Viruses of the eukaryotic plankton are predicted to increase carbon export efficiency in the global sunlit ocean

Romain Blanc-Mathieu[1,+], Hiroto Kaneko[1,+], Hisashi Endo[1], Samuel Chaffron[2,3], Rodrigo Hernández-Velázquez[1], Canh Hao Nguyen[1], Hiroshi Mamitsuka[1], Nicolas Henry[4], Colomban de Vargas[4], Matthew B. Sullivan[5], Curtis A. Suttle[6], Lionel Guidi[7] and Hiroyuki Ogata[1,*]

+ Equal contribution

* Corresponding author

Affiliations:

1: Bioinformatics Center, Institute for Chemical Research, Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan

2: Université de Nantes, Centrale Nantes, CNRS - UMR 6004, LS2N, Nantes, France.

3: Research Federation (FR2022) Tara Oceans GO-SEE, Paris, France

4: Sorbonne Universités, UPMC Université Paris 06, CNRS, Laboratoire Adaptation et Diversité en Milieu Marin, Station Biologique de Roscoff, 29680 Roscoff, France

5: Department of Microbiology and Department of Civil, Environmental and Geodetic Engineering, Ohio State University, Columbus, OH, United States of America

22 6: Departments of Earth, Ocean & Atmospheric Sciences, Microbiology & Immunology,
23 and Botany, and the Institute for the Oceans and Fisheries, University of British
24 Columbia, Vancouver, BC, V6T 1Z4, Canada
25
26 7: Sorbonne Université, CNRS, Laboratoire d'Océanographie de Villefanche, LOV, F-
27 06230 Villefranche-sur-mer, France
28
29
30 **Key words**: Biological carbon pump, viruses, shunt and pump, carbon export,
31 *Prasinovirus*, *Mimiviridae*, *Tara* Oceans.

Abstract

The biological carbon pump (BCP) is the process by which ocean organisms transfer carbon from surface waters to the ocean interior and seafloor sediments for sequestration. Viruses are thought to increase the efficiency of the BCP by fostering primary production and facilitating the export of carbon-enriched materials in the deep sea (the viral "shunt and pump"). A prior study using an oligotrophic ocean-dominated dataset from the *Tara* Oceans expedition revealed that bacterial dsDNA viruses are better associated with variation in carbon export than either prokaryotes or eukaryotes, but eukaryotic viruses were not examined. Because eukaryotes contribute significantly to ocean biomass and net production (> 40%), their viruses might also play a role in the BCP. Here, we leveraged deep-sequencing molecular data generated in the framework of *Tara* Oceans to identify and quantify diverse lineages of large dsDNA and smaller RNA viruses of eukaryotes. We found that the abundance of these viruses explained 49% of the variation in carbon export (compared with 89% by bacterial dsDNA viruses) and also substantially explained the variation in net primary production (76%) and carbon export efficiency (50%). Prasinoviruses infecting Mamiellales as well as *Mimivirus* relatives putatively infecting haptophytes are among the eukaryotic virus lineages predicted to be the best contributors to BCP efficiency. These findings collectively provide a first-level window into how eukaryotic viruses impact the BCP and suggest that the virus-mediated shunt and pump indeed plays a role.

Introduction

The oceanic biological carbon pump (BCP) is an organism-driven process by which atmospheric carbon (*i.e.* CO_2) is transferred to the ocean interior and seafloor for sequestration over periods ranging from centuries to those of geological time-scale durations. Between 15% and 20% of net primary production (NPP) is exported out of the euphotic zone, with 1% to 3% (ca. 0.2 gigatons) of fixed carbon reaching the seafloor annually (De La Rocha and Passow 2007; Herndl and Reinthaler 2013; Quéré et al. 2018; Zhang et al. 2018).

Three components of the BCP, namely, carbon fixation, export and remineralization, are governed by complex interactions between numerous members of planktonic communities (Falkowski et al. 1998). Among these organisms, diatoms (Tréguer et al. 2018) and zooplankton (Turner 2015) have been identified as important contributors to the BCP in nutrient-replete oceanic regions. In the oligotrophic ocean, Cyanobacteria and Collodaria (Lomas and Moran 2011), diatoms (Agusti et al. 2015; Leblanc et al. 2018) and other small (pico- to nano-) plankton (Richardson and Jackson 2007; Lomas and Moran 2011) have been implicated in the BCP. Overall, the composition of the planktonic community in surface waters, rather than a single species, is better associated with the intensity of the BCP (Boyd and Newton 1995; Guidi et al. 2009; Guidi et al. 2016).

A recent gene correlation network analysis based on *Tara* Oceans genomic data, ranging from viruses to zooplankton, outlined predicted species interactions associated with carbon export and revealed a remarkably strong association with bacterial dsDNA viruses relative to either prokaryotes or eukaryotes (Guidi et al. 2016). Cell lysis caused

75 by viruses promotes the production of dissolved organic matter and accelerates the
76 recycling of potentially growth-limiting nutrient elements (*i.e.* nitrogen and phosphorus)
77 in the photic zone (the "viral shunt") (Weinbauer and Peduzzi 1995; Gobler et al. 1997;
78 Wilhelm and Suttle 1999; Weinbauer 2004; Motegi et al. 2009). This recycling in turn
79 may increase primary production and carbon export (Brussaard et al. 2008; Weitz et al.
80 2015). Viruses have also been proposed to drive particle aggregation and transfer into the
81 deep sea *via* the release of sticky, carbon-rich viral lysate (the "viral shuttle") (Proctor
82 and Fuhrman 1991; Peduzzi and Weinbauer 1993; Shibata et al. 1997; Weinbauer 2004).
83 The combined effect of these two viral properties, coined "shunt and pump", is proposed
84 to enhance the magnitude and efficiency of the BCP (Suttle 2007). A study by Guidi et al.
85 (2016) revealed that populations of bacterial dsDNA viruses in the sunlit oligotrophic
86 ocean are strongly associated with variation in the magnitude of the flux of particulate
87 organic carbon. Although viruses of eukaryotes were not included in the above-
88 mentioned study, the significant contribution of their hosts to ocean biomass and net
89 production (Li 1995; Nelson et al. 1995; Worden et al. 2004; Liu et al. 2009) and their
90 observed predominance over prokaryotes in sinking materials of Sargasso Sea
91 oligotrophic surface waters (Fawcett et al. 2011; Lomas and Moran 2011) suggest that
92 eukaryotic viruses are responsible for a substantial part of the variation in exported
93 carbon. Furthermore, the mechanisms causing this association remain to be
94 dissected (Sullivan et al. 2017), as such an association might emerge through indirect
95 correlation with other parameters, such as NPP.
96 In this study, we explored the association between eukaryotic viruses and the BCP
97 as well as the viral mechanisms enhancing this process. We exploited comprehensive

98 organismal data from the *Tara* Oceans expedition (Sunagawa et al. 2015; Carradec et al.
99 2018) as well as related measurements of carbon export estimated from particle
100 concentration and size distributions observed *in situ* (Guidi et al. 2016). The investigation
101 of eukaryotic viruses based on environmental genomics has long been difficult because of
102 their lower concentration in the water column compared with prokaryotic dsDNA viruses
103 (Hingamp et al. 2013). Deep sequencing of planktonic community DNA and RNA, as
104 carried out in *Tara* Oceans, has enabled the identification of marker genes of major viral
105 groups infecting eukaryotes and begun to reveal that these groups represent a sizeable
106 fraction of the marine virosphere (Hingamp et al. 2013; Allen et al. 2017; Moniruzzaman
107 et al. 2017; Carradec et al. 2018; Mihara et al. 2018). In the present study, we identified
108 several hundred marker-gene sequences of nucleocytoplasmic large DNA viruses
109 (NCLDVs; so-called "giant viruses") in the prokaryotic size fraction. We also identified
110 RNA viruses in meta-transcriptomes of four eukaryotic size fractions. The resulting
111 profile of viral distributions was compared with the magnitude of carbon export (CE) and
112 its efficiency (CEE) to identify lineages of viruses predicted to contribute to the BCP.

