Microbial Ecology (2019) 78:388–408
https://doi.org/10.1007/s00248-018-01314-2

ENVIRONMENTAL MICROBIOLOGY

Diversity and Composition of Pelagic Prokaryotic and Protist Communities in a Thin Arctic Sea-Ice Regime

António Gaspar G. de Sousa[1,2] · Maria Paola Tomasino[1] · Pedro Duarte[3] · Mar Fernández-Méndez[3] · Philipp Assmy[3] · Hugo Ribeiro[1] · Jaroslaw Surkont[4] · Ricardo B. Leite[4] · José B. Pereira-Leal[4] · Luís Torgo[5,6] · Catarina Magalhães[1,2]

Received: 26 July 2018 / Accepted: 25 December 2018 / Published online: 8 January 2019

Abstract

One of the most prominent manifestations of climate change is the changing Arctic sea-ice regime with a reduction in the summer sea-ice extent and a shift from thicker, perennial multiyear ice towards thinner, first-year ice. These changes in the physical environment are likely to impact microbial communities, a key component of Arctic marine food webs and biogeochemical cycles. During the Norwegian young sea ICE expedition (N-ICE2015) north of Svalbard, seawater samples were collected at the surface (5 m), subsurface (20 or 50 m), and mesopelagic (250 m) depths on 9 March, 27 April, and 16 June 2015. In addition, several physical and biogeochemical data were recorded to contextualize the collected microbial communities. Through the massively parallel sequencing of the small subunit ribosomal RNA amplicon and metagenomic data, this work allows studying the Arctic's microbial community structure during the late winter to early summer transition. Results showed that, at compositional level, *Alpha-* (30.7%) and *Gammaproteobacteria* (28.6%) are the most frequent taxa across the prokaryotic N-ICE2015 collection, and also the most phylogenetically diverse. Winter to early summer trends were quite evident since there was a high relative abundance of thaumarchaeotes in the under-ice water column in late winter while this group was nearly absent during early summer. Moreover, the emergence of *Flavobacteria* and the SAR92 clade in early summer might be associated with the degradation of a spring bloom of *Phaeocystis*. High relative abundance of hydrocarbonoclastic bacteria, particularly *Alcanivorax*

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00248-018-01314-2) contains supplementary material, which is available to authorized users.

✉ António Gaspar G. de Sousa
antonio.sousa@ciimar.up.pt

Maria Paola Tomasino
mtomasino@ciimar.up.pt

Pedro Duarte
pedro.duarte@npolar.no

Mar Fernández-Méndez
mar.fernandez.mendez@npolar.no

Philipp Assmy
philipp.assmy@npolar.no

Hugo Ribeiro
hribeiro@ciimar.up.pt

Jaroslaw Surkont
jareksurkont@gmail.com

Ricardo B. Leite
rleite@igc.gulbenkian.pt

José B. Pereira-Leal
jleal@igc.gulbenkian.pt

Luís Torgo
ltorgo@dal.ca

Catarina Magalhães
cmagalhaes@ciimar.up.pt

[1] CIIMAR/CIMAR – Interdisciplinary Centre of Marine and Environmental Research, University of Porto, Terminal de Cruzeiros do Porto de Leixões, Av. General Norton de Matos s/n, 4450–208 Porto, Portugal

[2] Department of Biology, Faculty of Sciences, University of Porto, Rua Campo Alegre s/n, 4169-007 Porto, Portugal

[3] Norwegian Polar Institute, Fram Centre, N-9296 Tromsø, Norway

[4] Instituto Gulbenkian de Ciência, Rua da Quinta Grande, 6, 2780-156 Oeiras, Portugal

[5] LIAAD - Laboratory of Artificial Intelligence and Decision Support, INESC Tec, Porto, Portugal

[6] Faculty of Computer Science, Dalhousie University, Halifax, Canada, USA

(54.3%) and *Marinobacter* (6.3%), was also found. Richness showed different patterns along the depth gradient for prokaryotic (highest at mesopelagic depth) and protistan communities (higher at subsurface depths). The microbial N-ICE2015 collection analyzed in the present study provides comprehensive new knowledge about the pelagic microbiota below drifting Arctic sea-ice. The higher microbial diversity found in late winter/early spring communities reinforces the need to continue with further studies to properly characterize the winter microbial communities under the pack-ice.

Keywords Arctic Ocean · Microbiota · SSU rRNA amplicon · Diversity · Structure · Prokaryotes · Protists · Sea ice

Introduction

Over the last 30 years, the Arctic summer sea-ice extent and thickness have drastically decreased [1, 2]. As a consequence, the ice pack became much younger, and the older and thicker multiyear ice (MYI) that survives summer melt has largely disappeared and been replaced with first-year ice (FYI) [3–5]. The thinner and younger sea-ice regime that the Arctic is facing leads to changes in Arctic phytoplankton dynamics and biogeochemistry [6–10].

The biogeochemical and ecological implications of the changing Arctic sea-ice regime need to be monitored in detail at different trophic levels in order to assess its consequences for primary production and ecosystem sustainability. Microbial communities play a central role when evaluating the ecological impact of the Arctic's thinner ice regime, as they form the dietary basis of marine food webs and are central players in biogeochemical cycles.

Previous studies on the structure and diversity of microbial communities in sea-ice and the underlying water column showed comparable microbial diversity but distinct differences in the relative contribution of the major taxa [11–13]. At the compositional level, one of the most remarkable differences is the low abundance of the *Archaea* domain in sea-ice compared with polar surface waters where they are highly abundant [11, 14]. In addition, a transition from stable MYI to transient FYI bacterial communities is expected [13], reflected in the reduction of thaumarchaeotes observed in the surface waters after the sea-ice minimum in September 2007 [8]. Wilson et al. [15] identified the light regime and phytoplankton blooms as the main variables responsible for shaping marine prokaryotic communities throughout an annual cycle, with particular focus on the light-inhibited *Thaumarchaeota* class, which was abundant in winter surface waters but nearly absent during spring/summer seasons. Recently, this pattern was further linked to water mass characteristics [16].

Marine picoeukaryotes have been reported as the main photosynthetic protists in the Arctic Ocean [17, 18], mostly related to *Phaeocystis* sp. and *Micromonas* sp. (mixotrophic), and their distribution has been associated to water mass characteristics. Indeed, it has been reported that early *Phaeocystis pouchetii* spring blooms will deplete the nitrate surface inventory [8, 9] with possible negative effects on the magnitude of the diatom spring bloom [6]. As a direct consequence, a shift in dominance from diatoms towards *P. pouchetii* and other small-sized phytoplankton, which are much more competitive under nutrient-limiting conditions, has been noticed in recent years [19–22]. On the other hand, cyanobacteria are being described as nearly absent phototrophs in these ecosystems [11, 23]. Indeed, Fernández-Méndez et al. [24] found only one nitrogen-fixing cyanobacterial phylotype retrieved in sea ice, suggesting that nitrogen-fixing cyanobacteria are rare in the Central Arctic Ocean. Instead, most of the diversity of the *nifH* gene retrieved from sea ice, water column, and melt ponds of the Arctic Ocean was affiliated with non-cyanobacterial phylotypes such as *Proteobacteria* from *nifH* Cluster I and *Deltaproteobacteria* (including anaerobic diazotrophs), from *nifH* Cluster III [24]. Despite these observations, the Arctic Ocean's microbial communities are poorly characterized at a genetic level, especially north of Svalbard [25], which is well mirrored by the limited number of recently published works [15–18, 24]. Those few studies focused on specific functional groups of microbial communities based on their ecological role, i.e., photosynthetic picoeukaryotes, marine prokaryotes, and diazotrophs. In addition, the dynamics of winter microbial communities (prokaryotic and protist) underneath the Arctic ice pack are still poorly known [25].

This study describes the pelagic microbial communities below the drifting sea-ice during the Norwegian young sea ICE expedition 2015 (N-ICE2015) [26] that took place in the ice pack north of Svalbard, between January and June 2015. N-ICE2015 provided the environmental context for the high-throughput Illumina amplicon and metagenomic (regarding 16S rRNA genes) sequencing data for this study to improve current knowledge about the microbiota of the Arctic Ocean. It was also evaluated how the prokaryote and protist diversity is structured vertically and temporally during the late winter to early summer transition in order to get an insight about how microbial communities are being shaped by the changing sea-ice regime.

Methods

Sampling Sites

During the N-ICE2015 expedition, drifting ice camps in the southern Nansen Basin of the Arctic Ocean were established using RV Lance as a research platform and comprehensive data sets of the atmosphere-snow-ice-ocean system were obtained [26]. This study uses environmental data collected during N-ICE2015 that are described in detail in other papers and datasets [27–29].

Three microbial sample sets ($n = 9$) were collected on 9 March, 27 April, and 16 June during the drift of floes 2, 3, and 4 respectively [26]. Seawater was collected under snow-covered sea-ice at the surface (5 m), subsurface (20 or 50 m), and mesopelagic (250 m) depths. The northernmost site sampled on 9 March was situated over the deep Nansen Basin (NB) and the southernmost site sampled on 16 June was collected over the shallower Yermak Plateau (YP) (Fig. 1). We refer to the site sampled on 27 April over the slope towards YP as the Transition Region (TR) (Fig. 1). The area of investigation falls within the Yermak and Svalbard branches of the West Spitsbergen Current (WSC) that transports warmer and saltier Atlantic water along the West coast of Svalbard into the Arctic Ocean [29]. The advection of relatively warm Atlantic water masses has a profound effect on the Eastern European Arctic including the Svalbard Archipelago [31]. Three distinct water masses, regarding temperature and salinity properties, can be differentiated in our investigated area [29]: a surface layer of less saline and colder polar surface water (PSW), an intermediate layer of modified Atlantic water (MAW), separated from the fresher PSW by a sharp halocline, and lastly, a deeper layer of Atlantic Water (AW). Additionally, the melting sea-ice influences the formation of warm polar surface waters (PSWw) over the YP during springtime [29].

