Tight association between microbial eukaryote and *Imitervirales* communities in the Pacific Arctic Ocean

Jun Xia[1], Sohiko Kameyama[2], Florian Prodinger[1], Takashi Yoshida[3], Kyoung-Ho Cho[4], Jinyoung Jung[4], Sung-Ho Kang[4], Eun-Jin Yang[4], Hiroyuki Ogata[1], Hisashi Endo[1*]

[1] Institute for Chemical Research, Kyoto University, Gokasho, Uji 611-0011, Japan

[2] Faculty of Environmental Earth Science, Hokkaido University, N10W5 Sapporo, Hokkaido 060-0810, Japan

[3] Graduate School of Agriculture, Kyoto University, Kitashirakawa-Oiwake, Sakyo-ku, Kyoto 606-8502, Japan

[4] Korea Polar Research Institute, Incheon 21990, Republic of Korea

Running title: Association between eukaryote and *Imitervirales*

*For correspondence. E-mail endo@scl.kyoto-u.ac.jp; Tel. (+81) 0774-38-3272; Fax (+81) 0774-38-3269.

Summary

Viruses are important regulatory factors of marine microbial community including microeukaryotes. However, little is known about their role in the northern Chukchi Sea of the Arctic basin, which remains oligotrophic conditions in summer. To elucidate linkages of microbial eukaryotic community with viruses as well as environmental variables, we investigated the community structures of microeukaryotes (3–144 μm and 0.2–3 μm size fractions) and *Imitervirales* (0.2–3 μm size fraction), a major group of viruses infecting marine microeukaryotes. Surface water samples were collected at 21 ocean stations located in the northeastern Chukchi Sea (NECS), an adjacent area outside the Beaufort Gyre (Adjacent Sea; AS), and two melt ponds on sea ice in the summer of 2018. At the ocean stations, nutrient concentrations were low in most of the locations expect at the shelf in the AS. The community variations were significantly correlated between eukaryotes and *Imitervirales*, even within the NECS characterized by relatively homogeneous environmental conditions. The association of the eukaryotic community with the viral community was stronger than that with geographical and physicochemical environmental factors. These results suggest that *Imitervirales* are actively infecting their hosts even in cold and oligotrophic sea water in the Arctic Ocean.

Introduction

The Arctic Ocean is the smallest, shallowest, and coldest ocean on earth. It covers several seas including the Chukchi Sea and the Beaufort Sea. Sea ice cover the central area of the ocean slightly in summer and almost completely in winter, due to the extreme seasonality in the receipt of solar radiation in the polar area. Anthropogenic climate change is now accelerating and has a strong impact on the Arctic Ocean (Lannuzel et al., 2020; Stroeve et al., 2007). Due to the ice-albedo feedback mechanism, sea ice retreat to a greater extent in summer than before (Curry et al., 1995; Kashiwase et al., 2017; Lindsay et al., 2012). The highest surface air temperature during the past 120 years was recorded for the period between 2014 and 2019 (Perovich & Jones, 2019). Upper sea water was also freshened as a result of accelerating sea ice melting, and this situation is predicted to continue (Kwok & Cunningham, 2010; Münchow, 2016; Shu et al., 2018). In such changing environmental conditions, how microorganisms form and alter their community structures through intricate interactions between them and with surrounding environments is an important general issue.

Microbial eukaryotes play a fundamental role in the marine ecosystem (de Vargas et al., 2015; Worden et al., 2015). Being integrated in the food web, they drive biogeochemical cycles by contributing to primary production (Falkowski et al., 1998; Field et al., 1998) and transferring fixed carbon to the higher trophic levels (Sherr et al., 2007). Primary production in surface seawater of the Arctic Ocean estimated by

satellite remote sensing significantly increased from 1998 to 2018, presumably due to the climate change-induced environmental modifications such as sea ice loss and an increase of nutrient influx (Lewis et al., 2020). However, another study indicated that the abundance of nanophytoplankton (2–20 μm) decreased from 2004 to 2008 in the Canada Basin while that of picophytoplankton (i.e., cell size of 0.2–2 μm) increased because of the decrease in nutrient concentrations (Li et al., 2009).

A first molecular biological characterization of the microbial eukaryotic community in the Arctic Ocean has been reported fifteen years ago (Lovejoy et al., 2006). Since then, several groups have carried out molecular barcoding studies to investigate the microbial eukaryotic community in a variety of areas of the Arctic Ocean such as the Chukchi Sea, the Beaufort Sea, West Spitsbergen, the Amundsen Gulf, the Bering Strait, Greenland and the central Arctic Ocean (Comeau et al., 2011; Kilias et al., 2014; Marquardt et al., 2016; Monier et al., 2013; Onda et al., 2017; Terrado et al., 2009; Zhang et al., 2015). Strong seasonality has been revealed through the annual data of 18S rDNA derived from arctic surface water samples, with dominant microbial eukaryotic groups being largely different between summer and winter (Marquardt et al., 2016). Composition of the microbial eukaryotic community in the Arctic Ocean is also shown to vary across water masses and environments with different physicochemical parameters and nutrient concentrations (Hamilton et al., 2008; Joli et al., 2018; Kilias et al., 2014; Thaler & Lovejoy, 2013). These results

collectively suggest the importance of environmental conditions in constructing microeukaryotes at large time and spatial scales.

Apart from the physicochemical properties, viruses are thought to be a key effector of the communities of their microbial hosts in marine ecosystems (Middelboe & Brussaard, 2017; Rohwer & Thurber, 2009; Suttle, 2007). *Imitervirales*, belonging to the phylum *Nucleocytoviricota* (also known as nucleocytoplasmic large DNA viruses, NCLDVs) (Iyer et al., 2006), is one of the most dominant orders of viruses infecting diverse microbial eukaryotes in the ocean (Endo et al., 2020). A previous study showed a tight association between the community of NCLDVs and that of microbial eukaryotes based on a global metagenomic dataset (Endo et al., 2020). However, as the result was based on a global scale dataset, the observed association was expected from substantial differences in the host eukaryotes inhabiting geographically distant and environmentally distinct locations. We consider that investigating such virus-host associations at a smaller geographic and time scales would provide further insights into the possible regulatory role of viruses on the host community structure. However, such local studies are currently scarce for *Imitervirales* (or NCLDVs) (Clerissi et al., 2012; Sandaa et al., 2018), and to our knowledge, there is no study investigating both *Imitervirales* and eukaryotic communities in the Arctic Ocean. We reasoned that examining whether the two communities are associated with each other or not in geographically close locations with similar environmental conditions and in the same period would lead to a better

103 understanding of the interactions between *Imitervirales* and microbial eukaryotes. If
104 virions persistently remain in an environment for a long period of time, then we
105 would not expect a strong association between *Imitervirales* and microbial eukaryotes.
106 In contrast, if viruses actively infect various eukaryotes, then a strong eukaryote-viral
107 association would emerge.

108 In this study, we conducted a high spatial resolution sampling of microbial DNA
109 from the surface water during the 2018 cruise of the Korean ice breaking research
110 vessel (IBRV) Araon. We investigated the community structures of microeukaryotes
111 and *Imitervirales* in the basin region of the Chukchi Sea (the northeastern Chukchi
112 Sea; hereafter NECS) as well as in an adjacent sea (AS) outside the Beaufort Gyre
113 and melt ponds. The surface water of the NECS is characterized by the low salinity
114 and nutrients under the influence of the Beaufort Gyre system, making it distinct from
115 the AS areas. To gain insight into the interdependence of the eukaryotic and
116 *Imitervirales* communities, we analyzed the statistical relationships between the
117 eukaryotic and *Imitervirales* communities while controlling the effects of
118 environmental and geographical characteristics in the two different environmental
119 regimes (the "stable" NECS and the "dynamic" AS).

120

121 **Results**

122 **Water characteristics and environmental factors**

123 Twenty-one oceanic sampling stations (surface seawater samples) were classified into
124 two groups depending on the geographical locations and the temperature-salinity (TS)
125 diagram: the NECS (in the regions of Chukchi Plateau and Canada basin) and the AS
126 (Figs 1 and S2).

127 Of the measured physical parameters (Supplement Table S1), temperature
128 showed little difference among stations, but salinity showed relatively large
129 differences (Supplement Table S1, Fig. 2A and B). Sea surface temperatures (SST) in
130 the NECS (-0.99˚C on average) was slightly higher than those in the AS (-1.11˚C on
131 average), and the salinity in the NECS (27.98 psu on average) was substantially lower
132 than the AS (30.05 psu on average). Sea ice coverage in each station was 0% to
133 89.5% (Supplement Table S1).

134 Concentrations of nutrients (ammonia nitrogen, nitrate+nitrite, phosphate, and
135 silicate) as well as Chl *a* were measured for each location. Most of the sampled area
136 was oligotrophic, while water conditions of three "bloom" sites (stations A13, A14,
137 A15) in the AS presented high nutrient concentrations. The nutrient and Chl *a*
138 concentrations for the bloom stations (on average: nitrate + nitrite: 1.17 $\mu mol \cdot L^{-1}$;
139 phosphate: 1.02 $\mu mol \cdot L^{-1}$; silicate: 14.15 $\mu mol \cdot L^{-1}$; chlorophyll *a*: 7.23 $mg \cdot m^{-3}$) were
140 much higher than those in other stations (on average: nitrate+nitrite: 0.14 $\mu mol \cdot L^{-1}$;
141 phosphate: 0.52 $\mu mol \cdot L^{-1}$; silicate: 0.01 $\mu mol \cdot L^{-1}$; chlorophyll *a*: 0.17 $mg \cdot m^{-3}$).
142 Ammonia concentration was relatively high at the station A13 (0.11 $\mu mol \cdot L^{-1}$), while
143 it was less than 0.02 $\mu mol \cdot L^{-1}$ in other stations.

144

145 **Amplicon sequences**

146 Sequence information and the number of ASVs were summarized in Supplement
147 Table S2. The number of ASVs from each sample before subsampling is provided in
148 Fig. S1. For the 3-144 μm eukaryotic community, 45,588 to 214,775 reads obtained
149 from individual samples were subsampled at the minimum number of reads across
150 different samples (i.e., 45,588 reads), and then grouped into 107 to 390 eukaryotic
151 18S non-singleton ASVs with the mean proportion of raw read usage being 37%. For
152 the 0.2-3 μm eukaryotic community, subsampling was performed at 72,359 reads,
153 which was grouped into 106 to 385 eukaryotic 18S ASVs. For the *Imitervirales*
154 community, subsampling was also performed at 26,638 reads, which generated 244 to
155 525 *Imitervirales polB* ASVs per sample.

156

157 **Composition and local diversity of eukaryotic and *Imitervirales* communities**

158 We first investigated eukaryotic communities by excluding sequences from metazoa
159 and fungi, because they have different lifestyles and ecological functions from protists.
160 The community compositions were different between large (3-144 μm) and small
161 (0.2-3 μm) size-fractions (Fig. 3A and B). Eukaryotic communities of the large size
162 fraction were dominated by dinoflagellates (36.6% on average), diatoms (11.4%) and
163 other marine alveolates (29.7%, include ciliates and protaveolata), while those of
164 small size fraction were dominated by ciliates (28.5%), chlorophytes (19.8%) and

165 picozoa (10.8%). In the large size fractions, lower proportion of dinoflagellates
166 occurred in the AS sites than in the NECS sites (ANOVA followed by Tukey post hoc
167 tests, $p<0.01$), especially in the three bloom samples (ANOVA followed by Tukey
168 post hoc tests, $p<0.05$). On the contrary, diatoms tended to be more abundant in the
169 AS sites than in the NECS sites (ANOVA followed by Tukey post hoc tests, $p<0.01$).
170 In the small size fraction, although the dominant ciliates had little difference among
171 all the samples, chlorophytes showed higher proportion in the AS bloom samples and
172 the NECS samples than in other samples of the AS sites (ANOVA followed by Tukey
173 post hoc tests, $p<0.01$). Another unique feature of the bloom sites was that there was
174 almost no picozoa sequences in these samples, while the picozoa represented one of
175 the abundant phyla in the other samples.

