Extracellular rRNA profiling reveals the sinking and cell lysis dynamics of marine microeukaryotes

Hisashi Endo[1]*, Yuki Yamagishi[1], Thi Tuyen Nguyen[2], and Hiroyuki Ogata[1]

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

[2]Shimadzu Techno-Research, Inc., Nishinokyo-Shimoaicho, Nakagyo-ku, Kyoto, 604-8436, Japan

***Corresponding author**:
H. Endo, E-mail: endo@scl.kyoto-u.ac.jp, Phone: +81-774-38-3272

Abstract

Marine plankton communities consist of numerous species, and their composition and physiological states are closely linked to ecosystem functions. Understanding biogeochemical cycles requires measuring taxon-specific lytic mortality, as the dissolved organic matter released contributes to rapid nutrient recycling and long-term carbon sequestration following microbial transformation. This study introduces a pioneering quantitative and comprehensive analysis of microeukaryotes in the dissolved constituents of seawater by using a cell-free rRNA (cf-rRNA) profiling method. Our experimental pipeline successfully recovered 83% of dissolved rRNA. The ratio of cf-rRNA to cell-associated rRNA was more than 10-fold higher in the mesopelagic layer than in the upper epipelagic layer, suggesting the mesopelagic zone as a hotspot for eukaryotic cell lysis, likely due to viral infections. Many protist lineages, including phytoplankton such as haptophytes, are less susceptible to cell lysis in the epipelagic layer yet are actively lysed in the mesopelagic zone. Notably, over 86% of the significantly lysed species in the mesopelagic layer showed a habitat preference for the epipelagic layer. These findings indicate that sinking from the surface and lysis in the mesopelagic are prevalent dynamics for various eukaryotes, possibly driven by a “viral shuttle.”

Introduction

Protists, a diverse group of microscopic eukaryotes, are crucial components of marine ecosystems and biogeochemical cycles, playing roles in food webs, nutrient cycling, and carbon sequestration[1]. Studies on marine ecosystems and biogeochemical cycles have focused on specific lineages, such as diatoms and calcifying haptophytes, which can be distinguished by their microscopic morphology and chemotaxonomic pigment features. Recent studies based on high-throughput sequencing (HTS) have underscored the diversity of eukaryotic plankton, including protozooplankton, and the remarkable contributions of these species to natural assemblages[2,3]. Therefore, elucidating the complex interactions between diverse microorganisms is conducive to understanding marine systems[4,5].

Viruses are important members of marine microbial consortia as they control the abundance and composition of marine organisms via host cell lysis and biomolecule recycling[6]. According to field experiments, viral infection eliminates 20–40% of plankton populations daily[7-9]. Recent studies have uncovered a remarkable heterogeneity of eukaryotic viruses in marine environments, implying that plankton species suffer from environmental viral infections[10-12]. However, utilizing viral abundance and distribution profiles to assess the cell death of each plankton species is currently infeasible as most viruses detected in the environment lack host information.

The plankton mortality due to viral lysis has significant consequences for their biogeochemistry. The lytic processes cause cell contents to be released into the surrounding seawater as dissolved organic matter (DOM; defined as organic matter passing through a 0.2–0.7 μm filter). Besides, sloppy feeding by predators yields similar effects, as it converts large amounts of cell content to DOM[13,14], whereas zooplankton grazing incorporates whole prey cells into the biomass of higher trophic levels[15]. In this study, the term lytic cell degradation (LCD) was used to refer to all forms of DOM-producing mortality. The LCD can boost nutrient recycling on the surface water (a process termed "viral shunt") at the expense of reducing sinking carbon export to the deep ocean (biological carbon pump)[7,16]. Besides, viral infection can also trigger the aggregation and sinking export of infected cells without rupturing the cell walls, which results in the LCD in the deep ocean (a phenomenon known as "viral shuttle")[17-19]. However, the relationship between LCD and biological carbon export remains poorly understood.

Cell-free nucleic acids (DNA and RNA) are key DOM products released by LCD processes[20]. These nucleic acids exist in viral particles, extracellular vesicles, and free nucleotide molecules resulting from cell breakdown or excretion. RNA molecules comprise a major form of DOM that accounts for 60% of the dissolved organic phosphorus (DOP) in seawater, and their concentration exceeds that of DNA by up to 3–10 times[21-23]. Nucleic acids are seldom excreted spontaneously from living cells due to their high content of valuable elements such as nitrogen and phosphorus (C:N:P = 10:4:1)[20,24]. The high bioavailability and low structural stability render cf-RNA a labile DOM with a short half-life (a few hours to several days)[25]. These considerations imply that RNA in the DOM pool is continuously supplied from damaged cells via the LCD. Therefore, it provides information on recent or ongoing cell lysis events caused by viral infection and possibly sloppy feeding by zooplankton.

A recent milestone study demonstrated that viral lysis actively produces extracellular ribosomal RNA (cf-rRNA), which can serve as a proxy to evaluate the top-down control of prokaryotic populations[26]. Remarkably, when combined with HTS, this molecular marker allows for the simultaneous evaluation of the degree of viral lysis for hundreds to thousands of microbial species in an environment without information on virus–host relationships. Nevertheless, extracting sufficient RNA for quantification/sequencing from natural seawater filtrates remains methodologically challenging.

Concentration-based methods, such as tangential flow filtration and ultracentrifugation, are typically used to investigate cell-free nucleic acids in seawater[27,28]. These techniques require considerable time and effort, and the RNA is biased toward a specific molecular size or degraded during handling. A non-condensation method, which employs a lysis buffer and subsequent recovery onto a spin column, was also applied to extract RNA from a small volume (400 μL) of seawater filtrate[26]. However, as this method was explicitly designed for prokaryotic cf-rRNA profiling, it does not apply to eukaryotes due to the low density of eukaryotic cf-RNA, especially in oligotrophic and aphotic environments.

In this study, we established a novel method for extracting cell-free DNA/RNA from a large volume (> 40 mL) of dissolved seawater (filtrate of 0.2 μm pore filter) without concentration treatment. This method recovered > 80% of cf-rRNA, yielding sufficient rRNA pools for abundance and taxonomic analyses of eukaryotes, even from oligotrophic subtropical surface water and deep water characterized by a low density of plankton cells.

This method allowed us to obtain comprehensive profiles of the LCD of eukaryotic plankton based on cell-associated (living) and cell-free (dead) fractions of rRNA quantification and metabarcoding (Fig. 1).

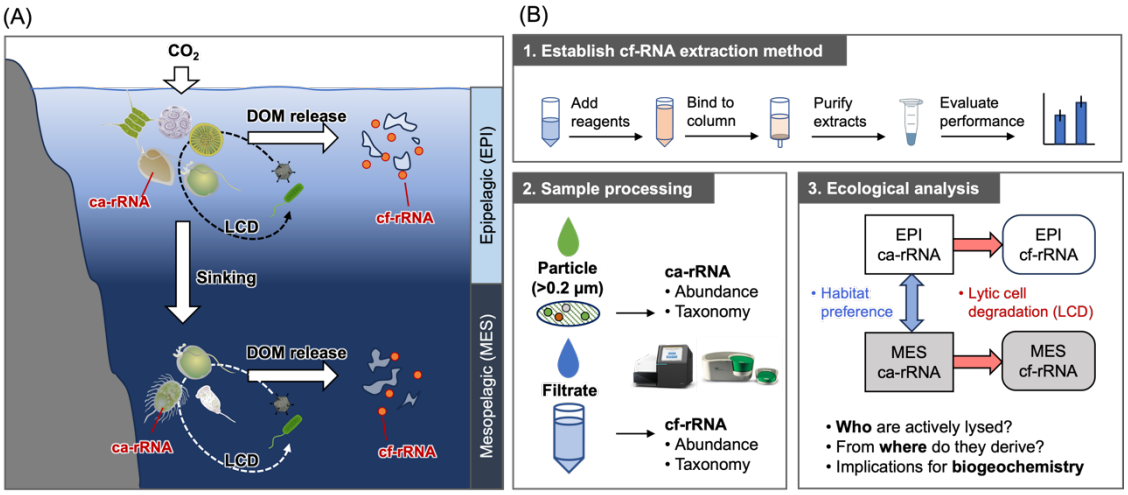


Figure 1. The concept of this study. (A) A schematic diagram showing the evaluation of taxon-specific mortality and vertical difference of microeukaryotes. (B) The workflow to evaluate the mortality of marine protists by cell-free 18S rRNA profiling. Abbreviations: EPI, Epipelagic (0–200 m depth); MES, Mesopelagic (200–1,000 m); DOM, Dissolved organic matter; LCD, Lytic cell degradation (cell degradation releasing cell contents to seawater); ca-rRNA, Cell-associated rRNA; cf-rRNA, Cell-free rRNA.

