



HAL
open science

Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean

Hiroto Kaneko, Romain Blanc-Mathieu, Hisashi Endo, Samuel Chaffron, Tom Delmont, Morgan Gaia, Nicolas Henry, Rodrigo Hernández-Velázquez, Canh Hao Nguyen, Hiroshi Mamitsuka, et al.

► **To cite this version:**

Hiroto Kaneko, Romain Blanc-Mathieu, Hisashi Endo, Samuel Chaffron, Tom Delmont, et al.. Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean. *iScience*, 2021, 24 (1), pp.102002. 10.1016/j.isci.2020.102002 . hal-03097258v2

HAL Id: hal-03097258

<https://hal.science/hal-03097258v2>

Submitted on 5 Jan 2021 (v2), last revised 10 Nov 2021 (v3)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal Pre-proof



Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean

Hiroto Kaneko, Romain Blanc-Mathieu, Hisashi Endo, Samuel Chaffron, Tom O. Delmont, Morgan Gaia, Nicolas Henry, Rodrigo Hernández-Velázquez, Canh Hao Nguyen, Hiroshi Mamitsuka, Patrick Forterre, Olivier Jaillon, Colomban de Vargas, Matthew B. Sullivan, Curtis A. Suttle, Lionel Guidi, Hiroyuki Ogata

PII: S2589-0042(20)31199-8

DOI: <https://doi.org/10.1016/j.isci.2020.102002>

Reference: ISCI 102002

To appear in: *ISCIENCE*

Received Date: 23 July 2020

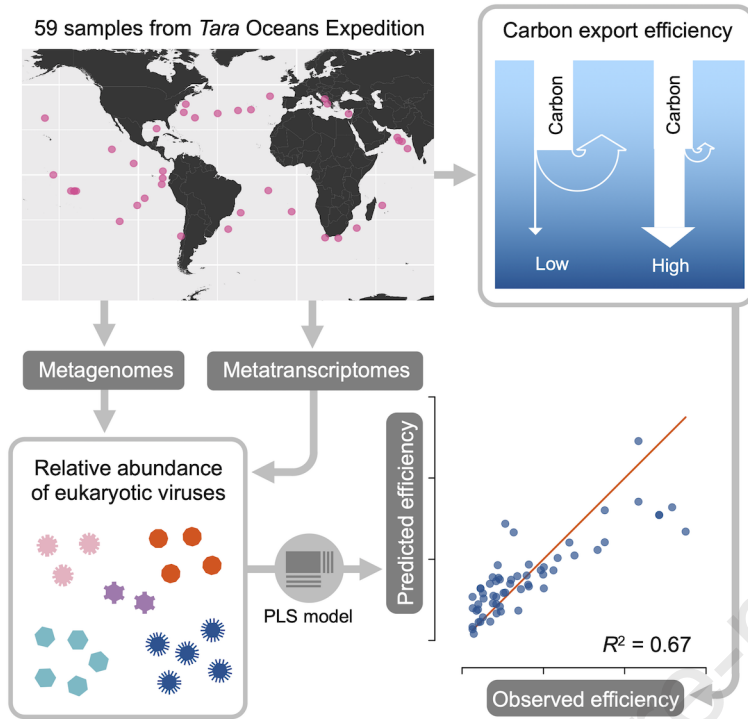
Revised Date: 13 November 2020

Accepted Date: 23 December 2020

Please cite this article as: Kaneko, H., Blanc-Mathieu, R., Endo, H., Chaffron, S., Delmont, T.O., Gaia, M., Henry, N., Hernández-Velázquez, R., Nguyen, C.H., Mamitsuka, H., Forterre, P., Jaillon, O., de Vargas, C., Sullivan, M.B., Suttle, C.A., Guidi, L., Ogata, H., Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean, *ISCIENCE* (2021), doi: <https://doi.org/10.1016/j.isci.2020.102002>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 The Author(s).



1 **Eukaryotic virus composition can predict the efficiency of carbon export in the**
2 **global ocean**

3 Hiroto Kaneko^{1,+}, Romain Blanc-Mathieu^{1,2,+}, Hisashi Endo¹, Samuel Chaffron^{3,4},
4 Tom O. Delmont^{4,5}, Morgan Gaia^{4,5}, Nicolas Henry⁶, Rodrigo Hernández-Velázquez¹,
5 Canh Hao Nguyen¹, Hiroshi Mamitsuka¹, Patrick Forterre⁷, Olivier Jaillon^{4,5},
6 Colomban de Vargas⁶, Matthew B. Sullivan⁸, Curtis A. Suttle⁹, Lionel Guidi¹⁰ and
7 Hiroyuki Ogata^{1,11,*}

8 + Equal contribution

9 * Correspondence: ogata@kuicr.kyoto-u.ac.jp

10 **Affiliations:**

11 1: Bioinformatics Center, Institute for Chemical Research, Kyoto University,
12 Gokasho, Uji, Kyoto 611-0011, Japan

13 2: Laboratoire de Physiologie Cellulaire & Végétale, CEA, Univ. Grenoble Alpes,
14 CNRS, INRA, IRIG, Grenoble, France.

15 3: Université de Nantes, CNRS UMR 6004, LS2N, F-44000 Nantes, France.

16 4: Research Federation (FR2022) *Tara* Oceans GO-SEE, Paris, France

17 5: Génomique Métabolique, Genoscope, Institut François Jacob, CEA, CNRS, 91000
18 Evry, France

19 6: Sorbonne Universités, CNRS, Laboratoire Adaptation et Diversité en Milieu Marin,
20 Station Biologique de Roscoff, 29680 Roscoff, France

21 7: Institut Pasteur, Department of Microbiology, 25 rue du Docteur Roux, 75015,
22 Paris, France

- 23 8: Department of Microbiology and Department of Civil, Environmental and Geodetic
24 Engineering, Ohio State University, Columbus, OH, United States of America
25 9: Departments of Earth, Ocean & Atmospheric Sciences, Microbiology &
26 Immunology, and Botany, and the Institute for the Oceans and Fisheries, University
27 of British Columbia, Vancouver, BC, V6T 1Z4, Canada
28 10: Sorbonne Université, CNRS, Laboratoire d'Océanographie de Villefranche, LOV,
29 F-06230 Villefranche-sur-mer, France
30 11: Lead contact
31

Journal Pre-proof

32 **Summary**

33 The biological carbon pump, in which carbon fixed by photosynthesis is exported to
34 the deep ocean through sinking, is a major process in Earth's carbon cycle. The
35 proportion of primary production that is exported is termed the carbon export
36 efficiency (CEE). Based on in-lab or regional scale observations, viruses were
37 previously suggested to affect the CEE (i.e., viral "shunt" and "shuttle"). In this study,
38 we tested associations between viral community composition and CEE measured at a
39 global scale. A regression model based on relative abundance of viral marker genes
40 explained 67% of the variation in CEE. Viruses with high importance in the model
41 were predicted to infect ecologically important hosts. These results are consistent with
42 the view that the viral shunt and shuttle functions at a large scale and further imply
43 that viruses likely act in this process in a way dependent on their hosts and ecosystem
44 dynamics.

45 **Introduction**

46 A major process in the global cycling of carbon is the oceanic biological carbon pump
47 (BCP), an organism-driven process by which atmospheric carbon (*i.e.*, CO₂) is
48 transferred and sequestered to the ocean interior and seafloor for periods ranging from
49 centuries to hundreds of millions of years. Between 15% and 20% of net primary
50 production (NPP) is exported out of the euphotic zone, with 0.3% of fixed carbon
51 reaching the seafloor annually (Zhang et al., 2018). However, there is wide variation
52 in estimates of the proportion of primary production in the surface ocean that is
53 exported to depth, ranging from 1% in the tropical Pacific to 35-45% during the North
54 Atlantic bloom (Buesseler and Boyd, 2009). As outlined below, many factors affect
55 the BCP.

56 Of planktonic organisms living in the upper layer of the ocean, diatoms
57 (Tréguer et al., 2018) and zooplankton (Turner, 2015) have been identified as
58 important contributors to the BCP in nutrient-replete oceanic regions. In the
59 oligotrophic ocean, cyanobacteria, collodarians (Lomas and Moran, 2011), diatoms
60 (Agusti et al., 2015; Karl et al., 2012; Leblanc et al., 2018), and other small (pico- to
61 nano-) plankton (Lomas and Moran, 2011) have been implicated in the BCP.

62 Sediment trap studies suggest that ballasted aggregates of plankton with biogenic
63 minerals contribute to carbon export to the deep sea (Iversen and Ploug, 2010; Klaas
64 and Archer, 2002). The BCP comprises three processes: carbon fixation, export, and
65 remineralization. As these processes are governed by complex interactions between
66 numerous members of planktonic communities (Zhang et al., 2018), the BCP is
67 expected to involve various organisms, including viruses (Zimmerman et al., 2019).