176 As for metazoa and fungal communities (Fig. S3), copepods were the most
177 dominant (20.0% on average) in the 3-144 μm size fraction samples. As for protist
178 community, the most abundant ASVs (>10% in at least one sample) in the large size
179 fraction belonged to *Heterocapsa* (dinoflagellate), *Chytriodinium* (dinoflagellate),
180 *Gyrodinium* (dinoflagellate), while those in the small size fraction were *Micromonas*
181 (chlorophyte), *Oligotrichia* (ciliate), *Chytriodinium* (dinoflagellate), *Phaeocystis*
182 (haptophyte), *Chaetoceros* (diatom) and *Carteria* (chlorophyte).

183 *Imitervirales* ASVs were mapped onto a larger set of *polB* sequences from the
184 *Tara* Oceans dataset and classified into 13 clades (Fig. S5). Clades 7 (28.6%) was the
185 most abundant clade, followed by clade 6 (20.2%) and clade 2 (17.0%) (Fig. 3C). In

186 the bloom sites, particularly high relative abundances were shown for clade 5 and 6.
187 In the other AS sites, clade 3 which includes the OLPVs (Organic Lake
188 Phycodnavirus 1 and 2) showed higher proportions than in the NECS. In the NECS
189 samples, clade 2 showed high proportions (28.0% on average). Clear difference was
190 found between the communities in the two aquatic habitats (sea water and melt pond
191 water) for the eukaryotic and *Imitervirales* communities (Fig. 3A–C and S6).
192 *Imitervirales* communities in the Arctic Ocean were clearly distinguished with those
193 obtained from subtropical coastal sea water and hotspring samples by the same
194 amplicon method (Fig. S6) (Li et al., 2019; Prodinger et al., 2020; Prodinger et al.,
195 unpublished). In the samples of the present study, *Imitervirales* communities were
196 classified into three groups: Arctic seawater, Arctic algae bloom related seawater and
197 melt pond water (Fig. S6). The sites in the NECS and AS shared 702 common
198 *Imitervirales* ASVs, while 357 and 871 unique ASVs were detected in the NECS and
199 AS sites, respectively (Fig. S7A). It was also shown that 515 *Imitervirales* ASVs were
200 shared between the bloom sites (A13, A14 and A15) and non-bloom sites (Fig. S7B),
201 while 319 and 1,096 ASVs were unique to the bloom sites and non-bloom sites,
202 respectively.

203 Shannon's diversity index was calculated for each community (Fig. S4).
204 Diversity of eukaryotic communities in the large size fraction showed the same
205 variation trend as those in the small fraction among different samples. The three
206 bloom sites in the AS had statistically lower diversity than others in both the large

(ANOVA followed by Tukey post hoc tests, $p<0.01$) and small eukaryotic communities (ANOVA followed by Tukey post hoc tests, $p<0.01$). The bloom sites had higher diversity of *Imitervirales* (4.34) than other sites (3.83) on average, although it was not statistically significant (ANOVA, $p=0.068$) (Fig. S4D).

Correlations with eukaryotic 18S community

Result of dbRDA (Fig. 4) and Spearman's rank correlation (Supplement Table S5) demonstrated that salinity and longitude were the two most significant variables in predicting the Chl *a* biomass. The Chl *a* concentration also showed positive correlations with phosphate and silicate (phosphate: $R = 0.73$, $p < 0.01$; silicate: $R = 0.598$, $p < 0.01$). Current velocity had no measurable influence on community variation in different waters ($p > 0.3$). Eukaryotic as well as *Imitervirales* communities in the AS and NECS were well separated from each other in a similar way as geographic distribution and TS diagram showed (Fig. S2).

According to the Mantel and partial Mantel tests, eukaryotic communities in both the large (3-144 μm) and small (0.2–3 μm) size fractions correlated significantly with *Imitervirales* communities in both the NECS and AS sites, even when the potential effects of spatial and environmental autocorrelations were removed ($q < 0.05$) (Table 1). Geographical distance was also a significant factor explaining the eukaryotic communities in the small fraction ($q < 0.05$), although no significant correlation was found for the large size fraction. For both the size fractions, environmental factors

228 were significant explanatory variables for the eukaryotic communities among the AS
229 sites, whereas no correlation was detected between environmental factors and
230 eukaryotic communities in the NECS sites. The Mantel test was also performed on the
231 eukaryotic 18S communities and each environmental factor (Table S7 and S8). All the
232 environmental factors in the NECS sites were not significantly correlated with the
233 eukaryotic 18S communities in the two size-fractions. In the AS sites, only phosphate
234 and silicate were significantly correlated with the eukaryotic 18S communities.
235
236 **Discussion**
237 **Basic environmental parameters and phytoplankton biomass**
238 Oligotrophy is one of the common features of surface sea water in the NECS.
239 Annual data (2008 to 2010) near the NECS indicated that concentration of nitrate plus
240 nitrite in surface sea water of the study area were mostly depleted in the summer with
241 the values between 0.01 and 0.1 μmol·L^{-1} (Fujiwara et al., 2014). In our study,
242 concentrations of nitrate plus nitrite also showed low values ($\leq$ 0.14 μmol·L^{-1}) except
243 for some bloom samples in the AS (0.27–3.12 μmol·L^{-1}) (Supplement Table S1). The
244 Chl *a* concentration, which is a proxy of phytoplankton biomass, was also low at the
245 nutrient-depleted stations (0.02–1.70 mg·m^{-3}) (Supplement Table S1), suggesting the
246 growth of phytoplankton was limited by nutrient availability (Ko et al., 2020). These
247 values were also consistent with recent Chl *a* data at the corresponding area obtained
248 from satellite (< 0.4 mg·m^{-3}) (Lee et al., 2019).

We separated the seawater sampling sites into two groups, NECS and AS, based on the geographical locations. The grouping was also supported by the TS diagram in which the NECS sites were characterized by lower salinity (Fig. S2). The Beaufort Gyre, which influences water properties in the NECS, is the greatest freshwater reservoir in the Arctic (Proshutinsky et al., 2019). On the other hands, samples having higher salinity were classified into the AS, because these locations would be more influenced by oceanic water masses and current regimes including the Pacific Water from the south and Atlantic Water from the west (Jones, 2001; Woodgate, 2013). It is also showed that the main reason of an increase of salinity and nutrient concentrations (resulting from summer algal blooms) in the oligotrophic northeastern Chukchi Sea surface water should be the intrusion of Atlantic cold saline water (Jung et al., 2021).

Community structures of microbial eukaryotes and *Imitervirales*

The eukaryotic communities were generally dominated by dinoflagellates, diatoms and other alveolates in the large size fraction (3-144 μm) and by ciliates and chlorophytes in the small size fraction (0.2–3 μm) (Fig. 3A and B). The dominance of these groups was roughly consistent with the previous studies that examined the microbial eukaryotic community structures in the Arctic Ocean by the molecular techniques (Comeau et al., 2011; Lovejoy et al., 2006; Marquardt et al., 2016; Onda et al., 2017; Xu et al., 2020) and the satellite ocean color remote sensing (Fujiwara et al., 2014; Lee et al., 2019). Diagnostic pigment signatures have indicated that

prasinophytes (Chlorophyta) were the dominant phytoplankton group in the northern Chukchi Sea, while diatoms and dinoflagellates were dominant in the southern Chukchi Sea (Fujiwara et al., 2014). Diatoms and chlorophytes are the common components of spring bloom in the Arctic Ocean (Von Quillfeldt, 2000). In our study, the phytoplankton communities in the bloom sites were dominated by unclassified marine alveolates (45.9% relative abundance) and diatoms (13.5%) in larger size fraction (Fig. 3A). The representative sequence of the unclassified marine alveolate ASV was best hit to the dinoflagellate *Heterocapsa rotundata* in the NCBI Reference RNA sequences database (2021/7/7 updated) (100% sequence similarity). Although this dinoflagellate species has been detected typically in the temperate estuaries (Kyeong et al., 2006; Millette et al., 2015), it was also found to be common near the study area by using microscopic technique (Ardyna et al., 2017).

Unanticipatedly high proportion of metazoan sequences were found in 3-144 μm eukaryotic group (Fig. S3). Most of them belong to copepods, which are predominant zooplankton in the Arctic Ocean (Kosobokova et al., 2011; Wang et al., 2019). However, body sizes of adult free-living copepods are usually above 200 μm, which cannot pass through the pre-filtration mesh. Although some of the copepod species (e.g., *Sphaeronellopsis monothrix*, 110 μm) are even smaller, they are the parasite of marine ostracods (Bowman & Kornicker, 1967). It is reported that smaller eggs of copepods are produced by the adults in spring and summer, and some of these may be float to the surface layer (Hirche & Niehoff, 1996). Thus, one possible explanation for

291 the dominance of metazoan sequences is the emergence of the larvae/eggs in the
292 seawater.

293 We detected significant differences in the eukaryotic community between the
294 NECS and AS for both size fractions by the dbRDA analysis (ANOSIM, $p < 0.01$)
295 (Fig. 4A and B). In the large size fraction, communities of the NECS sites were
296 consistently dominated by dinoflagellates, whereas the relative abundance of
297 dinoflagellates tended to be lower at the AS sites (Fig. 3A). In the small size fraction
298 (Fig. 3B), community difference between two sampling regimes was not obvious at
299 the phylum level, despite a clear separation by the dbRDA (Fig. 4B). These results
300 suggest that distinct ecosystem structures between the NECS and AS is likely caused
301 by the current systems and associated physicochemical characteristics. We also
302 evaluated eukaryotic communities from the two melt ponds on sea ices, which were
303 located nearby the two northernmost seawater sites (Fig. 3C). The communities were
304 largely distinct from the seawater communities, most likely reflecting the difference
305 in salinity between freshwater and seawater (Xu et al., 2020).

306 Besides eukaryotes, *Imitervirales* communities were analyzed in our study (Fig.
307 3C). Among *Imitervirales*, clades 2, 6, and 7 were abundant lineages at most of the
308 sampling sites (Figs 3C and S5). Intriguingly, these three dominant clades do not
309 include any reference species of *Imitervirales*. A previous study reported that the
310 Arctic Ocean is a hot spot of the endemic NCLDVs including *Imitervirales* (Endo et
311 al., 2020); the dominant phylotypes detected in our study may support the high

312 uniqueness of *Imitervirales* phylotypes in the study area. It is suggested that the
313 geographical distribution of viruses follow those of the host species (Ibarbalz et al.,
314 2019), the endemic feature is partly derived from the uniqueness of host eukaryotic
315 species. Community compositions of *Imitervirales* were also differentiated between
316 the NECS and AS stations by dbRDA analysis, as with the eukaryotic communities
317 (Fig. 4C). Expectedly, NMDS analysis (Fig. S6) clearly separated the *Imitervirales*
318 communities in the Arctic sites from those collected from coastal seawater and a hot
319 spring in Japan, which were evaluated using the same MEGAPRIMER method. This
320 separation would be due to the difference in host communities which are primarily
321 determined by the environmental conditions.
322
323 **Loose association between environmental variables and eukaryotic community**
324 In this study, salinity was the primary factor used for dividing the sites between
325 the NECS and AS (Fig. S2). Eukaryotic communities were also clearly separated
326 between the NECS and AS (Fig. 4A and B), indicating that the compositions of
327 eukaryotes were strongly influenced by the physical factors in the study area. Thus,
328 we separately assessed the relationship between eukaryotic community and
329 environmental variables or *Imitervirales* community for the NECS and AS to
330 eliminate possible autocorrelation caused by the difference of eukaryotic communities
331 among different water regimes.

In the AS sites, eukaryotic community was strongly correlated with environmental factor, but less correlated with geographical distance (Table 1). This suggests that the community was more affected by physicochemical environmental properties rather than dispersal events such as lateral advection among these sites. In fact, only the phosphate and silicate were significantly correlated with eukaryotic communities in the AS sites (Table S7 and S8). On the other hand, environmental factors (Table 1, S6, S7 and S8) did not show any association with eukaryotic communities in the NECS sites, whereas the effect of geographical distance was comparable to that detected in the AS sites. This indicates that other factors may be more important in making up the eukaryotic communities in the Beaufort Sea basin. In our study, all the sampling sites in the NECS were oligotrophic, and in some locations, the concentrations of nutrients were below the detection limit. Additionally, although temperature and salinity tend to be the key factors for microbial eukaryotic community structure and distribution in marine ecosystem (Caron et al., 2016; Sherr et al., 2007), these variables did not largely vary among the NECS sites. The low variation in environmental condition may cause the lack of correlation between environmental variables and the eukaryotic community.