The sampling does not allow a clear separation between temporal and spatial variability. However, following Meyer et al. [29], similar water masses were sampled in the three dates when microbial samples were collected (Supplementary Table S1 and Supplementary Fig. S1) with the exception of MAW, sampled in 9 March and 27 April,

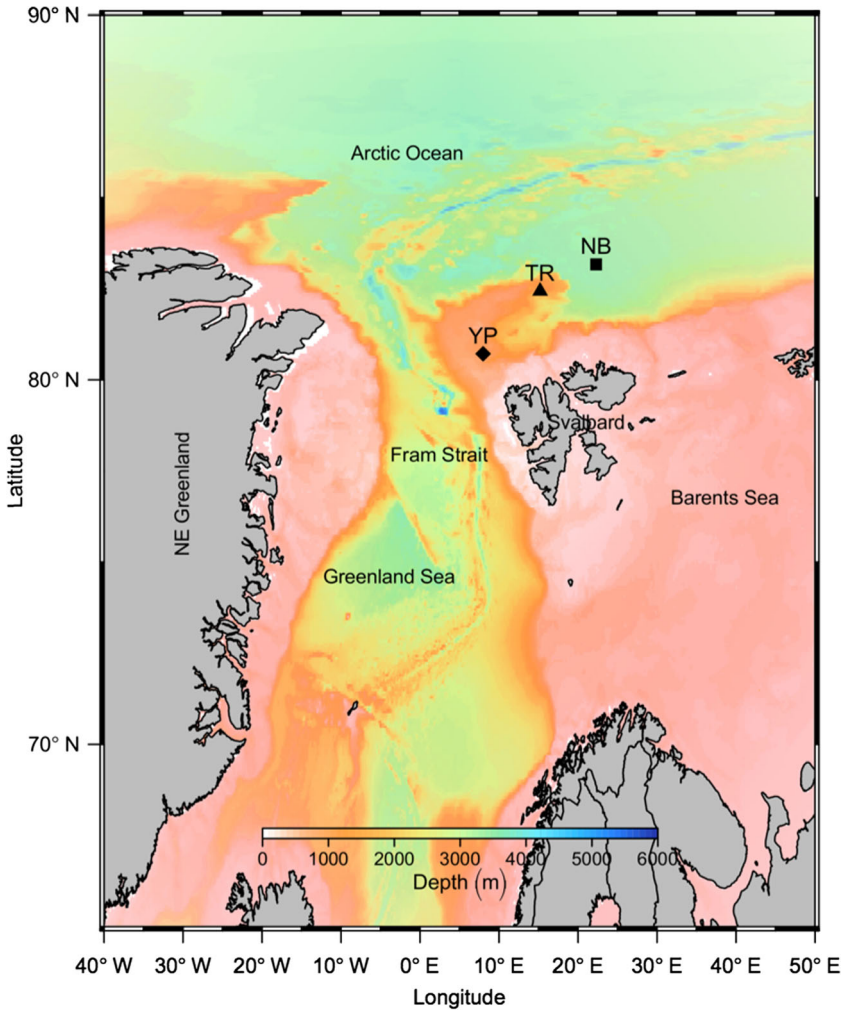

Fig. 1 Map highlighting the sampling sites north of Svalbard. Samples were collected at surface, subsurface, and mesopelagic depths over Nansen Basin (square, NB, 09.03.2015), Transition Region (triangle, TR, 27.04.2015) and Yermak Plateau (diamond, YP, 16.06.2015; see Supplementary Table S1). Map was done using the oceanmap R package [30] and with the public available bathymetry downloaded from the NOAA ETOPO1 repository

versus AW, sampled on 16 June. This similarity reduces the difficulties in interpreting the data and in contrasting observations done during in different seasons.

Water Column Sampling: Environmental Data and Microbial Collection

Details about the equipment used for physical, chemical, and biological water samples during N-ICE2015 are given in Meyer et al. [29]. This included a vessel-mounted CTD and a CTD operated through a hole made in the ice, both attached to a multi-bottle carousel water sampler. The former CTD (conductivity, temperature, and depth) included several sensors to measure, apart from salinity and temperature, chlorophyll, fluorescence (WETLabsECO-AFL model no. FLRTD-1547), and photosynthetically active radiation (PAR, model no. QCP2300-HP). Moreover, Ice-Atmosphere-Arctic Ocean Observing System (IAOOS) profilers equipped with a CTD and microstructure profilers with temperature, conductivity, and depth sensors were also employed. Water samples collected during CTD deployments were used to measure a large number of variables including concentrations of oxygen, ammonium (NH_4^+), nitrate (NO_3^-) plus nitrite (NO_2^-), phosphate (PO_4^{3-}), silicic acid (SiO_4^{4-}), total dissolved nitrogen (TDN), dissolved organic carbon (DOC), particulate organic carbon (POC) and nitrogen (PON), chlorophyll, and phaeopigments.

The water column physical and biogeochemical data mentioned in the previous paragraph are described and available in Peterson et al. [27] and Assmy et al. [28]. All the measurements made regarding the physical and biogeochemical data contextualizing the environment of the microbial N-ICE2015 collection are presented in Table 1.

In order to get reproducible, reliable, and compatible results, we adopted the scheme from the Ocean Sampling Day campaign, from sampling to sequencing the small subunit of ribosomal RNA (SSU rRNA) and the environmental DNA (shotgun metagenomics) of microbial communities [33]. The seawater was collected using a GoFlow bottle (20 L) and then filtered through a Sterivex® Filter with a 0.22 μm pore size, hydrophilic, PVDF, Durapore membrane (SVGV010RS, Merck Millipore, Portugal) with the help of a peristaltic pump. The filters were sealed and stored at − 80 °C until further analysis. See sample IDs assigned to each sample and their summarized description in Supplementary Table S1.

DNA Extraction, PCR, Library Preparation, and Sequencing of SSU rRNA Amplicons and Metagenomes

SterivexTM filters were thawed at room temperature and the plastic cover that surrounds the filter was cut off to remove the filter itself. Then, DNA extraction was performed using the PowerWater® DNA Isolation Kit protocol (MO BIO Laboratories, Inc.) following the manufacturer's instructions.

The 16S rRNA gene was amplified with the primer pair 515YF (5′ - GTGYCAGCMGCCGCGGTAA - 3′) and Y906R-jed (5′ - CCGYCAATTYMTTTRAGTTT - 3′), which was designed by Caporaso et al. [34, 35] and the latter modified by Apprill et al. [36] and Parada et al. [37]. The primer set has as target the V4-V5 hypervariable regions of the 16S rRNA gene. Both primers (515YF/Y906R-jed) have a degeneracy to cover a broad spectrum of diversity, specifically the *Crenarchaeota*/*Thaumarchaeota* (degeneracy at 515YF) phylum and the marine and freshwater clade SAR11 (alphaproteobacterial class; degeneracy at Y906R-jed) (http://www.earthmicrobiome.org) [36, 37]. The 18S rRNA gene was amplified with the primer set described in Stoeck et al. [38], TAReuk454FWD1 (5′ – CCAGCASCYGCGGTAATTCC – 3′) and TAReukREV3_modified (5′ – ACTTTCGTTCTTGATYRATGA – 3′), with the exception of reverse primer (TAReukREV3_modified) which had an additional TGA triplet added at the 3′ end compared to the original [39]. This primer set amplifies the V4 region of the 18S rRNA gene.

Both SSU rRNA genes independently amplified by PCR were used to build Illumina paired-end libraries sequenced on an Illumina MiSeq platform using 2 × 300 bp, V3 Chemistry (Illumina). These steps were performed by LGC Genomics (LGC Genomics GmbH, Berlin, Germany) and a detailed description is given in Ribeiro et al. [40].

The source of environmental DNA used to prepare the nine libraries for shotgun metagenomic sequencing was the same as previously used to sequencing the SSU rRNA amplicon genes according to the following steps: (i) the quality of environmental DNA was checked in an agarose gel; (ii) eDNA was sheared in small pieces (600 bp) using a Covaris sonicator and the fragments of selected size purified using AMPure XP beads (Agencourt); (iii) about 200 ng per sample of fragmented DNA was picked to construct Illumina libraries using the Ovation Rapid DR Multiplex System 1–96 (NuGEN); and (iv) the libraries were pooled and the size picked through gel electrophoresis. Finally, paired-end sequencing of Illumina libraries was performed on an Illumina MiSeq sequencer using V3 Chemistry (Illumina). All the steps mentioned before were performed by LGC Genomics (LGC Genomics GmbH, Berlin, Germany).