68 Viruses have been suggested to regulate the efficiency of the BCP. Lysis of
69 host cells by viruses releases cellular material in the form of dissolved organic matter

70 (DOM), which fuels the microbial loop and enhances respiration and secondary
71 production (Gobler et al., 1997; Weitz et al., 2015). This process, coined “viral shunt
72 (Wilhelm and Suttle, 1999)”, can reduce the carbon export efficiency (CEE) because
73 it increases the retention of nutrients and carbon in the euphotic zone and prevents
74 their transfer to higher trophic levels as well as their export from the euphotic zone to
75 the deep sea (Fuhrman, 1999; Weitz et al., 2015). However, an alternative process is
76 also considered, in which viruses contribute to the vertical carbon export (Weinbauer,
77 2004). For instance, a theoretical study proposed that the CEE increases if viral lysis
78 augments the ratio of exported carbon relative to the primary production-limiting
79 nutrients (nitrogen and phosphorous) (Suttle, 2007). Laboratory experimental studies
80 reported that cells infected with viruses form larger particles (Peduzzi and Weinbauer,
81 1993; Yamada et al., 2018), can sink faster (Lawrence and Suttle, 2004), and can lead
82 to preferential grazing by heterotrophic protists (Evans and Wilson, 2008) and/or to
83 higher growth of grazers (Goode et al., 2019). This process termed “viral shuttle
84 (Sullivan et al., 2017)” is supported by several field studies that reported association
85 of viruses with sinking material. Viruses were observed in sinking material in the
86 North Atlantic Ocean (Proctor and Fuhrman, 1991) and sediment of coastal waters
87 where algal blooms occur (Lawrence et al., 2002; Tomaru et al., 2007, 2011). In
88 addition, vertical transport of bacterial viruses between photic and aphotic zones was
89 observed in the Pacific Ocean (Hurwitz et al., 2015) and in *Tara* Oceans virome data
90 (Brum et al., 2015). A systematic analysis of large-scale omics data from oligotrophic
91 oceanic regions revealed a positive association between the magnitude of carbon flux
92 and bacterial dsDNA viruses (*i.e.*, cyanophages), which were previously unrecognized
93 as possible contributors to the BCP (Guidi et al., 2016).

94 More recently, viral infection of blooms of the photosynthetic eukaryote
95 *Emiliana huxleyi* in the North Atlantic were found to be accompanied by particle
96 aggregation and greater downward vertical flux of carbon, with the highest export
97 during the early stage of viral infection (Laber et al., 2018; Sheyn et al., 2018). Given
98 the significant contributions of eukaryotic plankton to ocean biomass and net
99 production (Hirata et al., 2011; Li, 1995) and their observed predominance over
100 prokaryotes in sinking materials of Sargasso Sea oligotrophic surface waters (Fawcett
101 et al., 2011; Lomas and Moran, 2011), various lineages of eukaryotic viruses may be
102 responsible for a substantial part of the variation in carbon export across oceanic
103 regions.

104 If the “viral shunt” and “shuttle” processes function at a global scale and if
105 these involve specific eukaryotic viruses, we expect to detect a statistical association
106 between eukaryotic viral community composition and CEE in a large-scale omics data.
107 To our knowledge, such an association has never been investigated. Although this test
108 per se does not prove that viruses regulate CEE, we consider the association is worth
109 being tested because such an association is a necessary condition for the global model
110 of viral shunt and shuttle and, under its absence, we would have to reconsider the
111 model. Deep sequencing of planktonic community DNA and RNA, as carried out in
112 *Tara* Oceans, has enabled the identification of marker genes of major viral groups
113 infecting eukaryotes (Hingamp et al., 2013; Carradec et al., 2018; Culley, 2018; Endo
114 et al., 2020). To examine the association between viral community composition and
115 CEE, we thus used the comprehensive organismal dataset from the *Tara* Oceans
116 expedition (Carradec et al., 2018; Sunagawa et al., 2015), as well as related
117 measurements of carbon export estimated from particle concentrations and size
118 distributions observed *in situ* (Guidi et al., 2016).

119 In the present study, we identified several hundred marker-gene sequences of
120 nucleocytoplasmic large DNA viruses (NCLDV) in metagenomes of 0.2–3 μm size
121 fraction. We also identified RNA and ssDNA viruses in metatranscriptomes of four
122 eukaryotic size fractions spanning 0.8 to 2,000 μm . The resulting profiles of viral
123 distributions were compared with an image-based measure of carbon export efficiency
124 (CEE), which is defined as the ratio of the carbon flux at depth to the carbon flux at
125 surface.

126 **Results and Discussion**

127 **Detection of diverse eukaryotic viruses in *Tara* Oceans gene catalogs**

128 We used profile hidden Markov model-based homology searches to identify marker-
129 gene sequences of eukaryotic viruses in two ocean gene catalogs. These catalogs were
130 previously constructed from environmental shotgun sequence data of samples
131 collected during the *Tara* Oceans expedition. The first catalog, the Ocean Microbial
132 Reference Gene Catalog (OM-RGC), contains 40 million non-redundant genes
133 predicted from the assemblies of *Tara* Oceans viral and microbial metagenomes
134 (Sunagawa et al., 2015). We searched this catalog for NCLDV DNA polymerase
135 family B (PolB) genes, as dsDNA viruses may be present in microbial metagenomes
136 because large virions ($> 0.2 \mu\text{m}$) have been retained on the filter or because viral
137 genomes actively replicating or latent within picoeukaryotic cells have been captured.
138 The second gene catalog, the Marine Atlas of *Tara* Oceans Unigenes (MATOU),
139 contains 116 million non-redundant genes derived from metatranscriptomes of single-
140 cell microeukaryotes and small multicellular zooplankton (Carradec et al., 2018). We
141 searched this catalog for NCLDV PolB genes, RNA-dependent RNA polymerase

142 (RdRP) genes of RNA viruses, and replication-associated protein (Rep) genes of
143 ssDNA viruses, since transcripts of viruses actively infecting their hosts, as well as
144 genomes of RNA viruses, have been captured in this catalog.

145 We identified 3,874 NCLDV PolB sequences (3,486 in metagenomes and 388
146 in metatranscriptomes), 975 RNA virus RdRP sequences, and 299 ssDNA virus Rep
147 sequences (Table 1). These sequences correspond to operational taxonomic units
148 (OTUs) at a 95% identity threshold. All except 17 of the NCLDV PolBs from
149 metagenomes were assigned to the families *Mimiviridae* ($n = 2,923$),
150 *Phycodnaviridae* ($n = 348$), and *Iridoviridae* ($n = 198$) (Table 1). The larger numbers
151 of PolB sequences assigned to *Mimiviridae* and *Phycodnaviridae* compared with other
152 NCLDV families are consistent with previous observations (Endo et al., 2020;
153 Hingamp et al., 2013). The divergence between these environmental sequences and
154 reference sequences from known viral genomes was greater in *Mimiviridae* than in
155 *Phycodnaviridae* (Figure 1A, S1A and S2). Within *Mimiviridae*, 83% of the
156 sequences were most similar to those from algae-infecting *Mimivirus* relatives.
157 Among the sequences classified in *Phycodnaviridae*, 93% were most similar to those
158 in *Prasinovirus*, whereas 6% were closest to *Yellowstone lake phycodnavirus*, which
159 is closely related to *Prasinovirus*. Prasinoviruses are possibly over-represented in the
160 metagenomes because the 0.2 to 3 μm size fraction selects their picoeukaryotic hosts.
161 RdRP sequences were assigned mostly to the order *Picornavirales* ($n = 325$),
162 followed by the families *Partitiviridae* ($n = 131$), *Narnaviridae* ($n = 95$),
163 *Tombusviridae* ($n = 45$), and *Virgaviridae* ($n = 33$) (Table 1), with most sequences
164 being distant (30% to 40% amino acid identity) from reference viruses (Figures 1B,
165 S1B and S3). These results are consistent with previous studies on the diversity of
166 marine RNA viruses, in which RNA virus sequences were found to correspond to

167 diverse positive-polarity ssRNA and dsRNA viruses distantly related to well-
168 characterized viruses (Culley, 2018). *Picornavirales* may be over-represented in the
169 metatranscriptomes because of the polyadenylated RNA selection. The majority ($n =$
170 201) of Rep sequences were annotated as *Circoviridae*, known to infect animals,
171 which is consistent with a previous report (Wang et al., 2018). Only eight were
172 annotated as plant ssDNA viruses (families *Nanoviridae* and *Gemnaviridae*) (Table 1).
173 Most of these environmental sequences are distant (40% to 50% amino acid identity)
174 from reference sequences (Figures 1C, S1C and S4). Additional 388 NCLDV PolBs
175 were detected in the metatranscriptomes. The average cosmopolitanism (number of
176 samples where an OTU was observed by at least two reads) for PolBs in
177 metagenomes was 23 samples against 2.9 for metatranscriptome-derived PolB
178 sequences, 5.5 for Repls, and 5.8 for RdRPs. Within metatranscriptomes, the average
179 gene-length normalized read counts for PolBs were respectively ten and three times
180 lower than those of RdRPs and Repls. Therefore, PolBs from metatranscriptomes were
181 not further used in our study.

182 **Composition of eukaryotic viruses can explain the variation of carbon** 183 **export efficiency**

184 Among the PolB, RdRP, and Rep sequences identified in the *Tara* Oceans gene
185 catalogs, 38%, 18%, and 11% (total = 1,523 sequences), respectively, were present in
186 at least five samples and had matching carbon export measurement data (Table 1). We
187 used the relative abundance (defined as the centered log-ratio transformed gene-length
188 normalized read count) profiles of these 1,523 marker-gene sequences at 59 sampling
189 sites in the photic zone of 39 *Tara* Oceans stations (Figure 2) to test for association
190 between their composition and a measure of carbon export efficiency (CEE, see
191 Transparent Methods, Figure S5). A partial least squares (PLS) regression model

192 explained 67% (coefficient of determination $R^2 = 67\%$) of the variation in CEE with a
193 Pearson correlation coefficient of 0.84 between observed and predicted values. This
194 correlation was confirmed to be statistically significant by permutation test ($P < 1 \times$
195 10^{-4}) (Figure 3A).