Tight association between *Imitervirales* and the microbial eukaryotic community

In contrast to environmental variables, *Imitervirales* communities were consistently correlated with eukaryotic communities in both the NECS and AS

353 regions (Table 1). Notably, the correlation coefficients were rarely influenced by the
354 geographical and environmental factors, suggesting that *Imitervirales* were associated
355 with the eukaryotes in both types of water independently from environmental factors.
356 This trend was most pronounced at the stations in the NECS, where environmental
357 variables were relatively stable and had no correlation with eukaryotic community
358 variations. Our results support the idea that the communities of *Imitervirales* and
359 eukaryotes are actively interacting and co-varying without detectable influence from
360 the environmental conditions even in oligotrophic and homogeneous environments.

361 It has been suggested that biological interactions, such as predator-prey and
362 symbiotic interactions, are responsible to determine community structure and the
363 dynamics of microbes (Chaffron et al., 2020; Lima-Mendez et al., 2015). Additionally,
364 viruses have been proposed as a key factor influencing the protist communities as
365 they can impose top-down controls on their specific host populations (Brussaard et al.,
366 1996; Nagasaki et al., 1994). Recent studies using Mantel statistics or co-occurrence
367 network analysis indicated that *Imitervirales* are tightly associated with a variety of
368 protist lineages at a global level (Endo et al., 2020; Meng et al., 2021), although only
369 little of them have been isolated (Mihara et al., 2018). Our 18S rDNA barcoding
370 revealed that chlorophytes and haptophytes, both of which are known host lineages of
371 *Imitervirales*, were major protists in the small size fraction. Although the dominating
372 clades in the large-sized eukaryotic communities such as dinoflagellates, ciliophora,
373 and diatoms have not yet been reported as host lineages, these groups were predicted

374 to be the most closely linked host group for *Imitervirales* from a global scale network
375 analysis (Meng et al., 2021). Considering the highest proportion of *Imitervirales*
376 among NCLDVs in the global ocean and their potential role as a top-down factor on
377 host populations, relative compositions of the host lineages may well result from the
378 combination of a variety of specific infections of NCLDVs and other viruses.

379 In the Arctic Ocean, an increase in sea surface temperature and decrease in sea
380 ice cover are progressing (Peng et al., 2020; Praetorius et al., 2018). These climate
381 change has been shown to be associated with the shift of eukaryotic community
382 structure as well as the increase of biomass and the potential loss of biodiversity in the
383 past decade (Arrigo & van Dijken, 2015; Li et al., 2009; Majaneva et al., 2012),
384 although another study suggests a decreasing tendency on biomass (Hill et al., 2013).
385 Increased temperature may provide competitive advantage to small
386 nanophytoplankton over larger phytoplankton, resulted in an increase of the
387 contribution of small phytoplankton in the community (Hare et al., 2007; Li et al.,
388 2009). Our study showed that the association with *Imitervirales* community was
389 generally higher for the small-sized plankton community than for the large-sized
390 community, implying the role of *Imitervirales* in structuring the eukaryotic
391 community in the study area may become increasingly important in a future. However,
392 it should be noted that virus-host interactions can be influenced by the environments,
393 especially temperature (Demory et al., 2017, 2021).
394

Experimental procedures

Sampling sites and processes

During the Arctic Ocean Cruise of the IBRV Araon 2018 of Korean Polar Research Institute (KOPRI), surface water samples were collected (SBE32 carousel water sampler) at 21 stations from 6th to 22nd of August 2018. Environmental parameters including salinity, temperature, Chl *a* and nutrient concentration were obtained in parallel. Salinity and temperature were measured by the CTD sensors in situ measurement of seawater. For Chl *a*, seawater samples were collected in the upper 100 m depth and filtered through 47 mm GF/F filters, then was extracted with 90% acetone (Jung et al., 2021). Chl *a* was measured by fluorometer (Trilogy, Turner Designs, USA) (Lee et al., 2016). For nutrient concentration measurement, 50 ml seawater sample for each site was collected by conical tube, stored at 4˚C. Nitrite, nitrate, ammonia, phosphate, and silicate were measured using a four-channel continuous auto-analyzer (QuAAtro, Seal Analytical) followed the Joint Global Ocean Flux Study (JGOFS) protocols (Gordon et al., 1993). Nutrient concentrations under detection limit and lower than 0.005 $\mu mol \cdot L^{-1}$ were considered 0.

Seawater (1 L) for the DNA analysis was collected from 2 m depth with Niskin-bottles attached to a CTD–CMS system for all stations except at two closed melt ponds (500 mL), where water samples were collected just below the surface by bucket. Collected seawater was prefiltered with a 144 μm pore-size mesh to remove large particles (prewashed with ultrapure water). Two liters of water were separated

416 into two replicates on average, then were filtered through 3 μm Millipore membrane
417 filter by air pump (< 0.03MPa) for larger size fraction, further filtered through 0.2 μm
418 Millipore membrane filter with the same method for smaller sized fraction. The
419 membrane filters were transfer to 1.5 mL microtubes and then stored in -20˚C on
420 board and then transferred to the laboratory while continuously kept at -20˚C.
421
422 **DNA extraction and purification**
423 DNA extraction and purification were performed following (Endo et al., 2013; Endo
424 et al., 2018). Briefly, each membrane filter was thawed at room temperature and was
425 put into the 1.5 mL microtubes with glass beads and XS buffer. The cells on filter
426 were crushed with a beads beater and the mixture was incubated at 70˚C for 60 min.
427 Glass beads were removed from mixture after centrifugation. 600 μL isopropanol
428 were added to the supernatant and mixed. The precipitated DNA was purified with
429 NucleoSpin gDNA Clean-up Kit (Macherey-Nagel). Finally, the purified DNA was
430 dissolved in low TE buffer and stored at -20˚C.
431
432 **Eukaryotic 18S gene amplification and purification**
433 Eukaryotic 18S rRNA gene V4 region fragments were amplified from extracted DNA
434 of both 3 μm and 0.2 μm size fractions using primer E572F
435 (5'-CYGCGGTAATTCCAGCTC-3') and E1009R
436 (5'-AYGGTATCTRATCRTCTTYG-3') (Comeau et al., 2011) with attached Illumina

437 MiSeq 300 PE overhang reverse adapters as described in Illumina metagenomic
438 sequencing library preparation protocols.

439 12.5 µL 2x KAPA HiFi HotStart ReadyMix was mixed with 5 µL 1 µmol·L^{-1}
440 amplicon PCR forward primer, 5 µL 1µmol·L^{-1} amplicon PCR reverse primer and 2.5
441 µL diluted DNA samples (0.25 ng·μl^{-1}), and were added into a PCR tube (final
442 volume 25 µL). The amplification was performed for each sample with the following
443 temperature cycling condition: initial denaturation at 98˚C for 30 sec was followed
444 by 30 cycles of denaturation at 98°C for 10 sec, annealing at 55°C for 30 sec and
445 72°C for 30 sec. A final extension step was at 72°C for 5 min.

446 Amplicons were purified with magnetic beads (Agencourt AMPure XP beads,
447 Beckman Coulter, Inc.). The purified DNA were dissolved in 25 µL ultrapure water
448 and stored at -20˚C.

449

450 ***Imitervirales polB* gene amplification and purification**

451 The degenerated 82 *polB* primer pairs (MEGAPRIMER, Supplement Table S3) were
452 used to amplify the *polB* gene of *Imitervirales* from 0.2 µm membrane filter DNA
453 samples (Li et al., 2018). A previously optimized amplification method named
454 "MP10" (Supplement Table S4) was performed. amplification protocol, materials and
455 temperature cycling condition were the same as a previous work (Prodinger et al.,
456 2020).

After amplification, we merged all the eight amplicons generated from the same DNA sample using ethanol precipitation (Prodinger et al., 2020). Finally, the DNA precipitation was air dried for around 10 min and resuspended in 25 µL ultrapure water. Gel extraction was performed to remove unspecific amplification products. Gel electrophoresis was made by 2% agarose gel. The gel was then stained in 5000x diluted SYBR gold buffer for 12 min. 400-500 bp visible bands were cut from the gel. The Promega's Wizard SV Gel and PCR Clean-Up System was used to perform gel extraction according to the marker's protocol. DNA was dissolved in 25 µL ultrapure water, stored at -20˚C.

Index PCR, library construction and sequencing

Index PCR was performed following the Illumina Miseq platform protocol. Produced amplicons of 3-144 µm eukaryotes and *Imitervirales* were purified with the magnetic beads (Agencourt AMPure XP beads, Beckman Coulter, Inc.). Finial DNA production was dissolved in 27.5 µL ultrapure water, stored at -20˚C less than 24h. Produced amplicons of 0.2-3 µm eukaryotes were purified by gel, performed by Macrogen Corp. Japan.

DNA concentration was measured by Qubit HS (high-sensitive) kit. Library was denatured following the standard MiSeq normalization method provided by Illumina. The MiSeq Reagent Kit v2 and NaOH were used for the library with final DNA concentration of 2 nM. Paired-end sequencing was performed on the MiSeq platform.

478

479 **Sequence processing and bioinformatic analysis**

480 Eukaryotic 18S sequences are processed with QIIME2 (version: 2019.10) (Bolyen et
481 al., 2019). 260 bp of left pair reads and 220 bp of right pair reads were trimmed.
482 DADA2 was used to cut primer sequences, merging the paired end reads, performing
483 quality control, dereplication, chimera check, and Amplicon Sequence Variants
484 (ASVs) generation (Callahan et al., 2016; Knight et al., 2018). Singleton ASVs were
485 removed. Taxonomic annotation was done with QIIME2's vsearch (Rognes et al.,
486 2016) plugin and the SILVA 132 small subunit with 97% similarity database (Quast
487 et al., 2013) at 97% identity for species eukaryotic 18S datasets. Dominant ASVs
488 (reads percentage over 0.50% of each size fraction) were again searched by blastn
489 (Altschul et al., 1990) in NCBI Reference RNA sequences dataset, result include
490 detailed linage information with highest identity value was selected.

491 For the *Imitervirales* sequences, MAPS2 (*Mimiviridae* Amplicon Processing
492 System) was used for sequence analysis (Prodinger et al., unpublished). DADA2 was
493 used to check and remove megaprimer sequences, merging, quality control,
494 dereplication, chimera check, and non-singleton ASV output. The ASVs were aligned
495 against *Imitervirales polB* amino acid sequence database (Li et al., 2018). Nucleotide
496 sequences were translated into amino acid sequences and then added to reference
497 alignment using mafft (version: 7.453, parameters: --thread -1 --genafpair
498 --maxiterate 1000) (Katoh et al., 2002; Katoh & Standley, 2013). Sequences which

were assigned to the *Imitervirales* were saved for the further analysis, while other sequences were removed. Translated ASVs were placed into the reference phylogenetic tree built by the *polB* sequences from *Tara* Ocean dataset (Endo et al., 2020) by pplacer (version: 1.1. alpha19) (Matsen et al., 2010). Thirteen *Imitervirales* clades were manually defined in the tree, and ASVs were assigned to each clade. Phylogenetic tree was edit and output by iTOL v5.7 (Letunic & Bork, 2019).

Ecological analysis

Community composition was evaluated based on number of reads of each ASV in every sample. ASVs were then subsampled by the rarefy function ("vegan" package) (Oksanen et al., 2018) in R (version 3.6.3). Relative abundance was represented by the rate of each ASV reads percentage in each sample. Shannon diversity index of eukaryotic and *Imitervirales* community was calculated by R ("vegan" package) based on the subsampled ASV table. ANOVA and Tukey post hoc tests were performed by R ("agricolae" package). Composition bar chart and diversity bar chart with error bars of standard deviation were calculated with Microsoft Excel (version 16.41). The map of sampling stations, temperature-salinity (TS) diagram and heatmap of environmental factors and Shannon diversity were generated by Ocean Data View (ODV, version 5.1.5) (Schlitzer, R., Ocean Data View, https://odv.awi.de, 2018). Biological correlation was performed by dbRDA (distance-based redundancy analysis) function (Legendre & Andersson, 1999), using R ("vegan" package) based

on Bray-Curtis dissimilarity. For dbRDA ordination, ASV composition was normalized by Hellinger transformation by decostand function. Spearman's rank correlation was performed by R (cor.test function) and *p* value was also calculated by R (cor.test function). ANOSIM with 9,999 permutation was performed for biological data grouping test. Results of dbRDA were plotted by "ggord" with 95% confidence interval circle contained samples in different water types. Non-metric multidimensional scaling (NMDS) analysis of *Imitervirales* community was performed by R (monoMDS function) based on Bray-Curtis dissimilarity matrix made with the subsampled ASV table.