Bioinformatics Pipeline for SSU rRNA Amplicons

Upstream Sequence Analysis: Raw OTU Table

The mothur pipeline was used to preprocess and assign taxonomy independently for each SSU rRNA library, i.e., 16S and 18S rRNA datasets, from the N-ICE2015 campaign following the MiSeq Standard Operating Procedure (SOP)

Table 1 Physical and biogeochemical conditions contextualizing the environment of microbial N-ICE2015 collection

Sample ID		NB_5	NB_50	NB_250	TR_5	TR_50	TR_250	YP_5	YP_20	YP_250	Dataset reference
Physical	Conservative temperature (°C)[a]	−1.86	−1.86	1.90	−1.82	−1.82	1.61	−1.19	−1.057	2.89	Meyer et al. [32]
	Absolute salinity[b]	34.5	34.5	35.1	34.5	34.5	35.1	33.3	33.8	35.2	Meyer et al. [32]
	Oxygen ($mL \cdot L^{-1}$)[c]	7.9	8.5	6.9	8.6	8.5	7.5	8.8	8.8	7.4	Assmy et al. [28]
	PAR ($\mu E \cdot m^{-2} \cdot s^{-1}$)[d]	ND	ND	ND	2.4023	0.13092	1×10^{-12}	117.7	4.3022	1×10^{-12}	Assmy et al. [28]
	Fluorescence ($mg \cdot m^{-3}$)[e]	ND	ND	ND	0.019	0.0086	−0.034	3.45	4.34	−0.041	Assmy et al. [28]
Biogeochemical	Ammonium (μM)	0	0	0.0166	0.133	0.0829	0.113	0.639	0.394	ND	Assmy et al. [28]
	Nitrite (μM)	0.060	0.060	0.060	0.060	0.060	ND	0.19	0.066	0.060	Assmy et al. [28]
	Nitrate (μM)	10.05	10.2	11.2	7.91	7.75	ND	1.12	8.90	12.1	Assmy et al. [28]
	Phosphate (μM)	0.782	0.841	0.870	0.612	0.611	ND	0.322	0.748	0.899	Assmy et al. [28]
	Silicate (μM)	4.178	4.252	4.689	3.138	3.124	ND	3.832	3.821	4.888	Assmy et al. [28]
	TDN (μM)[f]	8.8	8.8	ND	10	12	ND	2.6	5	ND	Assmy et al. [28]
	PON (μM)[g]	0.327	0.345	ND	0.327	0.303	ND	3.709	3.137	ND	Assmy et al. [28]
	DOC (μM)[h]	69	69	ND	58	57	ND	22	24	ND	Assmy et al. [28]
	POC (μM)[i]	2.58	1.95	ND	1.45	2.45	ND	18.03	20.2	ND	Assmy et al. [28]
	Chl *a* ($mg \cdot m^{-2}$)[j]	1.32	0.421	ND	0.202	0.657	ND	12.6	32.5	ND	Assmy et al. [28]
	Chl *a* ($mg \cdot m^{-3}$)[k]	0.0301	0.0226	ND	0.0652	0.0661	ND	3.15	1.21	ND	Assmy et al. [28]
	Phaeopigment ($mg \cdot m^{-3}$)	0.0371	0.0356	ND	0.0537	0.0521	ND	1.65	1.74	ND	Assmy et al. [28]

ND, not determined

The measurement dates given indicate the day of collection of the variables measured

[a] Conservative temperature measured in 09 March 2015, 27 April 2015, and 16 June 2015 days

[b] Absolute salinity measured in 09 March 2015, 27 April 2015, and 16 June 2015 days

[c] Oxygen measured using the Winkler method

[d] PAR, photosynthetically active radiation-measured in 26 April 2015 and 14 June 2015 days

[e] Fluorescence (WETLabsECO-AFL) measured in 26 April 2015 and 14 June 2015 days

[f] TDN, total dissolved nitrogen

[g] PON, particulate organic nitrogen

[h] DOC, dissolved organic carbon

[i] POC, particulate organic carbon

[j] Chl *a*, chlorophyll *a* by m^{-2}

[k] Chl *a*, chlorophyll *a* by m^{-3} (depth)

(https://www.mothur.org) [41]. Individually, the primer clipped forward and reverse reads of each library were joined from raw Illumina fastq files (mothur v. 1.39.5) [42]. Merged reads with ambiguities and sequences that were outside the 360–380 bp and 365–385 bp ranges, for the 16S and 18S rRNA amplicon datasets, respectively, were excluded. The remaining sequences were dereplicated and aligned against the SILVA database (v. 1.2.8) [43]. Those that did not align were excluded as well as the ones with homopolymers ($n > 8$). After undergoing thorough a dereplication step, the unique sequences that differ within 3 base pairs similarity from a more abundant one were clustered together. Chimeras were identified de novo and removed with UCHIME [44]. Then, the unique reads were assigned against SILVA (v. 1.2.8) using the RDP naïve Bayesian Classifier [45]. Undesirable lineages, "Chloroplast", "Mitochondria", "unknown", and "Eukaryota", were removed from the 16S dataset; while "Vertebrata", "Annelida" "Arthropoda", "Cnidaria", "Ctenophora", "Echinodermata", "Florideophycidae", "Mollusca", "Pav3", "D226", "FV18-2D11", and "Tunicata", were excluded from 18S. Afterward, a distance matrix was calculated and the sequences clustered into OTUs (operational taxonomic units) using 0.03 and 0.02 cut-off values for the 16S and 18S rRNA amplicon datasets, respectively, with OptiClust [46]. Finally, the 16S and 18S rRNA amplicon-based OTU tables were built.

Downstream Sequence Analysis: Composition

The OTU tables in biom format were imported to QIIME (v. MacQIIME 1.9.1) [47] to exclude rare OTUs (< 5 observations across samples) and rarefy at even sampling depth (to the sample with the lowest number of reads), 38,232 and 43,289 reads for 16S and 18S rRNA amplicon-based OTU tables, respectively. The number of sequences filtered during each upstream step for the 16S and 18S rRNA amplicon datasets is summarized in Supplementary Tables S2 and S3, respectively. The distribution of prokaryotic taxa across N-ICE2015 collection at phylum, class, order, family, genus, and OTU levels is provided in Supplementary Table S4, the raw prokaryotic OTU table (without excluding any taxa, rare OTUs neither rarefying) in Supplementary Table S5, and the distribution of eukaryotic taxa at different taxonomic ranks in Supplementary Table S6. All of the downstream analysis reported here was conducted in duplicate for the 16S and 18S rRNA datasets based on information retrieved from Supplementary Tables S4 and S6, respectively.

Downstream Sequence Analysis: Diversity and Structure

A multiple-sequence alignment with representative sequences of each OTU through 1000 iterations was performed using MAFFT (v. 7.310) [48] to build a maximum likelihood (ML) tree under the GTRGAMMA model with 1000 bootstrap replicates using RAxML (pthreads v. 8.0.26) [49]. The OTU table (without undesirable lineages) and the constructed ML tree were used to obtain alpha and beta diversity metrics (using QIIME). Three different alpha diversity metrics were calculated: the OTUs richness, the Shannon diversity [50], and Faith's phylogenetic diversity (PD) [51]. OTUs richness was based on the number of clustered OTUs (0.03 and 0.02 cut-off values for the 16S and 18S rRNA amplicon datasets retrieved from the Supplementary Tables S2 and S3), and the Faith's PD was calculated using the minimum distance that results from the sum of all branches within the community phylogeny necessary to span all phylotypes (=OTUs) that composed the community [51, 52]. Faith's PD was chosen because it is a suitable diversity metric for biodiversity assessment studies since it relies on the feature diversity inherent to the genetic sequences that compose one community instead of the distribution of taxonomic ranks per se [52]. In addition, the Pielou's evenness index was calculated to study the equitability of OTUs abundances [53]. For beta diversity, the unweighted UniFrac metric was estimated [54, 55] subsampling the 16S and 18S datasets at 38232 and 43,289 sequences, respectively, to obtain a distance matrix that was visualized using the Principal Coordinate Analysis method (PCoA). Results concerning beta and alpha diversity were obtained with the phyloseq R package (v.1.22.3) [56]. Finally, Pearson correlations were calculated using the Hmisc package (v. 4.1.1) [57] and visualized through the corrplot (v. 0.84) [58] package in R (v. 3.4.3) [59]. In addition, specific Pearson's correlations were calculated for microbial groups of interest.

All bar plots displayed herein were produced with ggplot2 (v. 2.2.1) [60].

Microbial Association Network Analysis

The SParse InversE Covariance estimation for Ecological Association and Statistical Inference pipeline (SpiecEasi R package v.0.1.4) was used to construct meaningful microbial ecological networks from the 16S and 18S OTU tables (Supplementary Tables S4 and S6) [61]. SpiecEasi was chosen because it can handle compositional data with higher p (OTUs) > n (samples) and it tries to reduce the spurious relationships often inferred by methods relying on correlations due to the erroneous assumption of independence between microbial OTUs [61]. Two microbial networks were built, one based on the top 50 and other on the top 100 most abundant prokaryotic and eukaryotic OTUs retrieved from Supplementary Tables S4 and S6. This OTU selection was done to reduce the computational costs and the complexity for the practical visualization of the networks. The top 50 prokaryotic OTUs (representing 80% of the 16S dataset) were joined together with the top 50 eukaryotic OTUs (74% of 18S dataset), as well as the top 100 prokaryotic OTUs (88%) with

the eukaryotic ones (83%) in order to infer prokaryotic-prokaryotic, eukaryotic-eukaryotic, and prokaryotic-eukaryotic positive or negative associations through network analyses. Then, both datasets were centered log-ratio transformed to apply the graphical model inference neighborhood selection (the Meinshausen-Buhlmann method), and the right model sparseness was estimated through 100 repetitions for StARS (Stability Approach to Regularization Selection) selection, with a minimum lambda threshold of 0.01. Taxonomic groups are represented by different colors and the specific OTUs identified. The top 50 microbial network is presented as Fig. 6 and supported by the top 100 microbial network included as Supplementary Fig. S2.

