Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Soil pH, not land-use type, predicts microbial motility and denitrification genes after rainforest conversion in Sumatra and Borneo

Mc Arthur Cababan

mcarthur.cababan@deped.gov.ph

Valencia National High School

Research Article

Keywords: Shotgun metagenomics, Oil palm, Soil pH, Denitrification, Microbial motility, Land-use change

Posted Date: November 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11596698/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Transpolar Drift and Atlantic inflow shape microbial biogeography and metabolic processes across the central Arctic Ocean

Authors:

Siddarthan Venkatachalam[1*], Mats A. Granskog[2], Puthiya Veettil Vipindas[1], Gunnar Bratbak[3], Philipp Assmy[2], Anna Vader[4], Rafael Gonçalves-Araujo[5], Thajudeen Jabir[1], Oliver Müller[3], Anju Chalari[1], Ahammed Shereef[1], Nuncio Murukesh[1], Anand Jain[1]

Affiliation:

[1]National Centre for Polar and Ocean Research, Ministry of Earth Sciences (Govt. of India), Headland Sada, Vasco da Gama, Goa, India

[2]Norwegian Polar Institute, Fram Centre, Tromsø, Norway

[3]University of Bergen, Department of Biosciences, Bergen, Norway

[4]University Centre in Svalbard, Longyearbyen, Norway

[5]National Institute for Aquatic Resources, Technical University of Denmark, Lyngby, Denmark

***Correspondence:** Siddarthan Venkatachalam, venkatachalam@ncpor.res.in

Abstract

The Arctic Ocean is changing rapidly. Sea-ice loss, advancing Atlantification and terrigenous inputs create contrasting ecoregions, yet their underlying microbial processes remain unresolved. Here, we combine high-resolution oceanographic observations and multi-omics data from the North Pole to the Svalbard shelf to determine how the shift from terrigenous influence to Atlantic-dominated waters shapes microbial biogeography, community assembly and metabolic processes. We demonstrate distinct, lifestyle-specific differences in microbial communities across four hydrographic regimes, with particle-attached communities showing a predominantly stochastic distribution while free-living communities were more regime-dependent. In the terrigenous-influenced Amundsen Basin, particle-attached microbial communities drive degradation of recalcitrant terrigenous organic matter with enrichment in denitrification, suggesting potential net nitrogen loss. Moving into Atlantic-dominated waters in the Nansen Basin during late summer, free-living communities increasingly rely on mixotrophic scavenging, trace gas oxidation and marine organic matter cycling as waters become progressively more nutrient-depleted. Surface-driven photoheterotrophy prevails in the warmest, ice-free waters. Comparative genomics across the Arctic Ocean further demonstrates highly connected yet regionally adapted microbial communities governed by these distinct hydrographic regimes. These results offer a mechanistic view of how microbial metabolic processes may restructure as the region becomes seasonally ice-free, Atlantification intensifies, and terrigenous inputs increase, thereby reshaping biogeochemical cycling across the Arctic Ocean.

Introduction

The central Arctic Ocean (CAO) is facing an unprecedented reduction of sea ice, largely driven by increased atmospheric and oceanic warming. Thick multi-year ice (MYI) has been replaced by younger, thinner first-year ice (FYI) (1, 2). Simultaneously, increased river runoff (3), net precipitation, and sea-ice melt are not only contributing to the freshening of the upper ocean (4, 5) but also delivering a high influx of terrestrial dissolved organic matter (DOM) directly into the Arctic Basin (6, 7). The inflow of warm, saline Atlantic water (AW) that enters through the Fram Strait via the West Spitsbergen Current is driving shoaling of AW, and a progressive weakening of the halocline in (parts of) the Eurasian Basin, which is a process referred to as Atlantification (8). These changes contribute to further sea-ice loss, altered ocean circulation and chemistry (7, 8) and shifts in nutrient availability (9), which profoundly impact the biodiversity of pelagic and sympagic habitats, potentially facilitating the introduction of boreal species into the Eurasian Basin of Arctic Ocean (10-12).

In contrast, the Transpolar Drift (TPD) current transports sea ice, riverine and terrestrial sourced material, including freshwater, coloured dissolved organic matter (CDOM) and nutrients, primarily from Siberian rivers and shelves, in the surface waters across the CAO. The TPD results in lower salinity and stronger stratification, influencing both biological productivity and ecosystem dynamics (5, 13-18). The stronger stratification and elevated humic-like CDOM compared to the AW-dominated waters likely result in reduced nutrient resupply and lower light availability due to thicker ice and higher in-water attenuation by CDOM, thereby limiting the overall primary production and biomass concentrations in the TPD (5, 14, 18).

However, the CAO remains one of the most poorly studied environments on Earth (19, 20). Thus, fundamental knowledge gaps in biodiversity, food web structure, and individual species-level contribution to ecosystem functioning still persist (16, 21). While extensive surveys across Arctic shelves have revealed a remarkable diversity of novel prokaryotic genomes (22), studies documenting biodiversity and their functional traits in the CAO remain

scarce (21). Previous studies revealed sea-ice associated microbial community assemblages linked to different ice types and seasons in Amundsen Basin (19, 23, 24), Nansen Basin (25) and across both basins (20, 26-29). Kohlbach et al. (16) showed that the distinct physicochemical conditions found in the western Eurasian Basin drive community composition of phytoplankton and zooplankton. Similarly, there have been a few studies in the past that focussed on pelagic prokaryotic and eukaryotic community structure in the Eurasian Basin and adjacent regions by using enzymatic assays, metabarcoding or time-series approaches (26, 30-33) and only very recently multi-omics studies from the MOSAiC drift expedition have begun to resolve how pelagic and sea ice microbiomes respond to rapidly changing physicochemical conditions across different eco regimes in the CAO (20, 27, 29).

Despite these recent insights, the distinct biogeochemical roles of particle-associated (PA) and free-living (FL) microbiomes remain elusive in the CAO. Most of the past studies in the CAO have largely relied on bulk water sampling, failing to resolve PA and FL niche partitioning which are known to exhibit fundamentally different genomic traits and ecological functions (34-36). As Atlantification on one hand restructures water-column dynamics (7, 8, 37) and the TPD modulates stratification and terrestrial inputs (5, 16, 18), quantifying the distribution and metabolic potential of different microbial groups are essential for predicting the future trajectory of the Arctic biological carbon pump (38, 39).

To address these gaps, we employed a comprehensive suite of physical and chemical measurements, flow cytometry, size-fractionated metabarcoding (16S/18S), gene centric and genome-resolved comparative metagenomic analysis along a 1200 km transect in the CAO (**Fig. 1**), spanning different hydrographic regimes (5), to delineate the holistic picture of prokaryote and unicellular eukaryote community structure (hereafter referred to as microbial communities) and associated ecosystem functions. This extensive and diverse dataset is first used to explore how contrasting hydrographic regimes in the western Eurasian Basin exert strong selective pressure on microbial lifestyle strategies, driving niche partitioning among PA and FL communities. We then describe how the TPD carrying terrigenous materials and the

AW inflow, enhancing physical mixing and marine productivity, create distinct microbial biogeography, community assembly and metabolic processes across the CAO. We also demonstrate that the microbial communities in the CAO are highly diverse, comprising unique regionally adapted taxa and a shared pool of taxa associated with the Arctic shelves, as revealed by comparative genomic analysis across the Arctic Ocean.

Results

Oceanographic regimes

Based on hydrography and biogeochemistry, the surface waters were categorized into four regimes along the transect following Z. Koenig et al. (5). The Transpolar Drift (TPD) regime is characterized by highly stratified waters with clear imprint from riverine and shelf sources, with a pronounced contribution of river-derived freshwater (f_{river}) and elevated concentrations of humic-like FDOM, dissolved organic carbon (DOC), phosphate (PO_4) and silicate (SiO_4) (**Fig. 2)**. South of the TPD and over the Gakkel Ridge (GR) we find a transition zone, where surface salinity increases and surface stratification, f_{river} and FDOM decreases, transitioning to the interior Nansen Basin (iNB) that is dominated by waters of Atlantic origin (modified Atlantic water) with higher salinity and low to no river water influence and much weaker stratification (5). Here, seasonal surface freshening is instead governed by sea-ice melt and Chlorophyll-a tends to be higher. Furthest south, the regime at the continental slope (AW-IF) has warm Atlantic water inflow, ice-free waters and the highest Chlorophyll-a (**Fig. 1, Fig. 2b, Fig. S1, Supplementary Table S1**).

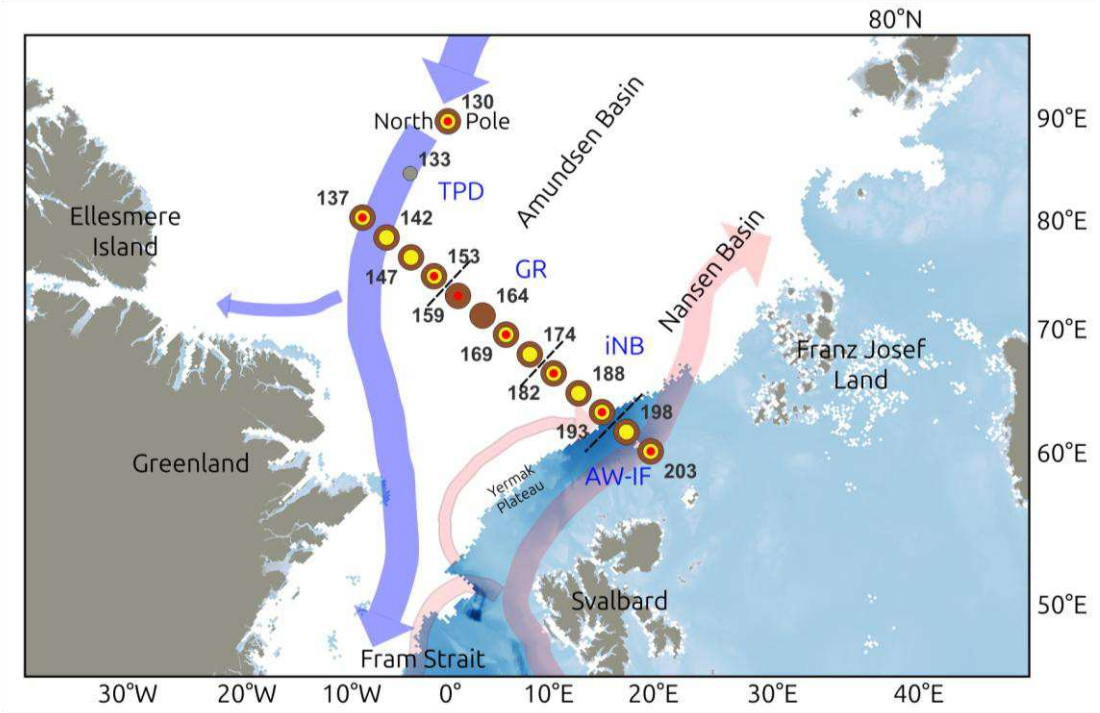


Fig. 1: Map of station locations and sea ice conditions during the AO2022 expedition. Brown circles indicate stations with microbial metabarcoding data, red dots indicate stations with metagenomic data, and yellow circles stations with flow cytometry (FCM) data. Ancillary hydrographic and hydrochemical data are available at all stations (see Methods). The gray dot indicates station 133 with only ancillary data. The dashed lines indicate the division of the transect into Transpolar Drift (TPD) in the Amundsen Basin, a transition zone over the Gakkel Ridge (GR), modified Atlantic water in the interior of the Nansen Basin (iNB) and Atlantic Water inflow and ice-free waters (AW-IF) hydrographic regimes (5). The sea-ice extent on 18 Aug 2022 (when the expedition exited the ice) is shown in white, based on AMSR2 data (40). Red arrows indicate major Atlantic Water pathways in the area, and blue arrows indicate the approximate location of the Transpolar Drift.

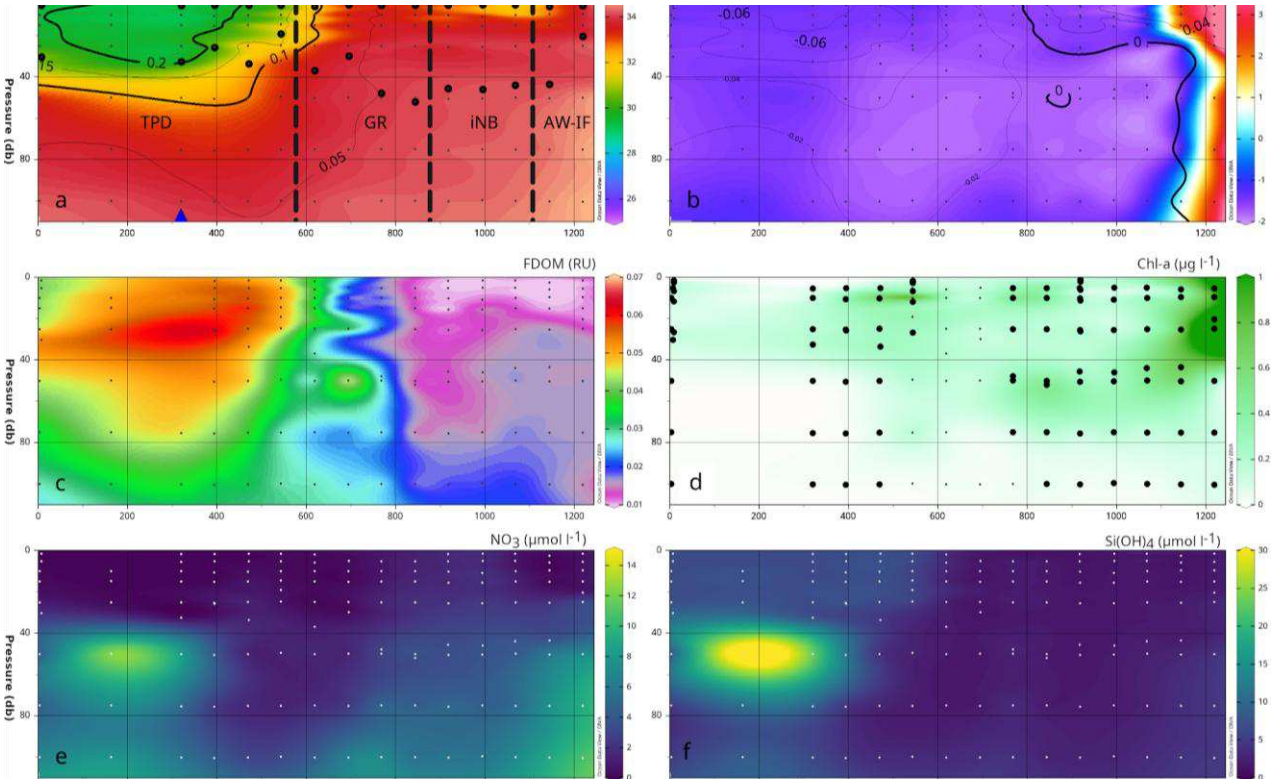


Fig. 2: Conditions along the sampled transect from the North Pole towards the Svalbard shelf. (a) Practical salinity and fraction of river water (f_{river}), (b) temperature and fraction of sea-ice melt (f_{sim}), (c) humic-like FDOM fluorescence, (d) Chl-a, (e) nitrate (NO_3), and (f) silicate ($Si(OH)_4$). The large black circles in panel (a) indicate locations where samples for microbial metabarcoding analysis were collected, the deeper sample is from the subsurface chlorophyll maxima (DCM) and thus indicates at what depth the DCM was located at a station. The large black circles in (d) indicate the location of available FCM data. The vertical dashed lines in (a) delineate the four hydrographic regimes TPD, GR, iNB and AW-IF (see Fig. 1 and text). The blue triangle in (a) indicates the location of the vertex along the transect in Fig. 1. Figure created with Ocean Data View, Schlitzer (41).