Mantel test and partial Mantel test (Mantel, 1967; Smouse et al., 1986) based on Pearson correlation coefficient were performed for calculating the correlation among geographic distance, environmental variables (i.e., a distance matrix combining temperature, salinity, dissolved inorganic nitrogen (DIN, nitrate + nitrite + ammonium nitrogen), phosphate, and silicate), eukaryotic community and *Imitervirales* community, using R ("ade4" package) (Bougeard & Dray, 2018) with permutations of 1,000. Geographic distance between each sampling station was calculated with R ("geosphere" package) from latitude and longitude data. Every environmental variable was normalized by $\log_{10}(x+1)$ function (x: the value of environmental factor). Euclidean distance of environmental factors and Bray-Curtis dissimilarity of subsampled relative abundances of eukaryotes and *Imitervirales* between sampling

sites were calculated with R. All *p* values were adjusted by the Holm's method (Holm, 1979) using R's p.adjust function.

Data availability

The raw reads generated in this study were uploaded to SRA (Sequence Read Archive) database on NCBI website. The accession numbers are from SRR12981736 to SRR12981758 under project ID PRJNA674408 (3-144 μm eukaryotic 18S), from SRR12981654 to SRR12981676 under project ID PRJNA674418 (0.2-3 μm eukaryotic 18S) and from SRR12981759 to SRR12981780 under project ID PRJNA674422 (0.2-3 μm *Imitervirales polB*).

Acknowledgments

We would like to thank colleagues from Korea Polar Research Institute for the help of sampling and physicochemical parameter determination; Tatsuhiro Isozaki, Kento Tominaga and Hiroaki Takebe from Laboratory of Marine Microbiology, Kyoto University, for helping with DNA sequencing and support of experiment. We also thank the captain and crew of the IBRV Araon Cruise for their support during the cruise. This work was supported by JSPS/KAKENHI (Nos. 18H02279 and 19H05667 to H.O., 17H03850 to T.Y. and H.O., and Nos. 19K15895 and 19H04263 to H.E.), and Scientific Research on Innovative Areas from the Ministry of Education, Culture, Science, Sports and Technology (MEXT) of Japan (Nos. 16H06429, 16K21723, and

561 16H06437 to H.O.). This research was also supported by the project
562 titled 'Korea-Arctic Ocean Warming and Response of Ecosystem (K-AWARE,
563 KOPRI, 1525011760)', funded by the Ministry of Oceans and Fisheries, Korea (KHC,
564 JJ, EJY, SHK). Computational work was completed at the SuperComputer System,
565 Institute for Chemical Research, Kyoto University. The authors declare no conflicts of
566 interest.

567

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.

Ardyna, M., Babin, M., Devred, E., Forest, A., Gosselin, M., Raimbault, P., & Tremblay, J. (2017). Shelf-basin gradients shape ecological phytoplankton niches and community composition in the coastal Arctic Ocean (Beaufort Sea). *Limnology and Oceanography*, *62*(5), 2113–2132.

Arrigo, K. R., & van Dijken, G. L. (2015). Continued increases in Arctic Ocean primary production. *Progress in Oceanography*, *136*, 60–70.

Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., … Caporaso, J. G. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nature Biotechnology*, *37*(8), 852–857.

Bougeard, S., & Dray, S. (2018). Supervised multiblock analysis in R with the ade4 package. *Journal of Statistical Software*, *86*(1), 1–17.

Bowman, T. E., & Kornicker, L. S. (1967). Two New Crustaceans: The Parasitic Copepod Sphaeronellopsis monothrix (Choniostomatidae) and lts Myodocopid Ostracod Host Parasterope pollex (Cylindroleberidae) from the Southern New

588 England Coast. Proceedings of the United States National Museum. *Proceedings*
589 *of the United States National Museum*, *123*(3613), 1– 28.
590 Brussaard, C. P. D., Kempers, R. S., Kop, A. J., Riegman, R., & Heldal, M. (1996).
591 Virus-like particles in a summer bloom of Emiliania huxleyi in the North Sea.
592 *Aquatic Microbial Ecology*, *10*(2), 105–113.
593 Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., &
594 Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina
595 amplicon data. *Nature Methods*, *13*(7), 581–583.
596 Caron, D. A., Alexander, H., Allen, A. E., Archibald, J. M., Armbrust, E. V., Bachy,
597 C., Bell, C. J., Bharti, A., Dyhrman, S. T., Guida, S. M., Heidelberg, K. B., Kaye,
598 J. Z., Metzner, J., Smith, S. R., & Worden, A. Z. (2016). Probing the evolution,
599 ecology and physiology of marine protists using transcriptomics. *Nature Reviews*
600 *Microbiology*, *15*(1), 6–20.
601 Chaffron, S., Delage, E., Budinich, M., Vintache, D., Henry, N., Nef, C., Ardyna, M.,
602 Zayed, A., Junger, P., Galand, P., Lovejoy, C., Murray, A., Sarmento, H., Oceans
603 coordinators, T., Acinas, S., Babin, M., Iudicone, D., Jaillon, O., Karsenti, E., …
604 Eveillard, D. (2020). Environmental vulnerability of the global ocean plankton
605 community interactome. *BioRxiv*, 2020.11.09.375295.
606 Christopher B. Field, Michael J. Behrenfeld, James T. Randerson, P. F. (1998).
607 Primary Production of the Biosphere: Integrating Terrestrial and Oceanic
608 Components. *Biochemical Society Transactions*, *281*, 237–240.

609 Clerissi, C., Desdevises, Y., & Grimsley, N. (2012). Prasinoviruses of the Marine
610 Green Alga Ostreococcus tauri Are Mainly Species Specific. *Journal of Virology*,
611 *86*(8), 4611–4619.
612 Comeau, A. M., Li, W. K. W., Tremblay, J. É., Carmack, E. C., & Lovejoy, C. (2011).
613 Arctic ocean microbial community structure before and after the 2007 record sea
614 ice minimum. *PLoS ONE*, *6*(11).
615 Curry, J. A., Schramm, J. L., & Ebert, E. E. (1995). Sea ice-albedo climate feedback
616 mechanism. In *Journal of Climate*. *8*(2), 240–247.
617 de Vargas, C., Audic, S., Nicolas Henry, Decelle, J., Mahé, F., Ramiro Logares,
618 Enrique Lara, & Cédric Berney. (2015). Eukaryotic plankton diversity in the
619 sunlit ocean. *Science*, *348*(6237), 1261605-1/11.
620 Demory, D., Arsenieff, L., Simon, N., Six, C., Rigaut-Jalabert, F., Marie, D., Ge, P.,
621 Bigeard, E., Jacquet, S., Sciandra, A., Bernard, O., Rabouille, S., & Baudoux, A.
622 C. (2017). Temperature is a key factor in Micromonas-virus interactions. *ISME*
623 *Journal*, *11*(3), 601–612.
624 Demory, D., Weitz, J. S., Baudoux, A. C., Touzeau, S., Simon, N., Rabouille, S.,
625 Sciandra, A., & Bernard, O. (2021). A thermal trade-off between viral
626 production and degradation drives virus-phytoplankton population dynamics.
627 *Ecology Letters*, *24*(6), 1133–1144.
628 Endo, H., Yoshimura, T., Kataoka, T., & Suzuki, K. (2013). Effects of CO2 and iron
629 availability on phytoplankton and eubacterial community compositions in the

630 northwest subarctic Pacific. *Journal of Experimental Marine Biology and*
631 *Ecology*, *439*, 160–175.
632 Endo, Hisashi, Blanc-Mathieu, R., Li, Y., Salazar, G., Henry, N., Labadie, K., de
633 Vargas, C., Sullivan, M. B., Bowler, C., Wincker, P., Karp-Boss, L., Sunagawa,
634 S., & Ogata, H. (2020). Biogeography of marine giant viruses reveals their
635 interplay with eukaryotes and ecological functions. *Nature Ecology and*
636 *Evolution*, *4*(12), 1639–1649.
637 Endo, Hisashi, Ogata, H., & Suzuki, K. (2018). Contrasting biogeography and
638 diversity patterns between diatoms and haptophytes in the central Pacific Ocean.
639 *Scientific Reports*, *8*(10916), 1–13.
640 Falkowski, P. G., Barber, R. T., & Smetacek, V. (1998). Biogeochemical controls and
641 feedbacks on ocean primary production. *Science*, *281*(5374), 200–206.
642 Fujiwara, A., Hirawake, T., Suzuki, K., Imai, I., & Saitoh, S. I. (2014). Timing of sea
643 ice retreat can alter phytoplankton community structure in the western Arctic
644 Ocean. *Biogeosciences*, *11*(7), 1705–1716.
645 Gordon, L. I., Jennings, J. C., Ross, A. A., & Krest, J. M. (1993). A suggested
646 protocol for continuous flow automated analysis of seawater nutrients (phosphate,
647 nitrate, nitrite and silicic acid) in the WOCE Hydrographic. *WOCE Operations*
648 *Manual*, *3*(3.1), Part 3.1.3.

649 Hamilton, A. K., Lovejoy, C., Galand, P. E., & Ingram, R. G. (2008). Water masses
650 and biogeography of picoeukaryote assemblages in a cold hydrographically
651 complex system. *Limnology and Oceanography*, *53*(3), 922–935.
652 Hare, C. E., Leblanc, K., DiTullio, G. R., Kudela, R. M., Zhang, Y., Lee, P. A.,
653 Riseman, S., & Hutchins, D. A. (2007). Consequences of increased temperature
654 and CO2 for phytoplankton community structure in the Bering Sea. *Marine*
655 *Ecology Progress Series*, *352*, 9–16.
656 Hill, V. J., Matrai, P. A., Olson, E., Suttles, S., Steele, M., Codispoti, L. A., &
657 Zimmerman, R. C. (2013). Synthesis of integrated primary production in the
658 Arctic Ocean: II. In situ and remotely sensed estimates. *Progress in*
659 *Oceanography*, *110*, 107–125.
660 Hirche, H. J., & Niehoff, B. (1996). Reproduction of the Arctic copepod Calanus
661 hyperboreus in the Greenland Sea-field and laboratory observations. *Polar*
662 *Biology*, *16*(3), 209–219.
663 Holm, S. (1979). A simple sequentially rejective multiple test procedure.
664 *Scandinavian Journal of Statistics*, *6*(2), 65–70.
665 Ibarbalz, F. M., Henry, N., Lombard, F., Bowler, C., Zinger, L., Busseni, G., & Byrne,
666 H. (2019). Global Trends in Marine Plankton Diversity across Kingdoms of Life
667 AR OCEANS EXPEDITION Article Global Trends in Marine Plankton
668 Diversity across Kingdoms of Life. *Cell*, *179*, 1084–1097.

669 Iyer, L. M., Balaji, S., Koonin, E. V., & Aravind, L. (2006). Evolutionary genomics of
670 nucleo-cytoplasmic large DNA viruses. *Virus Research*, *117*(1), 156–184.
671 Joli, N., Gosselin, M., Ardyna, M., Babin, M., Onda, D. F., Tremblay, J. É., &
672 Lovejoy, C. (2018). Need for focus on microbial species following ice melt and
673 changing freshwater regimes in a Janus Arctic Gateway. *Scientific Reports*,
674 *8*(9405), 1–11.
675 Jones, E. P. (2001). Circulation in the Arctic Ocean. *Polar Research*, *20*(2), 139–146.
676 Jung, J., Cho, K. H., Park, T., Yoshizawa, E., Lee, Y., Yang, E. J., Gal, J. K., Ha, S.
677 Y., Kim, S., Kang, S. H., & Grebmeier, J. M. (2021). Atlantic-Origin Cold Saline
678 Water Intrusion and Shoaling of the Nutricline in the Pacific Arctic. *Geophysical*
679 *Research Letters*, *48*(6), 1–10.
680 Kashiwase, H., Ohshima, K. I., Nihashi, S., & Eicken, H. (2017). Evidence for
681 ice-ocean albedo feedback in the Arctic Ocean shifting to a seasonal ice zone.
682 *Scientific Reports*, *7*(8170), 1–10.
683 Katoh, K., Misawa, K., Kuma, K. I., & Miyata, T. (2002). MAFFT: A novel method
684 for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic*
685 *Acids Research*, *30*(14), 3059–3066.
686 Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software
687 version 7: Improvements in performance and usability. *Molecular Biology and*
688 *Evolution*, *30*(4), 772–780.