196 We also tested for their association with estimates of carbon export flux at 150
197 meters (CE_{150}) and NPP. PLS regressions explained 54% and 64% of the variation in
198 CE_{150} and NPP with Pearson correlation coefficients between observed and predicted
199 values of 0.74 (permutation test, $P < 1 \times 10^{-4}$) and 0.80 (permutation test, $P < 1 \times$
200 10^{-4}), respectively (Figure S6). In these three PLS regression models, 83, 86, and 97
201 viruses were considered to be key predictors (*i.e.*, Variable Importance in the
202 Projection [VIP] score > 2) of CEE, CE_{150} , and NPP, respectively. PLS models for
203 NPP and CE_{150} shared a larger number of predictors (52 viruses) compared to the PLS
204 models for NPP and CEE (seven viruses) (two proportion Z-test, $P = 4.14 \times 10^{-12}$).
205 Consistent with this observation, CE_{150} was correlated with NPP (Pearson's $r = 0.77$;
206 parametric test, $P < 1 \times 10^{-12}$). This result implies that the magnitude of export in the
207 analyzed samples was partly constrained by primary productivity. However, CEE was
208 not correlated with NPP ($r = 0.16$; parametric test, $P = 0.2$) or CE_{150} ($r = 0.002$;
209 parametric test, $P = 0.99$). Thus, as expected, primary productivity was not a major
210 driver for the efficiency of carbon export.

211 To assess the sensitivity of the model to the definition of carbon export
212 efficiency, we employed an alternative measure of carbon export efficiency that
213 considers euphotic zone depth (T_{100} , see Transparent Methods). T_{100} was correlated
214 with CEE ($r = 0.66$; parametric test, $P < 1 \times 10^{-8}$) and PLS regression explained 44%
215 of the variation in T_{100} (permutation test, $P < 1 \times 10^{-4}$) (Figure S7). Of 72 predictors

216 of the PLS model for T_{100} , 30 were shared with that for CEE. This result demonstrates
217 the robustness of the PLS model to definitions of carbon export efficiency.

218 The 83 viruses (5% of the viruses included in our analysis) that were
219 associated with CEE with a VIP score > 2 are considered to be important predictors of
220 CEE in the PLS regression (Figure 3B, Supplemental Data 1), and these viruses are
221 hereafter referred to as VIPs (Viruses Important in the Prediction). Fifty-eight VIPs
222 had positive regression coefficient, and 25 had negative regression coefficient in the
223 prediction (Figure 3B). Most of the positively associated VIPs showed high relative
224 abundance in the Mediterranean Sea and in the Indian Ocean where CEE tends to be
225 high compared with other oceanic regions (Figure 4). Among them, 15 (red labels in
226 Figure 4) also had high relative abundance in samples from other oceanic regions,
227 showing that these viruses are associated with CEE at a global scale. In contrast,
228 negatively associated VIPs tend to have higher relative abundance in the Atlantic
229 Ocean and the Southern Pacific Ocean where CEE is comparatively lower. In the
230 following sections, we investigate potential hosts of the VIPs in order to interpret the
231 statistical association between viral community composition and CEE in the light of
232 previous observations in the literature.

233 **Viruses correlated with CEE infect ecologically important hosts**

234 Most of the VIPs (77 of 83) belong to *Mimiviridae* ($n = 34$ with 25 positive
235 VIPs and nine negative VIPs), *Phycodnaviridae* ($n = 24$ with 18 positive VIPs and six
236 negative VIPs), and ssRNA viruses of the order *Picornavirales* ($n = 19$ with 13
237 positive VIPs and six negative VIPs) (Figure 3B, Table S1). All the phycodnavirus
238 VIPs were most closely related to prasinoviruses infecting Mamiellales, with amino
239 acid sequence percent identities to reference sequences ranging between 35% and
240 95%. The six remaining VIPs were two NCLDV of the family *Iridoviridae*

241 negatively associated with CEE, three RNA viruses (two ssRNA viruses of the family
242 *Hepeviridae* negatively associated with CEE and one dsRNA virus of the family
243 *Partitiviridae* positively associated with CEE), and one ssDNA virus of the family
244 *Circoviridae* positively associated with CEE. A proportionally larger number of
245 PolBs were included in the model than RdRP and Rep sequences depending on their
246 representations in the input data. Therefore, the larger number of NCLDV VIPs
247 obtained does not necessarily mean that this group of viruses is more important than
248 others regarding their association with CEE.

249 Host information may help understand the relationship between these VIPs
250 and CEE. We performed genomic context analysis for PolB VIPs and phylogeny-
251 guided network-based host prediction for PolB and RdRP to infer putative virus–host
252 relationships (see Transparent Methods).

253 Taxonomic analysis of genes predicted in 10 metagenome-assembled genomes
254 (MAGs) from the eukaryotic size fractions and 65 genome fragments (contigs)
255 assembled from the prokaryotic size fraction encoding VIP PolBs further confirmed
256 their identity as *Mimiviridae* or *Phycodnaviridae* (Figure S8). The size of MAGs
257 ranged between 30 kbp and 440 kbp with an average of 210 kbp (Table S2). The
258 presence of genes with high sequence similarities to cellular genes in a viral genome
259 is suggestive of a virus–host relationship (Monier et al., 2009; Yoshikawa et al., 2019).
260 Two closely related *Mimiviridae* VIPs, PolB 000079111 (positively associated with
261 CEE) and PolB 000079078 (negatively associated with CEE), were phylogenetically
262 close to the pelagophyte virus *Aureococcus anophagefferens virus* (AaV). One MAG
263 (268 kbp in size) corresponding to PolB 000079111 encoded seven genes showing
264 high similarities to genes from Pelagophyceae, and another MAG (382 kbp in size),
265 corresponding to PolB 000079078, encoded five genes similar to genes from

266 Pelagophyceae. All but one of these 12 genes were encoded on a genome fragment
267 containing genes annotated as viral, including five NCLDV core genes (Supplemental
268 Data 2), excluding the possibility of contamination in these MAGs. Two closely
269 related *Phycodnaviridae* VIPs, PolB 001064263 and 010288541, were positively
270 associated with CEE. Both of these PolBs correspond to a MAG (134 kbp in size)
271 encoding one gene likely derived from Mamiellales. The genomic fragment harboring
272 this cellular gene was found to encode 10 genes annotated as viral (Supplemental
273 Data 2).

274 We conducted a phylogeny-guided, network-based host prediction analysis for
275 *Mimiviridae*, *Phycodnaviridae*, and *Picornavirales* (Figures S9 and S10). Only a
276 subset of the VIPs was included in this analysis because we kept the most reliable
277 sequences (n=44) to obtain a well-resolved tree topology. Within the *Prasinovirus*
278 clade, which contained thirteen VIPs (nine positive and four negative), seven different
279 eukaryotic orders were detected as predicted host groups for 10 nodes in the tree.
280 Mamiellales, the only known host group of prasinoviruses, was detected at eight
281 nodes (five of them had no parent-to-child relationships), whereas the other six
282 eukaryotic orders were found at only one node (or two in the case of Eutreptiales)
283 (Figure S9). The order Mamiellales includes three genera (*Micromonas*, *Ostreococcus*,
284 and *Bathycoccus*), which are bacterial-sized green microalgae common in coastal and
285 oceanic environments and are considered to be influential actors in oceanic systems
286 (Monier et al., 2016). Various prasinoviruses (fourteen with available genome
287 sequences) have been isolated from the three genera.

288 Within the family *Mimiviridae*, which contains fifteen VIPs (10 positive and
289 five negative), twelve different orders were predicted as putative host groups (Figure
290 S9). Collodaria was detected at 15 nodes (two of them had no parent-to-child

291 relationships), and Prymnesiales at six nodes (three of them had no parent-to-child
292 relationships), whereas all other orders were present at a maximum of one node each
293 with no parent-to-child relationships. The nodes enriched for Prymnesiales and
294 Collodaria fell within a monophyletic clade (marked by a red arrow in Figure S9)
295 containing four reference haptophyte viruses infecting Prymnesiales and two
296 reference haptophyte viruses infecting Phaeocystales. Therefore, the environmental
297 PolB sequences in this *Mimiviridae* clade (including five positive VIPs and one
298 negative VIP) are predicted to infect Prymnesiales or related haptophytes. The
299 detection of Collodaria may be the result of indirect associations that reflect a
300 symbiotic relationship with Prymnesiales, as some acantharians, evolutionarily related
301 to the Collodaria, are known to host Prymnesiales species (Mars Brisbin et al., 2018).
302 Known species of Prymnesiales and Phaeocystales have organic scales, except one
303 Prymnesiales species, *Prymnesium neolepis*, which bears siliceous scales (Yoshida et
304 al., 2006). Previous studies revealed the existence of diverse and abundant
305 noncalcifying picohaptophytes in open oceans (Endo et al., 2018; Liu et al., 2009).
306 Clear host prediction was not made for the other nine *Mimiviridae* VIPs shown in the
307 phylogenetic tree. Three VIPs (two positive and one negative) in the tree were
308 relatives of AaV. One negatively associated VIP was a relative of *Cafeteria*
309 *roenbergensis virus* infecting a heterotrophic protist. The five remaining *Mimiviridae*
310 VIPs are very distant from any known *Mimiviridae*.

311 Sixteen *Picornavirales* VIPs (eleven positive and five negative) were included
312 in the phylogeny-guided, network-based host prediction analysis (Figure S10). Nine
313 (seven positive and two negative) were grouped within *Dicistroviridae* (known to
314 infect insects) and may therefore infect marine arthropods such as copepods, the most
315 ubiquitous and abundant mesozooplankton groups involved in carbon export (Turner,

2015). Three other *Picornavirales* VIPs were placed within a clade containing known bacillarnaviruses. Two of them (35179764 and 33049404) were positively associated with CEE and had diatoms of the order Chaetocerotales as a predicted host group. The third one (107558617) was negatively associated with CEE and distant from other bacillarnaviruses, and had no host prediction. Diatoms have been globally observed in the deep sea (Agusti et al., 2015; Leblanc et al., 2018) and identified as important contributors of the biological carbon pump (Tréguer et al., 2018). One positively associated VIP (32150309) was in a clade containing *Aurantiochytrium single-stranded RNA virus* (AsRNAV), infecting a marine fungoid protist thought to be an important decomposer (Takao et al., 2005). The last three *Picornavirales* VIPs (59731273, 49554577, and 36496887) had no predicted host and were too distant from known *Picornavirales* to speculate about their putative host group.

