzyme family ribonucleotide reductase, reveals a high level of misannotation in sequences
1040 deposited to Genbank. *BMC Genomics* vol. 10 Preprint at https://doi.org/10.1186/1471-2164-10-589
1041 (2009).
1042 77. Sakowski, E. G. *et al.* Ribonucleotide reductases reveal novel viral diversity and predict biological and
1043 ecological features of unknown marine viruses. *Proceedings of the National Academy of Sciences*
1044 vol. 111 15786–15791 Preprint at https://doi.org/10.1073/pnas.1401322111 (2014).
1045 78. Wommack, K. E., Eri Wommack, K., Nasko, D. J., Chopyk, J. & Sakowski, E. G. Counts and
1046 sequences, observations that continue to change our understanding of viruses in nature. *Journal of*
1047 *Microbiology* vol. 53 181–192 Preprint at https://doi.org/10.1007/s12275-015-5068-6 (2015).
1048 79. Bidle, K. D. Elucidating marine virus ecology through a unified heartbeat. *Proceedings of the National*
1049 *Academy of Sciences of the United States of America* vol. 111 15606–15607 (2014).
1050 80. Stubbe, J. Ribonucleotide reductases: the link between an RNA and a DNA world? *Curr. Opin. Struct.*
1051 *Biol.* **10**, 731–736 (2000).
1052 81. Forterre, P. The two ages of the RNA world, and the transition to the DNA world: a story of viruses
1053 and cells. *Biochimie* **87**, 793–803 (2005).
1054 82. Lundin, D., Berggren, G., Logan, D. T. & Sjöberg, B.-M. The origin and evolution of ribonucleotide
1055 reduction. *Life* **5**, 604–636 (2015).
1056 83. Martinez-Hernandez, F. *et al.* Single-virus genomics reveals hidden cosmopolitan and abundant

1057 viruses. *Nat. Commun.* **8**, 15892 (2017).
1058 84. Sakowski, E. G. *et al.* Interaction dynamics and virus-host range for estuarine actinophages captured
1059 by epicPCR. *Nat Microbiol* **6**, 630–642 (2021).
1060 85. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489 (2021).
1061 86. Torrents, E. Ribonucleotide reductases: essential enzymes for bacterial life. *Frontiers in Cellular and*
1062 *Infection Microbiology* vol. 4 Preprint at https://doi.org/10.3389/fcimb.2014.00052 (2014).
1063 87. Sañudo-Wilhelmy, S. A., Gómez-Consarnau, L., Suffridge, C. & Webb, E. A. The Role of B Vitamins
1064 in Marine Biogeochemistry. *Annual Review of Marine Science* vol. 6 339–367 Preprint at
1065 https://doi.org/10.1146/annurev-marine-120710-100912 (2014).
1066 88. Doxey, A. C., Kurtz, D. A., Lynch, M. D. J., Sauder, L. A. & Neufeld, J. D. Aquatic metagenomes
1067 implicate Thaumarchaeota in global cobalamin production. *ISME J.* **9**, 461–471 (2015).
1068 89. Grossman, A. Nutrient Acquisition: The Generation of Bioactive Vitamin B12 by Microalgae. *Current*
1069 *biology: CB* vol. 26 R319–21 (2016).
1070 90. Heal, K. R. *et al.* Two distinct pools of B_{12} analogs reveal community interdependencies in the ocean.
1071 *Proceedings of the National Academy of Sciences* vol. 114 364–369 Preprint at
1072 https://doi.org/10.1073/pnas.1608462114 (2017).
1073 91. Sokolovskaya, O. M., Shelton, A. N. & Taga, M. E. Sharing vitamins: Cobamides unveil microbial
1074 interactions. *Science* **369**, eaba0165 (2020).
1075 92. Helliwell, K. E. *et al.* Cyanobacteria and Eukaryotic Algae Use Different Chemical Variants of Vitamin
1076 B12. *Current Biology* vol. 26 999–1008 Preprint at https://doi.org/10.1016/j.cub.2016.02.041 (2016).
1077 93. Poorvin, L., Rinta-Kanto, J. M., Hutchins, D. A. & Wilhelm, S. W. Viral release of iron and its
1078 bioavailability to marine plankton. *Limnol. Oceanogr.* **49**, 1734–1741 (2004).
1079 94. Giovannoni, S. J. Vitamins in the sea. *Proceedings of the National Academy of Sciences of the*
1080 *United States of America* vol. 109 13888–13889 (2012).
1081 95. Joglar, V. *et al.* Cobalamin and microbial plankton dynamics along a coastal to offshore transect in
1082 the Eastern North Atlantic Ocean. *Environmental Microbiology* vol. 23 1559–1583 Preprint at
1083 https://doi.org/10.1111/1462-2920.15367 (2021).
1084 96. Sañudo-Wilhelmy, S. A. *et al.* Multiple B-vitamin depletion in large areas of the coastal ocean. *Proc.*
1085 *Natl. Acad. Sci. U. S. A.* **109**, 14041–14045 (2012).
1086 97. Mitra, R. D., Shendure, J., Olejnik, J., Edyta-Krzymanska-Olejnik & Church, G. M. Fluorescent in situ
1087 sequencing on polymerase colonies. *Anal. Biochem.* **320**, 55–65 (2003).
1088 98. Mruwat, N. *et al.* A single-cell polony method reveals low levels of infected Prochlorococcus in
1089 oligotrophic waters despite high cyanophage abundances. *ISME J.* **15**, 41–54 (2021).
1090 99. Baran, N., Goldin, S., Maidanik, I. & Lindell, D. Quantification of diverse virus populations in the
1091 environment using the polony method. *Nat Microbiol* **3**, 62–72 (2018).
1092 100. Carlson, M. C. G. *et al.* Viruses affect picocyanobacterial abundance and biogeography in the
1093 North Pacific Ocean. *Nat Microbiol* **7**, 570–580 (2022).