689 Kilias, E., Kattner, G., Wolf, C., Frickenhaus, S., & Metfies, K. (2014). A molecular
690 survey of protist diversity through the central Arctic Ocean. *Polar Biology*, *37*(9),
691 1271–1287.
692 Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J.,
693 Gonzalez, A., Kosciolek, T., McCall, L. I., McDonald, D., Melnik, A. V.,
694 Morton, J. T., Navas, J., Quinn, R. A., Sanders, J. G., Swafford, A. D.,
695 Thompson, L. R., Tripathi, A., Xu, Z. Z., … Dorrestein, P. C. (2018). Best
696 practices for analysing microbiomes. *Nature Reviews Microbiology*, *16*(7),
697 410–422.
698 Ko, E., Gorbunov, M. Y., Jung, J., Joo, H. M., Lee, Y., Cho, K. H., Yang, E. J., Kang,
699 S. H., & Park, J. (2020). Effects of Nitrogen Limitation on Phytoplankton
700 Physiology in the Western Arctic Ocean in Summer. In *Journal of Geophysical*
701 *Research: Oceans*. *125*(11).
702 Kosobokova, K. N., Hopcroft, R. R., & Hirche, H. J. (2011). Patterns of zooplankton
703 diversity through the depths of the Arctic's central basins. *Marine Biodiversity*,
704 *41*(1), 29–50.
705 Kwok, R., & Cunningham, G. F. (2010). Contribution of melt in the Beaufort Sea to
706 the decline in Arctic multiyear sea ice coverage: 1993-2009. *Geophysical*
707 *Research Letters*, *37*(20), 1–5.

708 Kyeong, A. S., Hae, J. J., Kim, S., Gwang, H. K., & Jung, H. K. (2006). Bacterivory
709 by co-occurring red-tide algae, heterotrophic nanoflagellates, and ciliates.
710 *Marine Ecology Progress Series*, *322*(September), 85–97.
711 Lannuzel, D., Tedesco, L., van Leeuwe, M., Campbell, K., Flores, H., Delille, B.,
712 Miller, L., Stefels, J., Assmy, P., Bowman, J., Brown, K., Castellani, G., Chierici,
713 M., Crabeck, O., Damm, E., Else, B., Fransson, A., Fripiat, F., Geilfus, N. X., …
714 Wongpan, P. (2020). The future of Arctic sea-ice biogeochemistry and
715 ice-associated ecosystems. *Nature Climate Change*, *10*(11), 983–992.
716 Lee, J., Kang, S. H., Yang, E. J., Macdonald, A. M., Joo, H. M., Park, J., Kim, K., Lee,
717 G. S., Kim, J. H., Yoon, J. E., Kim, S. S., Lim, J. H., & Kim, I. N. (2019).
718 Latitudinal Distributions and Controls of Bacterial Community Composition
719 during the Summer of 2017 in Western Arctic Surface Waters (from the Bering
720 Strait to the Chukchi Borderland). *Scientific Reports*, *9*(16822), 1–10.
721 Lee, Y., Min, J. O., Yang, E. J., Cho, K. H., Jung, J., Park, J., Moon, J. K., & Kang, S.
722 H. (2019). Influence of sea ice concentration on phytoplankton community
723 structure in the Chukchi and East Siberian Seas, Pacific Arctic Ocean. *Deep-Sea*
724 *Research Part I: Oceanographic Research Papers*, *147*(June 2018), 54–64.
725 Lee, Y., Yang, E. J., Park, J., Jung, J., Kim, T. W., & Lee, S. H. (2016).
726 Physical-biological coupling in the Amundsen Sea, Antarctica: Influence of
727 physical factors on phytoplankton community structure and biomass. *Deep-Sea*
728 *Research Part I: Oceanographic Research Papers*, *117*, 51–60.

729 Legendre, P., & Andersson, M. J. (1999). Distance-based redundancy analysis:
730 Testing multispecies responses in multifactorial ecological experiments.
731 *Ecological Monographs*, *69*(1), 1–24.
732 Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and
733 new developments. *Nucleic Acids Research*, *47*(W1), 2–5.
734 Lewis, K. M., Van Dijken, G. L., & Arrigo, K. R. (2020). Changes in phytoplankton
735 concentration now drive increased Arctic Ocean primary production. *Science*,
736 *369*(6500), 198–202.
737 Li, W. K. W., McLaughlin, F. A., Lovejoy, C., & Carmack, E. C. (2009). Smallest
738 algae thrive as the arctic ocean freshens. *Science*, *326*(5952), 539.
739 Li, Y., Endo, H., Gotoh, Y., Watai, H., Ogawa, N., Blanc-Mathieu, R., Yoshida, T., &
740 Ogata, H. (2019). The earth is small for “leviathans”: Long distance dispersal of
741 giant viruses across aquatic environments. *Microbes and Environments*, *34*(3),
742 334–339.
743 Li, Y., Hingamp, P., Watai, H., Endo, H., Yoshida, T., & Ogata, H. (2018).
744 Degenerate PCR primers to reveal the diversity of giant viruses in coastal waters.
745 *Viruses*, *10*(9), 1–16.
746 Lima-Mendez, G., Faust, K., Henry, N., Decelle, J., Colin, S., Carcillo, F., Chaffron,
747 S., Cesar Ignacio-Espinosa, J., Roux, S., Vincent, F., Bittner, L., Darzi, Y., Wang,
748 J., Audic, S., Berline, L., Bontempi, G., Cabello, A. M., Coppola, L.,

749 Cornejo-Castillo, F. M., … Raes, J. (2015). 24 Silvia G. Acinas, 12 Shinichi
750 Sunagawa, 17 Peer Bork. *Science*, *10*(6237), 1–10.
751 Lindsay, R., Haas, C., Hendricks, S., Hunkeler, P., Kurtz, N., Paden, J., Panzer, B.,
752 Sonntag, J., Yungel, J., & Zhang, J. (2012). Seasonal forecasts of Arctic sea ice
753 initialized with observations of ice thickness. *Geophysical Research Letters*,
754 *39*(21), 1–6.
755 Lovejoy, C., Massana, R., & Pedrós-Alió, C. (2006). Diversity and distribution of
756 marine microbial eukaryotes in the arctic ocean and adjacent seas. *Applied and*
757 *Environmental Microbiology*, *72*(5), 3085–3095.
758 Majaneva, M., Rintala, J. M., Piisilä, M., Fewer, D. P., & Blomster, J. (2012).
759 Comparison of wintertime eukaryotic community from sea ice and open water in
760 the Baltic Sea, based on sequencing of the 18S rRNA gene. *Polar Biology*, *35*(6),
761 875–889.
762 Mantel, N. (1967). The Detection of Disease Clustering and a Generalized Regression
763 Approach. *Nature*, *27*(Part 1), 209–220.
764 Marquardt, M., Vader, A., Stübner, E. I., Reigstad, M., & Gabrielsen, T. M. (2016).
765 Strong seasonality of marine microbial eukaryotes in a high-Arctic fjord
766 (Isfjorden, in West Spitsbergen, Norway). *Applied and Environmental*
767 *Microbiology*, *82*(6), 1868–1880.

768 Matsen, F. A., Kodner, R. B., & Armbrust, E. V. (2010). pplacer: linear time
769 maximum-likelihood and Bayesian phylogenetic placement of sequences onto a
770 fixed reference tree. *BMC Bioinformatics*, *11*(1), 538.
771 Meng, L., Endo, H., Blanc-Mathieu, R., Chaffron, S., Hernández-Velázquez, R.,
772 Kaneko, H., & Ogata, H. (2021). Quantitative Assessment of Nucleocytoplasmic
773 Large DNA Virus and Host Interactions Predicted by Co-occurrence Analyses.
774 *MSphere*, *6*(2), 1–18.
775 Middelboe, M., & Brussaard, C. P. D. (2017). Marine viruses: Key players in marine
776 ecosystems. *Viruses*, *9*(10), 1–6.
777 Mihara, T., Koyano, H., Hingamp, P., Grimsley, N., Goto, S., & Ogata, H. (2018).
778 Taxon richness of "Megaviridae" exceeds those of bacteria and archaea in the
779 ocean. *Microbes and Environments*, *33*(2), 162–171.
780 Millette, N. C., Stoecker, D. K., & Pierson, J. J. (2015). Top-down control by micro-
781 and mesozooplankton on winter dinoflagellate blooms of Heterocapsa rotundata.
782 *Aquatic Microbial Ecology*, *76*(1), 15–25.
783 Monier, A., Terrado, R., Thaler, M., Comeau, A., Medrinal, E., & Lovejoy, C. (2013).
784 Upper Arctic Ocean water masses harbor distinct communities of heterotrophic
785 flagellates. *Biogeosciences*, *10*(6), 4273–4286.
786 Münchow, A. (2016). Volume and freshwater flux observations from Nares Strait to
787 the west of Greenland at daily time scales from 2003 to 2009. *Journal of*
788 *Physical Oceanography*, *46*(1), 141–157.

789 Nagasaki, K., Ando, M., Itakura, S., Imai, I., & Ishida, Y. (1994). Viral mortality in
790 the final stage of Heterosigma akashiwo (Raphidophyceae) red tide. *Journal of*
791 *Plankton Research*, *16*(11), 1595–1599.
792 Oksanen, A. J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D.,
793 Minchin, P. R., Hara, R. B. O., Simpson, G. L., Solymos, P., Stevens, M. H. H.,
794 & Szoecs, E. (2018). *Package ' vegan .' April*.
795 Onda, D. F. L., Medrinal, E., Comeau, A. M., Thaler, M., Babin, M., & Lovejoy, C.
796 (2017). Seasonal and interannual changes in ciliate and dinoflagellate species
797 assemblages in the Arctic Ocean (Amundsen Gulf, Beaufort Sea, Canada).
798 *Frontiers in Marine Science*, *4*(16), 1–14.
799 Peng, H. T., Ke, C. Q., Shen, X., Li, M., & Shao, Z. De. (2020). Summer albedo
800 variations in the Arctic Sea ice region from 1982 to 2015. *International Journal*
801 *of Climatology*, *40*(6), 3008–3020.
802 Perovich, D. K., & Jones, K. F. (2014). Arctic Report 2019. *Journal of Geophysical*
803 *Research: Oceans*, *119*(12), 8767–8777.
804 Praetorius, S., Rugenstein, M., Persad, G., & Caldeira, K. (2018). Global and Arctic
805 climate sensitivity enhanced by changes in North Pacific heat flux. *Nature*
806 *Communications*, *9*(1), 1–12.
807 Prodinger, F., Endo, H., Gotoh, Y., Li, Y., Morimoto, D., Omae, K., Tominaga, K.,
808 Blanc-Mathieu, R., Takano, Y., Hayashi, T., Nagasaki, K., Yoshida, T., & Ogata,

809 H. (2020). An optimized metabarcoding method for mimiviridae.
810 *Microorganisms*, *8*(4), 1–17.
811 Prodinger, F., Endo, H., Takano, Y., Li, Y., Tominaga, K., Isozaki, T., Blanc-Mathieu,
812 R., Gotoh, Y., Tetsuya, H., Taniguchi, E., Nagasaki, K., Yoshida, T., & Ogata, H.
813 (2021). Year-round dynamics of amplicon sequence variant communities differ
814 among eukaryotes, Mimiviridae, and prokaryotes in a coastal ecosystem.
815 *BioRxiv*.
816 Proshutinsky, A., Krishfield, R., Toole, J. M., Timmermans, M. L., Williams, W.,
817 Zimmermann, S., Yamamoto-Kawai, M., Armitage, T. W. K., Dukhovskoy, D.,
818 Golubeva, E., Manucharyan, G. E., Platov, G., Watanabe, E., Kikuchi, T.,
819 Nishino, S., Itoh, M., Kang, S. H., Cho, K. H., Tateyama, K., & Zhao, J. (2019).
820 Analysis of the Beaufort Gyre Freshwater Content in 2003–2018. *Journal of*
821 *Geophysical Research: Oceans*, *124*(12), 9658–9689.
822 Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., Peplies, J., &
823 Glöckner, F. O. (2013). The SILVA ribosomal RNA gene database project:
824 Improved data processing and web-based tools. *Nucleic Acids Research*, *41*(D1),
825 590–596.
826 Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: A
827 versatile open source tool for metagenomics. *PeerJ*, *2016*(10), 1–22.
828 Rohwer, F., & Thurber, R. V. (2009). Viruses manipulate the marine environment.
829 *Nature*, *459*(7244), 207–212.