Abundance and distribution of microorganisms

Total prokaryotic abundance ranged from 6.5×10^4 to 1.2×10^6 cells mL^{-1}, with lowest abundance observed in TPD, while over GR to the southernmost ice-free regime their abundances gradually increased **(Fig. 3a)**.

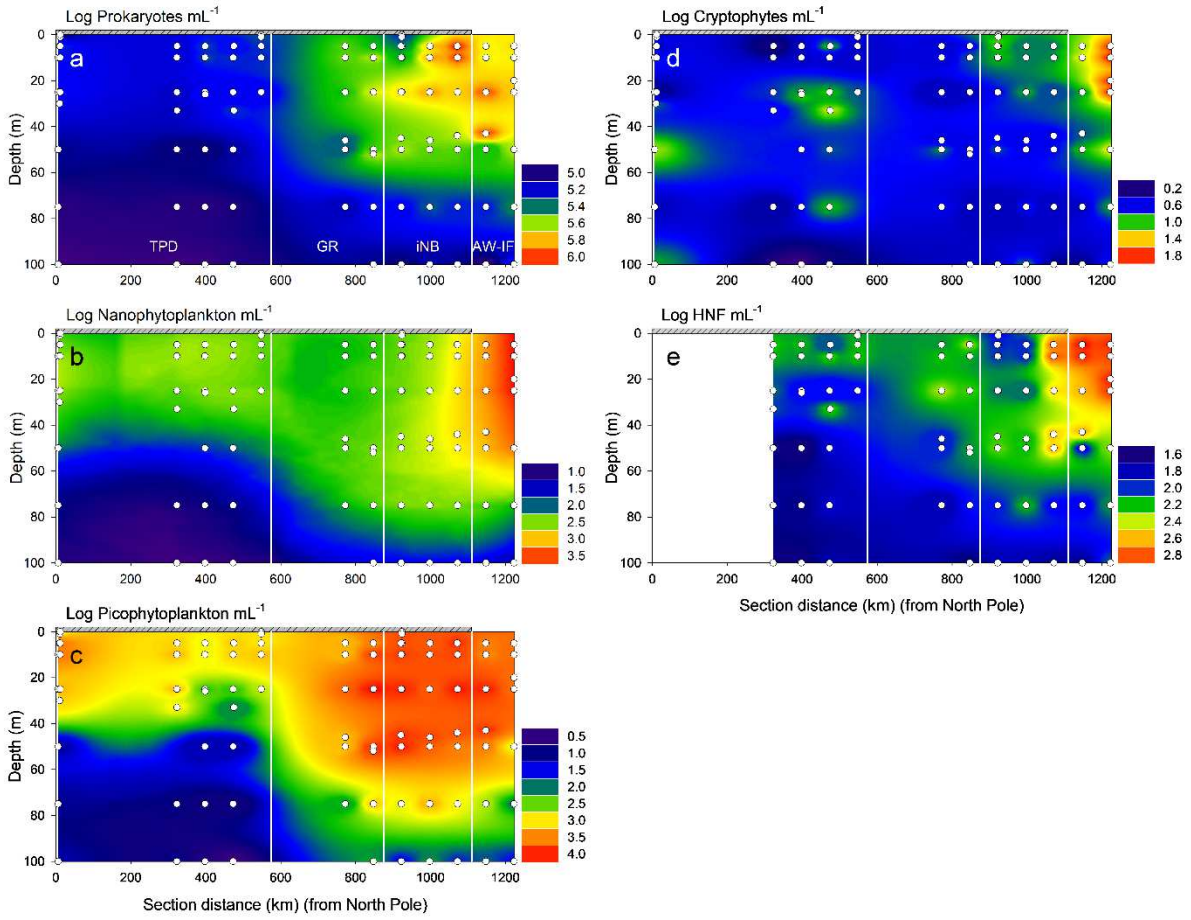


Fig. 3: Abundance of (a) prokaryotes (b) nano- and (c) pico-phytoplankton (2-20 µm and 0.2-2 µm, respectively), (d) cryptophytes and (e) heterotrophic nanoflagellates (HNF). White dots indicate sampling depths. White vertical lines delineate the four hydrographic regimes TPD, GR, iNB and AW-IF (see Fig. 1 and text). Contour plots were made using inverse distance for data smoothing. The grey bar at 0-meter indicates ice cover.

Different planktonic groups showed a similar trend of southward increase along the transect except picophytoplankton, which showed localized higher abundance near the geographic North Pole. Picophytoplankton abundances were generally lower in TPD, but increased again at GR and iNB surface (0–50 m). Nanophytoplankton abundances in TPD and GR were generally lower than in iNB (**Fig. 3b**), with communities dominated by smaller forms (**Supplementary Fig. S2 a,b**), while the abundance increased towards the ice edge in iNB and was highest in the ice-free AW-IF (up to 8.3 x10³ cells mL^{-1}) **(Supplementary Table S1)**. Similarly, heterotrophic nanoflagellates (HNF; **Fig. 3e**) were most abundant in surface

waters at or south of the ice edge, but remained overall low compared to nano- and picophytoplankton. Cyanobacteria (*Synechococcus*) were below the detection limit, except at the southernmost AW-IF stations where 30–240 cells mL^{-1} were detected in a few samples (data not shown).

Hydrographic regimes dynamically shape size-fractionated prokaryotic communities

The size-fractionated prokaryotic communities exhibited distinct spatial partitioning across the western Eurasian Basin, strongly constrained by hydrographic regimes **(Fig. 4)**. The terrestrial-influenced TPD harboured the highest number of unique, regime-specific Amplicon Sequence Variants (ASVs) in both >3 µm (1046 unique ASVs), hereafter referred to as particle attached (PA) fraction and 0.22-3 µm (1466 unique ASVs), hereafter referred to as free-living (FL) fraction **(Fig. 4a, b).** The other hydrographic regimes hosted fewer unique ASVs, while only a very small number of ASVs (PA: 81 ASVs; FL: 137 ASVs) were shared across all four hydrographic regimes in both size fractions. The beta diversity analysis revealed highly significant differences in community composition between the different zones, driven primarily by the hydrographic regime **(Fig. 4c, d)**. The PERMANOVA results demonstrated that hydrographic regimes were the primary driver of both PA (R^2 = 0.168, p = 0.001) and FL (R^2 = 0.166, p = 0.001) fractions, followed by sample depth, which played a secondary but significant role **(Supplementary Table S2)**.

Taxonomic composition and life style niches of prokaryotic communities

Taxonomic profiling of the 16S rRNA sequences revealed that prokaryotic communities were partitioned by lifestyle and dynamically shaped by the hydrographic regimes **(Fig. 4e, f; Supplementary Fig. S3)**. The PA fraction was co-dominated by Gammaproteobacteria (39.8–53.6%) and Bacteroidia (31.9–43.4%), along with a persistent presence of Alphaproteobacteria across all hydrographic regimes (8.7–12.6%). Notably, the PA fraction exhibited profound community shifts along the transect **(Supplementary Table S3)**. In the terrestrial-influenced TPD and GR regimes, the PA fraction was strongly structured

by Gammaproteobacteria specialists, particularly unclassified *Nitrincolaceae* (increasing from 14.6% in the TPD to 32.8% in the GR). However, as we move into the iNB regime, we observed a clear community shift in the *Gammaproteobacteria* clade *OM60 (NOR5)* from the TPD (1.6%) to the iNB (up to 21.8%) **(Fig. 4e, f; Supplementary Fig. S3)**. A similar community shift in the PA fraction was observed from TPD (<1%) to iNB (12.8%) for unclassified *Flavobacteriaceae* taxa. In the FL fraction, the Alphaproteobacteria class was the dominant group (22.2-39.3%), primarily composed of streamlined oligotrophs, such as the SAR11 clade (*Clade Ia*; 14.5-20.3%). The FL fraction also hosted prevalent depth-stratified archaeal populations, particularly at sea-ice-covered stations. The archaeal classes Thermoplasmata (17.2-19.1%) and Nitrososphaeria were highly abundant in the DCM across the TPD, GR, and iNB regimes, but declined in the AW-IF regimes. Furthermore, Verrucomicrobiia and Cyanobacteriia classes (*Synechococcus; 2.18%)* were uniquely restricted to the AW-IF regime in both size fractions respectively **(Fig. 4e, f; Supplementary Fig. S3)**.

To quantify these distinct lifestyle preferences, we calculated the Particle-Association Niche (PAN) index for the dominant taxa **(Supplementary Fig. S4).** At the class level, Gammaproteobacteria, Actinobacteria, Phycisphaerae and Bacteroidia exhibited distinct particle-association (PAN > 0.5). This particle-attached niche was occupied by highly specific aggregate-colonizing genera, such as *Polaribacter*, *Psychrobacter*, *Colwellia, and OM60(NOR5)* clade, etc*.,* **(Supplementary Fig. S4)**. Conversely, Alphaproteobacteria and Acidimicrobiia were overwhelmingly free-living (PAN < 0.5), represented by pelagic clades like SAR11 (*Clade Ia*). This free-living niche was even more pronounced among the Archaea, with taxa in the classes Thermoplasmata and Nitrososphaeria exhibiting near-exclusive free-living lifestyles. Notably, these fundamental lifestyle preferences remained largely conserved across different hydrographic regimes, but the intensity of particle association shifted dynamically along the transect for certain taxa. For example, Verrucomicrobiia exhibited strong particle attachment within the ice-covered interior basins (TPD and GR), but transitioned to a free-living lifestyle in the iNB and AW-IF hydrographic regimes **(Supplementary Fig. S5)**.

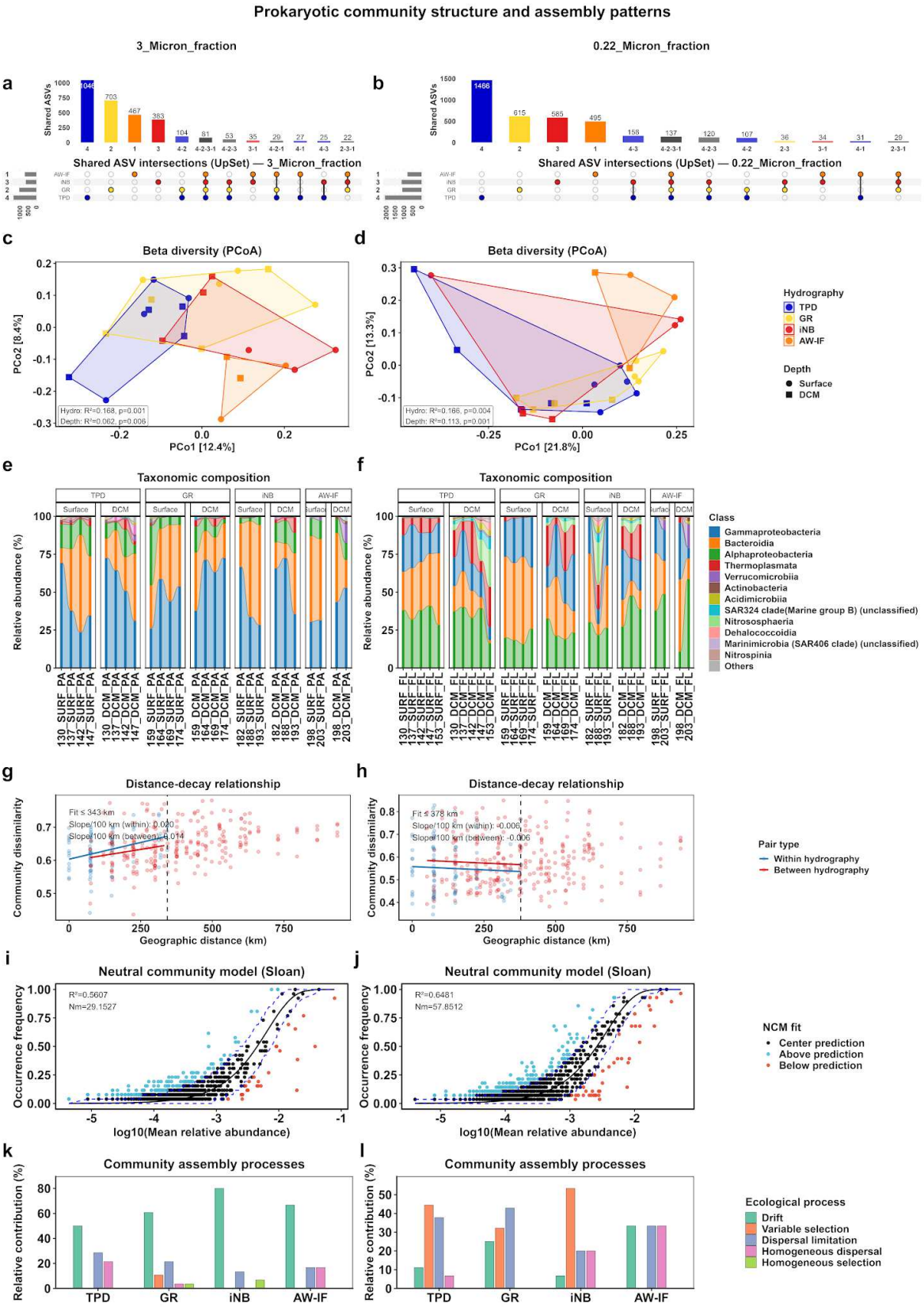


Fig. 4: Hydrographic regimes shape size-fractionated prokaryotic community structure and assembly mechanisms in the Central Arctic Ocean. The left-hand panels are from the >3 μm fraction, and the right-hand panels are from the 0.22-3 μm fraction. (a, b) UpSet plots

showing ASV intersections among four hydrographic regimes TPD, GR, iNB and AW-IF. (c, d) Principal Coordinates Analysis (PCoA) based on the unweighted UniFrac metric showing the community patterns of prokaryotes. PERMANOVA statistical results (R^2, p) were included in the plot. The hydrographic regimes and sampling depth were color and shape-coded. (e, f) Prokaryotic community composition showing the relative abundance (%) of the top 12 prokaryotic classes in the sample. (g, h) Distance-decay relationships on prokaryotic communities along the sampling transect between (red) and within (blue) hydrographic regimes. Linear regressions were fitted within the same geographic extent cutoff for both between and within hydrographic regimes, for comparative purposes. (i, j) Neutral community model (NCM) fit for both fractions with points indicating taxa falling within (black), above (blue), or below (red) 95% binomial prediction intervals around neutral expectations (dashed lines). (k, l) Relative contributions of different prokaryotic community assembly processes found across the hydrographic regimes.