Taxonomic Assignment of 18S rRNA Amplicon Libraries: Protist Ribosomal Reference Database

In order to improve the taxonomic assignment of 18S rRNA amplicon libraries classified against the SILVA database, the 18S rRNA amplicon dataset was additionally assigned using as a reference the Protist Ribosomal Reference database (PR^2, v. 4.5) [62]. The 18S rRNA amplicon libraries were processed in the same way described before, with few exceptions, particularly concerning the clustering step of the mothur MiSeq SOP pipeline. The sequences were clustered using the VSEARCH algorithm instead [63]. Since the PR^2 taxonomy is distinct from the SILVA database for some taxa, metazoans were removed in addition to the lineages mentioned above. The rare clusters were excluded (< 5 observations across samples) and rarefied at even sampling depth (43,647 sequences). The OTU table assigned up to species/strain level is presented in Supplementary Table S7.

Pipeline of Metagenomic 16S rRNA Genes: EBI Metagenomics

The forward and reverse raw fastq files ($n = 18$) were submitted and archived in the European Nucleotide Archive under the study accession PRJEB15043 in order to run the Illumina metagenomic reads with the EBI Metagenomics pipeline (v.3.0; https://www.ebi.ac.uk/metagenomics/) [64]. The workflow followed by the EMG automatic pipeline freely available online is described in detail at https://www.ebi.ac.uk/metagenomics/pipelines/3.0 and by Mitchell et al. [64]. The non-coding RNAs genes, i.e., rRNA, inclusive tRNA, were identified and extracted from metagenomic datasets using the HMMER (v.3.1b1; http://hmmer.org). The 16S rRNA genes selected were clustered into OTUs using the closed-reference OTU picking strategy with QIIME (v.1.9.1) [47]. Then, OTUs were classified against the Greengenes reference database (v.13.8) [65]. The OTU table with the classification of metagenomic 16S rRNA reads was used to compare the taxonomic profiles of prokaryotic communities obtained through 16S rRNA gene amplicon and metagenomic 16S rRNA reads. The absolute number of metagenomic 16S rRNA reads varied between 2716 in TR_250 and 6381 in YP_250. Supplementary Table S8 contains the full taxonomic path of metagenomic 16S rRNA genes across samples.

Phylogenetic Analysis of *Alcanivorax*

The two most abundant 16S rRNA amplicon sequences classified as *Alcanivorax*, "pOtu00002" and "pOtu00028," were compared with our metagenomic 16S rRNA reads and with publicly available metagenomic 16S rRNA sequences. The 16S rRNA gene sequence of *Alcanivorax borkumensis* SK2 (accession no.: AM286690.1:403189-404730) was used as query to retrieve the top blast [66] hits from both metagenomic datasets. The metagenomic 16S rRNA reads recovered from our metagenomic 16S rRNA reads dataset considered only the metagenome from sample NB_250. The publicly available metagenomic 16S rRNA sequences were retrieved from the 16S rRNA Public Assembled Metagenomes database, in IMG/M [67]. The blast hits retrieved had an *e*-value $\leq 1e^{-10}$ and a pairwise identity > 90% in the case of those sequences retrieved from our metagenome ($n = 6$), and $1e^{-50}$ and 80%, in the case of those hits recovered from 16S rRNA Public Assembled Metagenomes database ($n = 29$). The 29 *Alcanivorax* metagenomic sequences retrieved from the 16S rRNA Public Assembled Metagenomes database (IMG/M) were collected from diverse environments: freshwater (Ga0105052_100345103) and saline (Ga0075134_1043112, Ga0136574_1027101) lakes from Antarctica, coastal/intertidal (Ga0213839_10290031, Ga0181574_100261572, Ga0213858_100093481, Ga0136854_1201501), and oceanic (Ga0098057_10126854, Ga0105228_1041072, JGI11834J15748_10001651, Ga0206123_100480811, Ga0190303_10115492, Ga0206683_100768221, Ga0114916_10241841, Ga0163110_100314923, JGI12071J15745_1026451, Ga0114994_100173314, Ga0115564_100385703, JGI24650J20062_10146871, JGI24528J20060_10019162, JGI12031J13088_1020621, Ga0005504_10018661, Ga0031697_10007671, JGI24517J20059_10114001, JGI12001J15744_10163031, JGI11949J13268_10057921, Ga0031693_10103241, Ga0115002_100660851, Ga0114915_10314891) environments.

The sequence of *Rhodospirillum sulfurexigens* JA143T (NCBI accession no. AM710622.1) was included as an outgroup. The alignment made in MAFFT was manually curated and the overhangs that fail to align were trimmed resulting in 38 sequences with 357 bp positions. Then, the ML tree was built in RAxML. MAFFT and RAxML parameters were the same as used previously. The world map was

done with the maps R package [68] and the sequences tracked to the geographic locations from where they were collected (with the information retrieved from each project at IMG/M). Finally, the ML tree and the world map were edited in the Inkscape software (v.0.92, www.inkscape.org).

Availability of Data

Raw Illumina fastq files, concerning the SSU rRNA amplicon and metagenomic data used in this study, were deposited in the European Nucleotide Archive under the project accession number PRJEB21950 and PRJEB15043, respectively.

Results

Environmental Context of Microbial Collection North of Svalbard

All samples were collected north of Svalbard (Fig. 1 and Supplementary Table S1) under snow-covered sea-ice, but with distinct snow thickness depending on the season. The snow was thicker on the first-year (FYI) and second-year ice (SYI) during late winter and early spring resulting in very low levels of photosynthetically active radiation (PAR) in the underlying water column of the two northernmost sampling sites (e.g., 2.4-μmol photons·m^{-2}·s^{-1} at TR_5), and thinner on FYI in early summer corresponding to the collection of samples over the Yermak Plateau, with higher light intensities at the ocean surface (covered with almost snow-free ice, and PAR around 118-μmol photons·m^{-2}·s^{-1} at YP_5; Table 1).

Next to changes in the light regime, environmental variability was mainly imposed by the distinct water masses sampled along the drift (Table 1). Atlantic water and MAW characterized higher conservative temperature and absolute salinity values, 2.89 °C and 35.20 in AW, and 1.76 °C and 35.10 in MAW, than in the fresher PSW (− 1.71 °C and 34.26) and PSWw (− 1.06 °C and 33.80). Polar waters were in turn more oxygenated (PSW = 8.46 mL·L^{-1}, PSWw = 8.80 mL·L^{-1}) than Atlantic water masses (AW = 7.40 mL·L^{-1}, MAW = 7.20 mL·L^{-1}) and had higher concentrations of particulate organic nitrogen and carbon (Table 1). The last site sampled, under the Yermak Plateau, close to summer (16 June), was likely the only one with PAR values high enough to support an under-ice phytoplankton bloom resulting in much higher values of fluorescence (YP_5 = 3.45 mg·m^{-3}, YP_20 = 4.34 mg·m^{-3}), phaeopigments (YP_5 = 1.65 mg·m^{-3}, YP_20 = 1.74 mg·m^{-3}), and chlorophyll *a* (YP_5 = 3.15 mg·m^{-3}, YP_20 = 1.21 mg·m^{-3}) compared to chlorophyll *a* values < 0.1 mg·m^{-3} observed in subsurface waters sampled in the other sites (Table 1).

Alpha and Beta Diversity of Pelagic Microbial Communities

Faith's PD was higher for prokaryotic communities collected at 250 m, with the exception of YP station, and for eukaryotic communities retrieved at 5 and 20–50 m depth (Fig. 2a). The number of prokaryotic OTUs decreased from the deepest water layer sampled (250 m depth) towards subsurface and surface depths (< 50 m), with exception of the YP site (Fig. 2b). The number of eukaryotic OTUs followed an opposite trend (Fig. 2c). These results agreed with prokaryotic and eukaryotic PD trends (Fig. 2). The Shannon diversity index showed different patterns, with generally smaller differences between the various depths (Fig. 2b and c). Among the prokaryotic communities, those collected under the NB had higher OTUs at the 250-m sample but lower diversity (Shannon index) than those sampled at shallower depths which implies a prokaryotic community less evenly distributed (with the lowest Pielou's evenness value among the dataset - 0.35) (Fig. 2b). For eukaryotic communities, the alpha measures diverge for samples collected under the Plateau, with the lowest number of OTUs but the highest diversity (Shannon index) at the deepest sample suggesting a more even distribution than the epipelagic communities (with the highest Pielou's evenness value for eukaryotic communities sampled at YP station - 0.53) (Fig. 2c).

At beta diversity level, the unweighted UniFrac metric (considering the phylogenetic differences between communities, and not relative abundances) visualized through PCoA revealed a similar pattern for prokaryotic and eukaryotic microbial communities (Fig. 3a and b). Samples collected at YP (higher light levels) clustered together, while the samples collected at NB and TR displayed a distinct difference between the epipelagic and mesopelagic communities (Fig. 3a and b). This tendency is even more evident for the eukaryotic communities (Fig. 3b).