Results

Feasibility of quantitative analysis of cell-free RNA in seawater

The extraction efficiency of RNA was 83.1 ± 6.5% (mean ± standard deviation, SD) for the cf-fraction and 21.9 ± 5.7% (mean ± SD) for the cell-associated fraction (n = 3; Fig. S1A) as measured by rRNA abundance. We confirmed the linearity of the extraction volumes from 2.0×10^2 to 2.0×10^6 copy mL^{-1} of rRNA dissolved in seawater (log-log linear regression: $r^2 = 0.995$, $p < 0.001$, slope = 1.012, n = 15; Fig. S1B).

Hydrographic conditions in the study sites

The study sites were located across the Kuroshio axis (Fig. 2A). Water mass structures were clearly distinguished between the southern (stations 3 and E1) and northern sites (stations 5, 6, and 8) using the T–S diagram; northern stations were characterized by lower

surface temperature and salinity (Fig. 2B). Northern stations had relatively higher chlorophyll *a* fluorescence signals than southern stations (Fig. 2C).

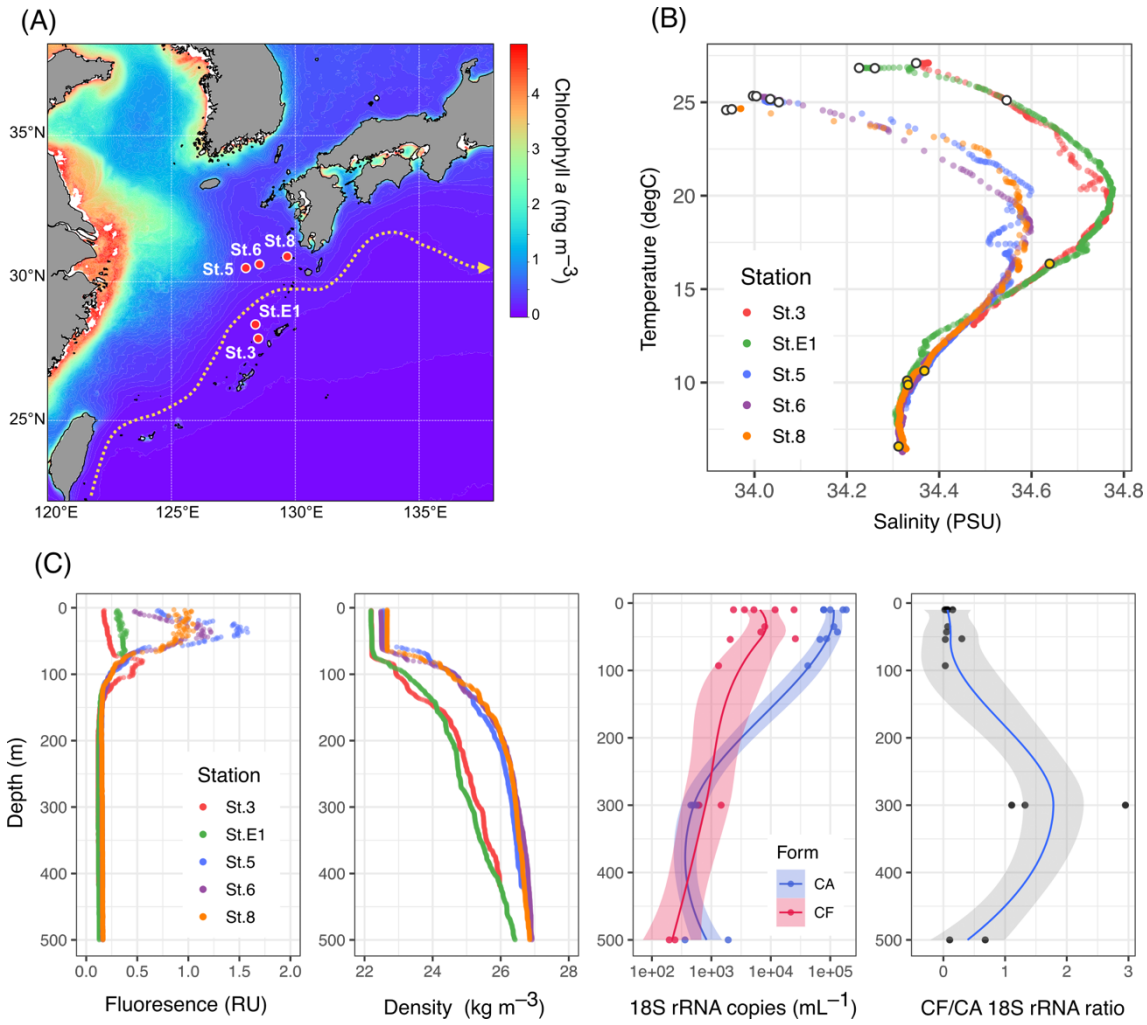


Figure 2. Location and environment of the study area. (A) A map of the sampling sites. The background image represents chlorophyll *a* concentration archived from MODIS-Aqua (monthly mean of October 2022, 4 km level 3 product). The dashed arrow indicates the position of the Kuroshio Current obtained from the Japan Coast Guard website (periods between October 10 and 17, 2022). (B) Temperature–salinity (T-S) diagram of all sampling sites. The sampling depths in the epipelagic and mesopelagic are indicated in the white and yellow circles, respectively. (C) Vertical profiles of fluorescence (relative unit of Seapoint chlorophyll fluorometer), seawater density (sigma-T), 18S rRNA abundance in different forms (cell-associated and cell-free fractions), and the abundance ratio of cell-free to cell-associated 18S rRNA (cf-rRNA/ca-rRNA).

Vertical abundance profiles of cell-associated and cell-free 18S rRNA

The abundance of 18S rRNA, as evidenced by ddPCR, varied across forms and depths (Fig. 2C). Cell-associated rRNA represented relatively high values in the euphotic (surface and SCM) layers ($1.05 \pm 0.45 \times 10^5$ copies mL^{-1}, mean ± SD, n = 10), but lower values in the mesopelagic layers ($7.57 \pm 6.54 \times 10^2$ copies mL^{-1}, n = 5). Cell-free rRNA showed higher values in the euphotic layers ($9.16 \pm 9.05 \times 10^3$ copies mL^{-1}) while lower values in the mesopelagic layers ($6.23 \pm 5.05 \times 10^2$ copies mL^{-1}). The ratio of the absolute abundances of cf-rRNA to ca-rRNA (cf-rRNA/ca-rRNA) ranged between 0.02–0.30 (0.08 on average) and 0.10–2.95 (1.23 on average) in euphotic and in mesopelagic layers, respectively.

Taxonomic composition of cell-associated and cell-free 18S rRNA

After removing rare and non-protist amplicon sequencing variants (ASVs), the sequencing analysis yielded 3,749 protist ASVs. The proportion of protist sequences to the total eukaryotic rRNA reads was 91.9% for cf-rRNA and 93.2% for ca-rRNA.