Contrasting assembly processes of particle-attached and free-living prokaryotes

To determine the relative influence of spatial separation and environmental filtering on prokaryotic community assembly, we integrated distance-decay relationships (DDR), the Sloan neutral community model (NCM), and phylogenetic null-model frameworks across the transect. The DDR revealed distinct spatial dynamics between particle-attached and free-living lifestyles **(Fig. 4g, h)**. The PA community dissimilarity increased with geographic distance across the transect (slope = 0.020/100 km; Mantel r = 0.37, $p < 0.001$), whereas FL communities showed no significant distance decay (slope = 0.006/100 km; Mantel r = 0.13, p = 0.08). The NCM, fitted across the full transect, indicated greater effective migration in the FL (Nm = 57.85, $R^2 = 0.648$) than in the PA fraction (Nm = 29.15, $R^2 = 0.561$) **(Fig. 4i, j),** consistent with the weaker distance decay of FL communities. Phylogenetic null modeling resolved the underlying processes within each regime **(Fig. 4k, l)**. The PA assembly was predominantly stochastic, governed by ecological drift (50–80% across regimes), with no detectable deterministic contribution at TPD or AW-IF. FL assembly by contrast was regime dependent, shifting between deterministic and stochastic processes across the transect. Variable selection, a deterministic process, contributed to FL assembly at TPD (44.4%), GR (32.1%) and highest at iNB (53.3%), and was not detected at AW-IF. Within regimes, dispersal limitation, a stochastic process, contributed more to FL than to PA assembly in every regime,

reaching its highest relative contribution at GR (42.9%). At AW-IF, FL assembly was entirely stochastic, with dispersal limitation, drift and homogeneous dispersal each contributing 33.3% **(Fig. 4k, l)**. Overall, PA assembly was predominantly stochastic throughout the transect, whereas the deterministic–stochastic balance in FL assembly varied by regime, with variable selection most prominent at iNB and dispersal limitation at GR.

Taxonomic composition of unicellular eukaryotes

Consistent with prokaryotic communities, the size fractionated unicellular eukaryotic communities showed strong spatial compartmentalisation across different hydrographic regimes **(Fig. 5).** For consistency, we retain the terms PA (>3 μm) and FL (0.22–3 μm) for the eukaryotic data, though some free-living eukaryotic taxa that were larger than 3 μm may also be retained in the PA fraction. The highest number of unique ASVs was found in the AW-IF for the PA fraction (255 ASVs) and in the TPD for the FL fraction (404 ASVs) **(Fig. 5a, b)**. The beta diversity and PERMANOVA analyses confirmed spatial structuring ($P < 0.05$) across different hydrographic regimes **(Fig. 5c, d)**, except for the FL fraction within iNB and AW-IF regimes, that had similar community structure **(Supplementary Table S2)**. Taxonomic profiling showed that the PA fraction maintained a stable composition across all hydrographic regimes **(Fig. 5e, f),** with communities overwhelmingly dominated by the class Dinophyceae (65.3–94.8%), particularly the genus *Gyrodinium* (e.g., *G. helveticum*) in the surface layers of TPD and other unclassified dinoflagellate genera in the GR and iNB regimes **(Supplementary Fig. S6, Supplementary Table S4)**. The diatoms (Bacillariophyceae) were also relatively abundant (up to 5%), especially in the TPD regime across both size fractions. Conversely, the community structure of the FL fraction was largely dominated by phototrophs and small heterotrophic flagellates across hydrographic regimes. Underneath the sea-ice-covered stations within the interior basins (TPD and GR), the FL community was dominated by the green algal class Mamiellophyceae, specifically the polar taxon *Micromonas polaris*, which comprised up to 46% of the assemblage **(Supplementary Fig. S6)**. In contrast, the iNB and AW-IF regimes hosted a diverse heterogeneous community characterized by a complex

mixture of phototrophic, heterotrophic, and parasitic lineages, which consisted of Dinophyceae, Syndiniales (e.g., *Dino-Group-I-Clade-1*), Spirotrichea (e.g., *Strombidiidae_X, Oligotrichida_XX*), and distinct algal clades including Prymnesiophyceae (*Chrysochromulina, Phaeocystis*) and Mamiellophyceae (*Micromonas, Crustomastigaceae-AB*).

Hydrographic boundaries and sea-ice cover govern unicellular eukaryotic assembly

The spatial turnover for both size fractions was strongly modulated by different hydrographic regimes rather than geographic distance **(Fig. 5. g,h)**. The community dissimilarity increased across different hydrographic regimes (slopes PA = 0.026 and FL = 0.030 per 100 km)**,** whereas both fractions exhibited minimal spatial turnover within individual regimes (slopes PA = 0.001 and FL = 0.009 per 100 km, respectively). Although both fractions showed dispersal constraints, the PA fraction was structured primarily by stochastic drift (R^2 = 0.651, *Nm* = 19.10), while FL fraction exhibited a weaker fit to the neutral model (R^2 = 0.372, *Nm* = 15.4), demonstrating that these populations are governed by a combination of random drift and strict environmental filtering across hydrographic regimes **(Fig. 5. i,j)**. Phylogenetic null modelling further resolved the specific mechanisms driving these structural patterns. PA fraction assembly beneath the ice-covered regimes (TPD, GR, and iNB) was governed primarily by stochasticity, specifically ecological drift (40–50%) and dispersal limitation (18-27%). However, in the AW-IF regime, dispersal limitation was absent, and homogeneous dispersal was instead the dominant mechanism (50%). Despite this overarching stochasticity, PA fraction also had contributions from deterministic processes like homogeneous selection (13.3-25%) and localized variable selection (4-20%) across hydrographic regimes **(Fig. 5. k,l)**. Within the FL fraction, deterministic processes played a vital role in TPD (48.9%) and dominated assembly in the GR regime (57.1%). In contrast, community assembly in the iNB regime was predominantly governed by stochastic processes, specifically by ecological drift (53.3%) and homogenous dispersal (20%). Stochastic processes remained the primary driver of community assembly in the AW-IF regime, with a notable contribution from deterministic processes such as variable selection (34%).

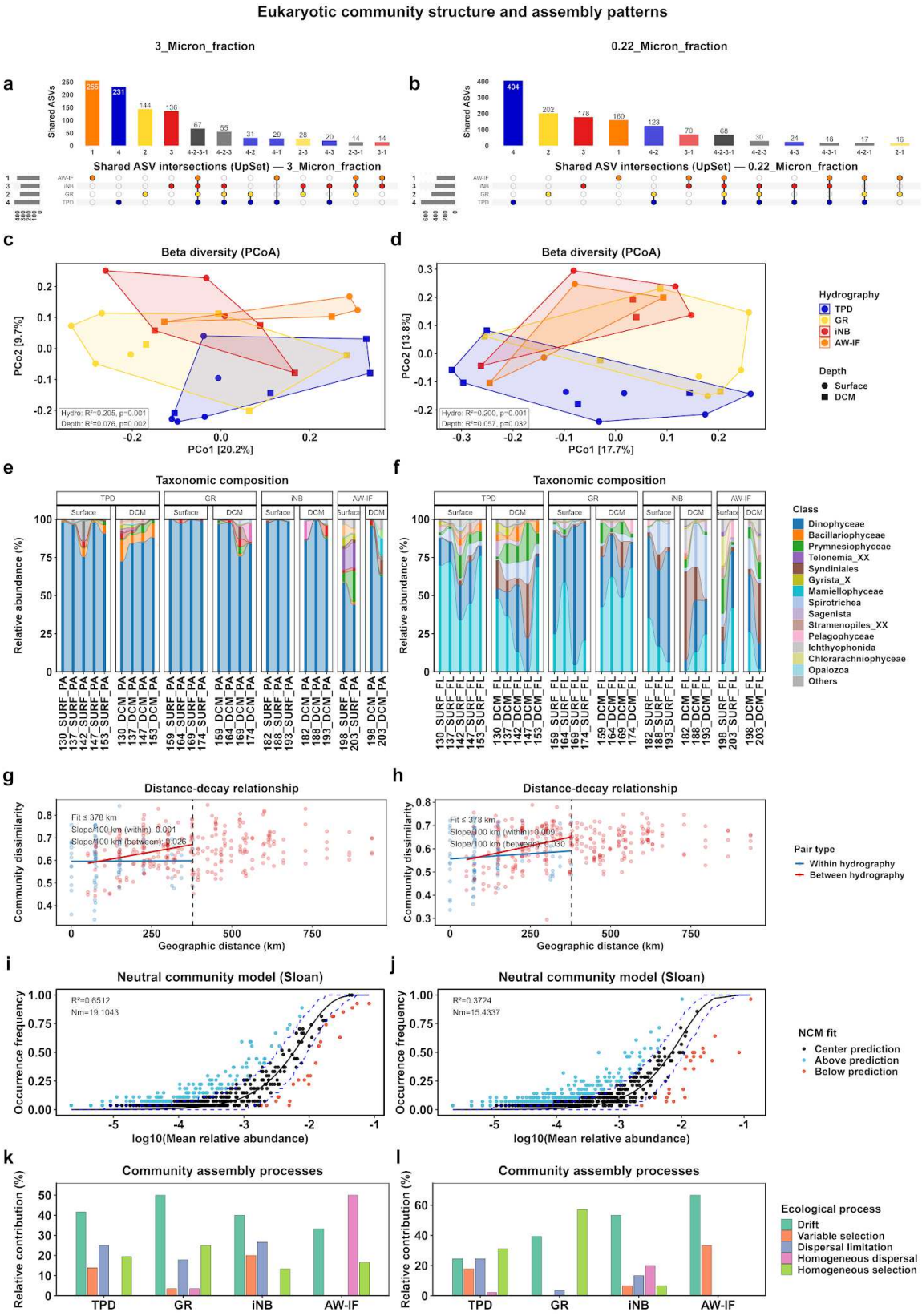


Fig. 5: Hydrographic regimes shape size-fractionated unicellular eukaryotic community structure and assembly mechanisms in the Central Arctic Ocean. The left-hand panels are from the 3 µm fraction, and the right-hand panels are from the 0.22-3 µm fraction. (a, b) UpSet plots showing eukaryotic ASV intersections among four hydrographic regimes TPD,

GR, iNB and AW-IF. (c, d) Principal Coordinates Analysis (PCoA) based on the unweighted UniFrac metric showing the community patterns of eukaryotes. PERMANOVA statistical results (R^2, p) were included in the plot. The hydrographic regimes and sampling depth were color and shape-coded. (e, f) Eukaryotic community composition showing the relative abundance (%) of the top 12 Eukaryotic classes in the sample. (g, h) Distance-decay relationships on Eukaryotic communities along the sampling transect between (red) and within (blue) hydrographic regimes. Linear regressions were fitted within the same geographic extent cutoff for both between and within hydrographic regimes, for comparative purposes. (i, j) Neutral community model (NCM) fit for both fractions with points indicating taxa falling within (black), above (blue), or below (red) 95% binomial prediction intervals around neutral expectations (dashed lines). (k, l) Relative contributions of different Eukaryotic community assembly processes found across the hydrographic regimes.

Metagenome Assembled Genomes (MAGs) across different hydrographic regimes

Genome-resolved metagenomics of size-fractionated samples across the four hydrographic regimes yielded a total of 469 MAGs, comprising 46 high-quality, 208 medium-quality, and 215 low-quality genomes **(Supplementary Fig. S7a-c)**. Taxonomic classification of this catalogue resolved 12 microbial phyla spanning domains of Bacteria (n=413; 68 unique genera), Archaea (n=44; 6 unique genera) and Eukarya (n=12; 2 unique genera) (**Supplementary Table S5)**. The vast majority of these MAGs were recovered from the FL fraction, possibly due to eukaryotic dominance in the 3 µm fraction, and standard metagenomic sequencing depths are often inadequate to resolve complex eukaryotic genomes. While the majority of recovered MAGs belonged to dominant microbial groups present across the transect, several regime-specific bacterial genera were also observed. For example, ubiquitous bacterial phyla Pseudomonadota, Bacteroidota, Actinomycetota and Cyanobacteriota were recovered in all four hydrographic regimes **(Supplementary Fig. S7b)**. In contrast, SAR 324 and Thermoplasmatota were recovered only in TPD, GR and iNB, and Verrucomicrobiota and Asgardarchaeota were exclusively recovered from AW-IF and TPD, respectively. At the genus level, *HTCC2207* (n=36), *Planktomarina* (n=22), *ASP10-02a* (n=19), *Pseudothioglobus* (n=17), *Polaribacter* (n=15), and *Arcticimaribacter* (n=13) were the most commonly recovered groups in all four regimes, while 6 genera (eg. *Pontimonas, ASP10-05a, Shewanella* etc) were exclusively recovered from TPD, 3 genera (*UBA9040, UBA9410,*

MGIIa-K1) from TZ, 7 genera (eg. *Marisimplicoccus, Abyssibacter, UBA9926*) from AW and a maximum of 12 genera (eg. *Ascidiaceihabitans, Ralstonia, Psychrobacter,* and *Luteolibacter* etc) from the AW-IF regime **(Supplementary Fig. S7c)**. Of the recovered eukaryotic MAGs, the majority belonged to the genus *Micromonas* (Chlorophyta; n=10), found in the FL fraction across TPD, GR, and iNB while the two remaining eukaryotic MAGs, assigned to Ochrophyta (raphid-pennate diatoms), were recovered exclusively from the PA fraction within TPD.

Lifestyle-driven distinct metabolic functions across different hydrographic regimes

To decipher community-wide core metabolic functions and taxon-specific contributions across different hydrographic regimes, we utilised both gene-centric (n=30 metagenomes; ~ 12 million non-redundant genes) and genome-resolved (n=255 MAGs; >70% completeness, <10% contamination) approaches. Multivariate statistical analyses showed that metabolic functions across CAO were highly distinct, fundamentally partitioned by both microbial lifestyle strategy and hydrographic regimes **(Fig. 6a, b, Supplementary Table S6)**. For example, metabolic genes associated with oxygenic photosynthesis in photosystem I (*psaA, psaC*) and photosystem II (*psbD*) were mainly detected in the PA fraction across all hydrographic regimes. In contrast, photoheterotrophy genes (*pufL, pufM, bop etc.*) were detected in both size fractions but were predominantly enriched in FL in the AW-IF regime. Similarly, aerobic respiration (*coxA, coxB*) was broadly detected in the FL fraction of the TPD, GR, and iNB regimes, while microaerobic respiration (*cydA, cydB, ccoN, etc.*) was strongly enriched in the TPD and GR regimes **(Fig. 6c)**.

Carbon fixation was mediated by multiple pathways, dominated by the Calvin–Benson–Bassham (CBB) cycle and the 3-hydroxypropionate/4-hydroxybutyrate (3HP/4HB) cycle. The CBB genes (*rbcL, rbcS, prkB,* and *cbbL*) were broadly distributed in the PA fraction across all hydrographic regimes, whereas the 3HP/4HB markers (*meH*, *accA*, *accC*) were strongly enriched in the FL fraction of the TPD and iNB regimes **(Fig. 6c)**.