113 Results and Discussion

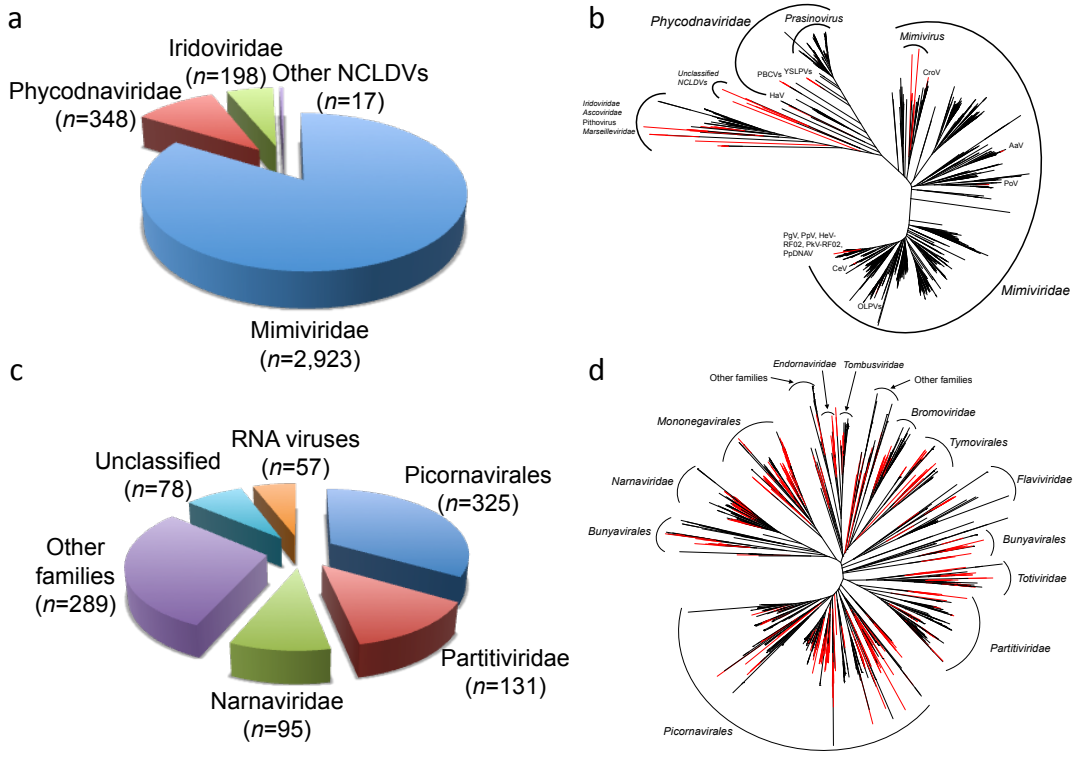
114 The discovery of diverse NCLDVs and RNA viruses in *Tara* Oceans gene
115 catalogs

116 We used profile hidden Markov model-based homology searching to identify marker-
117 gene sequences of viruses of eukaryotes in two ocean gene catalogs. These catalogs were
118 previously constructed from environmental shotgun sequence data of samples collected
119 during the *Tara* Oceans expedition. The first catalog, the Ocean Microbial Reference

120 Gene Catalog (OM-RGC), contains 40 million non-redundant genes predicted from the
121 assemblies of *Tara* Oceans viral and microbial metagenomes (Sunagawa et al. 2015). We
122 searched this catalog for NCLDV DNA polymerase family B (PolB) genes, as dsDNA
123 viruses may be present in microbial metagenomes because large virions (> 0.2 μm) have
124 been retained on the filter or viral genomes replicating within picoeukaryotic cells have
125 been captured. The second gene catalog, the Marine Atlas of *Tara* Oceans Unigenes
126 (MATOU), contains 116 million non-redundant genes predicted from meta-
127 transcriptomes of single-cell microeukaryotes and small multicellular zooplankton
128 (Carradec et al. 2018). We searched this catalog for RNA-dependent RNA polymerase
129 (RdRP) genes of RNA viruses because transcripts of viruses actively infecting their host
130 or genomes of RNA viruses have been captured in it.

131 We identified 3,486 NCLDV PolB sequences and 975 RNA virus RdRP
132 sequences. All except 17 of the NCLDV PolBs were assigned to the families *Mimiviridae*
133 (*n* = 2,923), *Phycodnaviridae* (*n* = 348) and *Iridoviridae* (*n* = 198) (Figure 1a). The large
134 number of PolB sequences assigned to *Mimiviridae* and *Phycodnaviridae* compared with
135 other NCLDV families is consistent with a previous observation based on a smaller
136 dataset (Hingamp et al. 2013). The divergence between these environmental sequences
137 and reference sequences from known viral genomes was greater in *Mimiviridae* than
138 *Phycodnaviridae* (Figure 1b, Supplementary Figure 1a). Within *Mimiviridae,* 83% of the
139 sequences were most similar to those from algae-infecting *Mimivirus* relatives. Among
140 the sequences classified in *Phycodnaviridae,* 92% were most similar to those in
141 *Prasinovirus*, while 5% were closest to *Yellowstone lake phycodnavirus*, which is closely
142 related to *Prasinovirus*. RdRP sequences were assigned mostly to the order

143 *Picornavirales* (*n* = 325) followed by the families *Partitiviridae* (*n* = 131), *Narnaviridae*
144 (*n* = 95), *Tombusviridae* (*n* = 45) and *Virgaviridae* (*n* = 33) (Figure 1c), with most
145 sequences being distant (30% to 40% amino acid identity) from reference viruses (Figure
146 1d, Supplementary Figure 1b). This result is consistent with previous studies on the
147 diversity of marine RNA viruses, in which RNA virus sequences were found to
148 correspond to diverse positive-polarity ssRNA and dsRNA viruses distantly related to
149 well-characterized viruses (reviewed in Culley [2018]).



150

151 **Figure 1: Numerous lineages of nucleocytoplasmic large DNA viruses (NCLDVs)**
152 **and RNA viruses were identified in environmental samples collected during the**
153 ***Tara* Oceans expedition (2009-2013). a** and **c** Taxonomic breakdown of environmental
154 sequences of NCLDV DNA polymerase family B and RNA virus RNA-dependent RNA
155 polymerase. **b** and **d** Unrooted maximum likelihood phylogenetic trees of environmental
156 (black) and reference (red) viral sequences.

157 **Eukaryotic viruses linked to carbon export efficiency in the sunlit ocean**

158 Among the PolB and RdRP sequences identified in the *Tara* Oceans gene catalogs, 37%
159 and 17% were respectively present in at least five samples and had accompanying carbon
160 export measurement data (Supplementary Table 1). Using abundance profiles of these
161 1,454 marker-gene sequences at 61 sampling sites in the photic zone of 40 *Tara* Oceans
162 stations (Figure 2a–c), we tested for associations with estimates of carbon export flux at
163 150 meters (CE_{150}) and measurements of carbon export efficiency (CEE). PLS regression
164 model explained 49% ($R^2 = 49\%$) of the variation in CE_{150}, with a Pearson correlation
165 coefficient between observed and predicted values of 0.67 ($P < 1 \times 10^{-4}$) (Figure 2d); this
166 result demonstrates that viruses of eukaryotes are associated with the magnitude of
167 carbon export. A comparison of viral abundances with CEE revealed that viruses are also
168 strongly associated with CEE ($R^2 = 50\%$, $r = 0.72$, $P < 1 \times 10^{-4}$) (Figure 2e and
169 Supplementary Figure 2a; see Supplementary Table 2 for details of PLS models and
170 Supplementary Figure 2b for details of the permutation tests). In these PLS regression
171 models, 62 and 26 viruses were considered to be important predictors (*i.e.* variable
172 importance in the projection [VIP] score > 2 and regression coefficient > 0) of CE_{150} and
173 CEE, respectively, and only two viruses were shared between them.

174 CE_{150} was found to be correlated with NPP ($r = 0.76$, $P < 1 \times 10^{-11}$), which
175 suggests that the association of viruses with CE_{150} is due to a viral shunt effect or to
176 primary-production enhancement of viral production. In line with this interpretation, we
177 found that viral abundances can predict NPP variations ($R^2 = 59\%$, $r = 0.78$, $P < 1 \times 10^{-4}$)
178 (Figure 2f); in addition, a larger number of shared predictors were obtained under the
179 PLS model for CE_{150} (44 viruses) than for CEE (3 viruses) ($P = 1.3 \times 10^{-7}$). In contrast,

CEE was not correlated with NPP ($r = 0.2$, $P = 0.1$) or CE_{150} ($r = 0.14$, $P = 0.3$). The association of some viruses with CEE is therefore not an indirect relationship caused by NPP; instead, viruses of eukaryotes may have a shuttling effect that enhances the efficiency of carbon export.