Based on the protist ASV composition, the original seawater samples were primarily clustered by sampling depth (surface epipelagic and deep mesopelagic zones; ANOSIM, $R = 0.45$, $p < 0.01$; Fig. 3). These groups were further categorized into cell-associated and cell-free rRNA forms (ANOSIM; $R = 0.25$, $p < 0.01$). Sampling locations were less important in explaining taxonomic composition differences (ANOSIM, $R = 0.03$, $p = 0.21$).

Eukaryotic taxa were classified into three groups based on their emergence patterns (Fig. 3). Group 1 comprised ubiquitously distributed and relatively abundant lineages, such as Prymnesiophyceae (Haptophyta), Dinophyceae (Dinoflagellata), Spirotrichea (Ciliophora), Choanoflagellatea (Choanoflagellata), and Acantharea (Radiolaria). Group 2 was characterized by lineages with moderate contributions, including Mediophyceae, Bacillariophyceae, Chrysophyceae (Gyrista), and Mamiellophyceae (Chlorophyta). Group 3 was marked by lineages representing low relative abundance yet emerged endemically or sporadically in specific samples.

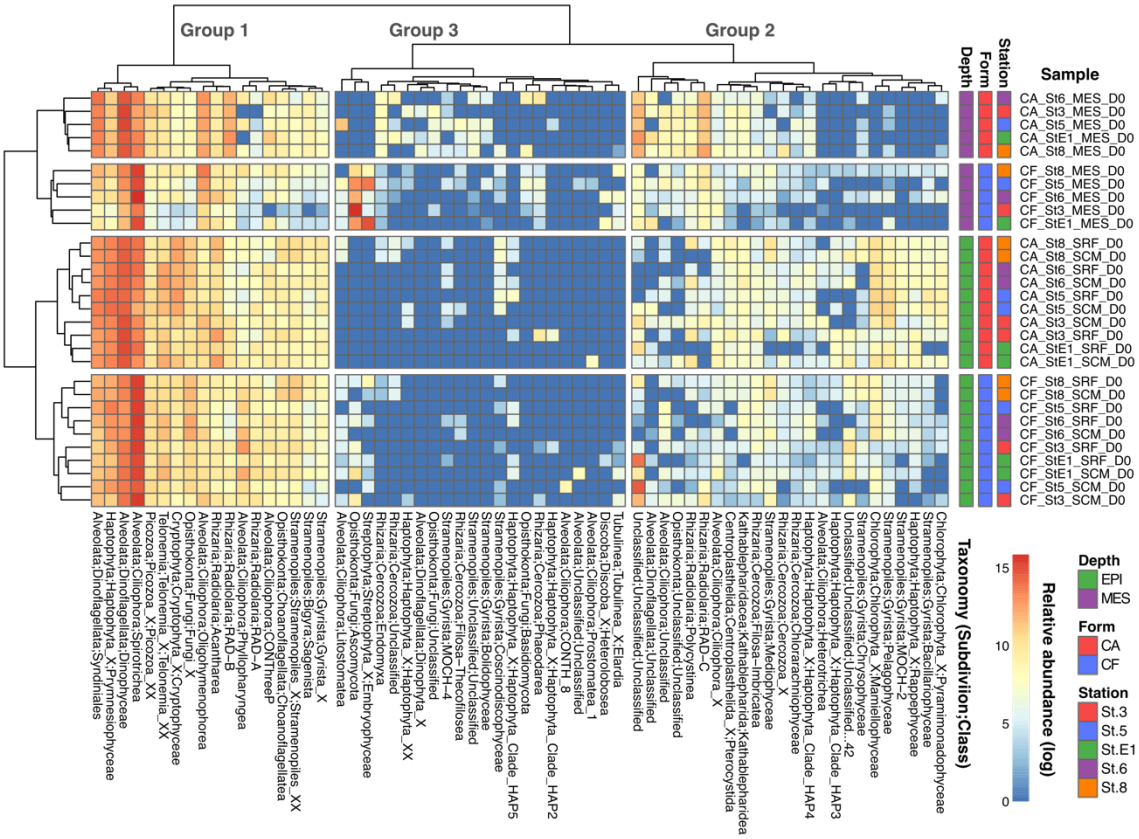


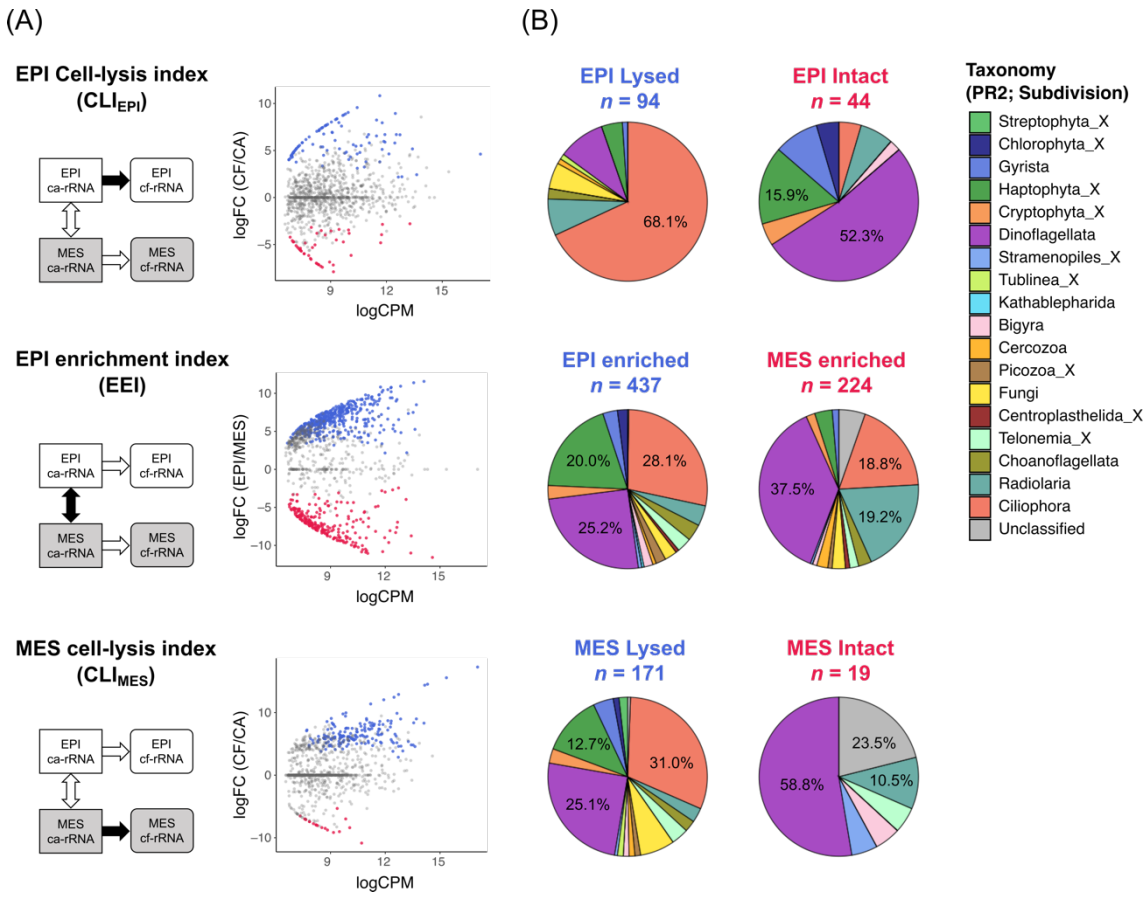
Figure 3. Heatmap and cluster analysis showing the taxonomic composition of protists. The heatmap indicates the log-transformed relative abundance of protist 18S rRNA at the class level (PR2_Lv5). Samples were clustered based on ASV-level taxonomic composition (not on the class level) in the left dendrogram. In the top dendrogram, protist clades were clustered based on their emerging patterns across samples at the class level. Protist classes with maximum contributions < 0.1% across all the samples were not shown.