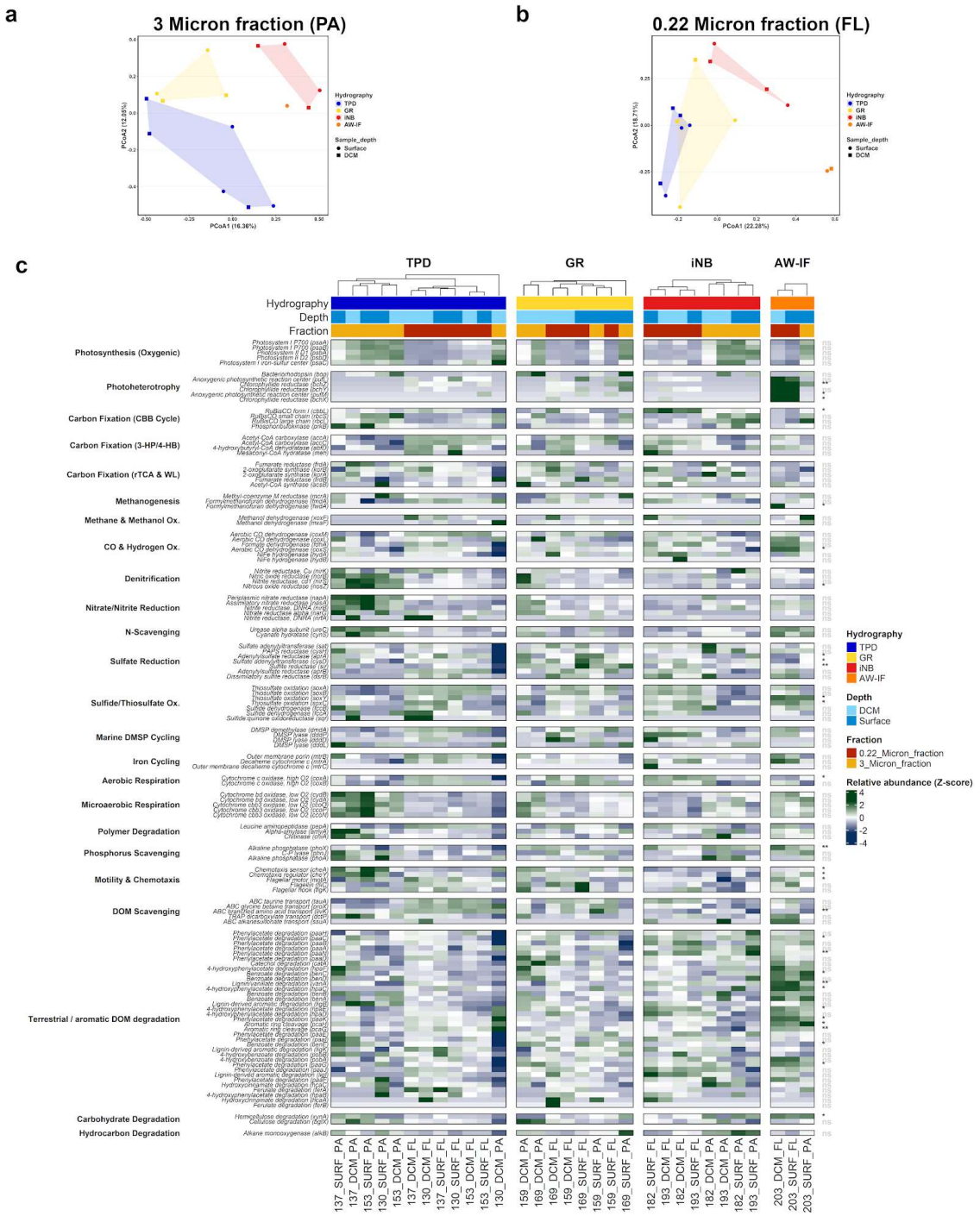


Fig. 6: Community wide metabolic functions across different hydrographic regimes in the CAO. PCoA analysis of metabolic functional gene coverage within the PA (a) and FL (b) size fractions. **(c)** Heatmap showing selected metabolic pathway gene coverage across different hydrographic regimes. The gene coverage values were normalized by the Z score.

At the genome-resolved level, the 3HP/4HB pathway genes (accA and accC) were prevalently identified among diverse prokaryotic groups including Pseudomonadota (*HTCC2207, Planktomarina, Pseudothioglobus*), Bacteroidota (*Polaribacter, UWMA-0277, MS024-2A*), Thermoplasmatota (*Poseidoniia*), Thermoproteota (*Nitrosopumilus*), Actinomycetota (*Aquiluna, UBA4592*), SAR324 (*Arctic96AD-7*), Marinisomatota (*Marinisoma, TCS55*), and Cyanobacteriota **(Supplementary Fig. S8)**. In contrast, CBB cycle markers were identified from a smaller subset of MAGs mainly affiliated with Pseudomonadota, Cyanobacteriota, Bacteroidota, and Marinisomatota. Reductive TCA cycle genes (*korA, korB, frdA, frdB*) and Wood–Ljungdahl-related pathway genes *(acsB, cooS, fwdA, fmdA)* were detected across the transect among archaeal (Thermoplasmatota, Thermoproteota) and bacterial (Pseudomonadota, Bacteroidota, Actinomycetota, SAR324, etc.,) phyla. Trace gas marker genes for methane, methanol, carbon monoxide (CO), and molecular hydrogen (H_2) oxidation were strongly enriched in the FL fraction and prevalently found across the GR, iNB, and AW-IF regimes **(Fig. 6c)**.

Pathways associated with denitrification (*nirK, nirS, norB, and nosZ*), nitrate/nitrite reduction (*napA, narG, nasA,* and *nirB*) and nitrogen scavenging (ureC and cynS) were most abundant in the PA fraction of the TPD and GR regimes, while dissimilatory nitrate reduction to ammonium (DNRA) were present across both the PA and FL fractions of TPD. These functions were mainly encoded by Bacteroidota (*Arcticimaribacter, Hel1-33-131, Marivariicella, etc*), Pseudomonadota (*ASP10-02a, Ascidiaceihabitans, Planktomarina, Ralstonia, etc*), Thermoplasmatota (*MGIIa-L1, MGIIb-O3*), Actinomycetota (*UBA4592 and Aquiluna*) and SAR324 (*Arctic96AD-7*).

Within sulfur cycling pathways, genes for thiosulfate oxidation (*soxA/B/C/Y*) were found in the FL fraction of all hydrographic regimes, while sulfide (*sqr, fccA*/*B*) oxidation pathways were highly enriched in both PA and FL fractions in the TPD. Similarly, sulfate reduction genes (*sat, cysD, aprA,* and *dsrB etc.*) were prevalent across the TPD, GR, iNB regimes, but were only sparsely detected in the ice-free AW-IF **(Fig. 6c)**. These sulfur cycling

pathway genes were mainly encoded by Bacteroidota (*Polaribacter, Arcticimaribacter, UWMA-0277*), Pseudomonadota (*HTCC2207, Marimicrobium, ASP10-02a*), Thermoplasmatota (*Poseidoniia, MGIIa-L1, MGIIb-O3*), Actinomycetota (*Aquiluna, UBA4592, UBA9040*), SAR324 (*Arctic96AD-7*), Marinisomatota (*Marinisoma, TCS55*), Cyanobacteriota (unclassified) and Thermoproteota (*Nitrosopumilus*), indicating broad phylogenetic coverage of sulfur metabolisms among the reconstructed MAGs **(Supplementary Fig. S8)**.

The complex organic matter degradation genes also showed hydrographic regime-specific distribution profiles. The terrestrial or aromatic DOM degradation markers (*ligK, ferA, paaH, etc.*) and DOM scavenging genes (*dctP, livK, proX, tauA*) were broadly distributed among both size fractions of all hydrographic regimes but most abundant in TPD and AW-IF. These functions were primarily encoded by Pseudomonadota (e.g. *Marimicrobium, Pseudothioglobus, Ralstonia, and ASP10-02a*, etc.), with additional contributions from Bacteroidota (*UWMA-0277, Polaribacter*) and Actinomycetota (*Aquiluna*). Similarly, polymer degradation markers (*chiA, amyA*) were enriched in the PA fraction of TPD and iNB, while marine DMSP cycling genes (*dmdA, dddP, dddD, etc.*) were broadly distributed but most abundant in GR and iNB, and mainly encoded among the phyla SAR324, Pseudomonadota, Thermoplasmatota, and Bacteroidota **(Supplementary Fig. S8)**.

Occurrence and biogeographic distribution of sMAGs across the Arctic Ocean

To elucidate the impact of local hydrographic and physicochemical conditions on basin-scale microbial biogeography, the reconstructed MAGs were dereplicated into 123 species-level clusters (hereafter, sMAGs). We performed a comprehensive read recruitment analysis of these 123 sMAGs against the publicly available size-fractionated TARA Arctic Ocean dataset (22), generated from lower-latitude Arctic Ocean shelf seas, along with unfractionated metagenomic datasets from the MOSAiC expedition that drifted through the CAO (27). These sMAGs recruited a median of 41% of reads in FL samples across different CAO hydrographic regimes, while in Tara Oceans, FL samples recruited 33% **(Fig. 7a)**.

In the bulk-water MOSAiC samples, read recruitment was relatively low, with a median of 17%. Overall, sMAGs recruited fewer than 10% of PA groups across the Arctic Ocean, consistent with the majority of sMAGs having been reconstructed from the FL fraction (116 of 123). Multivariate PCoA and PERMANOVA statistical analysis showed that sMAG abundance distribution patterns significantly varied between size fractions (R^2 = 0.39, P< 0.001) across different hydrographic regimes in the CAO and other parts of the Arctic Shelf seas (R^2 = 0.18, P< 0.001) **(Fig. 7b, Supplementary Table S7)**.

The biogeographic occurrence pattern showed that 30 sMAGs were widespread across all regimes in the Arctic Ocean, whereas another 15 sMAGs were found only in CAO regions **(Fig. 7c)**. Four sMAGs recovered from the PA fraction were consistently enriched in PA samples, but their abundance profiles differed. The archaeal taxon *MGIIb-O2* (193_DCM_PA_Bin1) was broadly detected from the PA fraction of the TPD, GR, and iNB regimes, as well as from the Tara Oceans PA fractions collected across Arctic shelves and outflow, but was rarely detected in the Atlantic sector **(Supplementary Fig. S9)**. Further, 20 sMAGs were identified as significant indicators of individual groups **(Fig. 7d)**, of which the majority were from AW-IF regimes (*CPC320, Ascidiaceihabitans, Planktomarina* etc.). TPD had 4 indicator sMAGs (*UWMA-0277, UBA4465, SGZJ01, HTCC2207*), GR with 2 (*ASP10-02a, UBA9040*), iNB with 3 (*unclassified cyanobacterium, Polaribacter* and *UBA974*) and Tara Arctic shelves with 2 (*MAG-121220-bin8, CAJJBL01*). The analysis also showed basin-scale lifestyle-specific associations with several sMAGs showing strong enrichment in the FL fraction (e.g., MGIIa-L1 -169_SURF_FL_Bin_9 and UBA974 -182_DCM_FL_Bin_30), while a distinct, smaller subset was consistently enriched in the PA fraction (e.g., D2472 -130_SURF_PA_Bin_4, Hel1-33-131- 137_DCM_PA_Bin_1). Interestingly, some genera, such as *Patiriisocius* (169_SURF_FL_Bin_21) and *Marivariicella* (159_SURF_FL_Bin_2), recruited reads substantially across both fractions of the Tara Oceans dataset, as well as in the unfractionated MOSAiC dataset **(Supplementary Fig. S9)**.

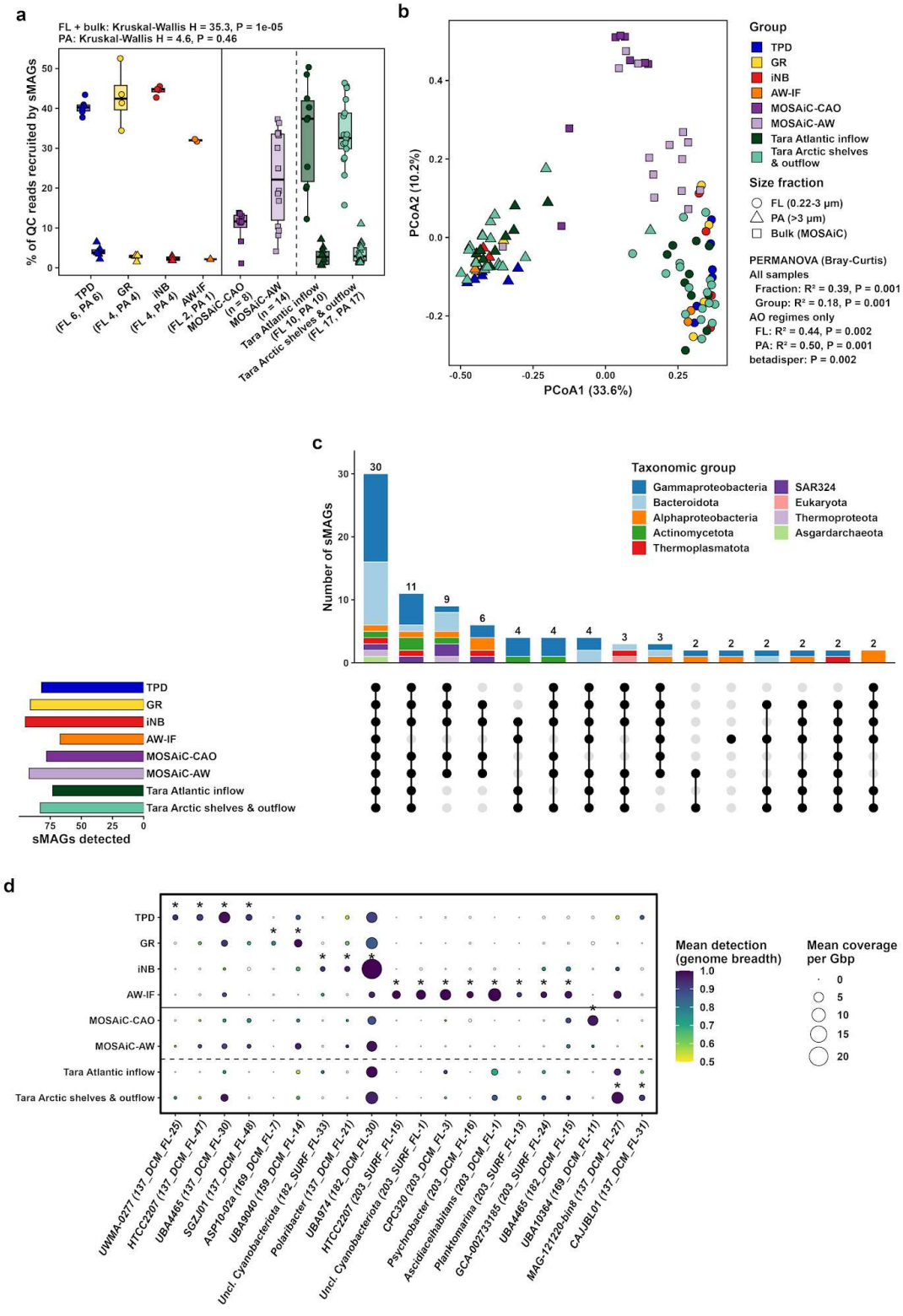


Fig. 7: Biogeographic distribution of species-level MAG clusters (sMAGs) across the Arctic Ocean. Sequence reads from 107 surface and DCM seawater metagenomes (AO-2022, n = 31; MOSAiC drift expedition, n = 22; Tara Arctic Oceans, n = 54) were mapped to the 123 sMAGs recovered in this study. Samples from the present study are grouped by hydrographic regime (TPD, GR, iNB, AW-IF), and MOSAiC Drift expedition samples are grouped into MOSAiC-CAO and MOSAiC-AW based on their sampling locations (27). Tara

Arctic Oceans samples, which cover the Arctic Shelf seas but not the central basin (22), are split into Atlantic inflow (Atlantic Arctic and North Atlantic) and Arctic shelves and outflow (Kara–Laptev, Pacific Arctic, Arctic Archipelago, Baffin Bay–Davis Strait). A sMAG was considered detected in a sample when at least 50% of its genome was covered by reads. **a.** Box plot showing the percentage of quality-filtered reads recruited by the sMAGs in PA, FL and MOSAiC bulk seawater samples. **b.** Principal coordinates analysis of Bray–Curtis dissimilarities based on sMAG coverage (Q_2Q_3 mean coverage per Gbp of quality-filtered sequence). **c.** UpSet plot showing the occurrence sMAGs across different regimes / groups in the Arctic Ocean. A sMAG was counted as present in a group when detected in 50% of that group samples. Vertical bars show the size of each intersection, coloured by taxonomic group (class for Pseudomonadota, phylum otherwise). Intersections with fewer than two sMAGs are not shown. **d**. Bubble plot showing the Indicator sMAGs (IndVal.g, 9,999 permutations, Benjamini–Hochberg-adjusted $P < 0.05$) across different regimes / groups. Circle size and fill show group mean coverage per Gbp and mean detection (open circles, <50%). Asterisks mark the group for which each sMAG is a significant indicator.