Outside *Picornavirales*, three RNA virus VIPs (two *Hepeviridae*, negatively associated, and one *Partitiviridae*, positively associated) were identified, for which no reliable host inferences were made by sequence similarity. Known *Hepeviridae* infect metazoans, and known *Partitiviridae* infect fungi and plants. The two *Hepeviridae*-like viruses were most closely related to viruses identified in the transcriptomes of mollusks (amino acid identities of 48% for 42335229 and 43% for 77677770) (Shi et al., 2016). The *Partitiviridae*-like VIP (35713768) was most closely related to a fungal virus, *Penicillium stoloniferum virus S* (49% amino acid identity).

One ssDNA virus VIP (38177659) was positively associated with CEE. It was annotated as a *Circoviridae*, although it groups with other environmental sequences as an outgroup of known *Circoviridae*. This VIP was connected with copepod, mollusk, and Collodaria OTUs in the co-occurrence network but no enrichment of predicted host groups was detected for its clade. *Circoviridae*-like viruses are known to infect

341 copepods (Dunlap et al., 2013) and have been reported to associate with mollusks
342 (Dayaram et al., 2015), but none have been reported for Collodaria.

343 Overall, we could infer hosts for 37 VIPs (Tables 2 and S3). Most of the
344 predicted hosts are known to be ecologically important as primary producers
345 (Mamiellales, Prymnesiales, Pelagophyceae, and diatoms) or grazers (copepods). Of
346 these, diatoms and copepods are well known as important contributors to the BCP but
347 others (*i.e.*, Mamiellales, Prymnesiales, Pelagophyceae) have not been recognized as
348 major contributors to the BCP. Our analysis also revealed that positive and negative
349 VIPs are not separated in either the viral or host phylogenies.

350 **Viruses positively correlated with CEE tend to interact with silicified** 351 **organisms**

352 The phylogeny-guided, network-based host prediction analysis correctly predicted
353 known virus–host relationships (for viruses infecting Mamiellales, Prymnesiales, and
354 Chaetocerotales) using our large dataset, despite the reported limitations of these co-
355 occurrence network-based approaches (Coenen and Weitz, 2018). This result
356 prompted us to further exploit the species co-occurrence networks (Table S4) to
357 investigate functional differences between the eukaryotic organisms predicted to
358 interact with positive VIPs, negative VIPs, and viruses less important for prediction of
359 CEE (VIP score < 2) (non-VIPs). For this purpose, we used literature-based
360 functional trait annotations associated with eukaryotic meta-barcodes (see Transparent
361 Methods). Positive VIPs had a greater proportion of connections with silicified
362 eukaryotes ($Q = 0.001$), but not with chloroplast-bearing eukaryotes ($Q = 0.16$) nor
363 calcifying eukaryotes ($Q = 1$), compared to non-VIPs (Table 3). No functional
364 differences were observed between negative VIPs and non-VIPs viruses (Table S5) or
365 positive VIPs (Table S6).

366 Multifarious ways viruses affect the fate of carbon

367 Our analysis revealed that eukaryotic virus composition was able to predict CEE in
368 the global sunlit ocean and 83 out of the 1,523 viruses had a high importance in the
369 predictive model. This association is not a proof that the viruses are the cause of the
370 variation of CEE. Viruses, especially those showing latent/persistent infections (Goic
371 and Saleh, 2012), may be found to be associated with CEE if their host affects CEE
372 regardless of viral infection. Organisms that preferentially grow in marine snow
373 (Bochdansky et al., 2017) may also bring associations between viruses infecting those
374 organisms and CEE. Alternatively, the observed associations between VIPs and CEE
375 may reflect a more direct causal relationship, which we attempt to explore in light of
376 the large body of literature on the mechanisms by which viruses impact the fate of
377 carbon in the oceans.

378 Among the 83 VIPs, 58 were positively associated with CEE. Such a positive
379 association is expected from the “viral shuttle” model, which states that viral activity
380 could facilitate carbon export to the deep ocean (Fuhrman, 1999; Sullivan et al., 2017;
381 Weinbauer, 2004) because a virus may induce secretion of sticky material that
382 contributes to cell/particle aggregation, such as transparent exopolymeric particles
383 (TEP) (Nissimov et al., 2018). We found that CEE (*i.e.*, $CE_{\text{deep}}/CE_{\text{surface}}$) increased
384 with the change of particles size from surface to deep ($\rho = 0.42$, $P = 8 \times 10^{-9}$) (Figure
385 S11). This positive correlation may reflect an elevated level of aggregation in places
386 where CEE is high, although it could be also due to the presence of large organisms at
387 depth.

388 Greater aggregate sinking along with higher particulate carbon fluxes was
389 observed in North Atlantic blooms of *Emiliana huxleyi* that were infected early by
390 the virus EhV, compared with late-infected blooms (Laber et al., 2018). In the same

391 bloom, viral infection stage was found to proceed with water column depth (Sheyn et
392 al., 2018). No EhV-like PolB sequences were detected in our dataset, which was
393 probably due to sampled areas and seasons.

394 Laboratory experiments suggest that viruses closely related to positive VIPs,
395 such as prasinoviruses, have infectious properties that may drive carbon export.
396 Cultures of *Micromonas pusilla* infected with prasinoviruses showed increased TEP
397 production compared with non-infected cultures (Lønborg et al., 2013). The hosts of
398 prasinoviruses (Mamiellales) have been proposed to contribute to carbon export in the
399 western subtropical North Pacific (Shiozaki et al., 2019). Some prasinoviruses encode
400 glycosyltransferases (GTs) of the GT2 family. The expression of GT2 family
401 members during infection possibly leads to the production of a dense fibrous
402 hyaluronan network and may trigger the aggregation of host cells (Van Etten et al.,
403 2017) with an increase in the cell wall C:N ratio. We detected one GT2 in a MAG of
404 two *Phycodnaviridae*-like positive VIPs (000200745 and 002503270) predicted to
405 infect Mamiellales, one in a MAG corresponding to the putative pelagophyte positive
406 VIP 000079111 related to AaV and six in two MAGs (three each) corresponding to
407 two *Mimiviridae*-like positive VIPs (000328966 and 001175669). *Phaeocystis*
408 *globosa virus* (PgV), closely related to the positive VIP PolB 000912507 (Figure S9),
409 has been linked with increased TEP production and aggregate formation during the
410 termination of a *Phaeocystis* bloom (Brussaard et al., 2007). Two closely related
411 bacillarnavirus VIPs were positively associated with CEE and predicted to infect
412 Chaetocerales. A previous study revealed an increase in abundance of viruses
413 infecting diatoms of *Chaetoceros* in both the water columns and the sediments during
414 the bloom of their hosts in a coastal area (Tomaru et al., 2011), suggesting sinking of
415 cells caused by viruses. Furthermore, the diatom *Chaetoceros tenuissimus* infected

416 with a DNA virus (CtenDNAV type II) has been shown to produce higher levels of
417 large-sized particles (50 to 400 μm) compared with non-infected cultures (Tomaru et
418 al., 2011; Yamada et al., 2018).

419 The other 25 VIPs were negatively associated with CEE. This association is
420 compatible with the “viral shunt,” which increases the amount of DOC (Wilhelm and
421 Suttle, 1999) and reduces the transfer of carbon to higher trophic levels and to the
422 deep ocean (Fuhrman, 1999; Weitz et al., 2015). Increased DOC has been observed in
423 culture of Mamiellales lysed by prasinoviruses (Lønborg et al., 2013). A field study
424 reported that PgV, to which the negative VIP PoIB 000054135 is closely related
425 (Figure S9), can be responsible for up to 35% of cell lysis per day during bloom of its
426 host (Baudoux et al., 2006), which is likely accompanied by consequent DOC release.
427 Similarly, the decline of a bloom of the pelagophyte *Aureococcus anophagefferens*
428 has been associated with active infection by AaV (to which one negative VIP is
429 closely related) (Moniruzzaman et al., 2017). Among RNA viruses, eight were
430 negative VIPs (six *Picornavirales* and two *Hepeviridae*). The higher representation of
431 *Picornavirales* in the viroplankton (Culley, 2018) than within cells (Urayama et al.,
432 2018) suggests that they are predominantly lytic, although no information exists
433 regarding the effect of *Picornavirales* on DOC release.

434 It is likely that the “viral shunt” and “shuttle” simultaneously affect and
435 modulate CEE in the global ocean (Zimmerman et al., 2019). The relative importance
436 of these two phenomena must fluctuate considerably depending on the host traits,
437 viral effects on metabolism, stages of infection, and environmental conditions.
438 Reflecting this complexity, viruses of a same host group could be found to be either
439 positively or negatively associated with CEE. We found that even two very closely
440 related *Mimiviridae* viruses (PoIBs 000079111 and 000079078 sharing 94%

441 nucleotide identity over their full gene lengths) most likely infecting pelagophyte
442 algae were positively and negatively associated with CEE.

443 Five percent of the tested viruses were associated with CEE in our study.
444 Similarly, four and two percent of bacterial virus populations were found to be
445 associated with the magnitude of carbon export (Guidi et al., 2016) and CEE (Figure
446 S12), respectively. These results suggest that viruses affecting CEE are rather
447 uncommon. It is plausible that such viruses affect CEE by infecting organisms that are
448 functionally important (abundant or keystone species), as we observed in host
449 prediction. The vast majority (95%) of non-VIPs may not have a significant impact on
450 CEE, because they do not strongly impact the host population, for instance, by stably
451 coexisting with their hosts. It is worth noting that experimental studies have reported
452 cultures of algae with viruses that reach a stable co-existence state after a few
453 generations (Yau et al., 2020).