Differentially abundant (DA) ASVs in each rRNA form and habitat

After removing less prevalent ASVs, 1,082 ASVs were used for DA analysis. We detected 138 DA ASVs in epipelagic samples, 94 of which were significantly enriched in the cf-fraction (significantly lysed) and 44 in the ca-fraction (significantly intact; Fig. 4A–B). A total of 190 DA ASVs (171 lysed and 19 intact) were observed in mesopelagic samples. Vertical habitat comparison identified 661 DA ASVs; 437 and 224 were significantly enriched in upper epipelagic and deep mesopelagic samples, respectively.

The subdivision Ciliophora was the most abundant in terms of the number of lysed ASVs at both epipelagic and mesopelagic depths (Fig. 4B). Dinoflagellata and

188 Haptophyta_X were the two most abundant lineages in intact ASVs in the epipelagic zone,
189 although they were the second- and third-most lysed lineages in the mesopelagic zone.
190



191

192

193 **Figure 4**. Differential abundance analysis assessing cell lysis and habitat characterization
194 of protist communities. (A) MA-plot showing the average abundance (x-axis)
195 and fold changes (y-axis) of all the ASVs. The upper and lower panels denote
196 the differential abundances between cell-free and cell-associated fractions (i.e.,
197 cell-lysis index, CLI) in the epipelagic and mesopelagic, respectively. The
198 middle panel indicates the abundance changes between epipelagic and
199 mesopelagic communities (i.e., epipelagic enrichment index, EEI) of the cell-
200 associated rRNA. (B) Pie charts summarizing the eukaryotic taxonomy
201 (PR2_Lv4; subdivision level) with significant differences in each comparison.
202

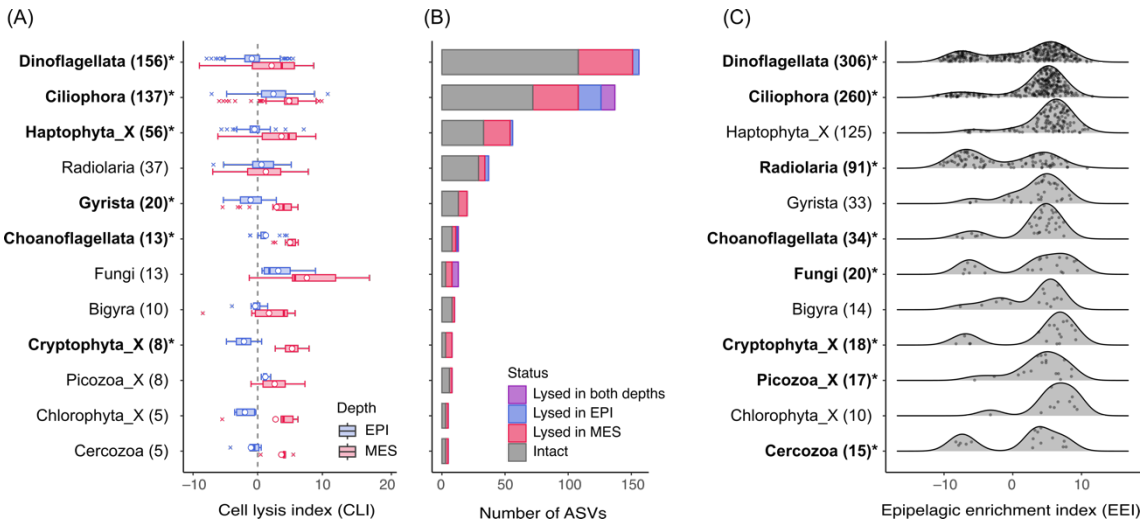
203 **Taxon specificity of prosperous and lytic depths**

204 When comparing the ASVs that emerged in both the epipelagic and mesopelagic
205 samples, 6 of the 12 eukaryotic subdivisions tested (Dinoflagellata, Ciliophora,
206 Haptophyta_X, Gyrista, Choanoflagellata, and Cryptophyta_X) displayed significantly

207 higher CLI (cell-lysis index) in mesopelagic samples (Wilcoxon rank-sum test, $p < 0.01$,
208 n = 8–156; Fig. 5A). The mean and median CLI values were higher in the mesopelagic
209 region for other lineages, although the difference was not statistically significant or could
210 not be estimated. All 12 major lineages showed a larger number of lysed DA ASVs in the
211 mesopelagic zone than in the epipelagic zone (Fig. 5B).

212 The distribution of EEI values, a proxy for the preferred habitat depth, differed among
213 eukaryotic groups. The protist lineages, which mainly consisted of phototrophic lineages
214 such as Haptophyta_X, Gyrista (50% were diatoms), and Chlorophyta_X, showed
215 unimodal density distributions with high ASV abundances in epipelagic layers (Fig. 5C).
216 Conversely, heterotrophs such as Ciliophora and Radiolaria showed bimodal distributions,
217 indicating that the lineages consisting of ASVs adapted to either mesopelagic or
218 epipelagic habitats. Bimodal patterns were also observed for Dinoflagellata and
219 Cryptophyta_X.

220



221

222

223 **Figure 5**. Taxon-specific lysis and habitat preference of protists. (A) Boxplot
224 summarizing the lysis index of protist ASVs in each lineage (subdivision
225 level) in the epipelagic (EPI) and mesopelagic (MES) zones (open circle,
226 average; centerline, median; box limits, 25%–75% quantiles; whiskers, 1.5
227 × interquartile range). The *p*-value of the difference between EPI and MES
228 < 0.01 is marked with an asterisk and bolded. (B) Number of ASVs and their
229 lytic status of each lineage at each depth. (C) Density charts summarizing
230 the epipelagic enrichment index (EEI) of protist ASVs in each clade. Values
231 in the brackets indicate the number of ASVs in each lineage. A lineage
232 having a bimodal distribution (i.e., comprising ASVs adapted to the

epipelagic and mesopelagic zones) is marked with an asterisk and bolded. ASVs with zero values in either CLI (A and B) or EEI (C) measures were removed from all plots and statistics. Protist lineages having fewer than five ASVs were not shown.

Characteristics of lysed eukaryotes in the deep sea

A significant positive correlation was detected between EEI (epipelagic enrichment index) and CLI_{MES} (CLI in the mesopelagic communities; Spearman's rank test, $\rho = 0.78$, $p < 0.001$, n = 670; Fig. 6A). By comparing the DA ASVs of the EEI and CLI_{MES} categories, we found that 147 ASVs were shared between the two measures, corresponding to 86.0% of the DA ASVs in CLI_{MES} and 33.6% in EEI. The intersecting ASVs were mainly Ciliophora (29.9%), Dinoflagellata (27.9%), and Haptophyta_X (14.3%; Fig. 6B).

The Sankey diagram shows an overall link between habitat depth and cell-lysis depth for protist ASVs (Fig. 6C). The majority (66.0%) of the ASVs significantly lysed in the epipelagic zone were originated from those ubiquitously distributed between the surface and deep (i.e., non-significant ASVs of the EEI measure), whereas 86.0% of the ASVs lysed in the mesopelagic zone were composed of epipelagic-specific ASVs.

The ASVs with a preference for epipelagic, mesopelagic, and both habitats comprised 67.9%, 1.4%, and 30.7% of the average relative contribution of cf-rRNA in the epipelagic zone, respectively (Fig. 6D). The average contribution of cf-rRNA in the mesopelagic zone was composed of ASVs in the epipelagic (7.8%), mesopelagic (33.6%), and both habitats (58.6 %).