Discussion

By integrating high-resolution hydrographic and physicochemical oceanographic measurements from four distinct hydrographic regimes **(Fig. 1-2)** with multi-omics datasets, we show, for the first time, lifestyle-specific biogeographic patterns of microbial communities, their underlying community assembly mechanisms, and their core metabolic functions across the surface waters of the CAO linked to different hydrographic regimes. Our findings reveal that both prokaryotic and unicellular eukaryotic communities form distinct community assemblages, governed by contrasting assembly mechanisms between PA and FL fractions across these regimes **(Fig. 4-5)**. To our knowledge, lifestyle-specific assembly processes have not been previously resolved in the CAO. Existing bulk-water surveys reported contradictory assembly mechanisms in polar waters, with homogeneous selection more pronounced in Arctic ecosystems than at lower latitudes (42), whereas another study identified dispersal limitation as the dominant process across polar environments (43). Our lifestyle-resolved data addresses this discrepancy, showing that PA communities were predominantly stochastic across all regimes, whereas FL communities exhibited regime- and domain-specific deterministic processes (higher in the GR for eukaryotes and in iNB for prokaryotes). Furthermore, unlike lower-latitude systems, where surface PA assemblages are environmentally structured (44), CAO surface PA communities remained uniformly stochastic across the transect from terrigenous to Atlantic-influenced regimes, suggesting that, unlike FL

communities, PA assembly is not shaped by the physiochemical gradients. Additionally, an earlier study in the Arctic Ocean suggested that prokaryotes assemble more stochastically than microeukaryotes (45). Our lifestyle-resolved data support this pattern in the PA fraction across all four regimes, and in the FL fraction at TPD, GR and AW-IF, where prokaryotic communities were consistently more stochastic than their eukaryotic counterparts **(Fig. 4)**. However, at iNB, the FL prokaryotic communities became more environmentally structured (53.3% variable selection) while eukaryotes remained predominantly stochastic (86.7%), contrary to the trend reported by Q. Liu et al. (45).

Metabolic functional analysis of 12 million non-redundant genes showed that 60% of our taxonomically annotated genes had unknown functions **(Fig. 6; Supplementary Fig. S8)**, in line with a recent study from the Southern Ocean (46). This gap likely reflects the underrepresentation of polar lineages in reference databases, as 24.5% of the 47 million genes in the Tara Oceans catalogue were assembled only from Arctic shelves (47), indicating that much of the Arctic gene pool remains functionally uncharacterized. Of the known functional gene categories, we show that different hydrographic regimes harbour distinct PA and FL communities that differ in biogeochemical potential (**Supplementary Table S6)**.

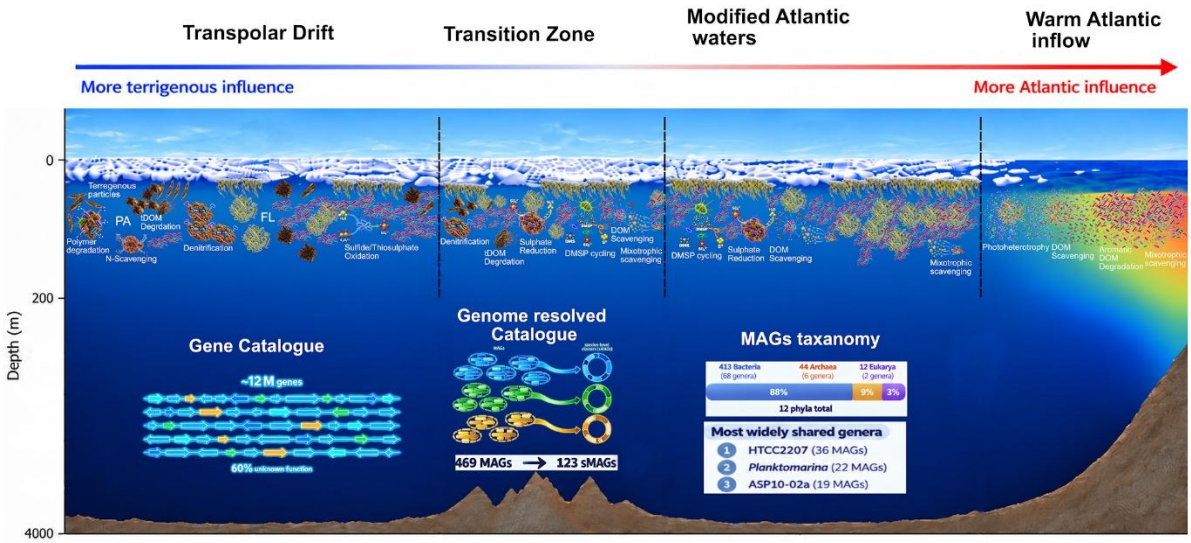


Fig. 8. Schematic illustration showing the most abundant biogeochemical metabolic functions observed across four hydrographic regimes in the surface waters of the Central Arctic Ocean. The illustration also provides an overview of the gene-centric and genome-resolved MAG catalogue generated in this study.

The prevalence of metabolic processes associated with denitrification, nitrate/nitrite reduction, nitrogen scavenging, microaerobic respiration, and polymer degradation suggests that the TPD is a potential hotspot for particle-associated organic matter processing (**Fig. 6, 8**). The comparatively higher DOC and terrestrial organic matter loads observed in the TPD, relative to the other regimes **(Fig. 2, Fig. S1)**, would support larger, carbon-dense aggregates that sustain the anoxic microenvironmental conditions conducive to a higher abundance of the above-mentioned processes in PA fractions (48). The high abundance of denitrification genes, along with lower abundances of DNRA pathway genes, suggests a potential net loss of nitrogen in the TPD **(Fig. 6c)**. Our physicochemical and biological observations further support the unique state of this ecoregion, which showed low nitrate and chlorophyll, elevated concentrations of phosphate and silicate in the surface waters (<40 m), along with the lowest bacterial and eukaryotic cell counts **(Fig. 2; Fig. 3)**. The low NO_3^-:PO_4^{3-} and high $Si(OH)_4$:NO_3^- ratios in the TPD are consistent with the decline of nitrate concentrations recently documented in polar surface waters along the wider TPD pathway, driven by the selective removal of fixed nitrogen via shelf denitrification (49). The strong stratification in the TPD (5) likely further restricts nitrate resupply from below the surface mixed layer, although summer biological uptake also contributes to the surface nitrate depletion. Productivity in the TPD is therefore constrained by low nitrate availability, which suppresses phytoplankton abundance relative to the other regimes, thereby limiting phosphate and silicate uptake and allowing both to accumulate. We also observed the highest number of ASVs in the TPD regime **(Fig. 4a)** dominated by specialist lineages (e.g., *Nitrincolaceae, Polaribacter*), which are known to process diverse organic matter pools, including terrestrial DOM and particle-associated material (50, 51). This was further confirmed by our genome-resolved data, which showed evidence of many terrestrial DOM degradation and nitrogen cycling pathways by heterotrophic lineages of Bacteroidota (e.g., *Polaribacter, Arcticimaribacter, Hel1-33-131*) and Pseudomonadota (e.g., *ASP10-02a, Marimicrobium, Ralstonia*) **(Supplementary Fig. S8)**.

The GR regime, a transition zone between TPD and iNB where the terrigenous influence is reduced, generally hosts core metabolic signatures observed in both adjacent regimes (**Fig. 6, 8**). Some stations retained high abundances of denitrification and DOM degradation genes in the PA fraction, resembling TPD, while others carried sulphate reduction, marine DMSP cycling and trace gas oxidation genes in the FL fraction, genes that are abundant in iNB. In the FL fraction specifically, trace gas oxidation genes increased in prevalence among recovered MAGs from GR (32.8%), iNB (37.5%), and AW-IF (38.5%), pointing to a mixotrophic, scavenging lifestyle by certain bacterial groups, which also encoded genes for heterotrophy in these regimes. A clear taxonomic and functional shift from TPD/GR to iNB **(Fig. 4e, f)** was supported by distinct physicochemical conditions (increases in salinity, chl-a, and sea ice melt and limited terrestrial influence) **(Fig. 2)**, along with high prokaryotic and phytoplankton cell counts **(Fig. 3)** and marine DMSP cycling **(Fig. 6c)**, suggesting a transition towards a marine phytoplankton-based food web structure (16). Taxonomically, this is reflected in the rise of bloom-associated bacterial taxa, Gammaproteobacterial clade OM60(NOR5) (1.6% in TPD to 21.8% in iNB) and unclassified Flavobacteriaceae (<1% to 12.8%) **(Fig. 4e, f)** known to participate in the degradation of algal-derived polysaccharides (52, 53).

Highly abundant photoheterotrophic functions in the FL fraction within the southernmost AW-IF regime was likely linked to increased light availability in this sea ice-free region **(Fig. 6c)**. Interestingly, this regime was also enriched with aromatic DOM degradation genes in the water column. In the absence of river-derived terrigenous DOM particles, the abundance of these pathways suggests decomposition of locally produced, algal-derived organic matter. This functional shift is coupled with the exclusive emergence of the Verrucomicrobiota phylum and the proliferation of copiotrophs, including MAGs affiliated with *Ascidiaceihabitans, Ralstonia, Psychrobacter*, and *Luteolibacter* **(Supplementary Fig. S8; Supplementary Table S5)**, which show evidence of aromatic DOM degradation and align with recent findings of carbohydrate-degrading populations in Atlantic-influenced Arctic waters

(54). Beyond this regime-specific partitioning, a subset of functions like sulphate reduction, marine DMSP cycling, DOM scavenging, and terrigenous DOM degradation was detected in both PA and FL fractions across all four regimes **(Fig. 6C, 8)**, pointing towards a core level of functional redundancy among CAO microbiomes for the utilisation of diverse physicochemical substrates present across different hydrographic regimes, thereby sustaining fundamental carbon and sulfur cycles across sharp physicochemical gradients (55-57). Together, these regime-specific signatures indicate that TPD and Atlantic-influenced waters differ not only in community composition but also in biogeochemical cycling (**Supplementary Tables S2, S6)**. Communities in the TPD are characterised by particle-mediated processing of terrigenous organic matter with a bias toward net nitrogen loss, consistent with persistent, stratification-driven nitrate limitation in Amundsen Basin (58, 59). In the more productive, Atlantic-influenced regimes (iNB, AW-IF), the microbial communities functional profiles shifted towards degrading phytoplankton-derived organic matter (54), and the widespread capacity for mixotrophic, trace-gas-oxidising scavenging.

Our read recruitment analysis sMAGs across the Arctic Ocean revealed a lifestyle-specific microbial distribution of taxa along with a combination of regionally adapted and highly connected pool of taxa to those of Arctic shelves **(Fig. 7, Supplementary Fig. S9)**. Thirty sMAGs, including *Patiriisocius, Marivariicella,* and *ASP10-02a* occurred in all eight groups, establishing a shared genomic baseline that spans across Eurasian and Amerasian Arctic **(Fig. 7c)**. Several other sMAGs showed strict biogeographic and lifestyle-specific partitioning. The archaeal lineage MGIIb-O2, largely confined to the PA fraction within our own dataset, recruited reads broadly from Arctic shelf seas and traced a continuous shelf-to-basin pathway into the TPD, GR, and iNB, but was rarely detected in the AW-IF and Atlantic Arctic sector. Some sMAGs affiliated with *UBA10364, D2472, MGIIa-K1* were mostly found endemic to CAO. Both indicator sMAGs of the Tara Arctic shelves and outflow group were recovered from TPD station 137, further linking TPD populations with shelf waters **(Fig. 7d)**. These findings demonstrate the significance of pan-arctic and large-scale baseline surveys (20, 22, 27), not

only used to benchmark the biodiversity of less explored marine habitats, but also to quantify biogeographic patterns of diverse microbial groups and their metabolic capacity in a rapidly changing Arctic Ocean.

As sea ice continues to retreat, this functional dichotomy is unlikely to resolve toward uniformly higher productivity. The nitrate-depleted state characterizing the TPD has been linked to a shelf-denitrification-driven regime shift that recent evidence suggests will not reverse even under partial sea-ice recovery (49) but partly also due to the strong freshwater-induced stratification that limits vertical resupply of nutrients (5). Increased terrigenous inputs are projected to reduce the efficiency of the Arctic Ocean's biological carbon pump by up to 40% by 2100, even as primary production increases (60). With ice-free summers projected before 2050 (61), two competing trajectories are therefore likely to develop concurrently. Atlantification weakens upper-ocean stratification (8), which could enhance vertical nitrate supply to the photic zone (59) and lead to increased productivity farther east and north of the Nansen Basin. At the same time, a persistent nitrate deficit under strong freshwater stratification sustains the particle-mediated, nitrogen-loss-biased strategy characteristic of the TPD. Which of these trajectories ultimately dominates the CAO carbon and nitrogen budget remains an open question. The lifestyle-specific, genome-resolved baseline established here addresses a critical knowledge gap in the CAO (20) and provides a basis for resolving this question as central basins continue to change.

Materials and Methods

Seawater sampling and physicochemical analysis

The material for this study was collected during the Norwegian Polar Institute's Arctic Ocean cruise (AO2022) on board FF *Kronprins Haakon* in July-August 2022 (62) **(Fig. 1; Supplementary Table S1)**. The sampling began at the North Pole and continued along a transect across western Amundsen and Nansen basins to the Svalbard shelf. The transect was densely ice-covered from the North Pole until about 82.4N in the NB **(Fig. 1)**, with only

the two southernmost stations in open water at the time of sampling (**Fig. 1**). Seawater samples were collected from Niskin bottles attached to a rosette with a conductivity-temperature-depth (CTD, SBE911+ units, SeaBird), which also provided vertical profiles of temperature, salinity and dissolved oxygen (62, 63). CTD profiles were collected every 10 nautical miles (nm) along the main transect (5), while water sampling for this study occurred approximately every 40 nm, as indicated in **Fig. 1**. Samples for salinity, inorganic nutrients (nitrate [NO_3], nitrite [NO_2], phosphate [PO_4], and silicic acid [$Si(OH)_4$]), stable oxygen isotopes ($\delta^{18}O$), fluorescent dissolved organic matter (FDOM), dissolved organic carbon (DOC), Chlorophyll-a (Chl-a) and flow cytometry (FCM) analysis of microorganisms were collected at standard depths (5, 10, 15, 25, 50, 75, and 100 dbar and Deep Chlorophyll Maximum (DCM) at a total of 14 stations across western Amundsen and Nansen basins, including the geographic North Pole **(Fig. 1, 2a).** Salinity samples were collected into thoroughly rinsed 500 mL glass bottles and sealed with a disposable plastic insert. Samples were stored at room temperature until analysis on board (within a few days of collection) using a Guildline Portasal 8410A salinometer calibrated with IAPSO P-Series standard seawater after every 24 samples (5, 62). Salinity is here reported in practical salinity (dimensionless). Oxygen isotope ratio ($\delta^{18}O$) samples were collected into glass vials and sealed with parafilm and analysed at the British Geological Survey (Keyworth, UK) using a dual-inlet mass spectrometer and standard equilibration with carbon dioxide following the procedure described by S. Epstein and T. Mayeda (64). Accuracy assessed by replicate measurements of 10% of the samples was better than ±0.04‰ (per mil) relative to Vienna standard mean ocean water (VSMOW). With salinity and $\delta^{18}O$ one can estimate the fractions of meteoric water (river runoff and precipitation), f_{river}, and sea-ice melt, f_{sim} (see e.g. (65, 66)). This can be done relative to Atlantic water (salinity 34.92 and $\delta^{18}O$ of 0.3‰; e.g. (66, 67) that is considered the marine endmember, and assigning endmember values of salinity and $\delta^{18}O$ for river runoff (0 and -18.8‰; (68)) and sea-ice melt (4 and -1.5‰), the latter based on sea-ice cores collected during the AO2022 expedition. Dissolved inorganic nutrients (nitrate [NO_3], nitrite [NO_2], phosphate [PO_4], and silicic acid [$Si(OH)_4$]) were sampled directly into 20 mL vials, preserved with 250 µL chloroform