830 Sandaa, R. A., Storesund, J. E., Olesin, E., Paulsen, M. L., Larsen, A., Bratbak, G., &
831 Ray, J. L. (2018). Seasonality drives microbial community structure, shaping
832 both eukaryotic and prokaryotic host–viral relationships in an arctic marine
833 ecosystem. *Viruses*, *10*(12), 1–22.
834 Sherr, B. F., Sherr, E. B., Caron, D. A., Vaulot, D., & Worden, A. Z. (2007). Oceanic
835 protists. *Oceanography*, *20*(2), 130–134.
836 Shu, Q., Qiao, F., Song, Z., Zhao, J., & Li, X. (2018). Projected Freshening of the
837 Arctic Ocean in the 21st Century. *Journal of Geophysical Research: Oceans*,
838 *123*(12), 9232–9244.
839 Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). *Multiple Regression and*
840 *Correlation Mantel Test of Matrix Correspondence*. *35*(4), 627–632.
841 Stroeve, J., Holland, M. M., Meier, W., Scambos, T., & Serreze, M. (2007). Arctic sea
842 ice decline: Faster than forecast. *Geophysical Research Letters*, *34*(9), 1–5.
843 Suttle, C. A. (2007). Marine viruses - Major players in the global ecosystem. *Nature*
844 *Reviews Microbiology*, *5*(10), 801–812.
845 Terrado, R., Vincent, W. F., & Lovejoy, C. (2009). Mesopelagic protists: Diversity
846 and succession in a coastal Arctic ecosystem. *Aquatic Microbial Ecology*, *56*(1),
847 25–40.
848 Thaler, M., & Lovejoy, C. (2013). Environmental selection of marine stramenopile
849 clades in the Arctic Ocean and coastal waters. *Polar Biology*, *37*(3), 347–357.

850 Von Quillfeldt, C. H. (2000). Common diatom species in Arctic spring blooms: Their
851 distribution and abundance. *Botanica Marina*, *43*(6), 499–516.
852 Wang, Y. G., Tseng, L. C., Lin, M., & Hwang, J. S. (2019). Vertical and geographic
853 distribution of copepod communities at late summer in the Amerasian Basin,
854 Arctic Ocean. *PLoS ONE*, *14*(7), 1–23.
855 Woodgate, R. (2013). Arctic Ocean circulation : going around at the top of the world.
856 *Nature Education Knowledge*, *4*(8), 8.
857 Worden, A. Z., Follows, M. J., Giovannoni, S. J., Wilken, S., Zimmerman, A. E., &
858 Keeling, P. J. (2015). Rethinking the marine carbon cycle: Factoring in the
859 multifarious lifestyles of microbes. *Science*, *347*(6223), 735–748.
860 Xu, D., Kong, H., Yang, E. J., Li, X., Jiao, N., Warren, A., Wang, Y., Lee, Y., Jung, J.,
861 & Kang, S. H. (2020). Contrasting Community Composition of Active Microbial
862 Eukaryotes in Melt Ponds and Sea Water of the Arctic Ocean Revealed by High
863 Throughput Sequencing. *Frontiers in Microbiology*, *11*(June), 1–15.
864 Zhang, F., He, J., Lin, L., & Jin, H. (2015). Dominance of picophytoplankton in the
865 newly open surface water of the central Arctic Ocean. *Polar Biology*, *38*(7),
866 1081–1089.

867
868

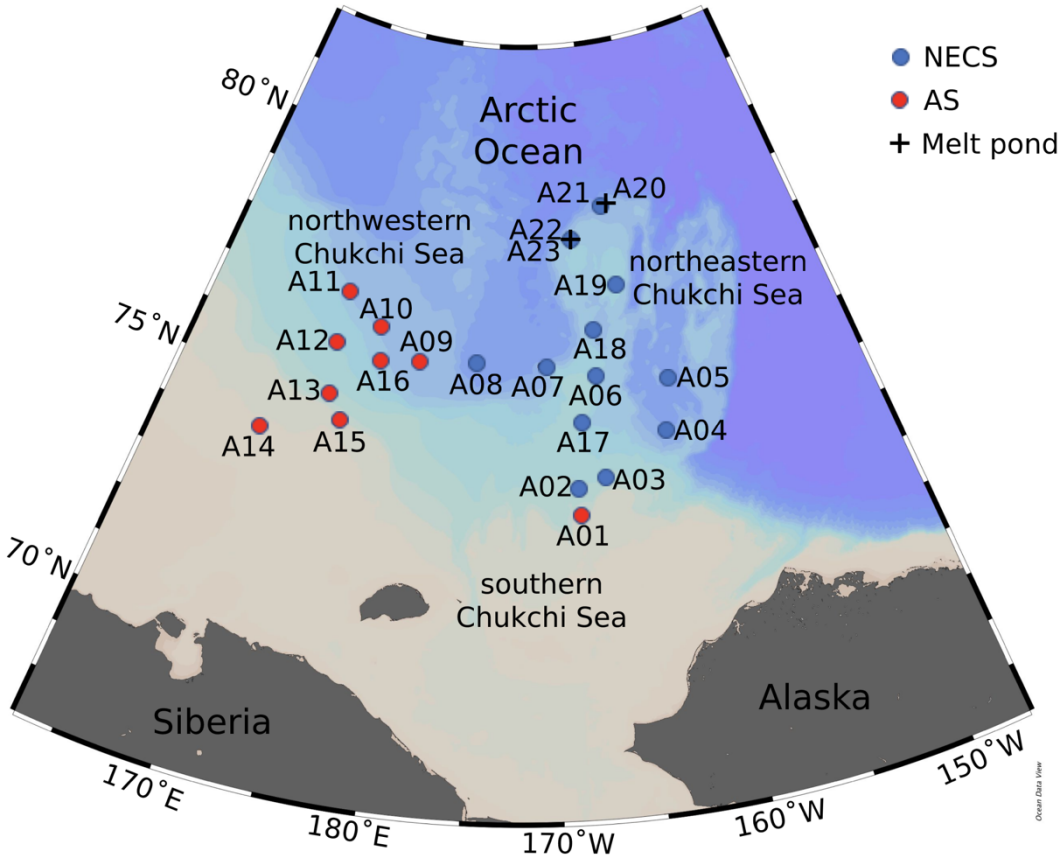
Table 1. R value of Mantel test among different eukaryotic 18S communities in two types of water (separated based on temperature-salinity diagram), geographic factors, environmental factors (including temperature, salinity, dissolved inorganic nitrogen (NO_2+NO_3+NH_4), phosphate and silicate) and *Imitervirales polB* gene community. "*Imitervirales*/geographic" and "*Imitervirales*/environmental" represent the R value of Partial Mantel test, influences of geographic or environmental factors were removed. (q-values were listed in Supplement Table S6)

	NECS (N=12)	AS (N=9)
3-144μm eukaryote		
geographic factor	0.30	0.34
environmental factor	-0.02	0.75**
Imitervirales	0.66**	0.70**
Imitervirales/geographic	0.63**	0.66**
Imitervirales/environmental	0.66**	0.50*
0.2-3μm eukaryote		
geographic factor	0.44**	0.30*
environmental factor	-0.06	0.86**
Imitervirales	0.65***	0.87***
Imitervirales/geographic	0.61***	0.86***
Imitervirales/environmental	0.65***	0.90***

*: q<0.05; **: q<0.01; ***: q<0.001.

879

880 Figures



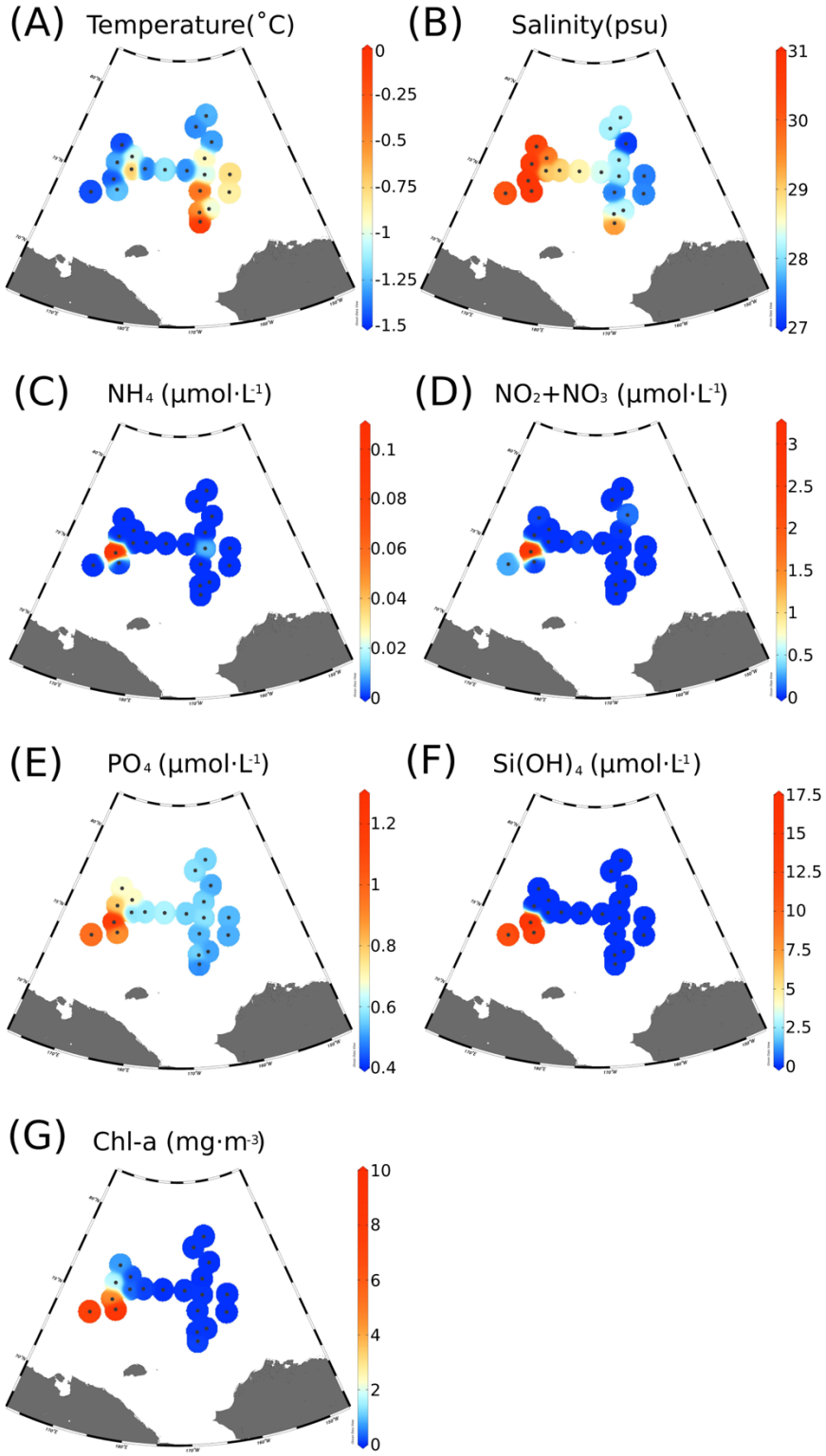
881

882 Fig. 1. Map of Arctic sampling stations of this study. Symbol colors represent water

883 types with different characteristics influenced by current system in this area. NECS:

884 northeastern Chukchi Sea; AS: adjacent sea outside Beaufort Gyre.