454 **Conclusions**

455 Eukaryotic virus community composition was able to predict CEE at 59 sampling
456 sites in the photic zone of the world ocean. This statistical association was detected
457 based on a large omics dataset collected throughout the oceans and processed with
458 standardized protocols. The predictability of CEE by viral composition is consistent
459 with the hypothesis that “viral shunt” and “shuttle” are functioning at a global scale.
460 Among 83 viruses with a high importance in the prediction of CEE, 58 viruses were
461 positively and 25 negatively correlated with carbon export efficiency. Most of these
462 viruses belong to *Prasinovirus*, *Mimiviridae*, and *Picornavirales* and are either new to
463 science or with no known roles in carbon export efficiency. Thirty-six of these “select”
464 viruses were predicted to infect ecologically important hosts such as green algae of
465 the order Mamiellales, haptophytes, diatoms, and copepods. Positively associated

466 viruses had more predicted interactions with silicified eukaryotes than non-associated
467 viruses did. Overall, these results imply that the effect of viruses on the “shunt” and
468 “shuttle” processes could be dependent on viral hosts and ecosystem dynamics.

469 **Limitations of the study**

470 The observed statistical associations between viral compositions and examined
471 parameters (*i.e.*, CEE, CE and NPP) do not convey the information about the direction
472 of their potential causality relationships, and they could even result from indirect
473 relationships as discussed above. Certain groups of viruses detected in samples may
474 be over- or under-represented because of the technical limitations in size fractionation,
475 DNA/RNA extraction and sequencing.

476 **Resource Availability**

477 **Lead Contact**

478 Further information and requests for resources should be directed to and will be
479 fulfilled by Lead Contact, Hiroyuki Ogata (ogata@kuicr.kyoto-u.ac.jp).

480 **Materials Availability**

481 This study did not generate unique reagent.

482 **Data and Code Availability**

483 The authors declare that the data supporting the findings of this study are available
484 within the paper and its supplemental files, as well as at the GenomeNet FTP:
485 ftp://ftp.genome.jp/pub/db/community/tara/Cpump/Supplementary_material/.

486 Our custom R script used to test for association between viruses and environmental
487 variables (CEE, CE₁₅₀, NPP and T₁₀₀) is available along with input data at the
488 GenomeNet FTP:
489 ftp://ftp.genome.jp/pub/db/community/tara/Cpump/Supplementary_material/PLSreg/.
490 The Taxon Interaction Mapper (TIM) tool developed for this study and used for virus
491 host prediction is available at <https://github.com/RomainBlancMathieu/TIM>.

492 **Acknowledgements**

493 We thank the *Tara* Oceans consortium, the projects Oceanomics and France
494 Genomique (grants ANR-11-BTBR-0008 and ANR-10-INBS-09), and the people and
495 sponsors who supported the *Tara* Oceans Expedition ([http://www.embl.de/tara-](http://www.embl.de/tara-oceans/)
496 [oceans/](http://www.embl.de/tara-oceans/)) for making the data accessible. This is contribution number 110 of the *Tara*
497 Oceans Expedition 2009–2013. Computational time was provided by the
498 SuperComputer System, Institute for Chemical Research, Kyoto University. We thank
499 Barbara Goodson, Ph.D., and Sara J. Mason, M.Sc., from Edanz Group ([https://en-](https://en-author-services.edanzgroup.com/)
500 [author-services.edanzgroup.com/](https://en-author-services.edanzgroup.com/)) for editing a draft of this manuscript. This work
501 was supported by JSPS/KAKENHI (Nos. 26430184, 18H02279, and 19H05667 to
502 H.O. and Nos. 19K15895 and 19H04263 to H.E.), Scientific Research on Innovative
503 Areas from the Ministry of Education, Culture, Science, Sports and Technology
504 (MEXT) of Japan (Nos. 16H06429, 16K21723, and 16H06437 to H.O.), the
505 Collaborative Research Program of the Institute for Chemical Research, Kyoto
506 University (2019-29 to S.C.), the Future Development Funding Program of the Kyoto
507 University Research Coordination Alliance (to R.B.M.), the ICR-KU International
508 Short-term Exchange Program for Young Researchers (to S.C.), and the Research
509 Unit for Development of Global Sustainability (to H.O. and T.O.D.).

510 **Author contributions**

511 H.O. and R.B.M. conceived the study. H.K. and R.B.M. performed most of the
 512 analyses. H.E. and L.G. designed carbon export analysis. R.H.V and S.C. performed
 513 network analysis. N.H. and C.d.V. analyzed eukaryotic sequences. T.O.D., M.G., P.F.
 514 and O.J. analyzed viral MAGs. C.H.N. and H.M. contributed to statistical analysis.
 515 M.B.S. and C.A.S. contributed to interpretations. All authors edited and approved the
 516 final version of the manuscript.

517 **Declaration of Interests**

518 The authors declare no competing interests.

519 **References**

- 520 Agusti, S., González-Gordillo, J.I., Vaqué, D., Estrada, M., Cerezo, M.I., Salazar, G.,
 521 Gasol, J.M., and Duarte, C.M. (2015). Ubiquitous healthy diatoms in the deep sea
 522 confirm deep carbon injection by the biological pump. *Nat. Commun.* *6*, 7608.
- 523 Baudoux, A., Noordeloos, A., Veldhuis, M., and Brussaard, C. (2006). Virally
 524 induced mortality of *Phaeocystis globosa* during two spring blooms in temperate
 525 coastal waters. *Aquat. Microb. Ecol.* *44*, 207–217.
- 526 Bochdansky, A.B., Clouse, M.A., and Herndl, G.J. (2017). Eukaryotic microbes,
 527 principally fungi and labyrinthulomycetes, dominate biomass on bathypelagic marine
 528 snow. *ISME J.* *11*, 362–373.
- 529 Brum, J.R., Ignacio-Espinoza, J.C., Roux, S., Doucier, G., Acinas, S.G., Alberti, A.,
 530 Chaffron, S., Cruaud, C., Vargas, C. de, Gasol, J.M., et al. (2015). Patterns and
 531 ecological drivers of ocean viral communities. *Science* *348*, 1261498.
- 532 Brussaard, C.P.D., Bratbak, G., Baudoux, A.-C., and Ruardij, P. (2007). *Phaeocystis*
 533 and its interaction with viruses. *Biogeochemistry* *83*, 201–215.
- 534 Buesseler, K.O., and Boyd, P.W. (2009). Shedding light on processes that control
 535 particle export and flux attenuation in the twilight zone of the open ocean. *Limnol.*
 536 *Oceanogr.* *54*, 1210–1232.
- 537 Carradec, Q., Pelletier, E., Silva, C.D., Alberti, A., Seeleuthner, Y., Blanc-Mathieu,
 538 R., Lima-Mendez, G., Rocha, F., Tirichine, L., Labadie, K., et al. (2018). A global
 539 ocean atlas of eukaryotic genes. *Nat. Commun.* *9*, 373.

- 540 Coenen, A.R., and Weitz, J.S. (2018). Limitations of Correlation-Based Inference in
541 Complex Virus-Microbe Communities. *MSystems* 3, e00084-18.
- 542 Culley, A. (2018). New insight into the RNA aquatic virosphere via viromics. *Virus*
543 *Res.* 244, 84–89.
- 544 Dayaram, A., Goldstien, S., Argüello-Astorga, G.R., Zawar-Reza, P., Gomez, C.,
545 Harding, J.S., and Varsani, A. (2015). Diverse small circular DNA viruses circulating
546 amongst estuarine molluscs. *Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet.*
547 *Infect. Dis.* 31, 284–295.
- 548 Dunlap, D.S., Ng, T.F.F., Rosario, K., Barbosa, J.G., Greco, A.M., Breitbart, M., and
549 Hewson, I. (2013). Molecular and microscopic evidence of viruses in marine
550 copepods. *Proc. Natl. Acad. Sci.* 110, 1375–1380.
- 551 Endo, H., Ogata, H., and Suzuki, K. (2018). Contrasting biogeography and diversity
552 patterns between diatoms and haptophytes in the central Pacific Ocean. *Sci. Rep.* 8,
553 10916.
- 554 Endo, H., Blanc-Mathieu, R., Li, Y., Salazar, G., Henry, N., Labadie, K., de Vargas,
555 C., Sullivan, M.B., Bowler, C., Wincker, P., et al. (2020). Biogeography of marine
556 giant viruses reveals their interplay with eukaryotes and ecological functions. *Nat.*
557 *Ecol. Evol.* <https://doi.org/10.1038/s41559-020-01288-w>
- 558 Evans, C., and Wilson, W.H. (2008). Preferential grazing of *Oxyrrhis marina* on virus
559 infected *Emiliana huxleyi*. *Limnol. Oceanogr.* 53, 2035–2040.
- 560 Fawcett, S.E., Lomas, M.W., Casey, J.R., Ward, B.B., and Sigman, D.M. (2011).
561 Assimilation of upwelled nitrate by small eukaryotes in the Sargasso Sea. *Nat. Geosci.*
562 4, 717–722.
- 563 Fuhrman, J.A. (1999). Marine viruses and their biogeochemical and ecological effects.
564 *Nature* 399, 541–548.
- 565 Gobler, C.J., Hutchins, D.A., Fisher, N.S., Cosper, E.M., and Sañudo- Wilhelmy, S.A.
566 (1997). Release and bioavailability of C, N, P Se, and Fe following viral lysis of a
567 marine chrysophyte. *Limnol. Oceanogr.* 42, 1492–1504.
- 568 Goic, B., and Saleh, M.-C. (2012). Living with the enemy: viral persistent infections
569 from a friendly viewpoint. *Curr. Opin. Microbiol.* 15, 531–537.
- 570 Goode, A.G., Fields, D.M., Archer, S.D., and Martínez, J.M. (2019). Physiological
571 responses of *Oxyrrhis marina* to a diet of virally infected *Emiliana huxleyi*. *PeerJ* 7,
572 e6722.
- 573 Guidi, L., Chaffron, S., Bittner, L., Eveillard, D., Larhlimi, A., Roux, S., Darzi, Y.,
574 Audic, S., Berline, L., Brum, J.R., et al. (2016). Plankton networks driving carbon
575 export in the oligotrophic ocean. *Nature* 532, 465.
- 576 Hingamp, P., Grimsley, N., Acinas, S.G., Clerissi, C., Subirana, L., Poulain, J.,
577 Ferrera, I., Sarmiento, H., Villar, E., Lima-Mendez, G., et al. (2013). Exploring