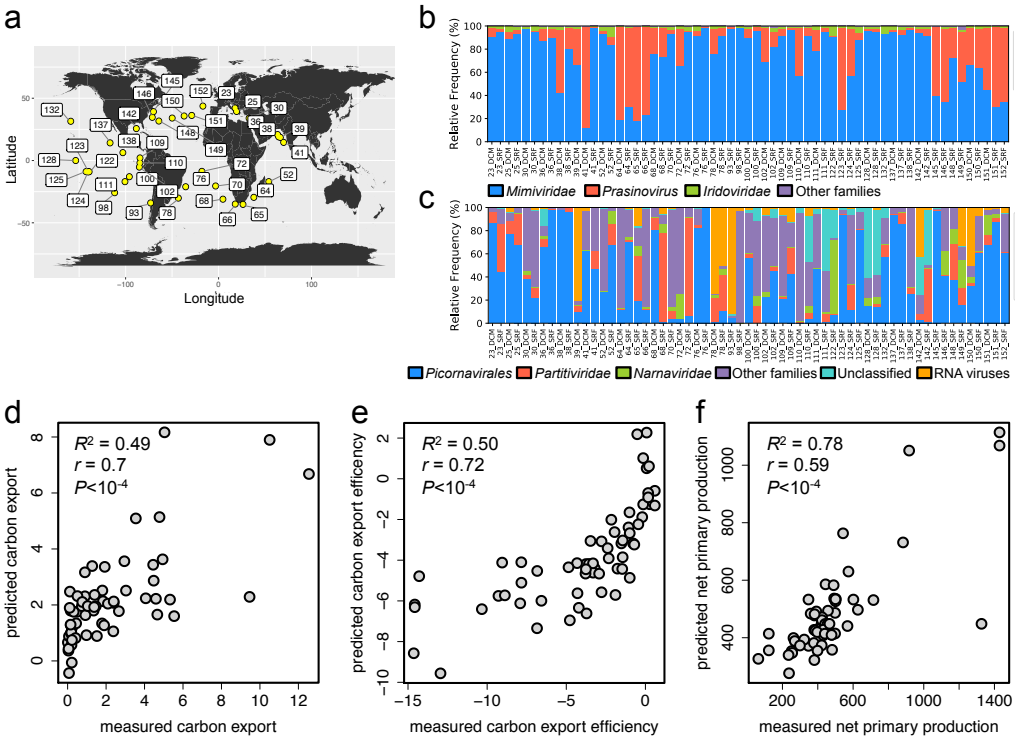


Figure 2: Abundance of nucleocytoplasmic large DNA viruses (NCLDVs) and RNA viruses is associated with carbon export efficiency and flux at 150m in the global ocean. a Location of 40 *Tara* Oceans stations that were the source of 61 prokaryote-enriched metagenomes and 244 eukaryotic meta-transcriptomes (61 sites x 4 organismal size fractions) from surface and DCM layers and measurements of carbon export. **b–c** Relative abundance of NLCVDs and RNA viruses in samples used for association analyses. **d–f** Plots showing the correlation between predicted and observed values in a leave-one-out cross-validation test for carbon export flux at 150 m (**d**), carbon export efficiency (**e**) and net primary production (**f**). Each PLS regression model was reconstructed using abundance profiles of 1,454 marker-gene sequences (1,282 PolBs and 172 RdRPs) derived from environmental samples. *r*, Pearson correlation coefficient; R^2, square of the correlation between measured response values and predicted response values. R^2, which was calculated as 1 – SSE / SST (sum of squares due to error and total), measures how successful the fit is in explaining the variation of the data. Model robustness was assessed using a permutation test ($n = 10{,}000$).

200 ## Giant viruses of small algae are predicted to enhance CEE

201 We considered 26 viruses positively associated with CEE with a VIP score > 2
202 (Supplementary File 1) to be important predictors of CEE in the PLS regression. These
203 viruses are hereafter referred to as VIPs. Eleven VIPs (nine NCLDVs and two RNA
204 viruses; red rectangle in Figure 3) were abundant in samples from Eastern India coast
205 (EAFR) and Benguela current coast (BENG) provinces, with some (*e.g.* polb 00248170
206 and 002035391) also relatively abundant in samples from different oceanic provinces
207 where CEE was also relatively high. This observation suggests that these viruses enhance
208 the BCP in different regions of the global ocean.

209 Most VIPs (23 of 26) belonged to *Mimiviridae* (*n* = 11) and *Phycodnaviridae* (*n* =
210 12). All the phycodnavirus sequences were most closely related to those of
211 prasinoviruses, with amino acid sequence identities to reference sequences ranging
212 between 64% and 88%. The three remaining VIPs were RNA viruses. The closest
213 homolog of one sequence (rdrp 36496887) was that of an octopus-associated member of
214 *Picornavirales* (Beihai picorna-like virus 106), while the other two (90355641 and
215 58847823) were most similar to that of a crustacean-associated *Bunyavirales* virus
216 (Wenling crustacean virus 9).

217 Taxonomic analysis of genes in genome fragments containing VIP PolBs further
218 confirmed their identity as *Mimiviridae* or *Phycodnaviridae* (Supplementary Figure 3).
219 One contig (CNJ01018797) was predicted to encode a protein having a high amino-acid
220 identity (94%) to a gene product from the transcriptome of the diatom *Triceratium*
221 *dubium* (Supplementary Figure 3b). NCLDVs infecting diatoms have not been reported

so far, but a previous co-abundance analysis has linked diatoms and *Mimiviridae* (Moniruzzaman et al. 2017).

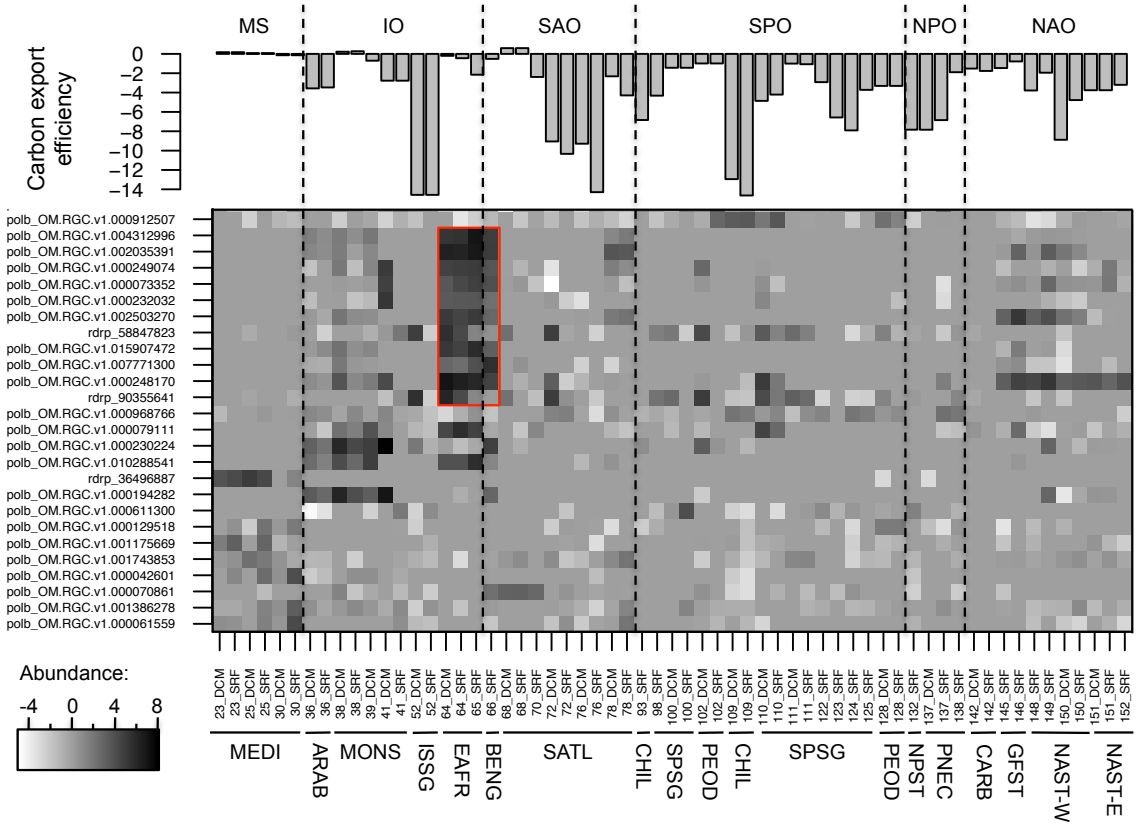


Figure 3: Biogeography of viral lineages associated with carbon export efficiency (CEE). The upper panel shows carbon export efficiency, ($CE_{deep} - CE_{surface}$) / CE_{deep}, for the 61 sampling sites. Sites with values above and below zero are those where carbon export flux was respectively higher and lower in deep than in surface waters; hence, they correspond to sites where the biological carbon pump was higher or lower compared with other sites. The bottom panel is a map reflecting abundances, expressed as center-log ratio transformed gene-length normalized read counts, of viruses positively associated with CEE and having a VIP score > 2. MS, Mediterranean Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SPO, South Pacific Ocean; NPO, North Pacific Ocean; NAO, North Atlantic Ocean. The bottom horizontal axis is labeled with *Tara* Oceans station numbers, sampling depth (SRF, surface; DCM, deep chlorophyll maximum) and abbreviations of biogeographic provinces.

Because most VIPs were only distantly related to isolated viruses, inferring characteristics such as host type and infection strategies was not straightforward. We therefore conducted a phylogeny-guided network-based host prediction analysis (Figure

241 4; see Methods). Within the *Prasinovirus* clade, which contained six VIPs, enrichment
242 analysis of predicted host groups along the phylogenetic tree uncovered 10 nodes
243 significantly enriched for five different eukaryotic orders. Mamiellales, the only known
244 host group of prasinoviruses, was detected at nine different nodes (seven of them had no
245 parent-to-children relationships), while the other four eukaryotic orders were found at
246 only one node (or two in the case of Eutreptiales) (Figure 4a). Within *Mimiviridae*,
247 significant enrichment was detected for 10 different orders; Collodaria was detected at 15
248 nodes (2 of them had no parent-to-children relationships) and Prymnesiales at 6 nodes (3
249 of them had no parent-to-children relationship), while all other orders were present at a
250 maximum of one node with no parent-to-children relationships. The nodes enriched in
251 Prymnesiales and Collodaria fell within or were sister to clades containing reference
252 viruses isolated from Prymnesiales and Phaeocystales (both are haptophytes orders)
253 species. This placement suggests that environmental PolB sequences in this clade belong
254 to *Mimiviridae* viruses that infect Prymnesiales. Interestingly four of these *Mimiviridae*
255 viruses were VIP viruses (Figure 4a). The detection of Collodaria may be the result of
256 indirect associations that reflect a symbiotic relationship with Prymnesiales, as some
257 acantharians, a lineage of Rhizaria related to Collodaria, are known to host Prymnesiales
258 species (Mars Brisbin et al. 2018). Nodes enriched in the remaining three other orders fell
259 within clades that were only distantly related to any reference *Mimiviridae*. The final
260 *Mimiviridae* VIP (polb 000079111) in the tree was a distant relative of *Aureococcus*
261 *anophagefferens virus* (AaV). No host groups were predicted for the clade containing the
262 Picornavirales VIP (Figure 4b). Overall, this host prediction analysis revealed that

lineages of prasinoviruses and putative prymnesioviruses are among the viral lineages best associated with CEE.