885

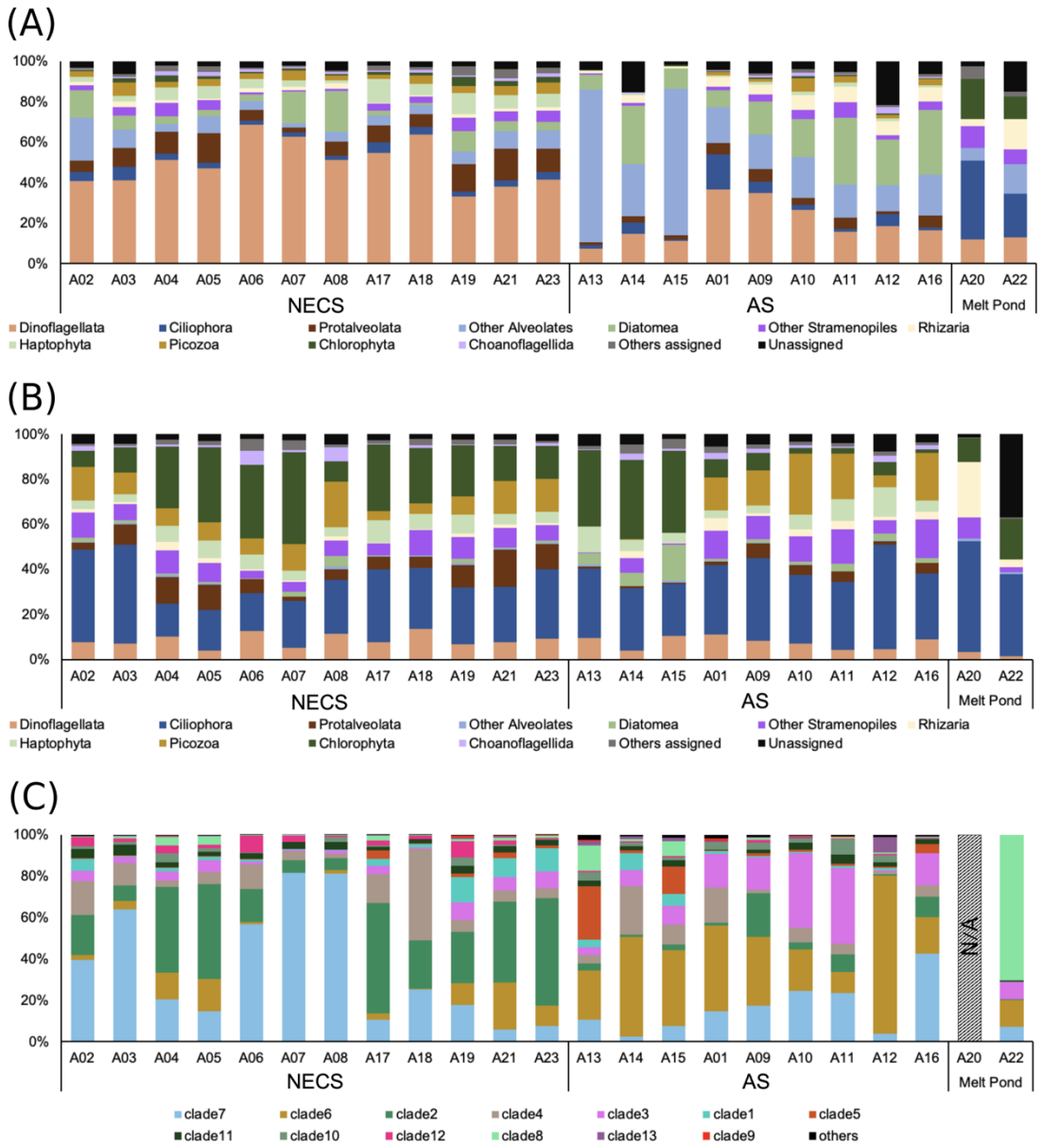


886

887 Fig. 2. Physical and chemical environmental variables among sampling sites: (A)

888 temperature (°C); (B) salinity (psu); (C) NH_4 (μmol·L^{-1}); (D) NO_2 and NO_3

889 (μmol·L^{-1}); (E) PO_4 (μmol·L^{-1}); (F) $Si(OH)_4$ (μmol·L^{-1}); (G) Chl *a* (mg·m^{-3}).



890

891 Fig. 3. Community compositions of eukaryotes and *Imitervirales*. Relative

892 compositions of eukaryotes at phylum level in (A) 3-144 μm fraction and (B) 0.2-3

893 μm fraction and (C) *Imitervirales* in the clade level. The color of each clade of the

894 *Imitervirales* is the same with phylogenetic tree (Fig. S5). Fungi and Metazoa

895 sequences were removed from eukaryotic sequences.

896

897

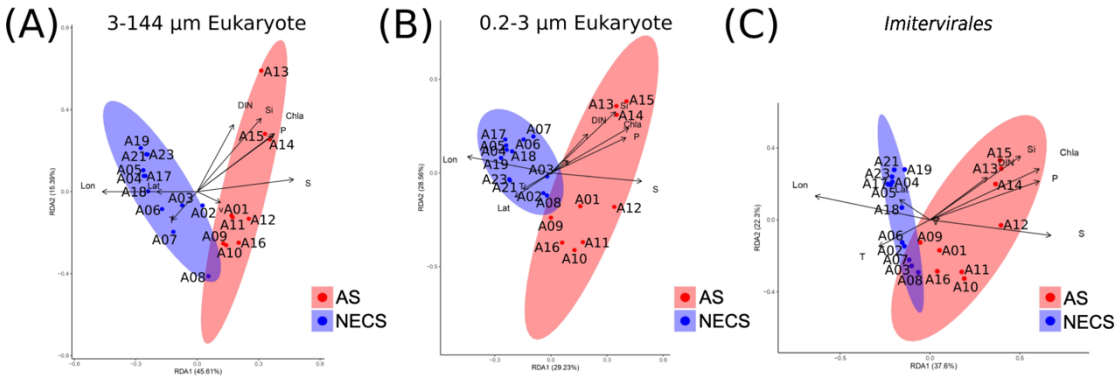
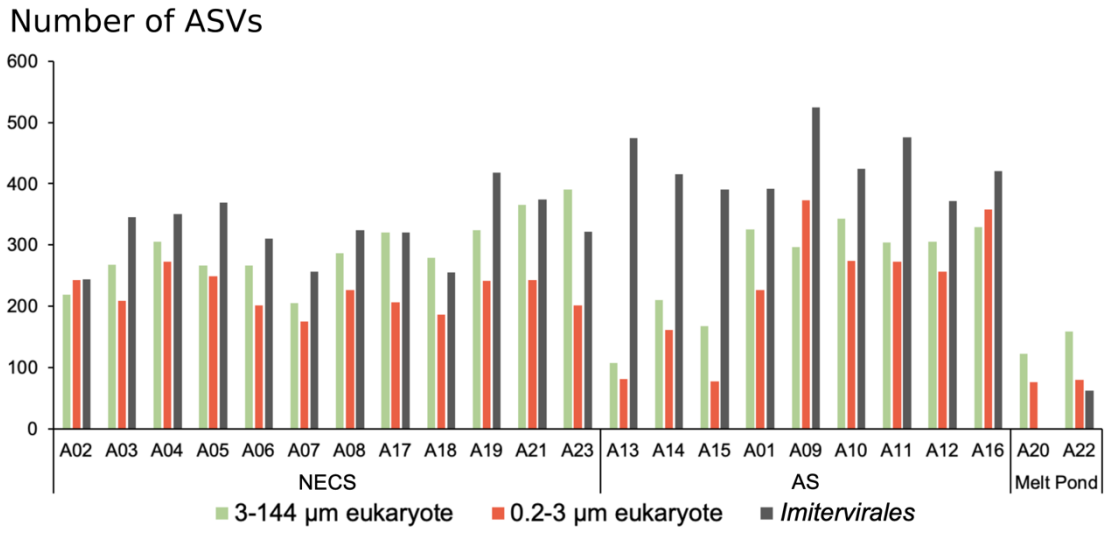


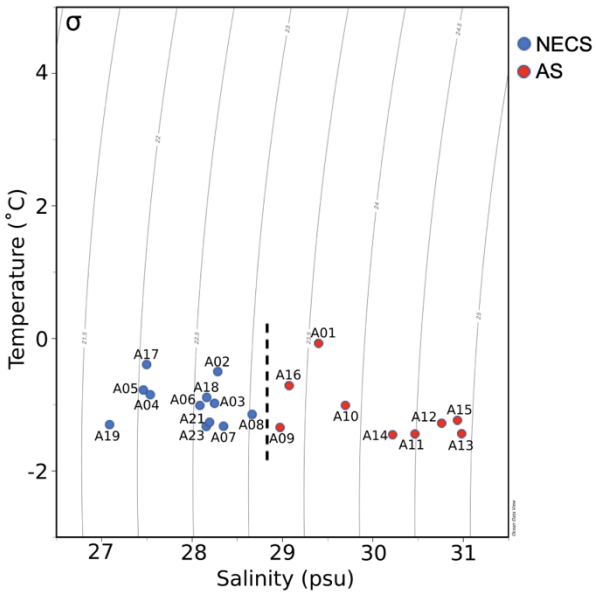
Fig. 4. dbRDA (distance-based redundancy analysis) ordination diagram of (A) 3-144 µm eukaryotic community based on 18S ASVs; (B) 0.2-3 µm eukaryotic community based on 18S ASVs; (C) *Imitervirales* community based on *polB* gene ASVs. Abbreviation of water types: NECS: northeastern Chukchi Sea; AS: adjacent sea outside Beaufort Gyre. Abbreviation of geographic and environmental factors: Lat: latitude; Lon: Longitude; T: temperature; S: salinity; DIN: dissolved inorganic nitrogen (sum of ammonia), nitrite and nitrate; P: phosphate; Si: silicate; Chl *a*: chlorophyll *a*; v: density current velocity.



909 Fig. S1. Number of non-singleton ASVs of each sample before subsampling. NECS:

910 northeastern Chukchi Sea; AS: adjacent sea outside Beaufort Gyre.

911



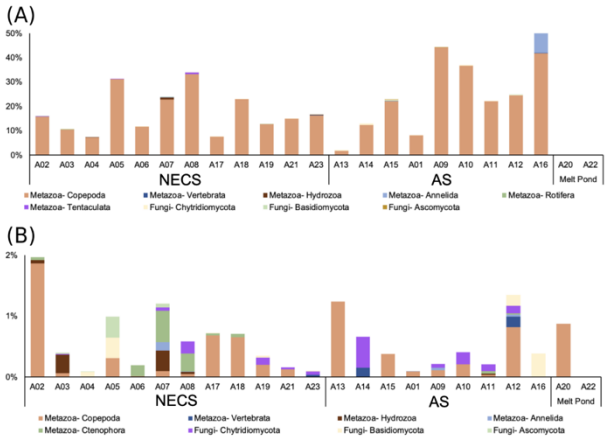
912

913 Fig. S2. TS diagram of all the sampling sites in this study. T: temperature; S: salinity;

914 σ: density. NECS: northeastern Chukchi Sea; AS: adjacent sea outside Beaufort Gyre.

915 Dashed line separates the different water types.

916



917

918 Fig. S3. Contribution of metazoa and fungi to the total eukaryotic communities in the

919 (A) 3-144 μm fraction and (B) 0.2-3 μm fraction.