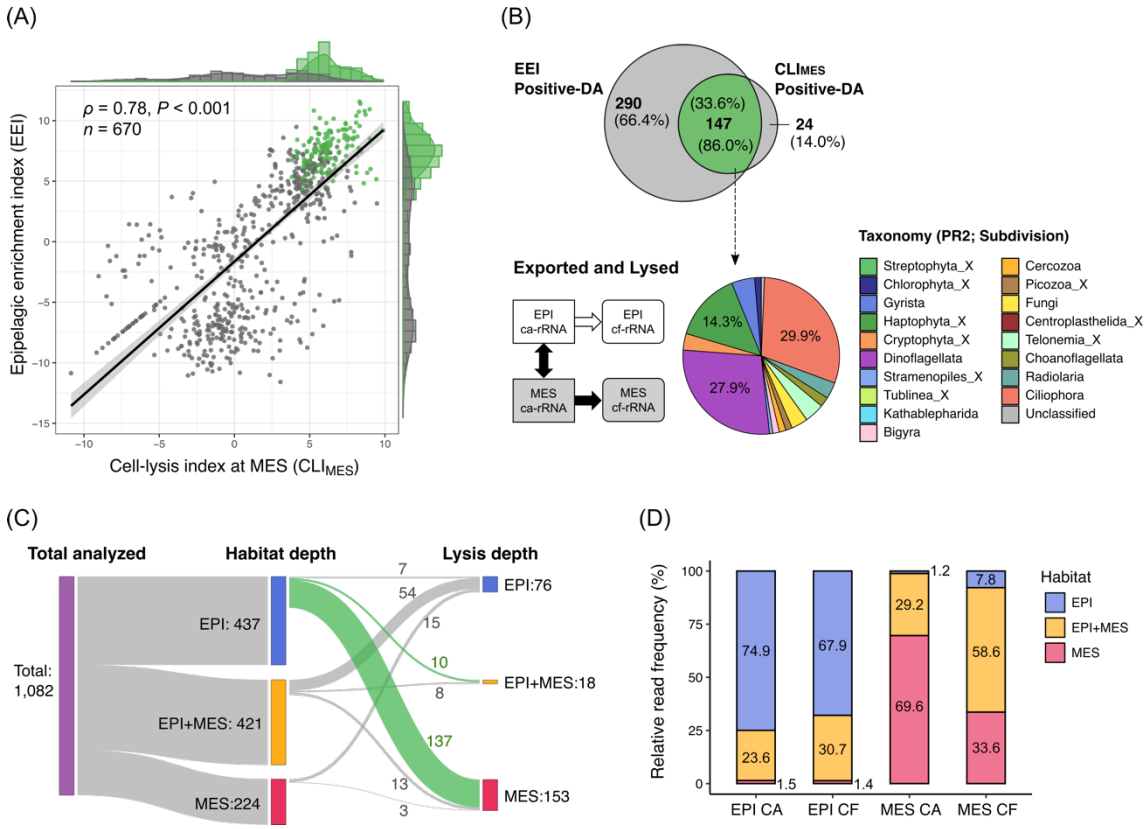


Figure 6. Detection of actively lysed microeukaryotes in the deep sea. (A) Relationship between the cell-lysis index in the mesopelagic communities (CLI_{MES}) and the epipelagic enrichment index (EEI), a prosperity measure of sunlit habitats. Differentially abundant ASVs in both the CLI_{MES} and EEI are shown in green. (B) Inter-relationship and the taxonomic composition (subdivision level) of ASVs preferring the epipelagic habitat and actively lysed in the mesopelagic zone. (C) Sankey diagram displaying the number of ASVs classified into each habitat depth and lysis depth. (D) Relative read frequency of protist ASVs having different habitats (EPI, EPI+MES, and MES) in each sample category (different depths and forms of 18S rRNA). ASVs with zero values in either measure were removed for all the plots.

Discussion

A quantitative assessment of intact and lysed eukaryotes

We established a simple method of extracting cell-free (pass through a 0.2 µm filter) RNA by directly capturing nucleic acids in a silica-membrane column. This method successfully recovered >80% of the rRNA dissolved in seawater over a wide concentration range, enabling quantitative and comprehensive analysis of dead and lysed microorganisms (Fig. 2A and 2B). This method demonstrated that over 90% of the cell-

free 18S-rRNA was derived from protists (single-celled organisms such as phytoplankton and protozoa) rather than from multicellular organisms such as zooplankton and fishes. RNA is a highly labile molecule with a turnover time of a few hours to a few days due to its high bioavailability [25]. Thus, cf-rRNA provides information on the taxonomy and degree of recent or ongoing cell lysis. Our results suggested that marine protists, likely formed via LCD processes, were the primary source of labile DOM in the study area.

rRNA metabarcoding of the cellular fraction revealed that Prymnesiophyceae (Haptophyta) and Dinophyceae (Dinoflagellata) were the most active and ubiquitous phytoplankton lineages in the surface zone of the study sites (Fig. 3). However, Dinophyceae may include heterotrophic species, and their contributions may be overestimated in rRNA-based phytoplankton communities. Microscopic analysis further confirmed the dominance of haptophytes and dinoflagellates, with average contributions of 35.4% and 40.0%, respectively, to the total identifiable phytoplankton cell count (Fig. S2). Our results are consistent with previous observations showing that haptophytes are highly abundant (up to 1.01×10^4 copies mL^{-1} of the 18S rRNA gene), contributing to 35–42% of the total chlorophyll *a* biomass in the Kuroshio ecosystem in autumn[29,30].

Enhanced eukaryotic cell lysis in the mesopelagic zone

The concentration of cf-rRNA was much lower than that of ca-rRNA in the surface but comparably high in the mesopelagic zone, resulting in apparent increases in the cf-rRNA:ca-rRNA ratio in the mesopelagic zone (Fig. 2C). The low cf-rRNA:ca-rRNA ratio in the surface zone may be partly explained by the rapid consumption of dissolved ribosomes due to high bacterial abundance. However, a previous study demonstrated comparable rates of plankton-derived DOM consumption between surface and deep seawaters[31]. Thus, the release of cellular contents, including ribosomes, via eukaryotic cell lysis is likely enhanced in the mesopelagic layer. Increases in lytic activity and the subsequent remineralization of organic molecules in the mesopelagic zone have also been reported in previous studies in different environments [18,32].

This assertion is supported by the DA analysis results, which evaluated the degree of cell lysis for each protist ASV at each depth by comparing the rRNA composition between the living (cell-associated) and dissolved (cell-free) fractions. We detected a higher number of significantly lysed ASVs and a lower number of significantly intact (i.e., less

310 susceptible to cell lysis) ASVs in the mesopelagic zone than in the surface ecosystems
311 (Fig. 4). We also uncovered a distinct taxonomic signature in CLI measurements between
312 epipelagic and mesopelagic zones, indicating depth-resolved differences in the cell-lysis
313 dynamics of protists. Notably, diverse lineages represented not only by the protozoan
314 Ciliophora but also by phytoplankton-including taxa Dinoflagellata and Haptophyta_X
315 comprised significantly lysed ASVs in the mesopelagic layer (Fig. 4B). A comprehensive
316 comparison revealed that both dominant and minor lineages, such as Choanoflagellata
317 and Cryptophyta, were more susceptible to cell lysis in the mesopelagic layer than in the
318 surface layer (Fig. 5A). These results consistently indicate that eukaryotic cell lysis was
319 promoted below the sunlit epipelagic zone.
320 The lytic activity also differed within the mesopelagic zone, as indicated by higher cf-
321 rRNA:ca-rRNA ratios in the upper (300 m) than in the middle mesopelagic layer (500 m;
322 Fig. 2C). This might be attributed to the accumulation of sinking particles in the upper
323 mesopelagic zone owing to the vertical density gradient of seawater. Approximately 1–
324 40% of photosynthetically fixed carbon is exported to the deep layer as cell aggregates or
325 fecal pellets (biological carbon pump)[33]. However, a substantial fraction (ca. 30–90%) of
326 the particle flux was shown to slow down or stop sinking in the upper mesopelagic zone,
327 presumably because of the steep increase in seawater density[34]. The CTD profile showed
328 steep density gradients of 100–300 m at the study sites (Fig. 2C). Therefore, organic
329 particles exported from the surface layer might have been retained and intensively lysed
330 in the upper mesopelagic zone.
331
332 **Surface eukaryotes were lysed in the deep ocean**
333 We showed that the mesopelagic zone (300–500 m) is a hotspot for viral lysis, and
334 various lineages were actively lysed in this layer. These results prompt the question: What
335 is the source of the protists intensively lysed in the deep sea? To address this, we
336 investigated the relationship between preferred habitat depth and lysis measures for each
337 ASV. Expectedly, heterotrophic lineages, such as Ciliophora, Radiolaria, and
338 Choanoglagellata, were clustered into epipelagic-preferred and mesopelagic-preferred
339 ASVs (Fig. 5C), indicating that these heterotrophic lineages are ubiquitous in the water
340 column at the subdivision or class level but exhibit distinct habitat depths at the ASV level.
341 Dinoflagellata and Cryptophyta also consist of surface- and deep-specific ASVs, likely