Taxonomic Profiles of Protist Communities

The protist N-ICE2015 dataset possess a repertoire of 35 taxa assigned at phylum level (Supplementary Table S6). Phylotypes affiliated with *Dinophyceae* and *Syndiniales* dominated most of the libraries. *Syndiniales* was the taxon most represented at 250 m depth in the NB and TR samples (Fig. 4a). The exceptions to this dominance were the YP samples at 20 and 250 m depth, which were predominated by *Chrysophyceae*, in the case of the YP_20 sample, and *Diatomea* as well as *Prymnesiophyceae*, in the case of the

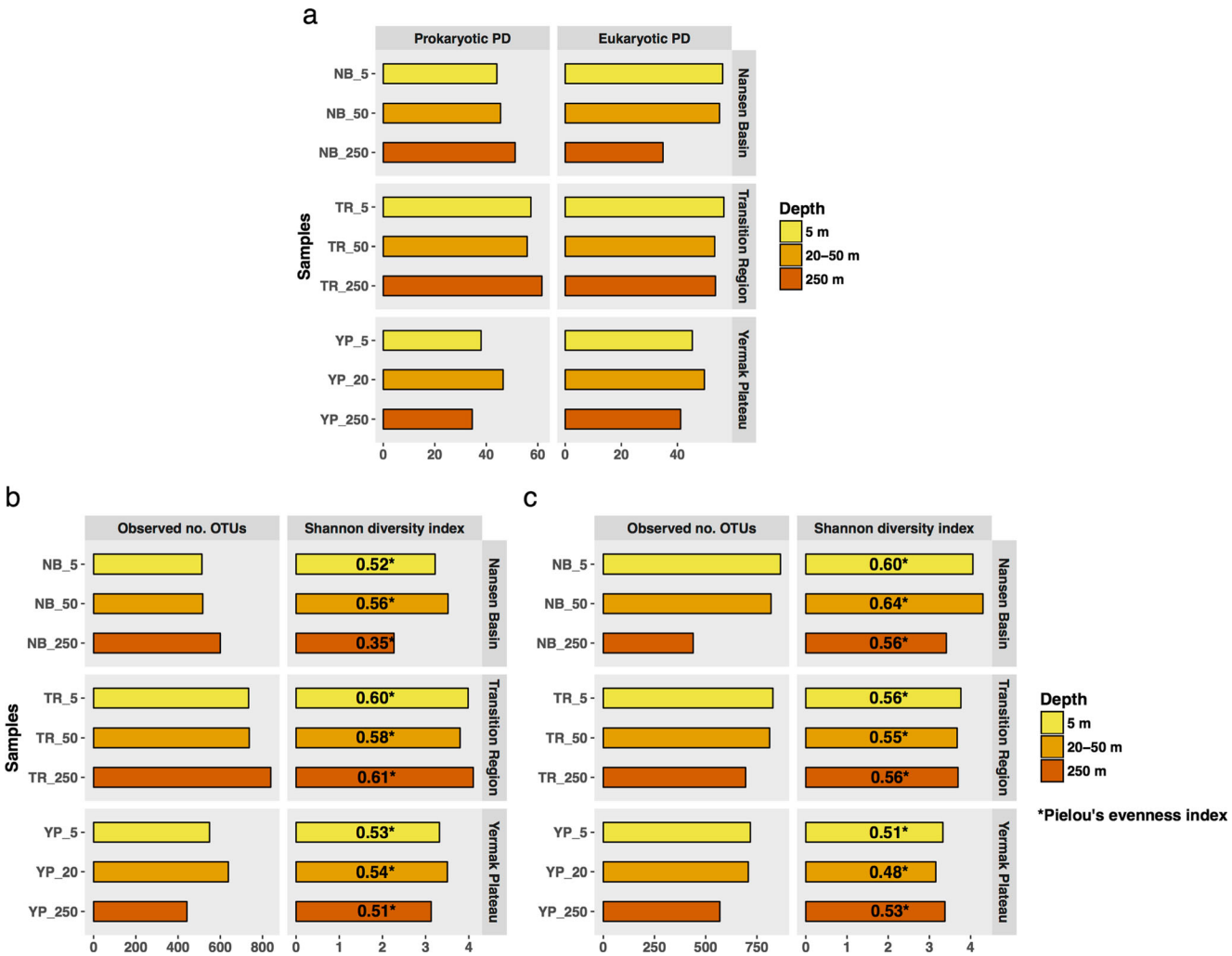

Fig. 2 Alpha diversity of microbial N-ICE2015 collection. **a** Faith's phylogenetic diversity (PD) of prokaryotic and eukaryotic samples. **b** Number of operational taxonomic units (OTUs) observed, Shannon diversity, and Pielou's evenness indices (bold values over the Shannon diversity index bars) within the prokaryotic samples. **c** Number of OTUs observed, Shannon diversity, and Pielou's evenness indices (bold values over the Shannon diversity index bars) within the eukaryotic samples. The OTU sequence similarity threshold was 97 and 98% for the prokaryotic and eukaryotic datasets, respectively. The Shannon diversity index is presented in natural digits (nats)

YP_250 sample (Fig. 4a). Despite the effort made to improve the taxonomic assignment of the 18S rRNA gene amplicon, with the use of PR^2 in addition to the SILVA database, we obtained few assignments at low taxonomic levels, with greater number of environmental clades and unclassified sequences (Fig. 4b, Supplementary Tables S6 and S7). Some photosynthetic unicellular eukaryotes, like the haptophyte *Phaeocystis* and polar centric diatoms, such as *Thalassiosira* and *Chaetoceros*, were classified at genus level (Fig. 4b), as well as some picoeukaryotes (0.2–3.0 μm in size), such as *Spumella* (affiliated with *Chromulinales*, Fig. 4b and Supplementary Table S6) and *Micromonas* (Fig. 4b). Unexpectedly, the higher relative abundance of photosynthetic algae, like *Phaeocystis* and diatoms (*Thalassiosira* and *Chaetoceros*), across the protist N-ICE2015 dataset did not appear at surface and subsurface depths over the YP but at 250 m (Fig. 4b).

Prokaryotic Communities Profiled with 16S rRNA Amplicon and Metagenomics

The taxonomic profile at a higher level established for prokaryotic communities (Fig. 4) agreed well with the metagenomic 16S rRNA data (Supplementary Fig. S3a) validating the metabarcoding protocol applied here concerning potential bias introduced by universal primers and amplification.

The *Alpha-* (30.6%) and *Gammaproteobacteria* (28.6%) classes contributed in almost equal proportions to the high abundance of *Proteobacteria* (Fig. 4c). These two classes were not just the most abundant across the prokaryotic N-ICE2015 collection but also the most phylogenetically diverse (i.e., higher number of phylotypes assigned to both classes than any other phylum or class). *Alphaproteobacteria* were mainly composed by one phylotype affiliated to

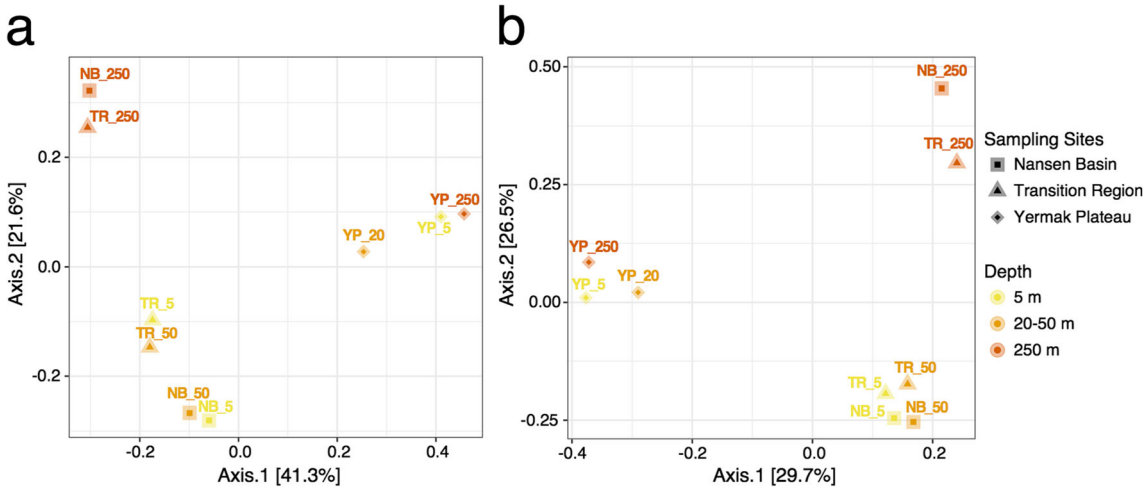

Fig. 3 Principal coordinates analysis (PCoA) of unweighted UniFrac distances across all samples. **a** Prokaryotic samples. **b** Eukaryotic samples. Samples retrieved from the surface (5 m), subsurface (20 or 50 m), and mesopelagic (250 m) seawater at Nansen Basin (NB), Transition Region (TR), and Yermak Plateau (YP) (see Supplementary Table S1). The two axes that explain most of the dataset variability are shown as well as the respective percentage variability in brackets

"*Candidatus* Pelagibacter" from the SAR11 clade (Fig. 4d). SAR11 represented on average 30.7% of the epipelagic communities sampled and 9.2% of the mesopelagic ones. Regarding the *Gammaproteobacteria* class, *Cellvibrionales* (2.7%), *Alteromonadales* (3.9%), and *Oceanospirillales* (19.1%) were the three most abundant orders. *Cellvibrionales* were mainly found on Yermak Plateau in early summer (Supplementary Table S4), while *Alteromonadales* were particularly abundant during springtime with a peak at 250-m depth over Nansen Basin (20.1%, NB_250; Supplementary Table S4). *Oceanospirillales* were highly abundant across all the samples, particularly in NB at mesopelagic depths (55.7%) (Supplementary Table S4). *Alteromonadales* and *Oceanospirillales* were dominated by three phylotypes that predominated at 250 m over Nansen Basin: *Alteromonadales*, *Marinobacter* ("pOtu00019", 6.3%) and *Glaciecola* ("pOtu00020", 12.8%), and *Oceanospirillales*, *Alcanivorax* ("pOtu00002", 54.3% at NB_250 and 9.9% across the entire prokaryotic N-ICE2015 dataset; see Fig. 4d and Supplementary Table S4). *Marinobacter* and *Alcanivorax* genera had also high frequencies in the TR samples (Fig. 4d and Supplementary Table S4).