- 578 nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. *ISME*
579 *J.* 7, 1678–1695.
- 580 Hirata, T., Hardman-Mountford, N.J., Brewin, R.J.W., Aiken, J., Barlow, R., Suzuki,
581 K., Isada, T., Howell, E., Hashioka, T., Noguchi-Aita, M., et al. (2011). Synoptic
582 relationships between surface Chlorophyll-*a* and diagnostic pigments specific to
583 phytoplankton functional types. *Biogeosciences* 8, 311–327.
- 584 Hurwitz, B.L., Brum, J.R., and Sullivan, M.B. (2015). Depth-stratified functional and
585 taxonomic niche specialization in the “core” and “flexible” Pacific Ocean Virome.
586 *ISME J.* 9, 472–484.
- 587 Iversen, M.H., and Ploug, H. (2010). Ballast minerals and the sinking carbon flux in
588 the ocean: carbon-specific respiration rates and sinking velocity of marine snow
589 aggregates. *Biogeosciences* 7, 2613–2624.
- 590 Karl, D.M., Church, M.J., Dore, J.E., Letelier, R.M., and Mahaffey, C. (2012).
591 Predictable and efficient carbon sequestration in the North Pacific Ocean supported
592 by symbiotic nitrogen fixation. *Proc. Natl. Acad. Sci.* 109, 1842–1849.
- 593 Klaas, C., and Archer, D.E. (2002). Association of sinking organic matter with
594 various types of mineral ballast in the deep sea: Implications for the rain ratio. *Glob.*
595 *Biogeochem. Cycles* 16, 1116.
- 596 Laber, C.P., Hunter, J.E., Carvalho, F., Collins, J.R., Hunter, E.J., Schieler, B.M.,
597 Boss, E., More, K., Frada, M., Thamatrakoln, K., et al. (2018). Coccolithovirus
598 facilitation of carbon export in the North Atlantic. *Nat. Microbiol.* 3, 537–547.
- 599 Lawrence, J.E., and Suttle, C.A. (2004). Effect of viral infection on sinking rates of
600 *Heterosigma akashiwo* and its implications for bloom termination. *Aquat. Microb.*
601 *Ecol.* 37, 1–7.
- 602 Lawrence, J.E., Chan, A.M., and Suttle, C.A. (2002). Viruses causing lysis of the
603 toxic bloom-forming alga *Heterosigma akashiwo* (Raphidophyceae) are widespread in
604 coastal sediments of British Columbia, Canada. *Limnol. Oceanogr.* 47, 545–550.
- 605 Leblanc, K., Quéguiner, B., Diaz, F., Cornet, V., Michel-Rodriguez, M., Durrieu de
606 Madron, X., Bowler, C., Malviya, S., Thyssen, M., Grégori, G., et al. (2018).
607 Nanoplanktonic diatoms are globally overlooked but play a role in spring blooms and
608 carbon export. *Nat. Commun.* 9, 953.
- 609 Li, W. (1995). Composition of Ultraphytoplankton in the Central North-Atlantic. *Mar.*
610 *Ecol. Prog. Ser.* 122, 1–8.
- 611 Liu, H., Probert, I., Uitz, J., Claustre, H., Aris-Brosou, S., Frada, M., Not, F., and de
612 Vargas, C. (2009). Extreme diversity in noncalcifying haptophytes explains a major
613 pigment paradox in open oceans. *Proc. Natl. Acad. Sci. U. S. A.* 106, 12803–12808.
- 614 Lomas, M.W., and Moran, S.B. (2011). Evidence for aggregation and export of
615 cyanobacteria and nano-eukaryotes from the Sargasso Sea euphotic zone.
616 *Biogeosciences* 8, 203–216.

- 617 Lønborg, C., Middelboe, M., and Brussaard, C.P.D. (2013). Viral lysis of
618 *Micromonas pusilla*: impacts on dissolved organic matter production and composition.
619 *Biogeochemistry* 116, 231–240.
- 620 Mars Brisbin, M., Mesrop, L.Y., Grossmann, M.M., and Mitarai, S. (2018). Intra-host
621 Symbiont Diversity and Extended Symbiont Maintenance in Photosymbiotic
622 *Acantharea* (Clade F). *Front. Microbiol.* 9, 1998.
- 623 Monier, A., Pagarete, A., de Vargas, C., Allen, M.J., Read, B., Claverie, J.-M., and
624 Ogata, H. (2009). Horizontal gene transfer of an entire metabolic pathway between a
625 eukaryotic alga and its DNA virus. *Genome Res.* 19, 1441–1449.
- 626 Monier, A., Worden, A.Z., and Richards, T.A. (2016). Phylogenetic diversity and
627 biogeography of the Mamiellophyceae lineage of eukaryotic phytoplankton across the
628 oceans. *Environ. Microbiol. Rep.* 8, 461–469.
- 629 Moniruzzaman, M., Wurch, L.L., Alexander, H., Dyhrman, S.T., Gobler, C.J., and
630 Wilhelm, S.W. (2017). Virus-host relationships of marine single-celled eukaryotes
631 resolved from metatranscriptomics. *Nat. Commun.* 8, 16054.
- 632 Nissimov, J.I., Vandzura, R., Johns, C.T., Natale, F., Haramaty, L., and Bidle, K.D.
633 (2018). Dynamics of transparent exopolymer particle production and aggregation
634 during viral infection of the coccolithophore, *Emiliana huxleyi*. *Environ. Microbiol.*
635 20, 2880–2897.
- 636 Peduzzi, P., and Weinbauer, M.G. (1993). Effect of concentrating the virus-rich 2-
637 2nm size fraction of seawater on the formation of algal flocs (marine snow). *Limnol.*
638 *Oceanogr.* 38, 1562–1565.
- 639 Proctor, L.M., and Fuhrman, J.A. (1991). Roles of viral infection in organic particle
640 flux. *Mar. Ecol. Prog. Ser.* 69, 133–142.
- 641 Sheyn, U., Rosenwasser, S., Lehahn, Y., Barak-Gavish, N., Rotkopf, R., Bidle, K.D.,
642 Koren, I., Schatz, D., and Vardi, A. (2018). Expression profiling of host and virus
643 during a coccolithophore bloom provides insights into the role of viral infection in
644 promoting carbon export. *ISME J.* 12, 704–713.
- 645 Shi, M., Lin, X.-D., Tian, J.-H., Chen, L.-J., Chen, X., Li, C.-X., Qin, X.-C., Li, J.,
646 Cao, J.-P., Eden, J.-S., et al. (2016). Redefining the invertebrate RNA virosphere.
647 *Nature* 540, 539–543.
- 648 Shiozaki, T., Hirose, Y., Hamasaki, K., Kaneko, R., Ishikawa, K., and Harada, N.
649 (2019). Eukaryotic Phytoplankton Contributing to a Seasonal Bloom and Carbon
650 Export Revealed by Tracking Sequence Variants in the Western North Pacific. *Front.*
651 *Microbiol.* 10, 2722.
- 652 Sullivan, M.B., Weitz, J.S., and Wilhelm, S. (2017). Viral ecology comes of age.
653 *Environ. Microbiol. Rep.* 9, 33–35.
- 654 Sunagawa, S., Coelho, L.P., Chaffron, S., Kultima, J.R., Labadie, K., Salazar, G.,
655 Djahanschiri, B., Zeller, G., Mende, D.R., Alberti, A., et al. (2015). Ocean plankton.
656 Structure and function of the global ocean microbiome. *Science* 348, 1261359.