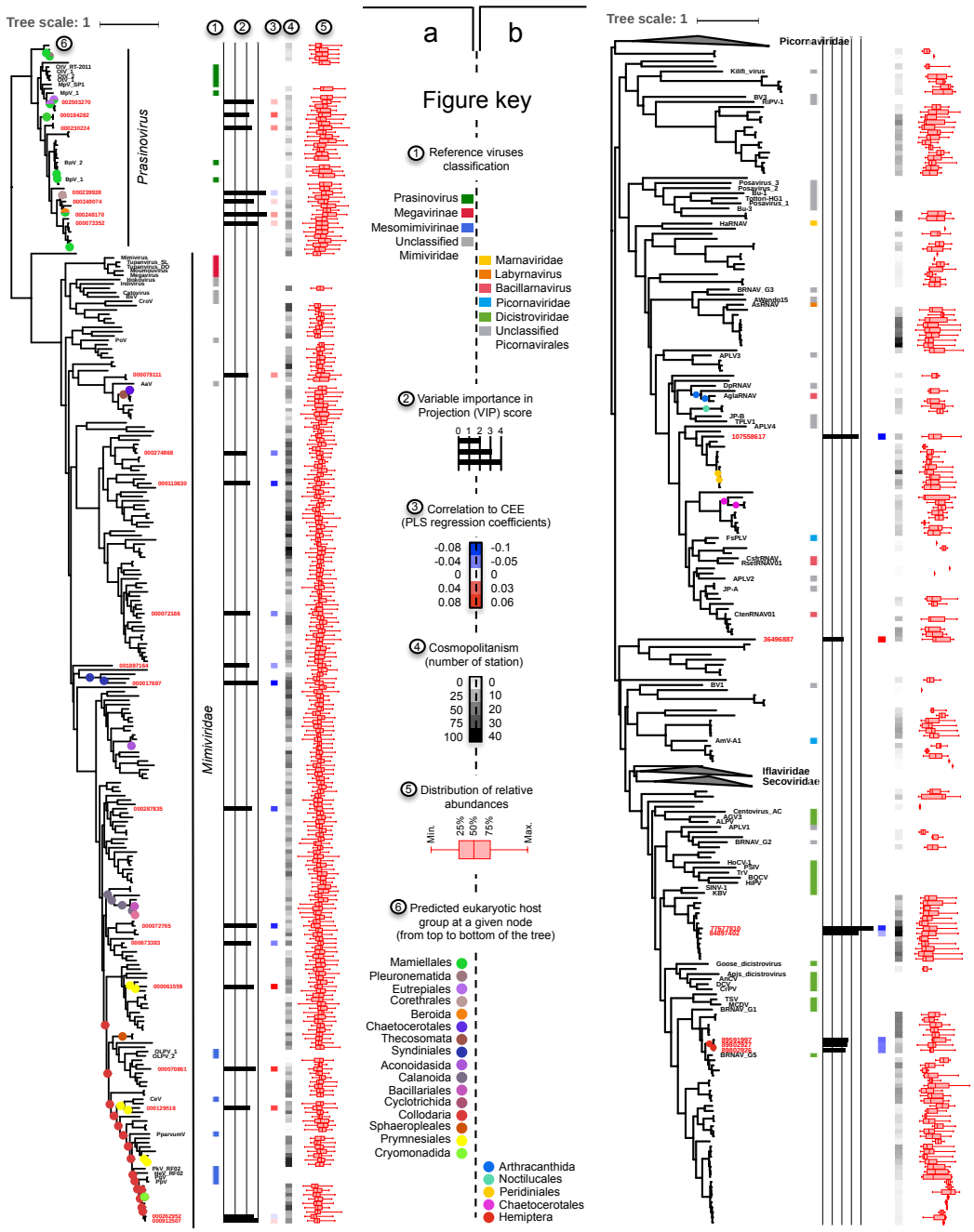


Figure 4: Phylogenetic position of viruses associated with the efficiency of carbon export. Details of the PLS model and viral global distribution, relative abundance and putative eukaryotic host group are superimposed on the trees. Phylogenetic trees contain environmental (labeled in red if VIP > 2) and reference (labeled in black) sequences of *Prasinovirus* and *Mimiviridae* DNA polymerase family B (**a**) and *Picornavirales* RdRP (**b**).

272 **Functional traits of eukaryotic organisms interacting with CEE-associated**
273 **viruses**

274 The host assignment of Mamiellales, Prymnesiales (Figure 4a) and Chaetocerotales
275 (Figure 4b) to viral clades containing reference viruses isolated from these organismal
276 groups suggests that the network-based approach was able to predict virus–host
277 relationships with a certain level of reliability using our large dataset despite the reported
278 limitations of similar types of analyses (Coenen and Weitz 2018). This positive signal
279 prompted us to use these species-association networks to investigate taxonomic and
280 functional differences between the eukaryotic organisms predicted to interact with viruses
281 that were either positively or negatively associated with CEE (Figure 5). At the functional
282 level, viruses positively associated with CEE had a greater number of connections with
283 chloroplast-bearing eukaryotes ($Q = 0.03$) and with silicified eukaryotes ($Q = 0.05$)
284 compared with viruses negatively associated with CEE (Supplementary Table 4); this
285 suggests that these virus–host systems contribute to CEE. This result also supports the
286 proposed idea that viral lysis enhances the effect of the BCP; that is, lysis of autotrophs is
287 expected to release growth-limiting nutrients (Gobler et al. 1997) that are recycled to
288 grow new cells, while carbon-rich cell debris, potentially ballasted with silica, tends to
289 aggregate and sink (Suttle 2007).

Viral sequence ID	Reg. Coeff.	Taxa	Chloroplast	Silicification	Calcification
polb_OM.RGC.v1.001386278	0.064	Dinophyceae	■		
polb_OM.RGC.v1.000070861	0.063	Collodaria			
polb_OM.RGC.v1.000042601	0.062	Dinophyceae	■		
polb_OM.RGC.v1.001743853	0.057	Trachymedusae			
polb_OM.RGC.v1.001175669	0.056	Rotaliida			■
polb_OM.RGC.v1.000129518	0.055	Dinophyceae	■		
polb_OM.RGC.v1.000611300	0.047	Ulvophyceae	■		
polb_OM.RGC.v1.000194282	0.047	Gymnodiniales	■		
rdrp_36496887	0.038	Poecilostomatoida			
polb_OM.RGC.v1.010288541	0.038	Bacillariophyta	■	■	
polb_OM.RGC.v1.000230224	0.033	Arthra_Symphy_F1			
polb_OM.RGC.v1.000079111	0.032	Cephaloidophorida			
polb_OM.RGC.v1.000968766	0.032	Calanoida			
polb_OM.RGC.v1.000248170	0.028	Beroida			
polb_OM.RGC.v1.007771300	0.027	Dinophyceae			
polb_OM.RGC.v1.015907472	0.024	Coscinodiscophytina	■	■	
polb_OM.RGC.v1.000232032	0.013	Gregarinomorphea			
polb_OM.RGC.v1.000073352	0.012	Mamiellophyceae	■		
polb_OM.RGC.v1.000249074	0.010	Dinophysiales			
polb_OM.RGC.v1.002035391	0.010	Coscinodiscophytina	■	■	
polb_OM.RGC.v1.004312996	0.009	Coscinodiscophytina	■	■	
polb_OM.RGC.v1.000912507	0.009	Acanthoecida			
polb_OM.RGC.v1.000062429	-0.009	Peridiniales			
polb_OM.RGC.v1.000239928	-0.018	Dinophyceae	■		
polb_OM.RGC.v1.001445888	-0.026	Copepoda			
polb_OM.RGC.v1.001897164	-0.035	Copepoda			
polb_OM.RGC.v1.001883961	-0.043	Ostracoda			
polb_OM.RGC.v1.000844241	-0.043	Peridiniales			
polb_OM.RGC.v1.000673383	-0.043	Calanoida			
polb_OM.RGC.v1.000072166	-0.043	Labyrinthulida			
polb_OM.RGC.v1.000274868	-0.045	Chlorophyceae	■		
polb_OM.RGC.v1.001561949	-0.048	Ostracoda			
polb_OM.RGC.v1.000205020	-0.048	Collodaria			
polb_OM.RGC.v1.004472615	-0.049	Calanoida			
polb_OM.RGC.v1.000228894	-0.051	Variosea			
polb_OM.RGC.v1.009636901	-0.051	Collodaria			
polb_OM.RGC.v1.000495602	-0.056	MAST_3-12			
polb_OM.RGC.v1.000074398	-0.061	Ichthyosporea			
polb_OM.RGC.v1.015629864	-0.062	Peridiniales			
polb_OM.RGC.v1.000287835	-0.064	Collodaria			
polb_OM.RGC.v1.000471455	-0.065	Collodaria			
polb_OM.RGC.v1.000110630	-0.073	Acantharea			
polb_OM.RGC.v1.002652025	-0.075	Dinophyceae	■		
rdrp_89802927	-0.075	Calanoida			
rdrp_89802926	-0.075	Insecta			
polb_OM.RGC.v1.002767869	-0.078	Collodaria			
rdrp_103795290	-0.082	Collodaria			
polb_OM.RGC.v1.000017697	-0.083	Calanoida			
rdrp_77677810	-0.099	Leptothecata			

Figure 5: Functional traits and taxonomy of putative hosts of viruses associated with the efficiency of carbon export. "Putative hosts" refers to eukaryotic V9-OTUs with the highest absolute edge weight in FlashWeave interaction networks. Taxa correspond to higher rank classification of the V9-OTU. Black squares indicate the presence of a functional trait for the V9-OTU best connected to the virus.