PCR-independent, metagenomic 16S rRNA reads dataset validated our 16S rRNA amplicon sequencing regarding the taxonomic profiles and frequencies of the two most abundant hydrocarbonoclastic bacteria, *Alcanivorax* and *Marinobacter* (Supplementary Fig. S3c and d). A phylogenetic analysis of the alkane-degrading bacteria *Alcanivorax* was performed with the two most abundant *Alcanivorax*-OTUs ("pOtu00002" and "pOtu00028") and metagenomic 16S rRNA reads ($n = 5$) as well as with similar sequences retrieved from environmental metagenomes from the global ocean ($n = 29$). Metagenomic *Alcanivorax* sequences are cosmopolitan and commonly retrieved from mesopelagic/bathypelagic depths (Supplementary Fig. S4). This is the case for many *Alcanivorax* sequences recovered from marine metagenomes of the Atlantic Ocean (Supplementary Fig. S4).

The phylum *Thaumarchaeota* comprised about 20% of the overall prokaryotic community in the samples collected in the epipelagic waters under Nansen Basin (NB_5 and NB_50) (Fig. 4c). The percentage of thaumarchaeotes decreased to half in samples collected at the same depths over the TR station. However, this decrease at subsurface depths (≤ 50 m) was accompanied by an increase at 250 m depth (from 5.9% at NB_250 sample to 25.4% at TR_250). Our results showed a progressive migration of thaumarchaeotes from surface waters, in late winter, to deeper mesopelagic waters, in early spring (Fig. 4c). These high frequencies of *Thaumarchaeota* in the two northernmost stations sampled contrasts with their low frequency in the YP samples (Fig. 4c). The taxonomic profile of *Thaumarchaeota* obtained from the 16S rRNA gene amplicon data (Fig. 4c) agreed well with the metagenomics 16S rRNA reads dataset (Supplementary Fig. S3a).

Correlations Between Environmental Variables and Taxa and Microbial Networks

Several correlations between specific microbial taxa relative abundances as well as correlations between microbial taxa relative abundances and environmental variables were identified (Fig. 5). The correlations between microbial taxa were further explored in detail at OTU level through networks of the top 50 microbial prokaryotic and eukaryotic OTUs (Fig. 6a). SAR11 represented on average one-third of the epipelagic communities sampled (Fig. 4d) showing a similar distribution with *Dinophyceae* (Fig. 4a). This apparent correlation between both taxa was further confirmed with a significantly positive Pearson's correlation between

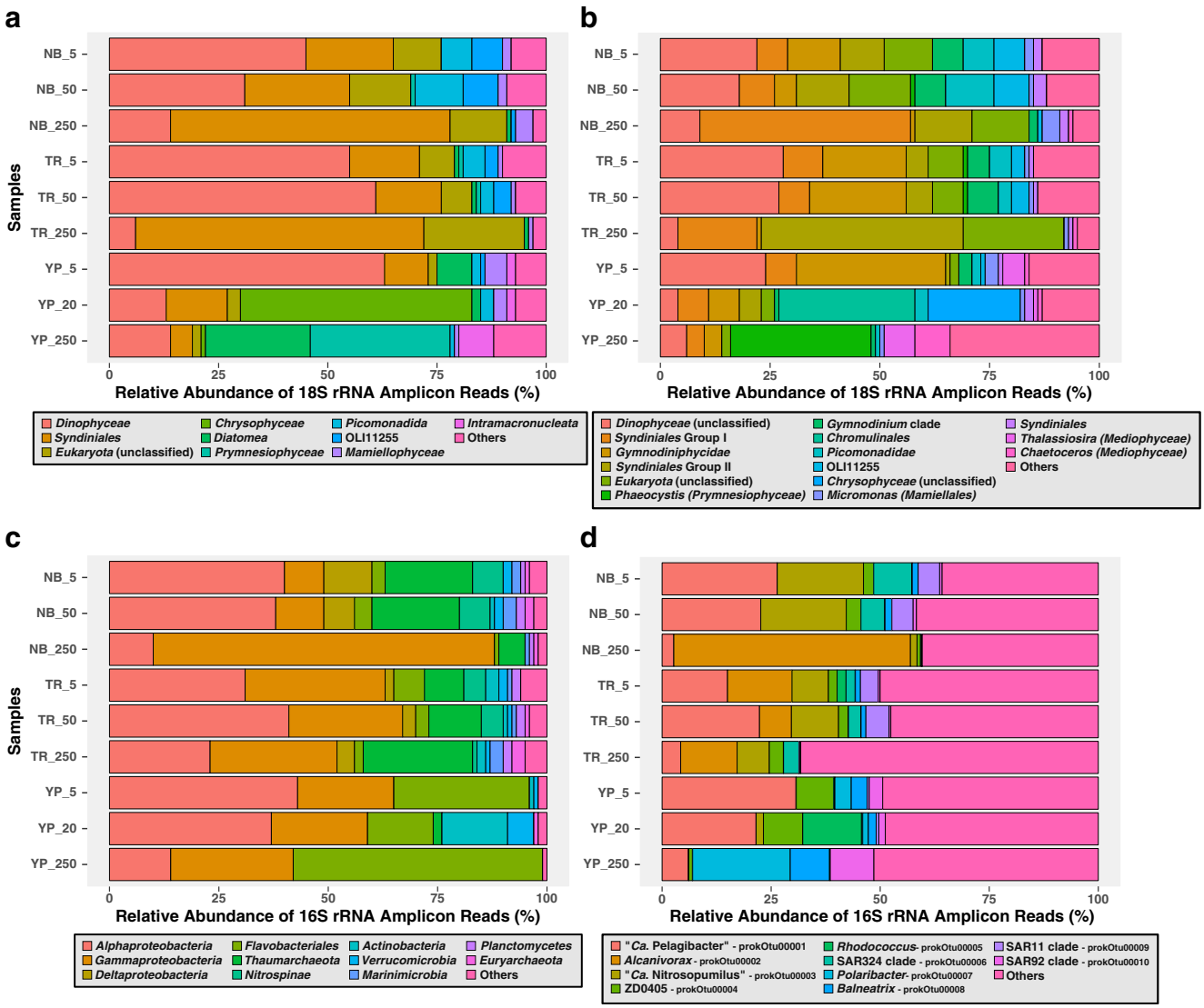

Fig. 4 Taxonomic profile of microbial N-ICE2015 collection. Relative abundance of bacterial, archaeal, and prostist sequences retrieved from the surface (5 m), subsurface (20 or 50 m), and mesopelagic (250 m) seawater at Nansen Basin (NB, 09.03.2015), Transition Region (TR, 27.04.2015), and Yermak Plateau (YP, 16.06.2015; see Supplementary Table S1). **a** Relative abundance of the most abundant taxa at higher level (≥ 1%) retrieved from the OTU table assigned at class level (SILVA v. 1.2.8 classification). **b** Relative abundance of the most abundant taxa at lower level (≥ 1%) retrieved from the OTU table at family (unclassified *Dinophyceae*) and genus (*Phaeocystis*, *Thalassiosira*, *Chaetoceros*, and *Micromonas*) levels. The remaining (*Syndiniales*, *Syndiniales* Group I and II, unclassified *Eukaryota* and *Chrysophyceae*, OLI11255, *Chromulinales*, *Gymnodinium* clade) belong to OTU table at family level but comprises unclassified groups/clades, with just environmental sequences known, or low similarity to be further classified at lower level. **c** Relative abundance of the most abundant taxa (≥ 1%) at higher taxonomic level, phylum (*Thaumarchaeota*, *Nitrospinae*, *Actinobacteria*, *Verrucomicrobia*, *Marinomicrobia* (SAR406 clade), *Planctomycetes*, *Euryarchaeota*), class (*Alpha-*, *Gamma-*, *Deltaproteobacteria*), and order (*Flavobacteriales*). **d** Relative abundance of the top 10 OTUs assigned at genus level; ZD0405, SAR324, SAR11, and SAR92 clades represent groups/clades with uncultured representatives. "Others" represents the sum of the frequency of occurrence of the rare taxa (< 1%)

them ($r = 0.77$, p value = 0.01612), and with the identification of a positive association between "pOtu00009" (SAR11 clade) and "eOtu00029" (unclassified *Dinophyceae*) (Fig. 6b). In addition, the great relative abundance of phylotypes affiliated to *Gammaproteobacteria*, mainly due to the hydrocarbon-degrading *Marinobacter* and *Alcanivorax* genera, was interpreted as a result of hydrocarbon availability from a biological source, like similar structural compounds from algae. Therefore, the possibility of a *Marinobacter-Alcanivorax-Dinophyceae* association was tested. However, no direct association was found between hydrocarbon-degrading genera and *Dinophyceae* (Fig. 6c). The environmental variables and microbial associations that could influence the notable shift of the highly abundant *Thaumarchaeota* were inspected (Fig. 5). *Thaumarchaeota* were positively correlated with DOC