342 reflecting their diverse trophic strategies (i.e., autotrophs, heterotrophs, and mixotrophs)
343 within lineages[35].
344 Notably, we found a significant positive correlation between cellular activity in the
345 epipelagic and lytic mortality in mesopelagic zone across all microeukaryotes (Fig. 6A).
346 Remarkably, most (86.0%) ASVs differentially (significantly) lysed in the mesopelagic
347 phase were epipelagic-preferred species (Fig. 6B and 6C). Although they contributed
348 33.6% of the total cf-rRNA reads in the mesopelagic region, mesopelagic-preferred ASVs
349 comprised only 1.8% of the differentially lysed ASVs (Fig. 6D). These results clearly
350 indicate that a large majority of eukaryotic species lysed in the mesopelagic zone were
351 not inhabiting it but were transported there from the upper surface layer via the biological
352 carbon pump.
353 Cell sinking and destruction dynamics can be explained by "viral shutte" and
354 subsequent cell lysis. For instance, previous *in situ* observations showed that viral
355 infection in the surface layer enhanced the sinking of infected cells of the haptophyte
356 *Gephyrocapsa huxleyi*, which were subsequently lysed in the mesopelagic zone[18]. A
357 laboratory experiment also demonstrated that viral infection facilitated the aggregate
358 formation of infected diatom cells, likely by producing proteinaceous extracellular
359 polymers[36]. However, these studies focused on limited phytoplankton species with known
360 viral isolates. Our analysis suggests that virus-induced vertical export ("viral shuttle")
361 and subsequent lysis at depth are prevalent dynamics for various protist species, including
362 heterotrophs. Additional information on virus–host interactions and other biological
363 factors, such as sloppy feeding, is required to confirm this theory. However, our
364 conclusion is supported by a previous study showing that plankton-infecting viruses in
365 the mesopelagic region are mainly transported from the surface zone with the sinking of
366 infected cells[10].
367 The study area is influenced by the subtropical Kuroshio Current, characterized by low
368 nutrient availability in the surface layer. Although oligotrophic regions typically have
369 lower export fluxes than productive eutrophic regions, the transfer efficiency of biological
370 carbon pumps is generally high in these regions, possibly due to low zooplankton
371 degradation and high viral activity[33,37]. A recent observation in the same area reported that
372 small phytoplankton and fecal pellets contributed to efficient carbon export[38]. Our results
373 and previous findings underscore the importance of surface microeukaryotic communities

in the biogeochemical cycles of oligotrophic oceans.

We revealed the importance of the mesopelagic zone in eukaryotic cell lysis and the epipelagic zone as the source of these cells using a newly established cell-free RNA analysis method. Dissolved RNA is a component of labile DOM that accounts for < 0.1% of the ocean DOC inventory yet 84% of the carbon turnover (flux of 15–25 Pg C yr^{-1})[39]. Viral lysis and sloppy feeding, which can be detected using our method, are estimated to contribute to approximately 45% of the labile DOC supply[39], highlighting the need to unravel the source organisms and the underlying mortality mechanisms. Most labile DOC produced in the surface mixed layer is oxidized to CO_2 via respiration and released into the atmosphere. In contrast, labile DOC released in the deep layer can be sequestered over a decadal timescale as dissolved inorganic carbon or refractory DOC because it is less susceptible to vertical mixing and photodegradation[39,40]. Thus, the export of plankton proliferating in the surface layer to the mesopelagic zone and their subsequent degradation may significantly impact the ocean carbon inventory on geological timescales. Ongoing ocean warming may modify the fate of fixed carbon in ocean biogeochemical cycles via shoaling the mixed layer depth, changes in the phytoplankton community, or decreases in viral infectivity[41,42]. Our findings provide a new perspective for evaluating the interplay between marine microorganisms and biogeochemical cycles in the ocean.

Material and methods

Sample collection

Field sampling was conducted at five stations near the Kuroshio off south Kyushu, Japan, during the KS-22-15 cruise (October 16–27, 2022) of the *R/V Shinsei Maru* (JAMSTEC). Seawater samples were collected from the surface (10 m), deep chlorophyll *a* maximum (SCM, 35–93 m), and mesopelagic (300 m or 500 m) layers using Niskin bottles attached to a CTD-RMS system (Table S1).

Samples for the cell-associated and cell-free rRNA analyses were prefiltered with a 144 µm pore-size mesh to remove metazoa. Then seawater (300 mL) was filtered through the 47 mm diameter 0.2 µm pore size polycarbonate filter (Isopore) with a gentle vacuum (< 0.013 MPa) to separate into cell-associated and cell-free rRNA fractions. The filter samples (cell-associated RNA) were soaked in 600 µL buffer RLT (Qiagen) containing 1% β-mercaptoethanol and 0.2 g of muddled 0.1 mm glass beads. Forty milliliters of the filtrates (cell-free rRNA) were subsampled into 50 mL Falcon tubes (Watson). The filter and filtrate samples were flash-frozen in liquid nitrogen and stored in a deep freezer.

Nucleic acid extraction and cDNA synthesis

Cell-associated RNA samples were extracted using the RNeasy Plus Mini Kit (Qiagen) following the manufacturer's protocol, with some modifications. We used buffer RLT instead of buffer RLT Plus (Qiagen), which includes detergents, as the lysis buffer because it yielded a higher RNA concentration in a preliminary experiment. The ice-cold lysis buffer was agitated with a bead beater at 2,500 rpm thrice for 50 s before starting the protocol.

To extract cell-free RNA from the seawater filtrate samples (< 0.22 µm), we used the Wizard Enviro Total Nucleic Acid (TNA) Kit (Promega), according to the manufacturer's instructions with an elution volume of 40 µL. This kit was developed to capture and concentrate viral TNA from a large volume of wastewater (> 40 mL) using a column-based aspiration system[43].

Genomic DNA removal and first-strand cDNA synthesis were conducted for the extracted nucleotides using SuperScript IV VILO Master Mix with the ezDNase enzyme kit (Thermo Fisher Scientific) following the manufacturer's specifications.

426 **Performance evaluation of cell-free RNA extraction method**

427 We conducted an RNA spike-in extraction experiment to assess the extraction
428 efficiency of cell-free (dissolved) rRNA using a Wizard Enviro TNA Kit. Purified
429 ribosomes from *Escherichia coli* strain B (New England Biosciences, cat. P0763S) were
430 suspended in artificial seawater, and the solution was subjected to TNA extraction and
431 cDNA synthesis using the abovementioned procedure. For comparison, the same
432 ribosomal solution was used directly for cDNA synthesis. We also evaluated the
433 performance of the ca-rRNA extraction protocol using the RNeasy Plus Mini Kit (Qiagen).
434 All experiments were performed in triplicate.