296 ## Integrating previous observations to model a virus-driven BCP

297 Our PLS models revealed that viruses of eukaryotes are associated not only with the
298 magnitude of carbon export, but also with the export efficiency of particles. This finding
299 suggests that viruses contribute to BCP enhancement by promoting aggregate formation
300 and subsequent sedimentation. Consistent with this view, prasinoviruses and putative
301 prymnesioviruses were found to be among the viruses best associated with the efficiency
302 of carbon export under our PLS model for CEE. Several prasinoviruses (fourteen with an
303 available genome sequence) have been isolated for three genera of green microalgae,
304 namely *Micromonas*, *Ostreococcus* and *Bathycoccus*. These genera belong to the order
305 Mamiellales, which are bacterial-sized phytoplankton common in coastal and oceanic
306 environments and considered to be influential actors in oceanic systems (Not et al. 2012;
307 Weynberg et al. 2017). *Chrysochromulina ericina* virus, *Prymnesium parvum* DNA virus,
308 *Prymnesium kappa* virus-RF02 and *Haptolina ericina* virus are *Mimiviridae* viruses
309 isolated from Prymnesiales (Haptophyta) cultures (Figure 4). They are closely related to
310 *Phaeocystis globosa* virus (PgV) and *Phaeocystis poutcheti* virus isolated from
311 Phaeocystales (Haptophyta) cultures. Both Prymnesiales and Phaeocystales species have
312 non-calcifying organic scales and some species can form blooms and colonies. The
313 detection of several putative prymnesioviruses in samples containing small cells is
314 consistent with the observation of diverse and abundant noncalcifying haptophytes in
315 open oceans (Liu et al. 2009; Endo et al. 2018). According to the estimates of Liu et al
316 (2009), these noncalcifying haptophytes contribute twice as much as diatoms or
317 prokaryotes to photosynthetic production, thereby highlighting their importance in the
318 BCP. Prymnesiales are an important mixotrophic group in oligotrophic ocean (Unrein et

319 al. 2014) and mixotrophy is proposed to increase vertical carbon flux by enabling the
320 uptake organic forms of nutrients (Ward and Follows 2016). Viral infection of such
321 organisms may thus help increase BCP efficiency.

322 As most other cultured algal viruses, prasinoviruses and prymnesioviruses are
323 lytic. Viral lysis not only generates the cellular debris used to build aggregates; it also
324 facilitates the process of aggregation (Weinbauer 2004) *via* viral-induced production of
325 agglomerative compounds, such as transparent exopolymer particles (TEPs). Lønborg et
326 al. (2013) proposed that the increased DOC and TEP production observed in cultures of
327 *Micromonas pusilla* infected with prasinoviruses (compared with non-infected cultures)
328 could stimulate particle aggregation and thus carbon export out of the photic zone. Some
329 prasinoviruses encode glycosyltransferases of the GT2 family. Similar to the a098r gene
330 (GT2) in *Paramecium bursaria Chlorella virus 1*, the expression of GT2 family members
331 during infection possibly leads to the production of a dense fibrous hyaluronan network at
332 the surface of infected cells. Such a network may trigger the aggregation of host cells,
333 facilitate viral propagation (Van Etten et al. 2017) and increase the cell wall C:N ratio.
334 PgV, closely related to prymnesioviruses (Figure 4), has been linked with increased TEP
335 production and aggregate formation during the termination of *Phaeocystis* bloom
336 (Brussaard et al. 2007).

337 Other previous experimental and *in situ* studies have linked viruses and viral
338 activity to vertical carbon flux, cell sinking, and aggregate formation. Several laboratory
339 experiments have revealed associations between viruses and sinking cells or between
340 viruses and aggregate formation. For example, sinking rates of the toxic bloom former
341 *Heterosigma akashiwo* are elevated following viral infection (Lawrence and Suttle 2004).

342 As early as 1993, an experimental study revealed increased particle aggregation and
343 primary productivity upon addition of virally enriched material to seawater samples
344 (Peduzzi and Weinbauer 1993). More recently, cultures of the diatom *Chaetoceros*
345 *tenuissimus* infected with a DNA virus (CtenDNAV type II) have been shown to produce
346 higher levels of large-sized particles (50 to 400 μm) compared with non-infected cultures
347 (Yamada et al. 2018). *In situ* studies have uncovered vertical transport of viruses between
348 photic and aphotic zones in the Pacific Ocean (Hurwitz et al. 2015) and in *Tara* Oceans
349 samples (Brum et al. 2015), which suggests their association with sinking material. In
350 addition, the level of active *Emiliania huxleyi* virus (EhV) infection in a coccolithophore
351 bloom has been found to be correlated with water column depth, which suggests that
352 infected cells are exported from the surface to deeper waters in aggregate form (Sheyn et
353 al. 2018). EhVs infecting *E. huxleyi* have also been found to facilitate particle
354 aggregation and downward vertical flux of both particulate organic and particulate
355 inorganic carbon in the North Atlantic (Laber et al. 2018). No *Phycodnaviridae* PolB had
356 their best hits to EhVs in our study, which is probably because of sampling regions and
357 periods. These previous observations come as potential mechanistic explanations that
358 provide support for our results suggesting eukaryotic viruses have a "shuttle effect" that
359 enhance carbon export.

360 Viruses *versus* hosts as key markers of CE and CEE variations

361 Our analysis used similar data and analytical methods as those of Guidi et al. (2016), who
362 explored the link between planktonic networks and the BCP. Those authors reported that
363 bacterial viruses are better associated with CE_{150} than are cellular organisms. In their
364 study, host lineages for prasinoviruses (Mamiellales) and prymnesioviruses

365 (Prymnesiales) were not found to be associated with CE_{150}. The fact that we detected
366 these viruses as predictors of CEE may mean that viruses are better predictors than their
367 hosts, most likely because viruses may better reflect the overall process and mechanisms
368 by which carbon is exported. Viral infection causes lysis and aggregation of cells,
369 enhances sinking rates and increases the export of organic carbon out of the sunlit ocean.
370 Among the eukaryotic organisms studied by Guidi et al. (2016), collodarians,
371 dinoflagellates and copepods were reported to be important contributors to CE_{150}. Some
372 of the viral taxa we found to be associated with CEE were connected with collodarians
373 and dinoflagellates in our host prediction analysis, although not as clearly as with
374 Mamiellales and Prymnesiales. Known copepod viruses have a ssDNA genome and
375 hence were not captured in our data. Only one genome sequence of a ssRNA virus and
376 one PolB sequence of a NCLDV are available for dinoflagellates viruses, and none have
377 been reported for collodarians. Viruses infecting these organisms may be among those we
378 found to be associated with CEE, but the lack of reference viruses hampered their
379 identification from the environmental samples.

380 Conclusions

381 In this study, we identified associations between CEE and diverse groups of planktonic
382 eukaryotic viruses collected in the photic ocean from 61 sampling sites during the *Tara*
383 Oceans expedition. Two lineages were predicted to be important groups facilitating
384 carbon export in the global ocean: prasinoviruses infecting tiny green algae of the order
385 Mamiellales, and *Mimivirus* relatives putatively infecting haptophytes of the order
386 Prymnesiales. This finding suggests that these eukaryotic viruses, like cyanophages

387 previously reported to be strongly associated with carbon export, are important factors in
388 the BCP. This idea is in agreement with observations that their hosts, despite being one to
389 two orders of magnitude less abundant than cyanobacteria, likely dominate carbon
390 biomass and net production in the ocean. The global-scale association between eukaryotic
391 viruses and CEE is empirical evidence supporting the shunt and pump hypothesis,
392 namely, that viruses enhance carbon pump efficiency by increasing the relative amount of
393 organic carbon exported from the surface to the deeper sea.

394 Methods

395 Data context

396 We used publicly available data generated in the framework of the *Tara* Oceans
397 expedition (Karsenti et al. 2011). Single-copy marker-gene sequences for NCLDVs and
398 RNA viruses were identified from two gene catalogs: the Ocean Microbial Reference
399 Gene Catalog (OM-RGC) and the Marine Atlas of Tara Oceans Unigenes (MATOU).
400 The viral marker-gene abundance profiles used in our study for prokaryotic-sized
401 metagenomes and eukaryotic-sized meta-transcriptomes were from Sunagawa et al.
402 (2015) and Carradec et al. (2018), respectively. For eukaryotic 18S rDNA V9 OTUs, we
403 used an updated version of the data of de Vargas et al. (2015) that included functional
404 trait annotations (chloroplast-bearing, silicified and calcified organisms) of V9-OTUs.
405 Abundance profiles are compositional matrices in which gene abundances are expressed
406 as unnormalized (V9 barcode data) or gene-length normalized (shotgun data) read counts.
407 Eukaryotic plankton samples were filtered for categorization into the following size
408 classes: piconano (0.8–5 μm), nano (5–20 μm), micro (20–180 μm) and meso (180–2000

409 μm) (Pesant et al. 2015). Indirect measurements of carbon export (mg^{-1} m^{-2} d^{-1}) in 5-m
410 increments from the surface to a 1000-m depth were taken from Guidi et al. (2016). The
411 original measurements were derived from the distribution of particle sizes and
412 abundances collected using an underwater vision profiler (Picheral et al. 2010). These
413 raw data are available from PANGEA (Picheral et al. 2014). Net primary production
414 (NPP) data were extracted from 8-day composites of the vertically generalized production
415 model (VGPM) (Behrenfeld and Falkowski 1997) at the week of sampling.