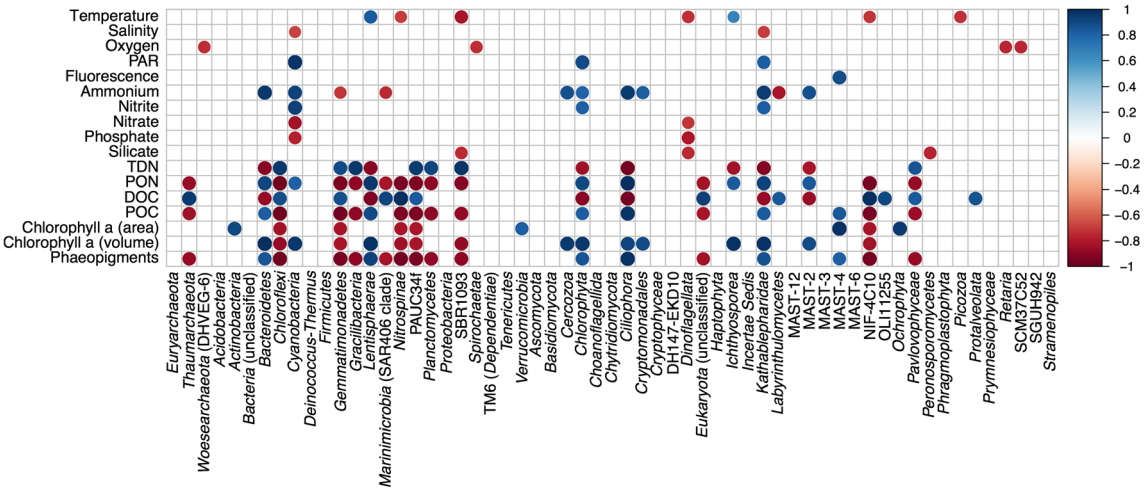

Fig. 5 Heatmap highlighting the significant Pearson's correlations (p value ≤ 0.05) between the environmental variables and the microbial phyla. Pairwise correlations without assigned color represent correlations that are not significant. Pearson's correlations between environmental variables and microbial phyla were performed on a subset of samples according to the availability of the environmental variable measured across samples: $n = 9$ for "latitude," "longitude," "depth," "temperature," "salinity," "oxygen"; $n = 8$ for "ammonium," "nitrite," "nitrate," "phosphate," "silicate"; $n = 6$ for "PAR," "fluorescence," "TDN," "PON," "DOC," "POC," "Chl*a* (area)," "Chl*a* (volume)," "phaeopigment" (see Table 1). DOC stands for dissolved organic carbon; TDN, total dissolved nitrogen; PAR, photosynthetically active radiation; chlorophyll *a* (volume, $mg \cdot m^{-3}$); PON, particulate organic nitrogen; chlorophyll *a* (area, $mg \cdot m^{-2}$); POC, particulate organic carbon (see in detail Table 1)

($r > 0.8$, p value = 0.007666) and negatively correlated with phaeopigments ($r > 0.8$, p value = 0.03333) (Fig. 5) and some photosynthetic groups such as *Cyanobacteria* ($r = 0.70$, p value = 0.03707) and *Chlorophyta* ($r = 0.79$, p value = 0.01084). Although *Thaumarchaeota* were positively correlated with *Nitrospinae* ($r = 0.60$), this correlation was not significant (p *value* = 0.08585). The network of *Thaumarchaeota-Nitrospinae-Chlorophyta* did not show any direct interaction between thaumarchaeal OTUs and OTUs affiliated with *Nitrospinae* or *Chlorophyta*. However, *Nitrospinae* OTUs, "pOtu000022" (an uncultured *Nitrospinaceae*) and "pOtu00025" (*Nitrospina*), were in the proximity of thaumarchaeal "pOtu00003" ("Ca. Nitrosopumilus"). In addition, the "pOtu00025" (*Nitrospina*), that belongs to the *Nitrospinae* phylum, showed a negative association with the green algae "eOtu00036" (*Pycnococcus*) (Fig. 6d).

Bacteroidetes (mainly comprised by *Flavobacteriales*) showed a positive correlation with the *Prymnesiophyceae* family ($r = 0.87$, p value = 0.00226), with this correlation being stronger in the case of the *Flavobacteriales* genus *Polaribacter* ($r = 0.99$, p value = $5.126e^{-07}$). The SAR92 clade was also positively related with the *Prymnesiophyceae* family ($r = 0.96$, p value = $5.466e^{-05}$) as well with *Diatomea* ($r = 0.99$, p value = $1.663e^{-07}$). The *Bacteroidetes-Diatomea-Phaeocystis*-SAR92 network pinpointed the OTUs involved in these correlations, such as "proOtu00010" (SAR92 clade) and "pOtu00030" (NS9 marine group, *Flavobacteriales*) positively connected with "eOtu00009" (*Phaeocystis*), and "pOtu00031" (*Formosa*, *Flavobacteriales*) positively linked to "eOtu00015" (*Thalassiosira*, *Diatomea*) as well as other *Bacteroidetes* OTUs in the proximity (Fig. 6e).

Discussion

Depth-Dependent Trends in Arctic's Microbial Diversity

A better understanding of the Arctic microbial diversity during the winter to summer transition is highly warranted against the backdrop of the changing Arctic sea-ice regime. While mesopelagic protistan communities showed lower PD and numbers of OTUs than those at surface and subsurface depths, prokaryotes displayed an opposite trend. This observation is in line with Wilson et al. [15] who found higher prokaryotic richness in mesopelagic than epipelagic waters north of Svalbard. The higher OTU richness found in prokaryotic libraries collected from MAW (NB_250 and TR_250) could be attributed to the higher temperatures associated with these waters. This result agrees with Fuhrman et al. [69] that found a latitudinal diversity gradient in marine bacterial communities correlated with latitude itself and temperature. However, Ghiglione et al. [70] explained the higher bacterial richness found in deep polar communities with the higher connectivity of deeper water masses and longer water mass residence time with stable environmental conditions.

The eukaryotic pattern described here was opposite to the one observed at two stations sampled over the YP in May 2010 [18] where a general trend of increasing diversity