435 The concentration of cDNA synthesized from *E. coli* 16S rRNA was quantified using
436 a digital PCR system (QIAcuity One 5plex, Qiagen) in a Nanoplate 8.5K 96-well plate.
437 A specific primer pair, 16S_EC442F and 16S_EC636R, was designed and used in the
438 QIAcuity EG PCR Kit (Qiagen) following the manufacturer's instructions (Table S2).

439

440 **Droplet digital PCR**

441 A QX200 Droplet Digital PCR (ddPCR) system (Bio-Rad) was used to quantify
442 eukaryotic rRNA. This system enables absolute quantification of the target sequence
443 without a standard curve by counting the nucleic acid molecules encapsulated in discrete
444 nanodroplets. We used the primer pair E572F and E1009R to target the V4 region of the
445 small subunit of the 18S rRNA gene (Table S2)[44]

446 The 20 μL reaction mixtures containing primers (0.4 μM in final concentration), the
447 cDNA template, and the ddPCR EvaGreen Supermix (2X; Bio-Rad) were prepared and
448 then loaded into a DG8 Cartridge, followed by 70 μL of Droplet generation Oil. Droplets
449 produced using a droplet generator (Bio-Rad) were transferred to 96-well PCR plates.
450 The amplifications were performed using a thermal cycler with the following conditions:
451 95 °C for 5 min, denaturation at 95 °C for 30 s, and annealing/extension at 54 °C for 2
452 min for 40 cycles. After amplification, the plate was analyzed using a QX200 Droplet
453 Reader (Bio-Rad).

454

455 **SSU rRNA metabarcoding**

456 Amplicon sequencing was performed as described previously[45]. Briefly, the V4 region
457 of the 18S rRNA gene was amplified in triplicate from the cDNA template using the

universal eukaryotic primer set E572F/E1009R (Comeau et al., 2011) attached to Illumina overhang adapters using KAPA HiFi HotStart ReadyMix. The thermal cycling condition was as follows: a pre-denaturation step at 98 °C for 30 s, followed by 30 cycles of denaturation at 98 °C for 10 s, annealing at 61 °C for 30 s, and extension at 72 °C for 30 s, with a final extension at 72 °C for 5 min. Triplicate PCR products of approximately 440 bp in length were mixed after purification using Agencourt AMPure XP beads (Beckman Coulter). Amplicon libraries were prepared using the Nextera XT Index Kit V2 and sequenced on an Illumina MiSeq platform using 300 bp paired-end sequencing.

Sequencing processing

Sequence data were analyzed using QIIME2 (ver. 2023.9)[46] with the DADA2 plugin[47]. Before being imported into the pipeline, raw reads were processed using fastp[48] to remove low-quality reads and adapter sequences. Primer sequences were removed from the quality-controlled reads with the "cutadapt trim-paired" command, then the paired-end reads were merged and further subjected to denoising, dereplication, and chimera removal to generate amplicon sequencing variants (ASVs) with the "dada2 denoise-paired" command. Rare ASVs that appeared in less than ten sequences across all samples or those found in only one sample were excluded. Taxonomy was assigned to remaining ASVs based on a pre-trained naive Bayes classifier (trained against the PR2 reference database version 4.14.0) using the "q2-feature-classifier" plugin[49]. ASVs classified as non-protist lineages (metazoans, prokaryotes, and unclassified) were excluded from the table. The read count of each sample was rarefied to 25,878 using random sampling.

Ecological analyses

To show the taxonomic composition of the microeukaryotes, a heatmap was created based on the log-transformed relative abundance of 18S rRNA at the class level (PR2_Lv5). Hierarchical clustering based on the Ward method with a Euclidean distance measure was performed to evaluate similarities across taxa. Samples were also clustered based on the Bray–Curtis dissimilarity of the ASV-level taxonomic composition (not at the class level) using Ward's D2 method. Statistical differences in sample groups (depths, forms, and sample sites) were assessed for ASV abundance using an analysis of similarities (ANOSIM) with 9,999 permutations. All statistical analyses were conducted

using the R package "vegan" (https://cran.ism.ac.jp/web/packages/vegan).

Analysis of DA ASVs

Differences in ASV abundance between the two sample categories were tested using edgeR (version 4.0.16)[50] in R. ASVs that appeared in fewer than four samples were removed from the dataset. Read counts of each ASV were adjusted based on the trimmed mean of M-values (TMM) normalization factor and the negative binomial dispersion as estimated through the 'calcNormFactors' and 'estimateDisp' functions. We fit a generalized linear mode to normalize read counts using the 'glmFit' command after adding a small prior count of 0.125 to avoid undefined ASVs when either of the counts is zero. Then, statistics of differential abundance were assessed using the 'glmLRT' command. ASVs with a False Discovery Rate (FDR) < 0.1 and an absolute value of $\log_2$-fold change (logFC) > 2 were considered significantly differentially abundant (or less abundant) ASVs (DA ASVs).

We performed DA analysis on three sets of comparisons: (1) cf- and ca-rRNA forms in epipelagic samples, (2) cf- and ca-rRNA forms in mesopelagic samples, and (3) epipelagic and mesopelagic samples of ca-rRNA. The logFC values between the cf- and ca-rRNA forms were defined as the cell-lysis index (CLI), and those between the epipelagic and mesopelagic layers were defined as the epipelagic enrichment index (EEI).

Habitat distribution analysis

To test the bimodality of the EEI measure (i.e., whether a protist lineage consists of ASVs adapted to different depths), the function 'Modes' in the R package "LaplacesDemon" (https://github.com/LaplacesDemonR) was used. This function determines the number and value of local maxima in the Gaussian kernel density distribution of variables with at least 10% of the distribution area. We regarded a protist lineage (subdivision level, PR2_Lv4) with at least one mode in each depth category (epipelagic and mesopelagic zones) as a multi-habitat lineage.

Acknowledgement

We gratefully acknowledge the captain, officers, and crew of the R/V *Shinsei-Maru* for their support during the cruise. We also thank Hideki Fukuda for the arrangement of the cruise. This work was supported by JSPS KAKENHI (19H05667, 21H05057, 22H00384, 22H00385, 22H03716, 22H02420, and 23K23685), JST PRESTO (JPMJPR23G3), and JST CREST (JPMJCR23J4). Computational work was performed at the SuperComputer System, Institute for Chemical Research, Kyoto University.

Author contributions

H.E. designed this study, joined the research cruise, and conducted the fieldwork. H.E., Y.Y., N.T. performed the laboratory experiments. H.E. performed the bioinformatics analysis. All the authors contributed to interpreting the data and writing the manuscript.

Competing interests

The authors declare no competing interests.

Data availability

The sequencing data have been deposited in DDBJ Sequence Read Archive (DRA) under accession number DRA018638 (BioSample accessions SAMD00780584–SAMD00780643). Custom scripts and associated data files used in this study are available on GitHub (https://github.com/HisashiENDO/Kuroshio_cf-rRNA).