and stored in the dark at 4℃. Post-cruise analysis was performed using a colorimetric method (69, 70) at the Institute of Marine Research, Bergen, Norway (5). Samples for the analyses of FDOM were filtered directly from Niskin bottles using gravity through a thoroughly rinsed filter cartridge (0.22 µm, Millipore Opticap XL4). Samples were stored at 4°C in pre-combusted amber borosilicate vials (40 mL) and measured onboard within a few days of collection. Excitation-emission matrices (EEMs) were collected using an AquaLog spectrofluorometer (Horiba Scientific) as described in U. J. Wünsch et al. (71). Here we use the intensity (Raman Units) of the humic-like peak C (72), representing terrestrial material as a measure of FDOM concentration, since it specifically targets material that originates from terrigenous sources. Samples for DOC were filtered similarly to FDOM into pre-combusted borosilicate vials (40~mL) through thoroughly rinsed filter cartridges (0.22 µm, Millipore Opticap XL4). Samples were acidified to pH 2 with phosphoric acid and stored in dark until analysis in the lab within 2-3 months. Carbon content was quantified using a Shimadzu TOC-VCPH analyzer after acidification to pH 2 and sparging oxygen (for details, see (71)). Seawater samples for Chl-a analysis were filtered shortly after each CTD cast onto Whatman® GF/F borosilicate filters (nominal pore size 0.7 µm). Filters were extracted overnight in 5 mL of methanol at 4°C. Acid-corrected Chl-a concentrations were measured in the dark, following O. Holm-Hansen and B. Riemann (72), using a Turner Trilogy fluorometer. Samples for flow cytometry (FCM) analysis of microorganisms (1.8 mL from each depth) were fixed with glutaraldehyde (0.5% final concentration) and incubated for 2 hours at 4 °C before freezing at -80 °C until analysis ashore. The abundance of pico- and nanophytoplankton and heterotrophic nanoflagellates (HNF) was determined using an Attune® Acoustic Focusing Flow cytometer (Applied Biosystems by Thermo Fisher Scientific, Waltham, MA, USA) equipped with a 20 mW 488-nm laser. Autotrophic microorganisms were counted immediately after thawing the samples, and the different groups were distinguished by their side-scatter and red and orange fluorescence (73). The HNF abundance was determined after staining with SYBR Green I (Molecular Probes, Eugene, OR, USA) according to the protocol (74). The prokaryotes (including both *Bacteria* and *Archaea*) were counted using a FACS Calibur (Becton Dickinson) flow cytometer with a

15 mW 488 nm laser. Prior to analysis, the samples were thawed, diluted 10-fold in 0.2-µm-filtered TE buffer (Tris 10 mM and EDTA 1 mM, pH 8), stained with SYBR Green I, and incubated for 10 min at 80 °C (75). Water samples for microbial metabarcoding and metagenomic studies were collected at the surface (5 m) and at the subsurface DCM at all stations (see **Fig. 2**a). Triplicate subsamples of about 4 L were size fractionated through 3 µm and 0.22 µm polycarbonate membrane filters (Merck-Millipore). The membrane filters containing microbial cell mass were immersed in RNAlater™ (Invitrogen) and stored at -20 °C until further analysis.

Genomic DNA extraction, microbial metabarcoding and metagenomic analysis

The total genomic DNA was extracted in duplicate from the size-fractionated frozen membrane filters (3 µm and 0.22 µm) using the DNeasy PowerWater Kit (Qiagen) according to the manufacturer's protocol. For the investigation of prokaryotic communities, the 16S rRNA gene was amplified using the primer pair (Pro341F/Pro805R) targeting the V3-V4 hypervariable region (76), following the PCR protocols described in our earlier study (77). For investigating eukaryotic microbial communities, the V4 region of the 18S rRNA gene was amplified using the primer pair E572F and E1009R (78). During PCR, a negative control reaction was included alongside the samples to verify the absence of contamination in the reagents. The amplified PCR products were purified using the Wizard® SV Gel and PCR Clean-Up System (Promega), and amplicon libraries were prepared using the Ion Plus Fragment Library kit and Ion Xpress Barcode Adapters (Thermo Fisher Scientific), which was described in our earlier study (79). The final pooled libraries (100 pmoles in 50 µL), along with mock ZymoBIOMICS microbial community DNA standards (D6306) from Zymo Research Corporation, USA, were sequenced using the GeneStudio S5 platform (Thermo Fisher Scientific). For the metagenomic analysis, representative samples were selected for each hydrographic regime, and sequence data were generated using a hybrid sequencing approach. The short-read sequencing libraries were prepared using Ion Xpress™ Plus Fragment Library Kit (Thermo Fisher Scientific) according to the protocol described in our

earlier study (79). The sequencing was carried out on the GeneStudio S5 platform (Thermo Fisher Scientific). The long-read libraries were prepared using the Native Barcoding Kit 96 V14 (SQK-NBD114.96) and sequencing was carried out on the Flow Cell (R10.4.1) on the MinION platform (Oxford Nanopore Technologies).

Bioinformatics and statistical analysis of metabarcoding sequence data

The generated prokaryotic (16S rRNA) and Eukaryotic (18S rRNA) sequence datasets were first quality-trimmed using fastp v.1.1.0 (80) and then further processed using the DADA2 v1.26 pipeline in R v4.2.0 (81) to generate amplicon sequence variants (ASVs). Low-quality and chimeric sequences were removed, and the taxonomic classification of ASVs was performed using the assignTaxonomy function of DADA2 against the SILVA v138.2 database (82) for prokaryotic datasets and the Protist Ribosomal Reference (PR2) database v5.1.0 for eukaryotic datasets (83), respectively. A maximum-likelihood phylogenetic tree was constructed by aligning ASV sequences with DECIPHER::AlignSeqs package (84) and inferring the tree with FastTree 2 under a GTR + Γ model (85) for both the 16S and 18S rRNA datasets. All statistical analyses and visualisations were performed in R v4.5.2 using microeco v2.0.0 package (86). For the 16S rRNA dataset, ASVs classified as unknown, Eukaryota, Chloroplast, and Mitochondria were removed, and the samples were subsampled to 8843 reads per sample **(Supplementary Table S8)**. For the 18S rRNA dataset, ASVs classified as unknown, Archaea, Bacteria, and Metazoa were removed, and the samples were subsampled to 16378 reads per sample **(Supplementary Table S9)**. The ASVs that were shared across hydrographic regimes were assessed using presence-absence matrices and visualised as UpSet plots using the ComplexUpset package v1.3.3 (87). Beta diversity (unweighted UniFrac distances) and prokaryotic community composition analyses were performed using microeco v2.0.0. Global and pairwise permutational multivariate analysis of variance (PERMANOVA) were computed across different hydrographic regimes using adonis2 in the vegan v2.7-2 package. Distance-decay relationships between unweighted UniFrac community dissimilarity and Haversine geographic distance were assessed using the geosphere package, with linear

regressions computed separately for within- and between-hydrography sample pairs, and regression fits restricted to the maximum within-regime geographic extent; slopes were reported per 100 km. The overall correlation between community dissimilarity and geographic distance was assessed using a Pearson correlation (Mantel test). The lifestyle preferences of Prokaryotic taxa were quantified using the particle-association niche (PAN) index (88). The contribution of neutral processes to community assembly across the transect was evaluated with the Sloan neutral community model (89). Community assembly processes within each hydrographic regime were quantified using the βNTI and RC_Bray framework (90) implemented in trans_nullmodel (microeco v2.0.0).

Bioinformatics analysis of metagenome sequence data

The base calling and demultiplexing short and long sequence reads were carried out by Torrent Suite software v5.18.1 and dorado v2.1.2, respectively. The sequencing yielded a total of 24 million short reads and about 2,68,218 long reads per sample on average **(Supplementary Table S10).** The short sequence reads were quality-checked using fastp v.1.1.0 (80) to remove low-quality and adapter sequences. For the long reads, base calling was performed using a dorado v1.3.2 and the further quality checked by fastplong v0.4.1 (91) to remove low-quality and adapter sequences. The high-quality short- and long-read sequences for each sample were assembled into longer, contiguous sequences (assemblies/contigs) using the de novo hybrid assembly approach in SPAdes v4.2.0 (92). The data generated by long-read sequencing were primarily used to fill gaps in the de novo assembly. The assemblies were further quality-checked using MetaQUAST v 5.0.2 (93), followed by automated binning was performed using four separate binning tools, namely, CONCOCT v1.1.0 (94), maxbin2 v2.2.7 (95), METABAT2 v2.12.1 (96) and BinSanity v.0.5.4 (97) to generate Metagenome Assembled Genomes (MAGs) in the Anvio v9 pipeline (98). All steps involved in reconstructing a non-redundant set of MAGs for each sample were previously described in detail (99). The MAGs were manually refined using the anvi-refine program via the anvi'o interactive interface, exploiting differential coverage, tetranucleotide frequency, and marker gene content to potentially remove contaminated sequences from the

genome. Then, DAS_Tool v1.1.2 (100) was employed to select a unique set of MAGs for each sample from all four binning tools. The reconstructed MAGs, abundance and coverage data were retrieved by using the anvi-summarize program. Further, prokaryotic MAGs completion and redundancy were calculated using CheckM2, and taxonomic classification was performed by using GTDB-Tk tool v2.5.2 against the reference database V226 (101). Eukaryotic MAG completeness, redundancy and taxonomic classification were identified using EukCC v2.1.3 (102) and EUKulele v2.0.9 (103). The reconstructed MAGs were classified into three tiers based on completeness and redundancy ratio into high-quality (HQ; ≥90% completeness, ≤5% contamination), medium-quality (MQ; ≥70%, ≤10%) and low-quality (LQ; ≥50% - ≤70%, ≤10%).

Community-wide gene-centric and genome-resolved metabolic functional analysis

To identify potential ecological functions across hydrographic regimes and resolve the metabolic contributions of individual microbial groups, metabolic functional characterisation was performed at both the community-wide (gene-centric) on metagenome assemblies and genome-resolved (MAG) levels. For the gene-centric analysis, open reading frames were predicted from all individual metagenomes using Pyrodigal v3.7.1 in metagenomic mode (104). Predicted proteins from all assemblies were pooled and clustered with MMseqs2 easy-linclust (105) at ≥95% amino-acid identity and ≥90% alignment coverage, yielding a non-redundant catalogue of about ~12 million genes. Representative sequences were functionally annotated with eggNOG-mapper v3.0 against the eggNOG v5 database (106) and with KofamScan v1.30 using KOfam profile HMMs (107), retaining only KO assignments above the KO-specific score thresholds. Gene abundances were quantified by mapping quality-filtered reads to the nucleotide sequences of the catalogue with Bowtie2 v2.4.3 using --very-sensitive-local mode (108). Ambiguously mapped reads (MAPQ = 0) were removed with SAMtools v1.21 (109), and read counts and mean coverage depth per gene were calculated with samtools coverage. The sequence read counts were normalised for sequencing depth using the median-of-ratios (relative log expression) method in DESeq2 v1.50.2 (110). To identify and visualize differences in metabolic functional composition among different hydrographic

regimes, we used PCoA and pairwise PERMANOVA analysis in the vegan v 2.7-2 package. To compare metabolic potential among regimes, we selected 128 marker genes representing 24 metabolic pathways and visualized using ComplexHeatmap v2.260 in R. Similar to gene-centric analysis, protein-coding genes were predicted from each MQ and HQ MAGs (Pyrodigal v3.7.1 for prokaryotic and MetaEuk for eukaryotic MAGs) and annotated with eggNOG-mapper v3.0 and visualized as described earlier.

Occurrence, distribution and prevalence of reconstructed MAGs across the Arctic Ocean

The reconstructed MAGs (n=469) were dereplicated using dRep v3.0.0 at 95% Average Nucleotide Identity (ANI) to obtain a non-redundant catalogue of species-level MAG clusters (sMAGs). To identify abundance, occurrence, distribution and prevalence of these sMAGs across Arctic Ocean, publicly available metagenomic sequence datasets from TARA Arctic Ocean (22) and Arctic MOSAiC drift expeditions (27) datasets were downloaded and analysed together with the metagenomes generated in this study. The competitive read recruitment analysis was carried out using the workflow (https://anvio.org/tutorials/competitive-read-recruitment/) on the Anvio V9.0 (98). For each sample, the percentage of reads recruited was calculated as the number of mapped reads divided by the number of quality-filtered reads. The detection and coverage per Gbp (Q2Q3 mean coverage/gigabases of quality-filtered sequence in the sample) were calculated. All statistical analyses including Kruskal–Wallis tests, PCoA, PERMANOVA were performed in R v 4.2.0 using vegan v2.7-2, indicspecies v1.8.0 packages.

Data availability

The raw metabarcoding sequence dataset (16S rRNA and 18S rRNA) and the metagenomic dataset (short and long reads) were submitted to the NCBI Sequence Read Archive under BioProject ID 1533043. The bioinformatics scripts used in the study are available on GitHub: https://github.com/srisvs33/CAO-microbial-biogeography . The reconstructed MAGs, metagenome assemblies, and all the other relevant metadata files were submitted to the public

repository and can be downloaded using the following link: https://zenodo.org/records/23602000

References

1. Sumata H, de Steur L, Divine DV, Granskog MA, Gerland S. 2023. Regime shift in Arctic Ocean sea ice thickness. Nature 615:443-449.
2. Babb DG, Kirillov S, Howell SEL, Landy JC, Glissenaar I, Stroeve J, Brady M, Ehn JK. 2023. The Complete Annual Record of Sea Ice Volume Export Through Fram Strait as Observed by Satellite from 2010-2022. ESS Open Archive 2023.
3. Feng D, Gleason CJ, Lin P, Yang X, Pan M, Ishitsuka Y. 2021. Recent changes to Arctic river discharge. Nature Communications 12:6917.
4. Smith MM, Fuchs N, Salganik E, Perovich DK, Raphael I, Granskog MA, Schulz K, Shupe MD, Webster M. 2025. Formation and fate of freshwater on an ice floe in the Central Arctic. The Cryosphere 19:619-644.
5. Koenig Z, Granskog MA, Muilwijk M, Randelhoff A, Renner AH, Chierici M, Fransson A, Gonçalves-Araujo R, Dodd PA. 2026. Identifying the extent of river-influenced waters of the transpolar drift with in situ measurements of dissolved organic matter fluorescence and seawater density. Journal of Geophysical Research: Oceans 131:e2024JC022200.
6. Mann PJ, Spencer RGM, Hernes PJ, Six J, Aiken GR, Tank SE, McClelland JW, Butler KD, Dyda RY, Holmes RM. 2016. Pan-Arctic Trends in Terrestrial Dissolved Organic Matter from Optical Measurements. Frontiers in Earth Science Volume 4 - 2016.
7. Terhaar J, Kwiatkowski L, Bopp L. 2020. Emergent constraint on Arctic Ocean acidification in the twenty-first century. Nature 582:379-383.
8. Polyakov IV, Pnyushkov AV, Alkire MB, Ashik IM, Baumann TM, Carmack EC, Goszczko I, Guthrie J, Ivanov VV, Kanzow T, Krishfield R, Kwok R, Sundfjord A, Morison J, Rember R, Yulin A. 2017. Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean. Science 356:285-291.
9. Tuerena RE, Mahaffey C, Henley SF, de la Vega C, Norman L, Brand T, Sanders T, Debyser M, Dähnke K, Braun J, März C. 2022. Nutrient pathways and their susceptibility to past and future change in the Eurasian Arctic Ocean. Ambio 51:355-369.
10. Polyakov IV, Alkire MB, Bluhm BA, Brown KA, Carmack EC, Chierici M, Danielson SL, Ellingsen I, Ershova EA, Gårdfeldt K, Ingvaldsen RB, Pnyushkov AV, Slagstad D, Wassmann P. 2020. Borealization of the Arctic Ocean in Response to Anomalous Advection From Sub-Arctic Seas. Frontiers in Marine Science Volume 7 - 2020.
11. Ingvaldsen RB, Assmann KM, Primicerio R, Fossheim M, Polyakov IV, Dolgov AV. 2021. Physical manifestations and ecological implications of Arctic Atlantification. Nature Reviews Earth & Environment 2:874-889.
12. Husson B, Bluhm BA, Cyr F, Danielson SL, Eriksen E, Fossheim M, Geoffroy M, Hopcroft RR, Ingvaldsen RB, Jørgensen LL, Lovejoy C, Meire L, Mueter F, Primicerio R, Winding M. 2024. Borealization impacts shelf ecosystems across the Arctic. Frontiers in Environmental Science Volume 12 - 2024.
13. Charette MA, Kipp LE, Jensen LT, Dabrowski JS, Whitmore LM, Fitzsimmons JN, Williford T, Ulfsbo A, Jones E, Bundy RM, Vivancos SM, Pahnke K, John SG, Xiang