416 Carbon export and carbon export efficiency

417 Carbon flux profiles (mg^{-1} m^{-2} d^{-1}) based on particle size distribution and abundances
418 were estimated according to Guidi et al. (2008). Carbon flux values from depths of 30 to
419 970 m were divided into 20-m bins, each obtained by averaging the carbon flux values
420 from the designated 20 m from profiles gathered during biological sampling within a 25-
421 km radius during 24 h and having less than 50% missing data (Supplementary Figure 4).
422 For compatibility with Guidi et al. (2016), carbon export was defined as the carbon flux
423 at 150 m. Carbon export efficiency was calculated as follows: CEE = (CE_{deep} −
424 $CE_{surface}$)/CE_{deep} (Sarmiento 2013) and CEE′ = CE_{deep}/$CE_{surface}$ (Francois et al. 2002). In
425 these formulas, $CE_{surface}$ is the maximum average carbon flux within the first 150 m (the
426 maximum layer depth varied between 20 to 130 m in this dataset), and CE_{deep} is the
427 average carbon flux 200 m below this maximum.

428 Acquisition of viral marker genes from ocean gene catalogs

429 Viral genes were collected from two gene catalogs: OM-RGC version 1 (Sunagawa et al.
430 2015) and MATOU (Carradec et al. 2018). The OM-RGC data were taxonomically re-

431 annotated as in Carradec et al. (2018). Importantly, the NCBI reference tree used to
432 determine the last common ancestor was modified to reflect the current NCLDVs
433 classification (see Carradec et al. [2018] for details). We classified viral gene sequences
434 as eukaryotic or prokaryotic according to their best BLAST score against viral sequences
435 in the Virus-Host DB (Mihara et al. 2016). DNA polymerase B (PolB) and RNA-
436 dependent RNA polymerase (RdRP) genes were used as markers for NCLDVs and RNA
437 viruses, respectively. For PolB, reference proteins from the NCLDV orthologous gene
438 cluster NCVOG0038 (Yutin et al. 2009) were aligned using *linsi* (Katoh and Standley
439 2013). A HMM profile was constructed from the resulting alignment using *hmmbuild*
440 (Eddy 1998). This PolB HMM profile was searched against OM-RGC amino acid
441 sequences annotated as NCLDVs, and sequences longer than 200 amino acids and having
442 hits with *E*-values $< 1 \times 10^{-5}$ were selected as putative PolBs. RdRP sequences were
443 chosen from the MATOU catalog as follows: sequences assigned to Pfam profiles
444 PF00680, PF00946, PF00972, PF00978, PF00998, PF02123, PF04196, PF04197 or
445 PF05919 and annotated as RNA viruses were retained as RdRPs. The resulting 3,486
446 PolB sequences and 975 RdRP sequences were used along with their abundance matrices
447 for PLS regression analyses, co-abundance network reconstructions and to build
448 phylogenies.

449 **Testing for associations of viral abundance with CE_{150} and CEE**

450 To test for an association between the abundance of viral marker genes and CEE, we
451 proceeded as follows. Only marker genes represented by at least two reads in five or
452 more samples were retained. To cope with the sparsity and composition of the data, gene-
453 length normalized read-count matrices were center-log transformed separately for RNA

454 viruses and NCLDVs. We next selected genes with a Spearman correlation coefficient to
455 CE_{150} or CEE greater than 0.2 or smaller than -0.2 (zero values were removed). To
456 assess the association between these marker genes and CEE, we used partial least square
457 (PLS) regression analysis as described in Guidi et al. (2016). The number of components
458 selected for the PLS model was chosen to minimize the root mean square error of
459 prediction (Supplementary Table 2). We assessed the strength of the association between
460 carbon export (the response variable) and viral abundance (the explanatory variable) by
461 correlating leave-one-out cross-validation predicted values with the measured carbon
462 export values. We tested the significance of the correlation by comparing the original
463 Pearson coefficient between explanatory and response variables with the distribution of
464 Pearson coefficients obtained from PLS models reconstructed on permutated data (10,000
465 iterations). We estimated the contribution of each gene (predictor) according to its
466 variable importance in the projection (VIP) (Chong and Jun 2005). The VIP measure of a
467 predictor estimates its contribution in the PLS regression. Predictors having high VIP
468 values (> 2) are assumed to be important for the PLS prediction of the response variable.

469 **Phylogenetic analysis**

470 Environmental PolB sequences annotated as *Phycodnaviridae* or *Mimiviridae* were
471 searched against reference PolB sequences using BLAST. Environmental sequences with
472 hits to a reference sequence having >40% identity and an alignment length greater than
473 400 amino acids were kept and aligned with reference sequences using *linsi* (Katoh and
474 Standley 2013). We further removed sequences that occurred in fewer than 10 samples to
475 reduce the size of the tree. Environmental RdRP sequences annotated as *Picornavirales*
476 were translated into six frames of amino acid sequences, and reference *Picornavirales*

477 RdRP sequences collected from the Virus-Host DB (Mihara et al. 2016) were searched
478 against the CDD database (Marchler-Bauer et al. 2015) using rpsBLAST. The resulting
479 alignment was used to trim reference and environmental RdRP sequences to the
480 conserved part corresponding to the domain, CDD: 279070, before alignment with *linsi*.
481 PolB and RdRP multiple sequence alignments were manually curated to discard poorly
482 aligned sequences. Trees were constructed using the *build* function of ETE3 (Huerta-
483 Cepas et al. 2016) as implemented at GenomeNet (https://www.genome.jp/tools-bin/ete).
484 Columns were automatically trimmed using *trimAl* (Capella-Gutiérrez et al. 2009), and
485 trees were constructed using FastTree with default settings (Price et al. 2009).

486 **Virus–eukaryote co-occurrence analysis**

487 We used FlashWeave (Tackmann et al. 2018) with Julia 1.0 to detect virus–host
488 associations on the basis of their co-abundance patterns. Read count matrices for the
489 3,486 PolB, 975 RdRP and 18S V9 DNA barcodes obtained from samples collected at the
490 same location were fed into FlashWeave. 18S-V9 data were filtered to retain OTUs with
491 an informative taxonomic annotation. 18S-V9 OTUs and viral marker sequences were
492 further filtered to conserve only those present in at least five samples. FlashWeave
493 analyses were run separately for each of the four eukaryotic size fractions. The number of
494 samples per size fraction ranged between 51 and 99 for NCLDVs and between 36 and 62
495 for RNA viruses. The number of retained OTUs per size fraction varied between 1,775
496 and 2,269 for NCLDVs and between 48 and 125 for RNA viruses (Supplementary Table
497 3). FlashWeave was run under the settings heterogenous = false and sensitive = true, and
498 PolB/RdRP-V9-OTU edges receiving a weight > 0.2 and a *Q*-value < 0.01(the default)
499 were retained.

500 ## Mapping of putative hosts onto viral phylogenies

501 In our association networks, individual viral sequences were often associated with
502 multiple 18S-V9 OTUs belonging to drastically different eukaryotic groups, a situation
503 that can reflect interactions among multiple organisms but also noise associated with this
504 type of analysis (Coenen and Weitz 2018). To extract meaningful information from these
505 networks we reasoned as follow: We assumed that evolutionarily related viruses infect
506 evolutionary related organisms, similar to the case of phycodnaviruses (Clasen and Suttle
507 2009). In the interaction networks, the number of connections between viruses in a given
508 clade and its eukaryotic host group should accordingly be enriched compared with the
509 number of connections with non-host organisms arising by chance. Following this
510 reasoning, we assigned the most likely eukaryotic host group as follows. The tree
511 constructed from viral marker-gene sequences (PolB or RdRP) was traversed from root to
512 tips to visit every node. We counted how many connections existed between leaves of
513 each node and the V9-OTUs of a given eukaryotic group (order level). We then tested
514 whether the node was enriched compared with the rest of the tree using Fischer's exact
515 test and applied the Benjamini-Hochberg procedure to control the FDR among
516 comparisons of each eukaryotic taxon (order level). To avoid the appearance of
517 significant associations driven by a few highly connected leaves, we required half of the
518 leaves within a node to be connected to a given eukaryotic group. Significant enrichment
519 of connections between a virus clade and a eukaryotic order was considered to be
520 indicative of a possible virus–host relationship. We refer to the above approach, in which
521 taxon interactions are mapped onto a phylogenetic tree of a target group using the
522 organism's associations predicted from a species co-abundance-based network, as TIM,

523 for Taxon Interaction Mapper. The script is available upon request. This approach can be
524 extended to interactions other than virus–host relationships.

525 Acknowledgements

526 This work was supported by JSPS/KAKENHI (Nos. 26430184, 18H02279, 19H05667 to
527 H.O.), Scientific Research on Innovative Areas from the Ministry of Education, Culture,
528 Science, Sports and Technology (MEXT) of Japan (Nos. 16H06429, 16K21723,
529 16H06437 to H.O.), the Collaborative Research Program of the Institute for Chemical
530 Research, Kyoto University (2019-29 to S.C.), the Future Development Funding Program
531 of Kyoto University Research Coordination Alliance (to R.B.M.), ICR-KU International
532 Short-term Exchange Program for Young Researchers (to S.C.). Computational time was
533 provided by the SuperComputer System, Institute for Chemical Research, Kyoto
534 University. We thank the *Tara* Oceans consortium, the projects Oceanomics and France
535 Genomique (grants ANR-11-BTBR-0008 and ANR-10-INBS-09), people and sponsors
536 who supported the *Tara* Oceans Expedition (http://www.embl.de/tara-oceans/) for
537 making the data accessible. This is contribution number XXX of the *Tara* Oceans
538 Expedition 2009–2013.