References

1. Worden, A.Z. *et al.* Environmental science. Rethinking the marine carbon cycle: factoring in the multifarious lifestyles of microbes. *Science* **347**, 1257594 (2015).
2. de Vargas, C. *et al.* Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**, 1261605 (2015).
3. Cordier, T. *et al.* Patterns of eukaryotic diversity from the surface to the deep-ocean sediment. *Sci Adv* **8**, eabj9309 (2022).
4. Guidi, L. *et al.* Plankton networks driving carbon export in the oligotrophic ocean. *Nature* **532**, 465-470 (2016).
5. Chaffron, S. *et al.* Environmental vulnerability of the global ocean epipelagic plankton community interactome. *Science Advances* **7**(2021).
6. Vincent, F. & Vardi, A. Viral infection in the ocean-A journey across scales. *PLoS Biol* **21**, e3001966 (2023).
7. Suttle, C.A. Marine viruses--major players in the global ecosystem. *Nat Rev Microbiol* **5**, 801-12 (2007).
8. Vincent, F., Sheyn, U., Porat, Z., Schatz, D. & Vardi, A. Visualizing active viral infection reveals diverse cell fates in synchronized algal bloom demise. *Proc Natl Acad Sci U S A* **118**(2021).
9. Baudoux, A.C., Veldhuis, M.J.W., Noordeloos, A.A.M., van Noort, G. & Brussaard, C.P.D. Estimates of virus- vs. grazing induced mortality of picophytoplankton in the North Sea during summer. *Aquatic Microbial Ecology* **52**, 69-82 (2008).
10. Endo, H. *et al.* Biogeography of marine giant viruses reveals their interplay with eukaryotes and ecological functions. *Nat Ecol Evol* **4**, 1639-1649 (2020).
11. Zayed, A.A. *et al.* Cryptic and abundant marine viruses at the evolutionary origins of Earth's RNA virome. *Science* **376**, 156-162 (2022).
12. Meng, L. *et al.* Genomic adaptation of giant viruses in polar oceans. *Nat Commun* **14**, 6233 (2023).
13. Saba, G.K., Steinberg, D.K. & Bronk, D.A. The relative importance of sloppy feeding, excretion, and fecal pellet leaching in the release of dissolved carbon and nitrogen by Acartia tonsa copepods. *Journal of Experimental Marine Biology and Ecology* **404**, 47-56 (2011).
14. Møller, E.F. Production of dissolved organic carbon by sloppy feeding in the copepods Acartia tonsa, Centropages typicus, and Temora longicornis. *Limnology and Oceanography* **52**, 79-84 (2007).
15. Brum, J.R., Morris, J.J., Décima, M. & Stukel, M.R. Mortality in the oceans: causes and consequences. *Association for the Sciences of Limnology and Oceanography*, 16-48 (2014).
16. Zimmerman, A.E. *et al.* Metabolic and biogeochemical consequences of viral infection in aquatic ecosystems. *Nat Rev Microbiol* **18**, 21-34 (2020).
17. Kranzler, C.F. *et al.* Impaired viral infection and reduced mortality of diatoms in iron-limited oceanic regions. *Nature Geoscience* **14**, 231-237 (2021).
18. Laber, C.P. *et al.* Coccolithovirus facilitation of carbon export in the North Atlantic. *Nat Microbiol* **3**, 537-547 (2018).
19. Vincent, F. *et al.* Viral infection switches the balance between bacterial and eukaryotic recyclers of organic matter during coccolithophore blooms. *Nature Communications* **14**(2023).
20. Karl, D.M. & Björkman, K.M. Dynamics of Dissolved Organic Phosphorus. in *Biogeochemistry of Marine Dissolved Organic Matter* 233-334 (2015).
21. Bell, D.W., Pellechia, P.J., Ingall, E.D. & Benitez-Nelson, C.R. Resolving marine dissolved organic phosphorus (DOP) composition in a coastal estuary. *Limnology and Oceanography* **65**, 2787-2799 (2020).
22. Karl, D.M. & Bailiff, M.D. The measurement and distribution of dissolved nucleic acids in aquatic environments. *Limnology and Oceanography* **34**, 543-558 (1989).
23. Sakano, S. & Kamatani, A. Determination of dissolved nucleic acids in seawater by the fluorescence dye, ethidium bromide. *Marine Chemistry* **37**, 239-255 (1992).

24. Thornton, D.C.O. Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean. *European Journal of Phycology* **49**, 20-46 (2014).

25. Marshall, N.T., Vanderploeg, H.A. & Chaganti, S.R. Environmental (e)RNA advances the reliability of eDNA by predicting its age. *Sci Rep* **11**, 2769 (2021).

26. Zhong, K.X., Wirth, J.F., Chan, A.M. & Suttle, C.A. Mortality by ribosomal sequencing (MoRS) provides a window into taxon-specific cell lysis. *ISME J* **17**, 105-116 (2023).

27. Brum, J.R. & Sullivan, M.B. Rising to the challenge: accelerated pace of discovery transforms marine virology. *Nat Rev Microbiol* **13**, 147-59 (2015).

28. Linney, M.D., Schvarcz, C.R., Steward, G.F., DeLong, E.F. & Karl, D.M. A method for characterizing dissolved DNA and its application to the North Pacific Subtropical Gyre. *Limnology and Oceanography: Methods* **19**, 210-221 (2021).

29. Endo, H. & Suzuki, K. Spatial Variations in Community Structure of Haptophytes Across the Kuroshio Front in the Tokara Strait. *Kuroshio Current: Physical, Biogeochemical, and Ecosystem Dynamics*, 207-221 (2019).

30. Endo, H., Umezawa, Y., Takeda, S. & Suzuki, K. Haptophyte communities along the Kuroshio current reveal their geographical sources and ecological traits. *Mol Ecol* **32**, 110-123 (2023).

31. Shen, Y. & Benner, R. Molecular properties are a primary control on the microbial utilization of dissolved organic matter in the ocean. *Limnology and Oceanography* **65**, 1061-1071 (2019).

32. Lara, E. *et al.* Unveiling the role and life strategies of viruses from the surface to the dark ocean. *Sci Adv* **3**, e1602565 (2017).

33. Herndl, G.J. & Reinthaler, T. Microbial control of the dark end of the biological pump. *Nat Geosci* **6**, 718-724 (2013).

34. Giering, S.L.C. *et al.* Vertical imbalance in organic carbon budgets is indicative of a missing vertical transfer during a phytoplankton bloom near South Georgia (COMICS). *Deep Sea Research Part II: Topical Studies in Oceanography* **209**(2023).

35. Schneider, L.K. *et al.* A dataset on trophic modes of aquatic protists. *Biodivers Data J* **8**, e56648 (2020).

36. Yamada, Y., Tomaru, Y., Fukuda, H. & Nagata, T. Aggregate Formation During the Viral Lysis of a Marine Diatom. *Frontiers in Marine Science* **5**(2018).

37. Kaneko, H. *et al.* Eukaryotic virus composition can predict the efficiency of carbon export in the global ocean. *iScience* **24**, 102002 (2021).

38. Yamada, Y. *et al.* Functions of extracellular polymeric substances in partitioning suspended and sinking particles in the upper oceans of two open ocean systems. *Limnology and Oceanography* (2024).

39. Moran, M.A. *et al.* The Ocean's labile DOC supply chain. *Limnology and Oceanography* **67**, 1007-1021 (2022).

40. Jiao, N. *et al.* The microbial carbon pump and climate change. *Nat Rev Microbiol* (2024).

41. Demory, D. *et al.* A thermal trade-off between viral production and degradation drives virus-phytoplankton population dynamics. *Ecol Lett* **24**, 1133-1144 (2021).

42. Lewandowska, A.M. *et al.* Effects of sea surface warming on marine plankton. *Ecol Lett* **17**, 614-23 (2014).

43. Mondal, S. *et al.* A direct capture method for purification and detection of viral nucleic acid enables epidemiological surveillance of SARS-CoV-2. *Science of The Total Environment* **795**(2021).

44. Comeau, A.M., Li, W.K., Tremblay, J.E., Carmack, E.C. & Lovejoy, C. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS One* **6**, e27492 (2011).

45. Yang, Q. *et al.* Characterization of sinking and suspended microeukaryotic communities in spring Oyashio waters. *bioRxiv* (2023).

46. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852-857 (2019).

47. Callahan, B.J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat*

638 *Methods* **13**, 581-3 (2016).

639 48. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*
640 **34**, i884-i890 (2018).

641 49. Bokulich, N.A. *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with
642 QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).

643 50. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential
644 expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).

645