Y, Hatta M, Petrova MV, Heimbürger-Boavida L-E, Bauch D, Newton R, Pasqualini A, Agather AM, Amon RMW, Anderson RF, Andersson PS, Benner R, Bowman KL, Edwards RL, Gdaniec S, Gerringa LJA, González AG, Granskog M, Haley B, Hammerschmidt CR, Hansell DA, Henderson PB, Kadko DC, Kaiser K, Laan P, Lam PJ, Lamborg CH, Levier M, Li X, Margolin AR, Measures C, Middag R, Millero FJ, Moore WS, Paffrath R, Planquette H, Rabe B, et al. 2020. The Transpolar Drift as a Source of Riverine and Shelf-Derived Trace Elements to the Central Arctic Ocean. Journal of Geophysical Research: Oceans 125:e2019JC015920.

14. Nöthig EM, Lalande C, Fahl K, Metfies K, Salter I, Bauerfeind E. 2020. Annual cycle of downward particle fluxes on each side of the Gakkel Ridge in the central Arctic Ocean. Philos Trans A Math Phys Eng Sci 378:20190368.

15. Liguori BTP, Ehlert C, Nöthig E-M, van Ooijen JC, Pahnke K. 2021. The Transpolar Drift Influence on the Arctic Ocean Silicon Cycle. Journal of Geophysical Research: Oceans 126:e2021JC017352.

16. Kohlbach D, Fransson A, Amargant-Arumi M, Assmann KM, Assmy P, Bratbak G, Chierici M, Cristea A, Divine DV, Down E, Eikrem W, Foss Ø, Gardner J, Gradinger RR, Granskog MA, Hess S, Ingvaldsen RB, Koenig Z, Kwaśniewski S, Majaneva S, Marquardt M, Müller O, Ntniou IV, Ormańczyk M, Raffel B, Renaud P, Renner AHH, de Freitas TR, Sen A, Smoła Z, Svensen C, Vader A, Våge S, Wiktor J, Wold A, Zabłocka M, Ziegler A, Bluhm BA. 2025. Nansen and Amundsen basins: Gradients of physico-chemical properties and biota composition with implications for future resource management of the central Arctic Ocean. Elementa: Science of the Anthropocene 13.

17. Laukert G, Bauch D, Rabe B, Krumpen T, Damm E, Kienast M, Hathorne E, Vredenborg M, Tippenhauer S, Andersen N. 2025. Dynamic ice–ocean pathways along the Transpolar Drift amplify the dispersal of Siberian matter. Nature Communications 16:3172.

18. Sandven H, McKee D, Gonçalves-Araujo R, Röttgers R, Opdal AF, Assmy P, Granskog MA. 2026. Terrestrial Runoff Divides the Central Arctic Ocean and Shapes the Biologically Relevant Marine Lightscape. Geophysical Research Letters 53:e2025GL121318.

19. Fernández-Gómez B, Díez B, Polz MF, Arroyo JI, Alfaro FD, Marchandon G, Sanhueza C, Farías L, Trefault N, Marquet PA, Molina-Montenegro MA, Sylvander P, Snoeijs-Leijonmalm P. 2019. Bacterial community structure in a sympagic habitat expanding with global warming: brackish ice brine at 85–90 °N. The ISME Journal 13:316-333.

20. Mock T, Boulton W, Balmonte J-P, Barry K, Bertilsson S, Bowman J, Buck M, Bratbak G, Chamberlain EJ, Cunliffe M, Creamean J, Ebenhöh O, Eggers SL, Fong AA, Gardner J, Gradinger R, Granskog MA, Havermans C, Hill T, Hoppe CJM, Korte K, Larsen A, Müller O, Nicolaus A, Oldenburg E, Popa O, Rogge S, Schäfer H, Shoemaker K, Snoeijs-Leijonmalm P, Torstensson A, Valentin K, Vader A, Barry K, Chen IMA, Clum A, Copeland A, Daum C, Eloe-Fadrosh E, Foster B, Foster B, Grigoriev IV, Huntemann M, Ivanova N, Kuo A, Kyrpides NC, Mukherjee S, Palaniappan K, Reddy TBK, Salamov A, et al. 2022. Multiomics in the central Arctic Ocean for benchmarking biodiversity change. PLOS Biology 20:e3001835.

21. Geoffroy M, Polyakov I, Reigstad M, Acinas SG, Amiraux R, Angot H, Ardyna M, Babin M, Bowler C, Couet D, Eddy TD, Falciatore A, Guidi L, Hoppmann M, Houssais M-N, Karp-Boss L, Lovejoy C, Marechal E, Pelletier E, Ortega-Retuerta E, Ghiglione J-F, Pohnert G, Rabe B, Salazar G, Schmale J, Schuback N, Sullivan M, Tippenhauer S, Troublé R, Trudnowska E, Vardi A, Vincent F. 2026. Oceanography, biogeochemical cycles, and biodiversity in the Central Arctic Ocean: Current state of knowledge and

research directions for the Tara Polaris expeditions. Elementa: Science of the Anthropocene 14.

22. Royo-Llonch M, Sánchez P, Ruiz-González C, Salazar G, Pedrós-Alió C, Sebastián M, Labadie K, Paoli L, M. Ibarbalz F, Zinger L, Churcheward B, Babin M, Bork P, Boss E, Cochrane G, de Vargas C, Gorsky G, Grimsley N, Guidi L, Hingamp P, Iudicone D, Jaillon O, Kandels S, Not F, Ogata H, Pesant S, Poulton N, Raes J, Sardet C, Speich S, Setmmann L, Sullivan MB, Chaffron S, Eveillard D, Karsenti E, Sunagawa S, Wincker P, Karp-Boss L, Bowler C, Acinas SG, Tara Oceans C. 2021. Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. Nature Microbiology 6:1561-1574.

23. Bowman JS, Rasmussen S, Blom N, Deming JW, Rysgaard S, Sicheritz-Ponten T. 2012. Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. The ISME Journal 6:11-20.

24. Hop H, Vihtakari M, Bluhm BA, Assmy P, Poulin M, Gradinger R, Peeken I, von Quillfeldt C, Olsen LM, Zhitina L, Melnikov IA. 2020. Changes in Sea-Ice Protist Diversity With Declining Sea Ice in the Arctic Ocean From the 1980s to 2010s. Frontiers in Marine Science Volume 7 - 2020.

25. de Sousa AGG, Tomasino MP, Duarte P, Fernández-Méndez M, Assmy P, Ribeiro H, Surkont J, Leite RB, Pereira-Leal JB, Torgo L, Magalhães C. 2019. Diversity and Composition of Pelagic Prokaryotic and Protist Communities in a Thin Arctic Sea-Ice Regime. Microbial Ecology 78:388-408.

26. Rapp JZ, Fernández-Méndez M, Bienhold C, Boetius A. 2018. Effects of Ice-Algal Aggregate Export on the Connectivity of Bacterial Communities in the Central Arctic Ocean. Frontiers in Microbiology Volume 9 - 2018.

27. Boulton W, Salamov A, Grigoriev IV, Calhoun S, LaButti K, Riley R, Barry K, Fong AA, Hoppe CJM, Metfies K, Oetjen K, Eggers SL, Müller O, Gardner J, Granskog MA, Torstensson A, Oggier M, Larsen A, Bratbak G, Toseland A, Leggett RM, Moulton V, Mock T. 2025. Metagenome-assembled-genomes recovered from the Arctic drift expedition MOSAiC. Scientific Data 12:204.

28. Venkatachalam S, Granskog MA, Gonçalves-Araujo R, Divine DV, Vipindas PV, Jabir T, Shereef A, Jain A. 2025. Distinct bacterial community structures with abundant carbon degradation and sulfur metabolisms found in different sea-ice types from the Central Arctic Ocean. Microbiology Spectrum 13:e01291-25.

29. Müller O, Gardner J, Olsen LM, Salganik E, Assmy P, Gradinger R, Bratbak G, Hoppe CJM, Lange BA, Muilwijk M, Divine DV, Aberle N, Krause JW, Reigstad M, Leu E, Tsagaraki TM, Larsen A, Høyland KV, Balmonte JP, Boulton W, Dahle H, Eggers L, Fong AA, Guillou G, Lebreton B, Metfies K, Mock T, Petelenz E, Tatarek A, Torres-Valdés S, Torstensson A, Wiktor J, Granskog MA. 2026. Arctic sea-ice ridges are biomass hotspots harboring diverse microbial communities. Communications Earth & Environment 7:385.

30. Balmonte JP, Teske A, Arnosti C. 2018. Structure and function of high Arctic pelagic, particle-associated and benthic bacterial communities. Environmental Microbiology 20:2941-2954.

31. Rego A, Fernandez-Guerra A, Duarte P, Assmy P, Leão PN, Magalhães C. 2021. Secondary metabolite biosynthetic diversity in Arctic Ocean metagenomes. Microbial Genomics 7.

32. Oldenburg E, Kronberg RM, Metfies K, Wietz M, von Appen W-J, Bienhold C, Popa O, Ebenhöh O. 2024. Beyond blooms: the winter ecosystem reset determines

microeukaryotic community dynamics in the Fram Strait. Communications Earth & Environment 5:643.

33. Chamberlain EJ, Rokitta S, Rost B, D'Angelo A, Creamean JM, Loose B, Ulfsbo A, Fong AA, Hoppe CJ, Droste ES. 2025. Predictive links between microbial communities and biological oxygen utilization in the Arctic Ocean. Limnology and Oceanography 70:2315-2331.

34. Smith MW, Zeigler Allen L, Allen AE, Herfort L, Simon HM. 2013. Contrasting genomic properties of free-living and particle-attached microbial assemblages within a coastal ecosystem. Frontiers in Microbiology Volume 4 - 2013.

35. Jain A, Balmonte JP, Singh R, Bhaskar PV, Krishnan KP. 2021. Spatially resolved assembly, connectivity and structure of particle-associated and free-living bacterial communities in a high Arctic fjord. FEMS Microbiology Ecology 97.

36. Jain A, Krishnan KP, Singh A, Thomas FA, Begum N, Tiwari M, Bhaskar VP, Gopinath A. 2019. Biochemical composition of particles shape particle-attached bacterial community structure in a high Arctic fjord. Ecological Indicators 102:581-592.

37. Polyakov IV, Pnyushkov AV, Charette M, Cho K-H, Jung J, Kipp L, Muilwijk M, Whitmore L, Yang EJ, Yoo J. 2025. Atlantification advances into the Amerasian Basin of the Arctic Ocean. Science Advances 11:eadq7580.

38. Manizza M, Menemenlis D, Zhang H, Miller CE. 2019. Modeling the Recent Changes in the Arctic Ocean CO2 Sink (2006–2013). Global Biogeochemical Cycles 33:420-438.

39. Fadeev E, Rogge A, Ramondenc S, Nöthig E-M, Wekerle C, Bienhold C, Salter I, Waite AM, Hehemann L, Boetius A, Iversen MH. 2021. Sea ice presence is linked to higher carbon export and vertical microbial connectivity in the Eurasian Arctic Ocean. Communications Biology 4:1255.

40. Spreen G, Kaleschke L, Heygster G. 2008. Sea ice remote sensing using AMSR-E 89-GHz channels. Journal of Geophysical Research: Oceans 113.

41. Schlitzer R. 2022. Ocean data view.

42. Pan Y, Tao Y, Yang X, Du S, Ding H, Li J, Jia H, Chen H. 2025. Underlying mechanisms of spatial distribution of prokaryotic community in surface seawater from Arctic Ocean to the Sea of Japan. Microbiology Spectrum 13:e00517-25.

43. Ravikumar P, Ravindran A, Raman K. 2026. Deciphering global patterns of marine microbial community assembly and network stability. mSystems 11:e00470-26.

44. Milke F, Wagner-Doebler I, Wienhausen G, Simon M. 2022. Selection, drift and community interactions shape microbial biogeographic patterns in the Pacific Ocean. The ISME Journal 16:2653-2665.

45. Liu Q, Li Y, Wang H, Yang G, Kan J, Yang M, Yu X, Guo C, Wang M, Wang W, Zhang Q, Zhu J, Zhao X, Jiang Y. 2023. Assembly and Network Stability of Planktonic Microorganisms under the Influence of Salinity Gradient: an Arctic Case Study from the Lena River Estuary to the Laptev Sea. Microbiology Spectrum 11:e02115-22.

46. Faure E, Pommellec J, Noel C, Cormier A, Delpech L-M, Eren AM, Fernandez-Guerra A, Vanni C, Fourquez M, Houssais M-N, Guyet U, Da Silva C, Gavory F, Perdereau A, Labadie K, Wincker P, Poulain J, Hassler C, Lin Y, Cassar N, Maignien L. 2026. Water mass specific genes dominate the Southern Ocean microbiome. Nature Communications 17:2025.

47. Salazar G, Paoli L, Alberti A, Huerta-Cepas J, Ruscheweyh H-J, Cuenca M, Field CM, Coelho LP, Cruaud C, Engelen S, Gregory AC, Labadie K, Marec C, Pelletier E, Royo-Llonch M, Roux S, Sánchez P, Uehara H, Zayed AA, Zeller G, Carmichael M, Dimier

C, Ferland J, Kandels S, Picheral M, Pisarev S, Poulain J, Acinas SG, Babin M, Bork P, Boss E, Bowler C, Cochrane G, de Vargas C, Follows M, Gorsky G, Grimsley N, Guidi L, Hingamp P, Iudicone D, Jaillon O, Kandels-Lewis S, Karp-Boss L, Karsenti E, Not F, Ogata H, Pesant S, Poulton N, Raes J, Sardet C, et al. 2019. Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome. Cell 179:1068-1083.e21.

48. Bianchi D, Weber TS, Kiko R, Deutsch C. 2018. Global niche of marine anaerobic metabolisms expanded by particle microenvironments. Nature Geoscience 11:263-268.