539 References

540 Agusti S, González-Gordillo JI, Vaqué D, Estrada M, Cerezo MI, Salazar G, Gasol JM,
541 Duarte CM. 2015. Ubiquitous healthy diatoms in the deep sea confirm deep
542 carbon injection by the biological pump. Nat. Commun. 6:7608.

543 Alberti A, Poulain J, Engelen S, Labadie K, Romac S, Ferrera I, Albini G, Aury J-M,
544 Belser C, Bertrand A, et al. 2017. Viral to metazoan marine plankton
545 nucleotide sequences from the Tara Oceans expedition. Sci. Data [Internet] 4.
546 Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5538240/

547 Allen LZ, McCrow JP, Ininbergs K, Dupont CL, Badger JH, Hoffman JM, Ekman M,
548 Allen AE, Bergman B, Venter JC. 2017. The Baltic Sea Virome: Diversity and
549 Transcriptional Activity of DNA and RNA Viruses. mSystems 2:e00125-16.

550 Behrenfeld MJ, Falkowski PG. 1997. Photosynthetic rates derived from satellite-
551 based chlorophyll concentration. Limnol. Oceanogr. 42:1–20.

552 Boyd P, Newton P. 1995. Evidence of the potential influence of planktonic
553 community structure on the interannual variability of particulate organic
554 carbon flux. Deep Sea Res. Part Oceanogr. Res. Pap. 42:619–639.

555 Brum JR, Ignacio-Espinoza JC, Roux S, Doulcier G, Acinas SG, Alberti A, Chaffron S,
556 Cruaud C, Vargas C de, Gasol JM, et al. 2015. Patterns and ecological drivers of
557 ocean viral communities. Science 348:1261498.

558 Brussaard CPD, Bratbak G, Baudoux A-C, Ruardij P. 2007. Phaeocystis and its
559 interaction with viruses. Biogeochemistry 83:201–215.

560 Brussaard CPD, Wilhelm SW, Thingstad F, Weinbauer MG, Bratbak G, Heldal M,
561 Kimmance SA, Middelboe M, Nagasaki K, Paul JH, et al. 2008. Global-scale

562 processes with a nanoscale drive: the role of marine viruses. ISME J. 2:575–
563 578.

564 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for
565 automated alignment trimming in large-scale phylogenetic analyses.
566 Bioinforma. Oxf. Engl. 25:1972–1973.

567 Carradec Q, Pelletier E, Silva CD, Alberti A, Seeleuthner Y, Blanc-Mathieu R, Lima-
568 Mendez G, Rocha F, Tirichine L, Labadie K, et al. 2018. A global ocean atlas of
569 eukaryotic genes. Nat. Commun. 9:373.

570 Chong I-G, Jun C-H. 2005. Performance of some variable selection methods when
571 multicollinearity is present. Chemom. Intell. Lab. Syst. 78:103–112.

572 Clasen JL, Suttle CA. 2009. Identification of freshwater Phycodnaviridae and their
573 potential phytoplankton hosts, using DNA pol sequence fragments and a
574 genetic-distance analysis. Appl. Environ. Microbiol. 75:991–997.

575 Coenen AR, Weitz JS. 2018. Limitations of Correlation-Based Inference in Complex
576 Virus-Microbe Communities. mSystems 3:e00084-18.

577 Culley A. 2018. New insight into the RNA aquatic virosphere via viromics. Virus Res.
578 244:84–89.

579 De La Rocha CL, Passow U. 2007. Factors influencing the sinking of POC and the
580 efficiency of the biological carbon pump. Deep Sea Res. Part II Top. Stud.
581 Oceanogr. 54:639–658.

582 Eddy SR. 1998. Profile hidden Markov models. Bioinformatics 14:755–763.

583 Endo H, Ogata H, Suzuki K. 2018. Contrasting biogeography and diversity patterns
584 between diatoms and haptophytes in the central Pacific Ocean. Sci. Rep.
585 8:10916.

586 Falkowski PG, Barber RT, Smetacek V. 1998. Biogeochemical Controls and
587 Feedbacks on Ocean Primary Production. Science 281:200–206.

588 Fawcett SE, Lomas MW, Casey JR, Ward BB, Sigman DM. 2011. Assimilation of
589 upwelled nitrate by small eukaryotes in the Sargasso Sea. Nat. Geosci. 4:717–
590 722.

591 Francois R, Honjo S, Krishfield R, Manganini S. 2002. Factors controlling the flux of
592 organic carbon to the bathypelagic zone of the ocean. Glob. Biogeochem.
593 Cycles 16:34-1-34–20.

594 Gobler CJ, Hutchins DA, Fisher NS, Cosper EM, Saňudo‐Wilhelmy SA. 1997. Release
595 and bioavailability of C, N, P Se, and Fe following viral lysis of a marine
596 chrysophyte. Limnol. Oceanogr. 42:1492–1504.

597 Guidi L, Chaffron S, Bittner L, Eveillard D, Larhlimi A, Roux S, Darzi Y, Audic S,
598 Berline L, Brum JR, et al. 2016. Plankton networks driving carbon export in
599 the oligotrophic ocean. Nature 532:465.

600 Guidi L, Jackson GA, Stemmann L, Miquel JC, Picheral M, Gorsky G. 2008.
601 Relationship between particle size distribution and flux in the mesopelagic
602 zone. Deep Sea Res. Part Oceanogr. Res. Pap. 55:1364–1374.

603 Guidi L, Stemmann L, Jackson GA, Ibanez F, Claustre H, Legendre L, Picheral M,
604 Gorskya G. 2009. Effects of phytoplankton community on production, size,
605 and export of large aggregates: A world-ocean analysis. Limnol. Oceanogr.
606 54:1951–1963.

607 Herndl GJ, Reinthaler T. 2013. Microbial control of the dark end of the biological
608 pump. Nat. Geosci. 6:718–724.

609 Hingamp P, Grimsley N, Acinas SG, Clerissi C, Subirana L, Poulain J, Ferrera I,
610 Sarmento H, Villar E, Lima-Mendez G, et al. 2013. Exploring nucleo-
611 cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. ISME
612 J. 7:1678–1695.

613 Huerta-Cepas J, Serra F, Bork P. 2016. ETE 3: Reconstruction, analysis and
614 visualization of phylogenomic data. Mol. Biol. Evol.:msw046.

615 Hurwitz BL, Brum JR, Sullivan MB. 2015. Depth-stratified functional and taxonomic
616 niche specialization in the “core” and “flexible” Pacific Ocean Virome. ISME J.
617 9:472–484.

618 Karsenti E, Acinas SG, Bork P, Bowler C, Vargas CD, Raes J, Sullivan M, Arendt D,
619 Benzoni F, Claverie J-M, et al. 2011. A Holistic Approach to Marine Eco-
620 Systems Biology. PLOS Biol. 9:e1001177.

621 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version
622 7: improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

623 Laber CP, Hunter JE, Carvalho F, Collins JR, Hunter EJ, Schieler BM, Boss E, More K,
624 Frada M, Thamatrakoln K, et al. 2018. Coccolithovirus facilitation of carbon
625 export in the North Atlantic. Nat. Microbiol. 3:537–547.

626 Lawrence JE, Suttle CA. 2004. Effect of viral infection on sinking rates of
627 Heterosigma akashiwo and its implications for bloom termination. Aquat.
628 Microb. Ecol. 37:1–7.

629 Leblanc K, Quéguiner B, Diaz F, Cornet V, Michel-Rodriguez M, Madron XD de,
630 Bowler C, Malviya S, Thyssen M, Grégori G, et al. 2018. Nanoplanktonic
631 diatoms are globally overlooked but play a role in spring blooms and carbon
632 export. Nat. Commun. 9:953.

633 Li W. 1995. Composition of Ultraphytoplankton in the Central North-Atlantic. Mar.
634 Ecol. Prog. Ser. 122:1–8.

635 Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou S, Frada M, Not F, de Vargas C. 2009.
636 Extreme diversity in noncalcifying haptophytes explains a major pigment
637 paradox in open oceans. Proc. Natl. Acad. Sci. U. S. A. 106:12803–12808.

638 Lomas MW, Moran SB. 2011. Evidence for aggregation and export of cyanobacteria
639 and nano-eukaryotes from the Sargasso Sea euphotic zone. Biogeosciences
640 8:203–216.

641 Lønborg C, Middelboe M, Brussaard CPD. 2013. Viral lysis of Micromonas pusilla:
642 impacts on dissolved organic matter production and composition.
643 Biogeochemistry 116:231–240.

644 Marchler-Bauer A, Derbyshire MK, Gonzales NR, Lu S, Chitsaz F, Geer LY, Geer RC, He
645 J, Gwadz M, Hurwitz DI, et al. 2015. CDD: NCBI's conserved domain database.
646 Nucleic Acids Res. 43:D222–D226.