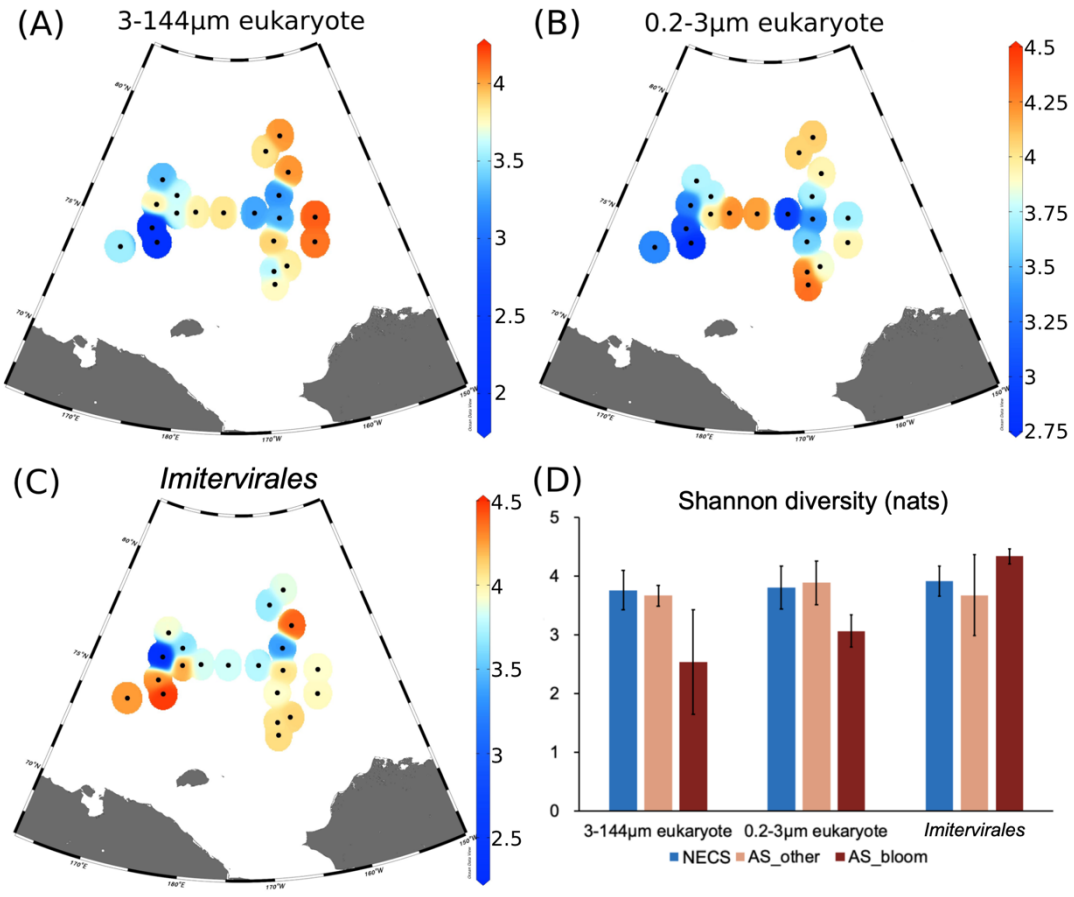
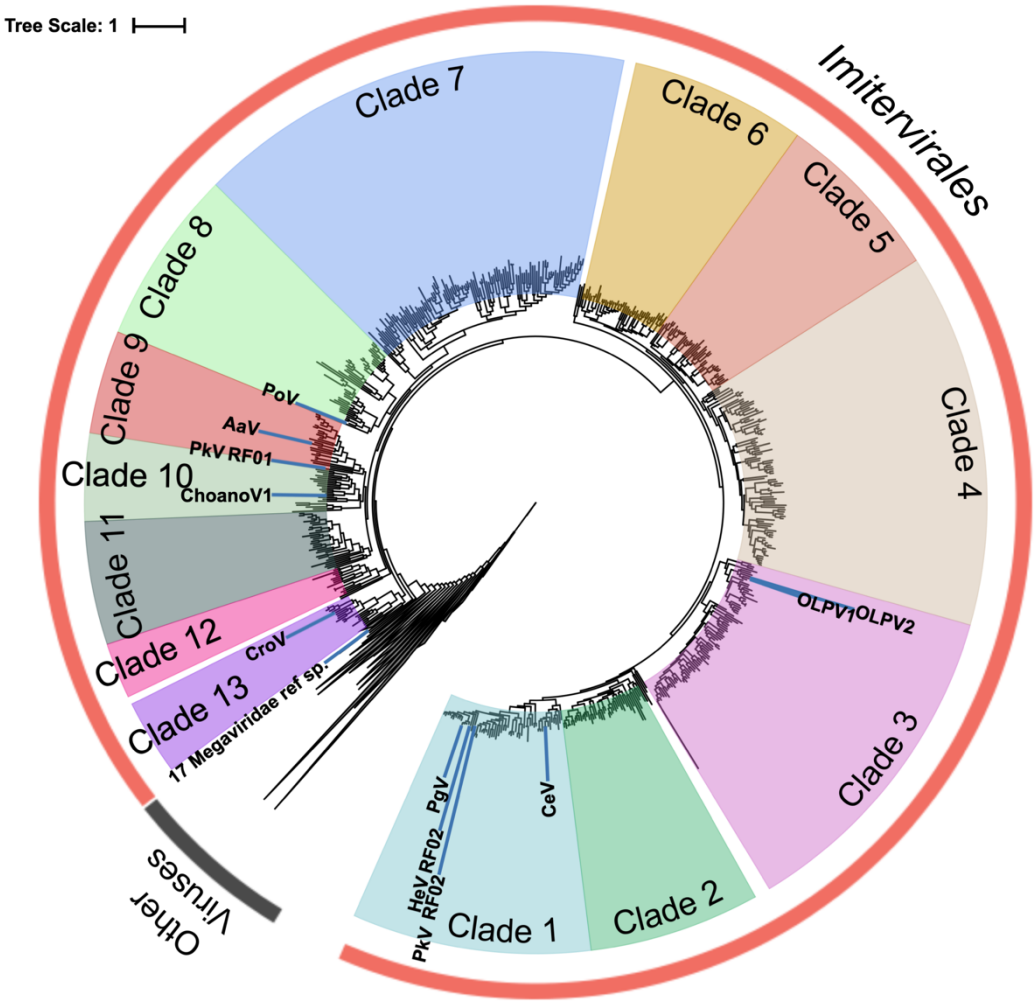


Fig. S4. Distribution of Shannon diversity index across sampling sites: (A) eukaryotes in the 3-144 μm fraction; (B) eukaryotes in the 0.2-3 μm fraction; (C) *Imitervirales*. (D) Bar plots summarizing the Shannon diversity in water type for the three communities (error bar of ± one standard deviation). Abbreviation of water types: NECS: northeastern Chukchi Sea; AS: adjacent sea outside Beaufort Gyre.



927

928 Fig. S5. Phylogenetic tree built based on marine virus *polB* gene from *Tara* Ocean

929 dataset. Clade 1 to 13 are belonging to *Imitervirales*. Reference sequence in the

930 thirteen clades: PgV (clade 1) - *Phaeocystis globosa* virus; HeV (clade 1)- *Haptolina*

931 *ericina* virus; PkV (clade 1 and clade 9) - *Prymnesium kappa* viruses; CeV (clade 1) -

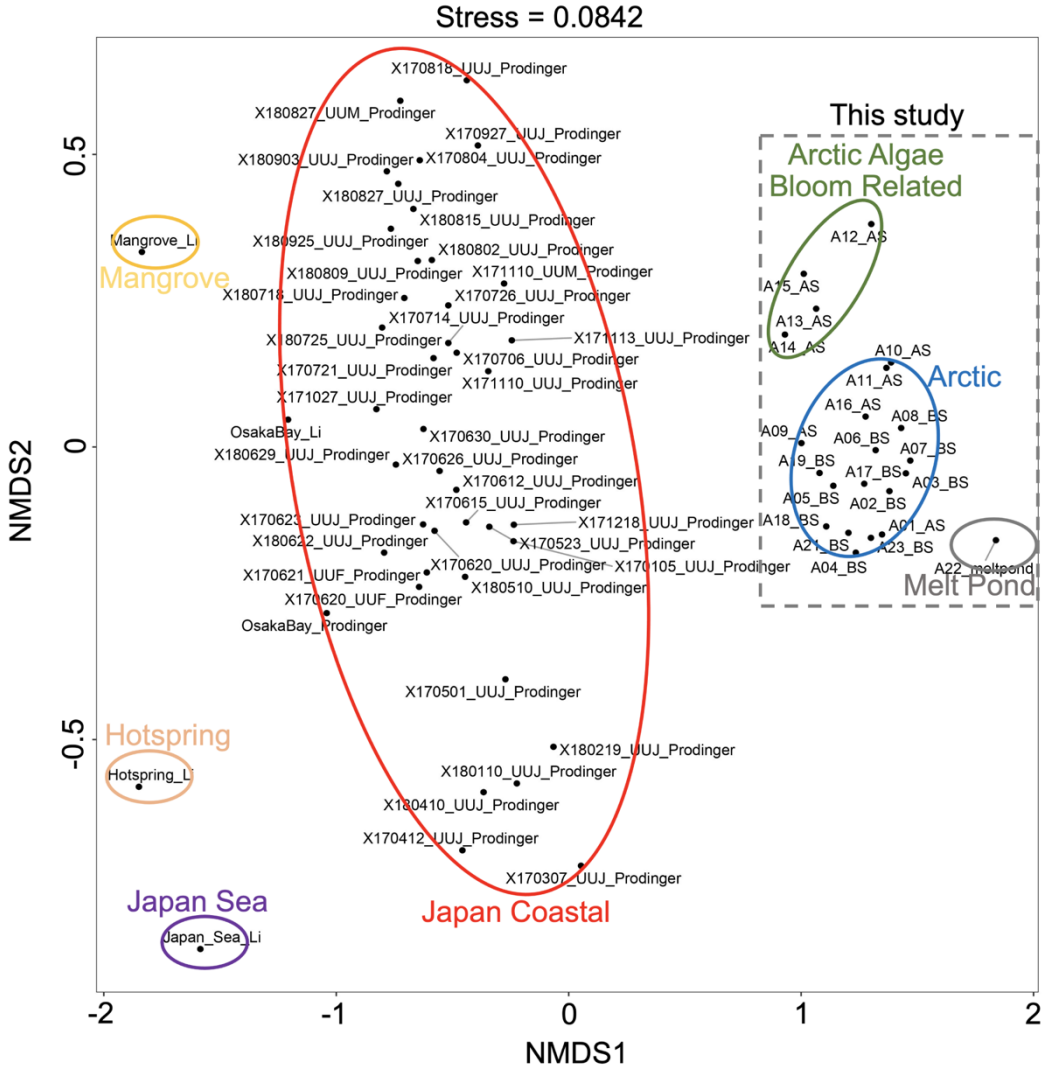
932 *Chrysochromulina ericina* virus; OLPV (clade 3) - Organic Lake Phycodnaviruses;

933 PoV (clade 8) - *Pyramimonas orientalis* virus; AaV (clade 9) - *Aureococcus*

934 *anophagefferens* virus; ChoanoV1 (clade 10) – Choanovirus1; CroV (clade 13) -

935 *Cafeteria roenbergensis* virus.

936

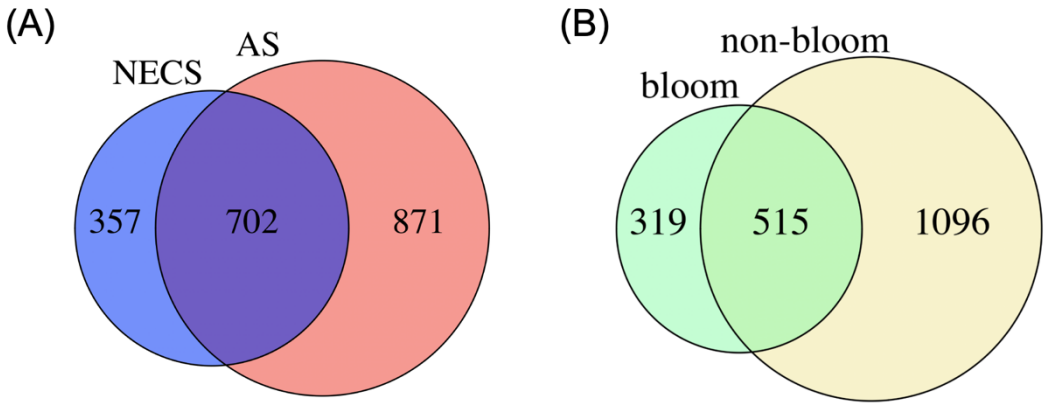


937

938 Fig. S6. Non-metric multidimensional scaling (NMDS) plot with *Imitervirales* ASVs
939 of Arctic samples in this study, hotspring samples, subtropical coastal sea water
940 samples (Li et al., 2019; Prodinger et al., 2020, Prodinger et al., unpublished). ASVs
941 were firstly subsampled by the minimum number of each sample, and relative
942 abundance was calculated by percentage of ASVs in each sample. Bray-Curtis
943 dissimilarity was used to calculate the distance matrix.

944

945



946

947 Fig. S7. Number of unique and common non-singleton *Imitervirales* ASVs between

948 (A) NECS and AS and between (B) bloom and non-bloom sites.

949

950 Dataset S1. Microsoft Excel file includes 10 supplementary tables of sampling data,

951 sequencing data, primer data and secondary result of analysis.