49. Santos-García M, Ganeshram RS, Oziel L, Dodd PA, de Steur L, Tuerena RE, Stedmon CA. 2026. Sea ice loss drives a regime shift in Arctic Ocean nitrogen biogeochemistry. Communications Earth & Environment 7:442.

50. Wietz M, Engel A, Ramondenc S, Niwano M, von Appen W-J, Priest T, von Jackowski A, Metfies K, Bienhold C, Boetius A. 2024. The Arctic summer microbiome across Fram Strait: Depth, longitude, and substrate concentrations structure microbial diversity in the euphotic zone. Environmental Microbiology 26:e16568.

51. Ruben M, Marchant H, Wietz M, Gentz T, Strauss J, Koch BP, Mollenhauer G. 2024. Microbial Communities Degrade Ancient Permafrost-Derived Organic Matter in Arctic Seawater. Journal of Geophysical Research: Biogeosciences 129:e2023JG007936.

52. Teeling H, Fuchs BM, Becher D, Klockow C, Gardebrecht A, Bennke CM, Kassabgy M, Huang S, Mann AJ, Waldmann J, Weber M, Klindworth A, Otto A, Lange J, Bernhardt J, Reinsch C, Hecker M, Peplies J, Bockelmann FD, Callies U, Gerdts G, Wichels A, Wiltshire KH, Glöckner FO, Schweder T, Amann R. 2012. Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. Science 336:608-611.

53. Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L, Reintjes G, Waldmann J, Quast C, Glöckner FO, Lucas J, Wichels A, Gerdts G, Wiltshire KH, Amann RI. 2016. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. eLife 5:e11888.

54. Priest T, von Appen W-J, Oldenburg E, Popa O, Torres-Valdés S, Bienhold C, Metfies K, Boulton W, Mock T, Fuchs BM, Amann R, Boetius A, Wietz M. 2023. Atlantic water influx and sea-ice cover drive taxonomic and functional shifts in Arctic marine bacterial communities. The ISME Journal 17:1612-1625.

55. Louca S, Polz MF, Mazel F, Albright MBN, Huber JA, O'Connor MI, Ackermann M, Hahn AS, Srivastava DS, Crowe SA, Doebeli M, Parfrey LW. 2018. Function and functional redundancy in microbial systems. Nature Ecology & Evolution 2:936-943.

56. Azam F, Malfatti F. 2007. Microbial structuring of marine ecosystems. Nature Reviews Microbiology 5:782-791.

57. Curson ARJ, Liu J, Bermejo Martínez A, Green RT, Chan Y, Carrión O, Williams BT, Zhang S-H, Yang G-P, Bulman Page PC, Zhang X-H, Todd JD. 2017. Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. Nature Microbiology 2:17009.

58. Lund-Hansen LC, Bendtsen J, Stratmann T, Tonboe R, Olsen SM, Markager S, Sorrell BK. 2020. Will low primary production rates in the Amundsen Basin (Arctic Ocean) remain low in a future ice-free setting, and what governs this production? Journal of Marine Systems 205:103287.

59. Randelhoff A, Holding J, Janout M, Sejr MK, Babin M, Tremblay J-É, Alkire MB. 2020. Pan-Arctic Ocean Primary Production Constrained by Turbulent Nitrate Fluxes. Frontiers in Marine Science Volume 7 - 2020.

60. Oziel L, Gürses Ö, Torres-Valdés S, Hoppe CJM, Rost B, Karakuş O, Danek C, Koch BP, Nissen C, Koldunov N, Wang Q, Völker C, Iversen M, Juhls B, Hauck J. 2025. Climate change and terrigenous inputs decrease the efficiency of the future Arctic Ocean's biological carbon pump. Nature Climate Change 15:171-179.

61. Jahn A, Holland MM, Kay JE. 2024. Projections of an ice-free Arctic Ocean. Nature Reviews Earth & Environment 5:164-176.

62. Dodd PA, Nikolopoulos A, Buckley S, Campbell K, Divine D, Eggen V, Gonçalves-Araujo R, Granskog M, Hop H, Kern Y. 2022. Arctic Ocean 2022 cruise report.

63. Dodd P, Buckley S, Campbell K, Chierici M, Divine D, Eggen V, Fransson A, Gonçalves-Araujo R, Granskog M, Hop H. 2022. CTD profiles from NPI cruise AO-2022 across the Nansen and Amundsen Basins of the Arctic Ocean and core parameters measured from niskin bottle samples.[Dataset]. Norwegian Polar Institute.

64. Epstein S, Mayeda T. 1953. Variation of O18 content of waters from natural sources. Geochimica et Cosmochimica Acta 4:213-224.

65. Östlund HG, Hut G. 1984. Arctic Ocean water mass balance from isotope data. Journal of Geophysical Research: Oceans 89:6373-6381.

66. Bauch D, Schlosser P, Fairbanks RG. 1995. Freshwater balance and the sources of deep and bottom waters in the Arctic Ocean inferred from the distribution of H218O. Progress in Oceanography 35:53-80.

67. Dodd PA, Rabe B, Hansen E, Falck E, Mackensen A, Rohling E, Stedmon C, Kristiansen S. 2012. The freshwater composition of the Fram Strait outflow derived from a decade of tracer measurements. Journal of Geophysical Research: Oceans 117.

68. Cooper L, McClelland J, Holmes R, Raymond P, Gibson J, Guay C, Peterson B. 2008. Flow-weighted values of runoff tracers (δ18O, DOC, Ba, alkalinity) from the six largest Arctic rivers. Geophysical Research Letters 35.

69. Grasshoff K, Ehrhardt M, Kremling K. 1983. Sea Water Analysis. Verlag Chemie: Weinheim, Germany.

70. Gundersen K, Møgster JS, Lien VS, Ershova E, Lunde LF, Arnesen H, Olsen A-K. 2022. Thirty Years of Nutrient Biogeochemistry in the Barents Sea and the adjoining Arctic Ocean, 1990–2019. Scientific Data 9:649.

71. Wünsch UJ, Gonçalves-Araujo R, Granskog MA, Gry Ebbesen L, Papadimitraki M, Stedmon CA. 2025. Fluorescence as a tracer of the susceptibility of dissolved organic matter to photodegradation in the Arctic Ocean. Limnology and Oceanography 70:2299-2314.

72. Coble PG. 2007. Marine Optical Biogeochemistry: The Chemistry of Ocean Color. Chemical Reviews 107:402-418.

73. Paulsen ML, Doré H, Garczarek L, Seuthe L, Müller O, Sandaa R-A, Bratbak G, Larsen A. 2016. Synechococcus in the Atlantic Gateway to the Arctic Ocean. Frontiers in Marine Science Volume 3 - 2016.

74. Zubkov MV, Burkill PH, Topping JN. 2007. Flow cytometric enumeration of DNA-stained oceanic planktonic protists. Journal of Plankton Research 29:79-86.

75. Marie D, Partensky F, Vaulot D, Brussaard C. 1999. Enumeration of phytoplankton, bacteria, and viruses in marine samples. Current protocols in cytometry 10:11.11. 1-11.11. 15.

76. Takahashi S, Tomita J, Nishioka K, Hisada T, Nishijima M. 2014. Development of a Prokaryotic Universal Primer for Simultaneous Analysis of Bacteria and Archaea Using Next-Generation Sequencing. PLOS ONE 9:e105592.

77. Venkatachalam S, Kannan VM, Saritha VN, Loganathachetti DS, Mohan M, Krishnan KP. 2021. Bacterial diversity and community structure along the glacier foreland of Midtre Lovénbreen, Svalbard, Arctic. Ecological Indicators 126:107704.

78. Comeau AM, Li WKW, Tremblay J-É, Carmack EC, Lovejoy C. 2011. Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. PLOS ONE 6:e27492.

79. Vipindas PV, Venkatachalam S, Jabir T, Yang EJ, Cho KH, Jung J, Lee Y, Moon JK, Jain A. 2025. Distribution of Surface-Layer Prokaryotes in the Western Arctic Ocean: Responses to Pacific Water Inflow and Sea Ice Melting. Environmental Microbiology 27:e70154.

80. Chen S. 2025. Fastp 1.0: An ultra-fast all-round tool for FASTQ data quality control and preprocessing. Imeta 4:e70078.

81. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13:581-583.

82. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41:D590-D596.

83. Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, Boutte C, Burgaud G, de Vargas C, Decelle J, del Campo J, Dolan JR, Dunthorn M, Edvardsen B, Holzmann M, Kooistra WHCF, Lara E, Le Bescot N, Logares R, Mahé F, Massana R, Montresor M, Morard R, Not F, Pawlowski J, Probert I, Sauvadet A-L, Siano R, Stoeck T, Vaulot D, Zimmermann P, Christen R. 2013. The Protist Ribosomal Reference database (PR2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Research 41:D597-D604.

84. Wright ES. 2016. Using DECIPHER v2. 0 to analyze big biological sequence data in R.

85. Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE 5:e9490.

86. Liu C, Cui Y, Li X, Yao M. 2021. microeco: an R package for data mining in microbial community ecology. FEMS Microbiology Ecology 97.

87. Krassowski M, Arts M, Lagger C. Max. krassowski/complex-upset: v1. 3.5. Zenodo; 2020.

88. Salazar G, Cornejo-Castillo FM, Borrull E, Díez-Vives C, Lara E, Vaqué D, Arrieta JM, Duarte CM, Gasol JM, Acinas SG. 2015. Particle-association lifestyle is a phylogenetically conserved trait in bathypelagic prokaryotes. Molecular Ecology 24:5692-5706.

89. Sloan WT, Lunn M, Woodcock S, Head IM, Nee S, Curtis TP. 2006. Quantifying the roles of immigration and chance in shaping prokaryote community structure. Environmental Microbiology 8:732-740.

90. Stegen JC, Lin X, Fredrickson JK, Chen X, Kennedy DW, Murray CJ, Rockhold ML, Konopka A. 2013. Quantifying community assembly processes and identifying features that impose them. The ISME Journal 7:2069-2079.

91. Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta 2:e107.

92. Prjibelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes De Novo Assembler. Current Protocols in Bioinformatics 70:e102.

93. Mikheenko A, Saveliev V, Gurevich A. 2016. MetaQUAST: evaluation of metagenome assemblies. Bioinformatics 32:1088-1090.

94. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Loman NJ, Andersson AF, Quince C. 2013. CONCOCT: clustering contigs on coverage and composition. arXiv preprint arXiv:13124038.

95. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32:605-607.

96. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7:e7359.

97. Graham ED, Heidelberg JF, Tully BJ. 2017. BinSanity: unsupervised clustering of environmental microbial assemblies using coverage and affinity propagation. PeerJ 5:e3035.

98. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD, Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A, Scott JJ, Vázquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE, Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nature Microbiology 6:3-6.

99. Venkatachalam S, Vipindas PV, Jabir T, Jain A, Krishnan KP. 2024. Metagenomic insights into novel microbial lineages with distinct ecological functions in the Arctic glacier foreland ecosystems. Environmental Research 241:117726.

100. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature Microbiology 3:836-843.

101. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. Bioinformatics 38:5315-5316.

102. Saary P, Mitchell AL, Finn RD. 2020. Estimating the quality of eukaryotic genomes recovered from metagenomic analysis with EukCC. Genome Biology 21:244.

103. Krinos AI, Hu SK, Cohen NR, Alexander H. 2020. EUKulele: Taxonomic annotation of the unsung eukaryotic microbes. arXiv preprint arXiv:201100089.

104. Larralde M. 2022. Pyrodigal: Python bindings and interface to Prodigal, an efficient method for gene prediction in prokaryotes. Journal of Open Source Software 7:4296.

105. Steinegger M, Söding J. 2018. Clustering huge protein sequence sets in linear time. Nature Communications 9:2542.

106. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. Molecular Biology and Evolution 38:5825-5829.

107. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics 36:2251-2252.

108. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nature methods 9:357-359.
109. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10.
110. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15:550.

Acknowledgements

The authors (S.V., P.V.V., T.J., A.C., A.S., N.M., A.J.) wish to express their sincere gratitude to the Director, National Centre for Polar and Ocean Research (NCPOR), for providing the necessary facilities to carry out this research. This work was supported by the Polar Science and Cryosphere Research (PACER) program implemented by the Ministry of Earth Sciences (Govt. of India) through the NCPOR, India. A.V, M.A.G and P.A were partly supported by the Fram Centre program for Sustainable Development of the Arctic Ocean (SUDARCO) Cristin-ID: 2551323. P.A was partly supported by iC3: Centre for ice, Cryosphere, Carbon and Climate supported by the Research Council of Norway through its Centres of Excellence funding scheme (grant number 332635). We thank the Arka supercomputing facility by the Ministry of Earth Sciences (Govt. of. India), located at the Indian Institute of Tropical Meteorology (IIT), Pune, for providing computational resources to carry out this research work. We would like to thank the captain and crew of FF Kronprins Haakon and the science team onboard for their support. S.V. thanks Dr. K. P. Krishnan (late), Group Director of Arctic Operations and Logistics, NCPOR, for his support and guidance during the initial phase of the work. The present research work is a part of the Indian scientific expedition to the Arctic in 2022, Horizon 2020 Arctic PASSION and Norwegian Polar Institutes Arctic Ocean Cruise 2022 (AO2022). NCPOR contribution for This is NCPOR contribution no: XXXX.

Author contribution

Siddarthan Venkatachalam: Conceptualization, Methodology, Formal analysis, Investigation, Bioinformatics analysis, Data curation, Writing – Original draft preparation, Visualization,

Writing – Review & Editing, Project administration; Mats A. Granskog: Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing – Original draft preparation, Visualization, Writing – Review & Editing, Funding acquisition; Puthiya Veettil Vipindas: Methodology, Formal analysis, Investigation, Data curation, Writing – Original draft preparation, Visualization, Writing – Review & Editing; Gunnar Bratbak: Methodology, Formal analysis, Investigation, Data curation, Writing – Original draft preparation, Visualization, Writing – Review & Editing, Funding acquisition; Philipp Assmy: Formal analysis, Investigation, Writing – Review & Editing, Funding acquisition; Anna Vader: Formal analysis, Investigation, Writing – Review & Editing; Rafael Gonçalves-Araujo: Formal analysis, Investigation, Writing – Review & Editing; Thajudeen Jabir: Formal analysis, Investigation, Writing – Review & Editing;Oliver Müller: Formal analysis, Investigation, Writing – Review & Editing;Anju Chalari: Formal analysis, Investigation, Writing – Review & Editing;Ahammed Shereef: Formal analysis, Investigation, Writing – Review & Editing; Nuncio Murukesh: Formal analysis, Investigation, Writing – Review & Editing; Anand Jain: Conceptualization, Methodology, Formal analysis, Investigation, Writing – Review & Editing, Funding acquisition, Project administration

Competing interests

The authors declare no competing interests.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- TableS5envfitsoil.csv
- TableS4soillanduseeffects.csv
- TableS7normalizationsizefactors.csv
- validationvspublished.csv
- TableS3samplesummary.csv
- TableS1soilvaluecorrections.csv
- mainTable8pHassociationsbothregions.csv
- TableS8genesexcludedlowdetection.csv
- bacterialorders.csv
- TableS6Burkholderialesfunctionlinks.csv
- samplemetadataanddiversity.csv
- permanova.csv
- mainTable1datasetandsoilsummary.csv
- functionaldiversitytests.csv
- permanovapairwise.csv
- TableS10sabahsoilsummary.csv
- TableS12markergenespathwaysfull.csv
- mainTable7pooledoilpalmeffects.csv
- crossstudymeta.py
- sumatrareanalysis.py
- SubmissionChecklist.docx