- 657 Suttle, C.A. (2007). Marine viruses--major players in the global ecosystem. *Nat. Rev.*
658 *Microbiol.* *5*, 801–812.
- 659 Takao, Y., Nagasaki, K., Mise, K., Okuno, T., and Honda, D. (2005). Isolation and
660 characterization of a novel single-stranded RNA Virus infectious to a marine fungoid
661 protist, *Schizochytrium* sp. (Thraustochytriaceae, Labyrinthulea). *Appl. Environ.*
662 *Microbiol.* *71*, 4516–4522.
- 663 Tomaru, Y., Hata, N., Masuda, T., Tsuji, M., Igata, K., Masuda, Y., Yamatogi, T.,
664 Sakaguchi, M., and Nagasaki, K. (2007). Ecological dynamics of the bivalve-killing
665 dinoflagellate *Heterocapsa circularisquama* and its infectious viruses in different
666 locations of western Japan. *Environ. Microbiol.* *9*, 1376–1383.
- 667 Tomaru, Y., Fujii, N., Oda, S., Toyoda, K., and Nagasaki, K. (2011). Dynamics of
668 diatom viruses on the western coast of Japan. *Aquat. Microb. Ecol.* *63*, 223–230.
- 669 Tréguer, P., Bowler, C., Moriceau, B., Dutkiewicz, S., Gehlen, M., Aumont, O.,
670 Bittner, L., Dugdale, R., Finkel, Z., Iudicone, D., et al. (2018). Influence of diatom
671 diversity on the ocean biological carbon pump. *Nat. Geosci.* *11*, 27–37.
- 672 Turner, J.T. (2015). Zooplankton fecal pellets, marine snow, phytodetritus and the
673 ocean's biological pump. *Prog. Oceanogr.* *130*, 205–248.
- 674 Urayama, S., Takaki, Y., Nishi, S., Yoshida- Takashima, Y., Deguchi, S., Takai, K.,
675 and Nunoura, T. (2018). Unveiling the RNA virosphere associated with marine
676 microorganisms. *Mol. Ecol. Resour.* *18*, 1444–1455.
- 677 Van Etten, J., Agarkova, I., Dunigan, D., Tonetti, M., De Castro, C., and Duncan, G.
678 (2017). Chloroviruses Have a Sweet Tooth. *Viruses* *9*, 88.
- 679 Wang, H., Wu, S., Li, K., Pan, Y., Yan, S., and Wang, Y. (2018). Metagenomic
680 analysis of ssDNA viruses in surface seawater of Yangshan Deep-Water Harbor,
681 Shanghai, China. *Mar. Genomics* *41*, 50–53.
- 682 Weinbauer, M.G. (2004). Ecology of prokaryotic viruses. *FEMS Microbiol. Rev.* *28*,
683 127–181.
- 684 Weitz, J.S., Stock, C.A., Wilhelm, S.W., Bourouiba, L., Coleman, M.L., Buchan, A.,
685 Follows, M.J., Fuhrman, J.A., Jover, L.F., Lennon, J.T., et al. (2015). A multitrophic
686 model to quantify the effects of marine viruses on microbial food webs and ecosystem
687 processes. *ISME J.* *9*, 1352–1364.
- 688 Wilhelm, S.W., and Suttle, C.A. (1999). Viruses and Nutrient Cycles in the
689 SeaViruses play critical roles in the structure and function of aquatic food webs.
690 *BioScience* *49*, 781–788.
- 691 Yamada, Y., Tomaru, Y., Fukuda, H., and Nagata, T. (2018). Aggregate Formation
692 During the Viral Lysis of a Marine Diatom. *Front. Mar. Sci.* *5*, 167.
- 693 Yau, S., Krasovec, M., Benites, L.F., Rombauts, S., Groussin, M., Vancaester, E.,
694 Aury, J.-M., Derelle, E., Desdevises, Y., Escande, M.-L., et al. (2020). Virus-host
695 coexistence in phytoplankton through the genomic lens. *Sci. Adv.* *6*, eaay2587.

- 696 Yoshida, M., Noël, M.-H., Nakayama, T., Naganuma, T., and Inouye, I. (2006). A
697 haptophyte bearing siliceous scales: ultrastructure and phylogenetic position of
698 *Hyalolithus neolepis* gen. et sp. nov. (Prymnesiophyceae, Haptophyta). *Protist* 157,
699 213–234.
- 700 Yoshikawa, G., Blanc-Mathieu, R., Song, C., Kayama, Y., Mochizuki, T., Murata, K.,
701 Ogata, H., and Takemura, M. (2019). Medusavirus, a Novel Large DNA Virus
702 Discovered from Hot Spring Water. *J. Virol.* 93, e02130-18.
- 703 Zhang, C., Dang, H., Azam, F., Benner, R., Legendre, L., Passow, U., Polimene, L.,
704 Robinson, C., Suttle, C.A., and Jiao, N. (2018). Evolving paradigms in biological
705 carbon cycling in the ocean. *Natl. Sci. Rev.* 5, 481–499.
- 706 Zimmerman, A.E., Howard-Varona, C., Needham, D.M., John, S.G., Worden, A.Z.,
707 Sullivan, M.B., Waldbauer, J.R., and Coleman, M.L. (2019). Metabolic and
708 biogeochemical consequences of viral infection in aquatic ecosystems. *Nat. Rev.*
709 *Microbiol.* 18, 21–34.
- 710
- 711

712 **Figure legends**

713 **Figure 1. Viruses of eukaryotic plankton identified in *Tara* Oceans samples are**
714 **distantly related to characterized viruses.** Unrooted maximum likelihood
715 phylogenetic trees containing environmental (black) and reference (red) viral
716 sequences for NCLDV DNA polymerase family B (A), RNA virus RNA-dependent
717 RNA polymerase (B), and ssDNA virus replication-associated protein (C). See also
718 Figures S1-S4

719

720 **Figure 2. Carbon export efficiency and relative marker-gene occurrence of**
721 **eukaryotic plankton viruses along the sampling route.** (A) Carbon export
722 efficiency (CEE) estimated at 39 *Tara* Oceans stations where surface and DCM layers
723 were sampled for prokaryote-enriched metagenomes and eukaryotic
724 metatranscriptomes. See also Figures S5 and S11. (B, C) Relative marker-gene
725 occurrence of major groups of viruses of eukaryotic plankton for NCLDVs in
726 metagenomes (B) and for RNA and ssDNA viruses in metatranscriptomes (C) at 59
727 sampling sites.

728

729 **Figure 3. Relative abundance of eukaryotic plankton viruses associated with**
730 **carbon export efficiency in the global ocean.** (A) Bivariate plot between predicted
731 and observed values in a leave-one-out cross-validation test for carbon export
732 efficiency. The PLS regression model was constructed using occurrence profiles of
733 1,523 marker-gene sequences (1,309 PolBs, 180 RdRPs and 34 Reps) derived from
734 environmental samples. r , Pearson correlation coefficient; R^2 , the coefficient of
735 determination between measured response values and predicted response values. R^2 ,
736 which was calculated as $1 - \text{SSE}/\text{SST}$ (sum of squares due to error and total)

737 measures how successful the fit is in explaining the variance of the response values.
738 The significance of the association was assessed using a permutation test ($n = 10,000$)
739 (grey histogram in (A)). The red diagonal line shows the theoretical curve for perfect
740 prediction. (B) Pearson correlation coefficients between CEE and occurrence profiles
741 of 83 viruses that have VIP scores > 2 (VIPs) with the first two components in the
742 PLS regression model using all samples. PLS components 1 and 2 explained 83% and
743 11% of the variance of CEE, respectively. Fifty-eight VIPs had positive regression
744 coefficients in the model (shown with circles), and 25 had negative regression
745 coefficients (shown with triangles). See also Figures S6, S7 and S12, Table S1, and
746 Supplemental Data 1.

747

748 **Figure 4. Biogeography of viruses associated with carbon export efficiency.** The
749 upper panel shows carbon export efficiency ($CEE = CE_{\text{deep}}/CE_{\text{surface}}$) for 59 sampling
750 sites. The bottom panel is a map reflecting relative abundances, expressed as centered
751 log-ratio transformed, gene-length normalized read counts of viruses positively and
752 negatively associated with CEE that have VIP scores > 2 (VIPs). MS, Mediterranean
753 Sea; IO, Indian Ocean; SAO, South Atlantic Ocean; SPO, South Pacific Ocean; NPO,
754 North Pacific Ocean; NAO, North Atlantic Ocean. The bottom horizontal axis is
755 labeled with *Tara* Oceans station numbers, sampling depth (SRF, surface; DCM, deep
756 chlorophyll maximum), and abbreviations of biogeographic provinces. Viruses
757 labeled in red correspond to positive VIPs that tend more represented in more than
758 one biogeographic province.

759

760 **Tables**761 **Table 1. Taxonomic breakdown of viral marker genes.**

Viruses		Identified	Used in PLS regression ^a
NCLDVs	Mimiviridae	2,923	1,148
	Phycodnaviridae	348	99
	Iridoviridae	198	59
	Other NCLDVs ^b	17	3
	Total	3,486	1,309
RNA viruses	Picornavirales (ssRNA+)	325	80
	Partitiviridae (dsRNA)	131	22
	Narnaviridae (ssRNA+)	95	6
	Other families	289	53
	Unclassified	78	9
	RNA viruses	57	10
Total	975	180	
ssDNA viruses	Circoviridae	201	22
	Geminiviridae	4	0
	Nanoviridae	4	0
	Unclassified	39	2
	ssDNA viruses	51	10
Total	299	34	
All		4,760	1,523

762 ^aThe marker genes had to occurred in at least five samples and harbor a Spearman correlation
763 coefficient > |0.2| with carbon export efficiency.

764 ^bThere was no unclassified NCLDV.

765

766 **Table 2. Host predictions per viral and host group for viruses associated with**
 767 **carbon export efficiency.** See also Figures S8-S10, Tables S2 and S3, and
 768 Supplemental Data 2.

Virus-host relationship	Positive VIPs ^a	Negative VIPs ^a	Total
NCLDV-Mamiellales	10	4	15
NCLDV-Prymnesiales	5	1	6
NCLDV-Pelagophyceae	2	1	3
NCLDV-No prediction	26	11	36
RNA virus-Copepoda	7	2	9
RNA virus-Chaetocerotales	2	0	2
RNA virus-Labyrinthulomycetes	1	0	1
RNA virus-No prediction	4	6	10
ssDNA virus-Copepoda	1	0	1
Total	58	25	83

769 ^aVIPs refers to viruses having VIP scores > 2. Positive and negative VIPs had positive and negative
 770 regression coefficients in the PLS model, respectively.
 771

772 **Table 3. Functional differences between eukaryotes found to be best**
 773 **connected to positively associated and not associated with carbon export**
 774 **efficiency.** See also Tables S4-S6.