647 Mars Brisbin M, Mesrop LY, Grossmann MM, Mitarai S. 2018. Intra-host Symbiont
648 Diversity and Extended Symbiont Maintenance in Photosymbiotic
649 Acantharea (Clade F). Front. Microbiol. [Internet] 9. Available from:
650 https://www.frontiersin.org/articles/10.3389/fmicb.2018.01998/full

651 Mihara T, Koyano H, Hingamp P, Grimsley N, Goto S, Ogata H. 2018. Taxon Richness
652 of "Megaviridae" Exceeds those of Bacteria and Archaea in the Ocean.
653 Microbes Environ. 33:162–171.

654 Mihara T, Nishimura Y, Shimizu Y, Nishiyama H, Yoshikawa G, Uehara H, Hingamp P,
655 Goto S, Ogata H. 2016. Linking Virus Genomes with Host Taxonomy. Viruses
656 8:66.

657 Moniruzzaman M, Wurch LL, Alexander H, Dyhrman ST, Gobler CJ, Wilhelm SW.
658 2017. Virus-host relationships of marine single-celled eukaryotes resolved
659 from metatranscriptomics. Nat. Commun. 8:16054.

660 Motegi C, Nagata T, Miki T, Weinbauer MG, Legendre L, Rassoulzadegand F. 2009.
661 Viral control of bacterial growth efficiency in marine pelagic environments.
662 Limnol. Oceanogr. 54:1901–1910.

663 Nelson DM, Tréguer P, Brzezinski MA, Leynaert A, Quéguiner B. 1995. Production
664 and dissolution of biogenic silica in the ocean: Revised global estimates,
665 comparison with regional data and relationship to biogenic sedimentation.
666 Glob. Biogeochem. Cycles 9:359–372.

667 Not F, Siano R, Kooistra WHCF, Simon N, Vaulot D, Probert I. 2012. Chapter One -
668 Diversity and Ecology of Eukaryotic Marine Phytoplankton. In: Piganeau G,
669 editor. Advances in Botanical Research. Vol. 64. Genomic Insights into the
670 Biology of Algae. Academic Press. p. 1–53. Available from:
671 http://www.sciencedirect.com/science/article/pii/B9780123914996000013
672

673 Peduzzi P, Weinbauer MG. 1993. Effect of concentrating the virus-rich 2-2nm size
674 fraction of seawater on the formation of algal flocs (marine snow). Limnol.
675 Oceanogr. 38:1562–1565.

Pesant S, Not F, Picheral M, Kandels-Lewis S, Bescot NL, Gorsky G, Iudicone D, Karsenti E, Speich S, Troublé R, et al. 2015. Open science resources for the discovery and analysis of *Tara* Oceans data. Sci. Data 2:150023.

Picheral M, Guidi L, Stemmann L, Karl DM, Iddaoud G, Gorsky G. 2010. The Underwater Vision Profiler 5: An advanced instrument for high spatial resolution studies of particle size spectra and zooplankton. Limnol. Oceanogr. Methods 8:462–473.

Picheral M, Searson S, Taillandier V, Bricaud A, Boss E, Stemmann L, Gorsky G, Tara Oceans Consortium C, Tara Oceans Expedition P. 2014. Vertical profiles of environmental parameters measured from physical, optical and imaging sensors during Tara Oceans expedition 2009-2013. Available from: https://doi.pangaea.de/10.1594/PANGAEA.836321

Price MN, Dehal PS, Arkin AP. 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Mol. Biol. Evol. 26:1641–1650.

Proctor LM, Fuhrman JA. 1991. Roles of viral infection in organic particle flux. Mar. Ecol. Prog. Ser. 69:133–142.

Quéré CL, Andrew RM, Friedlingstein P, Sitch S, Hauck J, Pongratz J, Pickers PA, Korsbakken JI, Peters GP, Canadell JG, et al. 2018. Global Carbon Budget 2018. Earth Syst. Sci. Data 10:2141–2194.

696 Richardson TL, Jackson GA. 2007. Small Phytoplankton and Carbon Export from the
697 Surface Ocean. Science 315:838–840.

698 Sarmiento JL. 2013. Ocean Biogeochemical Dynamics. Princeton University Press

699 Sheyn U, Rosenwasser S, Lehahn Y, Barak-Gavish N, Rotkopf R, Bidle KD, Koren I,
700 Schatz D, Vardi A. 2018. Expression profiling of host and virus during a
701 coccolithophore bloom provides insights into the role of viral infection in
702 promoting carbon export. ISME J.:1.

703 Shibata A, Kogure K, Koike I, Ohwada K. 1997. Formation of submicron colloidal
704 particles from marine bacteria by viral infection. Mar. Ecol. Prog. Ser.
705 155:303–307.

706 Steinberg DK, Mooy BASV, Buesseler KO, Boyd PW, Kobari T, Karl DM. 2008.
707 Bacterial vs. zooplankton control of sinking particle flux in the ocean's
708 twilight zone. Limnol. Oceanogr. 53:1327–1338.

709 Sullivan MB, Weitz JS, Wilhelm S. 2017. Viral ecology comes of age. Environ.
710 Microbiol. Rep. 9:33–35.

711 Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B,
712 Zeller G, Mende DR, Alberti A, et al. 2015. Ocean plankton. Structure and
713 function of the global ocean microbiome. Science 348:1261359.

714 Suttle CA. 2007. Marine viruses--major players in the global ecosystem. Nat. Rev.
715 Microbiol. 5:801–812.

716 Tackmann J, Rodrigues JFM, Mering C von. 2018. Rapid inference of direct
717 interactions in large-scale ecological networks from heterogeneous microbial
718 sequencing data. bioRxiv:390195.

719 Tréguer P, Bowler C, Moriceau B, Dutkiewicz S, Gehlen M, Aumont O, Bittner L,
720 Dugdale R, Finkel Z, Iudicone D, et al. 2018. Influence of diatom diversity on
721 the ocean biological carbon pump. Nat. Geosci. 11:27–37.

722 Turner JT. 2015. Zooplankton fecal pellets, marine snow, phytodetritus and the
723 ocean's biological pump. Prog. Oceanogr. 130:205–248.

724 Unrein F, Gasol JM, Not F, Forn I, Massana R. 2014. Mixotrophic haptophytes are key
725 bacterial grazers in oligotrophic coastal waters. ISME J. 8:164–176.

726 Van Etten JL, Agarkova I, Dunigan DD, Tonetti M, De Castro C, Duncan GA. 2017.
727 Chloroviruses Have a Sweet Tooth. Viruses [Internet] 9. Available from:
728 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5408694/

729 de Vargas C, Audic S, Henry N, Decelle J, Mahé F, Logares R, Lara E, Berney C, Le
730 Bescot N, Probert I, et al. 2015. Ocean plankton. Eukaryotic plankton
731 diversity in the sunlit ocean. Science 348:1261605.

732 Ward BA, Follows MJ. 2016. Marine mixotrophy increases trophic transfer efficiency,
733 mean organism size, and vertical carbon flux. Proc. Natl. Acad. Sci. 113:2958–
734 2963.

735 Weinbauer MG. 2004. Ecology of prokaryotic viruses. FEMS Microbiol. Rev. 28:127–
736 181.

737 Weinbauer MG, Peduzzi P. 1995. Effect of virus-rich high molecular weight
738 concentrates of seawater on the dynamics of dissolved amino acids and
739 carbohydrates. Mar. Ecol. Prog. Ser. 127:245–253.

740 Weitz JS, Stock CA, Wilhelm SW, Bourouiba L, Coleman ML, Buchan A, Follows MJ,
741 Fuhrman JA, Jover LF, Lennon JT, et al. 2015. A multitrophic model to
742 quantify the effects of marine viruses on microbial food webs and ecosystem
743 processes. ISME J. 9:1352–1364.

744 Weynberg KD, Allen MJ, Wilson WH. 2017. Marine Prasinoviruses and Their Tiny
745 Plankton Hosts: A Review. Viruses 9.

746 Wilhelm SW, Suttle CA. 1999. Viruses and Nutrient Cycles in the SeaViruses play
747 critical roles in the structure and function of aquatic food webs. BioScience
748 49:781–788.

749 Worden AZ, Nolan JK, Palenik B. 2004. Assessing the dynamics and ecology of
750 marine picophytoplankton: The importance of the eukaryotic component.
751 Limnol. Oceanogr. 49:168–179.

752 Yamada Y, Tomaru Y, Fukuda H, Nagata T. 2018. Aggregate Formation During the
753 Viral Lysis of a Marine Diatom. Front. Mar. Sci. [Internet] 5. Available from:
754 https://www.frontiersin.org/articles/10.3389/fmars.2018.00167/full

755 Yutin N, Wolf YI, Raoult D, Koonin EV. 2009. Eukaryotic large nucleo-cytoplasmic
756 DNA viruses: Clusters of orthologous genes and reconstruction of viral
757 genome evolution. Virol. J. 6:223.

758 Zhang C, Dang H, Azam F, Benner R, Legendre L, Passow U, Polimene L, Robinson C,
759 Suttle CA, Jiao N. 2018. Evolving paradigms in biological carbon cycling in the
760 ocean. Natl. Sci. Rev. 5:481–499.

761 Supplementary material

762 • Supplement.docx: supplementary figures and tables.
763 • Supplementary_file_1: Information on the predictors (viruses) in the PLS model
764 predicting carbon export efficiency. Predictors are sorted by VIP score.
765 • Supplementary data available at GenomeNet FTP:
766 ftp://ftp.genome.jp/pub/db/community/tara/Cpump/