1094 101. Zuo, Y. & Deutscher, M. P. Exoribonuclease superfamilies: structural analysis and phylogenetic
1095 distribution. *Nucleic Acids Res.* **29**, 1017–1026 (2001).
1096 102. Mathews, C. K. Deoxyribonucleotides as genetic and metabolic regulators. *FASEB J.* **28**, 3832–
1097 3840 (2014).
1098 103. Varik, V., Oliveira, S. R. A., Hauryliuk, V. & Tenson, T. HPLC-based quantification of bacterial
1099 housekeeping nucleotides and alarmone messengers ppGpp and pppGpp. *Sci. Rep.* **7**, 11022 (2017).
1100 104. Wheeler, L. J., Rajagopal, I. & Mathews, C. K. Stimulation of mutagenesis by proportional
1101 deoxyribonucleoside triphosphate accumulation in Escherichia coli. *DNA Repair* **4**, 1450–1456
1102 (2005).
1103 105. Dedeo, C. L., Cingolani, G. & Teschke, C. M. Portal Protein: The Orchestrator of Capsid
1104 Assembly for the dsDNA Tailed Bacteriophages and Herpesviruses. *Annu Rev Virol* **6**, 141–160
1105 (2019).
1106 106. Hatfull, G. F. Bacteriophage genomics. *Curr. Opin. Microbiol.* **11**, 447–453 (2008).
1107 107. Hurwitz, B. L. & U'Ren, J. M. Viral metabolic reprogramming in marine ecosystems. *Curr. Opin.*
1108 *Microbiol.* **31**, 161–168 (2016).
1109 108. Adriaenssens, E. M. & Cowan, D. A. Using signature genes as tools to assess environmental viral
1110 ecology and diversity. *Appl. Environ. Microbiol.* **80**, 4470–4480 (2014).
1111 109. Shapiro, J. W. & Putonti, C. Gene Co-occurrence Networks Reflect Bacteriophage Ecology and
1112 Evolution. *MBio* **9**, (2018).
1113 110. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems
1114 Technology (RAST). *Nucleic Acids Res.* **42**, D206–14 (2014).

1115 111. Katoh, K., Misawa, K., Kuma, K.-I. & Miyata, T. MAFFT: a novel method for rapid multiple sequence
1116 alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).

1117 112. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the
1118 organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649 (2012).

1119 113. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat.*
1120 *Methods* **7**, (2010).

1121 114. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search
1122 programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

1123 115. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements
1124 in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).