Functional trait	Positive VIPs ^a (n = 50)		Non-VIPs ^a (n = 983)		P-value (Fisher's exact test, two sided)	Adjusted P-value (BH) (Q)
	Presence	Absence	Presence	Absence		
Chloroplast	20	30	276	690	0.109	0.164
Silicification	11	39	60	920	0.000	0.001
Calcification	1	49	30	950	1.000	1.000

775 ^aVIPs refers to viruses having VIP scores > 2. Positive VIPs had positive regression coefficients in the
 776 PLS model.
 777

Journal Pre-proof

778 **Supplemental Data titles**

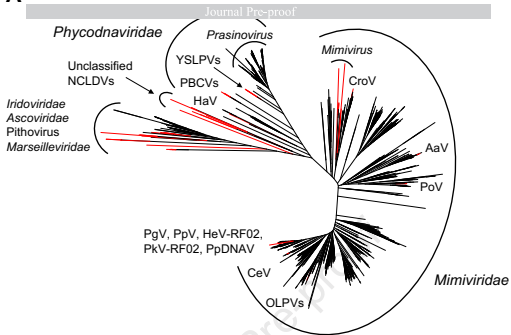
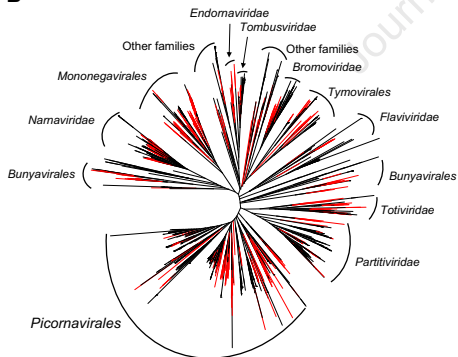
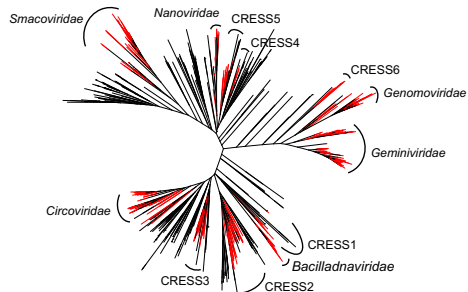
779 **Supplemental Data 1. VIP scores and regression coefficients in the PLS model of**
780 **CEE for NCLDV polB, RNA virus RdRP and ssDNA virusRep along with their**
781 **taxonomic assignment, Related to Figure 3.**

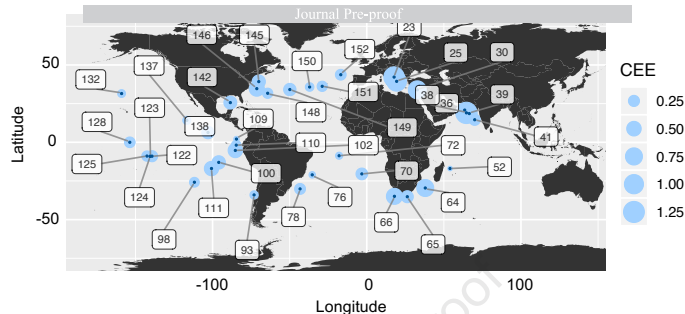
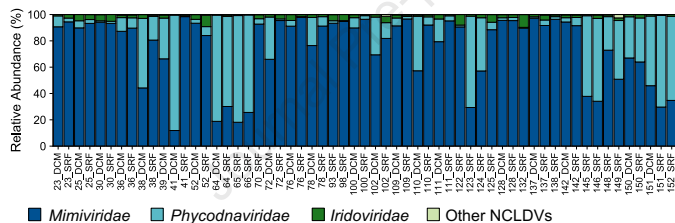
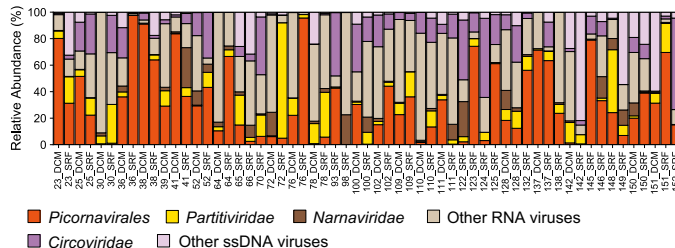
782

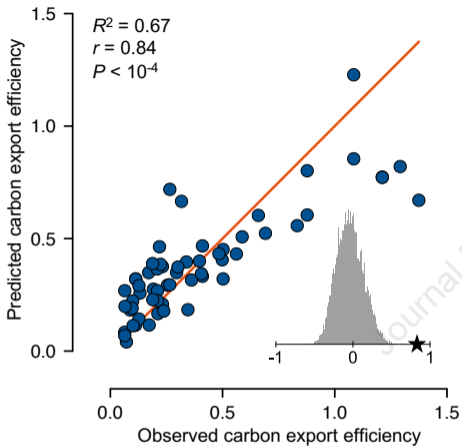
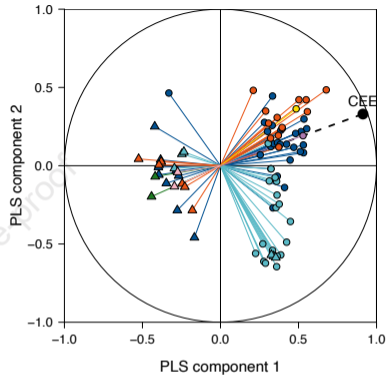
783 **Supplemental Data 2. Genomic context-based host prediction, Related to Table 2.**

784

Journal Pre-proof

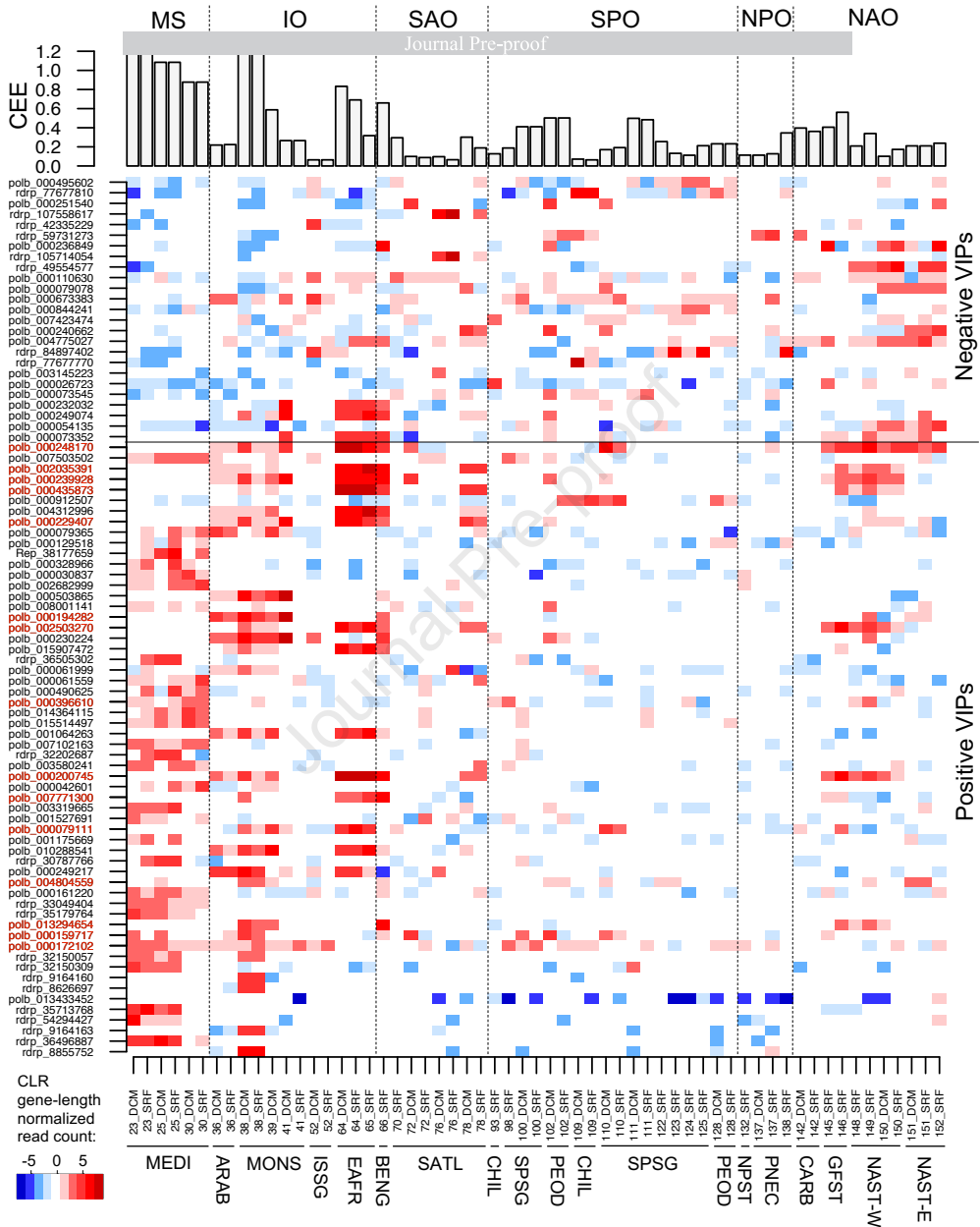
A**B****C**

A**B****C**

A**B**

Sign of regression coefficient : ○ Positive (pos)
 △ Negative (neg)

Taxonomy : ● *Mimiviridae* (n=34, 25 pos/9 neg)
 ● *Phycodnaviridae* (n=24, 18 pos/6 neg)
 ● *Iridoviridae* (n=2, 0 pos/2 neg)
 ● *Picornavirales* (n=19, 13 pos/6 neg)
 ● *Hepeviridae* (n=2, 0 pos/2 neg)
 ● *Partitiviridae* (n=1, 1 pos/0 neg)
 ● *Circoviridae* (n=1, 1 pos/0 neg)



Highlights

Eukaryotic virus community composition is shown to predict carbon export efficiency

Tens of viruses are highly important in the prediction of the efficiency

These viruses are inferred to infect ecologically important hosts

Journal Pre-proof