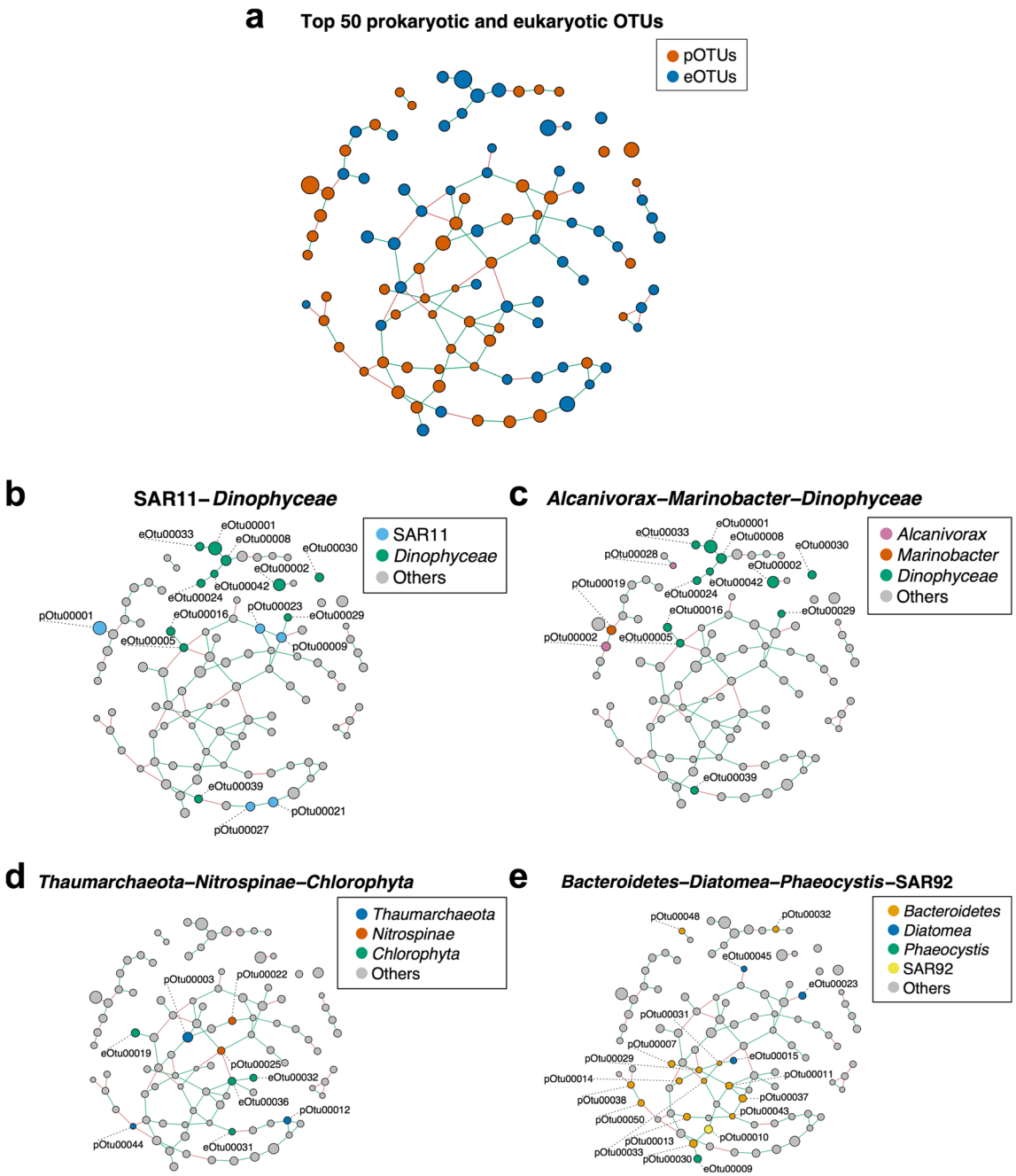

Fig. 6 Microbial ecological networks for the top 50 prokaryotic and eukaryotic OTUs. **a** Network of the top 50 prokaryotic and eukaryotic OTUs. **b** Network highlighting the SAR11-*Dinophyceae* associations. **c** Network highlighting the *Alcanivorax-Marinobacter-Dinophyceae* associations. **d** Network highlighting the *Thaumarchaeota-Nitrospinae-Chlorophyta* associations. **e** Network highlighting the *Bacteroidetes-Diatomea-Phaeocystis*-SAR92 associations. Positive associations are represented by green edges and negative associations by red edges. Node size is proportional to the geometric mean of the relative abundance of OTUs. The "Otu" prefix "p" means prokaryotic and the "e" means eukaryotic

with depth was found. Unfortunately, the lack of vertically resolved studies of eukaryotic diversity (e.g., Metfies et al. [17]) hampers comparison. Nevertheless, the few vertically resolved studies [71–73] showed a depth-dependent trend for eukaryotic richness as shown herein.

The microbial plankton collected during N-ICE2015 covered a period of gradual melting of snow on top of the sea-ice and a simultaneous increase in the light availability beneath the ice pack. Therefore, the similarity of NB and TR epipelagic communities in comparison with mesopelagic ones could be interpreted as a consequence of a deeper mixed layer depth (> 50 m, see the mean mixed layer depth for each floe on Meyer et al. [29]). This layer favors a homogenous environment, during the late winter and early spring period, for the

epipelagic microbial plankton inhabiting the PSW rather than the low light levels below sea ice. The progressive sea ice melting induced shoaling of the surface mixed layer (≈ 5-m depth [29]) which we would expect to result in shifts in the microbial communities collected at 5 and 20 m depth over YP. Contrary to what we anticipated, the communities at 5 and 250-m depth were most similar at the YP station. This result together with the composition of microbial communities is in agreement with the hypothesis of deep export of *Phaeocystis* aggregates explained below that could have ballasted other microbial taxa together. Ultimately, the congruent correspondence pattern, regarding the clustering of prokaryotic communities with the protistan ones, can be explained by the distinct water masses, which is according to previous works [17, 70, 74].

The Dominant *Proteobacteria* and Its Co-occurrence Patterns with Eukaryotes

The SAR11 clade is the most abundant phylotype in the surface of the oceans representing up to one-third of the bacterioplankton [75]. In the N-ICE2015 dataset, SAR11 represented on average 30.7% of the epipelagic communities sampled and 9.2% of the mesopelagic ones. Previous studies reported a relative abundance of this group in Arctic surface waters of 15% on average [8, 76]. The difference to the results presented herein could be attributable to the primers used in our study that were designed to efficiently capture the diversity of marine and freshwater alphaproteobacteria from the SAR11 clade (see the "Methods" section) (http://www.earthmicrobiome.org) [37]. This could also explain why Wilson et al. [15] found ≤ 2.5% of SAR11 clade in their libraries.

The genome of the SAR11 alphaproteobacterial isolate "*Ca.* Pelagibacter ubique", with a representativeness of 16.8% in our 16S rRNA amplicon dataset ("pOtu00001"), unveils great metabolic ability to cope with an oligotrophic lifestyle [77]. Physiological studies (e.g., Steindler et al. [78]) demonstrated efficient heterotrophic growth under both dark and light. It has a specific requirement of reduced sulfur, like the phytoplankton-derived osmolyte 3-dimethylsulphoniopropionate (DMSP) [79], and it is responsible for the consumption of 30% of the assimilated DMSP in the North Atlantic [80]. Several phylogenetically distant algae have the ability to produce the osmoprotectant DMSP, such as diatoms, dinoflagellates, and prymnesiophytes [81]. Particularly, *Phaeocystis pouchetii* produces high levels of DMSP and together with diatoms contribute to the DMSP pool in Arctic surface waters [82]. Despite this evidence, dinoflagellates, that are much more powerful producers of DMSP than diatoms, [83] were positively related with SAR11 clade, with the identification of "eOtu00029" (unclassified *Dinophyceae*) potentially a DMSP-producer [84]. Therefore, the high relative abundance of SAR11 clade in Arctic under-ice surface waters may be sustained by DMSP production by dinoflagellates.

Gammaproteobacteria were also a prominent component of the proteobacteria assemblage in the 16S libraries of the N-ICE2015 dataset. The greater contribution of *Cellvibrionales* in early summer over YP was confirmed by a proteorhodopsin-harboring SAR92 clade (*Porticoccaceae* family) which depends on a photoheterotrophic metabolism [85]. Previous reports successively observed an increased abundance and activity of this clade during the phytoplankton spring blooms in the southern North Sea [86–88] as it has been suggested herein, with SAR92 clade being positive correlated either with diatoms or *Prymnesiophyceae*, and with the positive interaction with the NS9 marine group ("pOtu00030") in the probable degradation of compounds derived from *Phaeocystis* ("eOtu00009"). Intriguingly, the highest frequency of *Phaeocystis* phylotypes occurred at 250-m depth over Yermak Plateau, which might be the consequence of ballasting by cryogenic gypsum (see below). As far as we know, this is the first time that the SAR92 clade can be associated with the degradation of phytoplankton-derived DOM in the Arctic Ocean. A recent study on the bacterial composition in first-year drift ice and under-ice seawater in the Fram Strait did not report the contribution of this gammaproteobacterial clade [89]. However, it was already detected in surface waters of the Canadian Arctic [90].

High Frequencies of Hydrocarbon-Degrading Bacteria in the Arctic

Although hydrocarbon-degrading bacteria have been previously retrieved from Arctic sea-ice communities amended with oil [23] and unamended seawater [91], *Alcanivorax* and *Marinobacter* are not expected to represent a significant component of Arctic microbial communities [23], as it was observed at NB and TR stations. Nevertheless, our 16S rRNA gene amplicon and metagenomic datasets corroborated the high frequencies of *Marinobacter* and *Alcanivorax*. Indeed, phylogenetic analysis of the alkane-degrading bacteria *Alcanivorax* from worldwide environmental metagenomes showed a cosmopolitan distribution of metagenomic *Alcanivorax* sequences [92]. In addition, some of these metagenomic *Alcanivorax* sequences were recovered from marine metagenomes collected at mesopelagic/bathypelagic depths in the Atlantic Ocean (Supplementary Fig. S4). Therefore, these results agree with the presence of *Alcanivorax* sequences at 250-m depth in MAW at NB and TR stations (Supplementary Fig. S3b and c). Actually, metagenomic 16S rRNA sequences from these alkane-degrading bacteria were already retrieved from the western Arctic Ocean (Canada Basin) at 67 and 190-m depth (IMG Genome ID-3300009420). The tight clade formed by

Alcanivorax 16S rRNA gene sequences from amplicon and metagenomic approaches (Supplementary Fig. S4) supports the consistency of the two independent studies, in spite of different library preparation methods and distinct sequencing runs.

In addition, the highest occurrence of *Alcanivorax* was detected at 250 m over Nansen Basin with a simultaneous peak of "pOtu00020" (12.8%) affiliated to the *Glaciecola* genus. In agreement, North Sea strains of *Glaciecola* spp. showed a predominance in oil-contaminated ice core experiments in the Arctic sea-ice [93]. Notwithstanding our observations, hydrocarbon content has not been measured and we do not have evidence that ice-covered waters north of Svalbard have been contaminated with crude oil or of any recent seep activity in the Southern Nansen Basin and the Eastern Yermak Plateau. Alkane-degrading bacteria like *Alcanivorax* and *Marinobacter* had been shown to have the ability to degrade *n*-alkanes produced by the dinoflagellate *Gymnodinium catenatum* [94] and *Cyanobacteria* [95]. Although *Cyanobacteria* only represents < 1% of sequencing reads in the prokaryotic N-ICE2015 dataset (Supplementary Table S4), dinoflagellates (comprising *Dinophyceae*, *Syndiniales* order, and *Gymnodiniphycidae*) sequencing reads were found in higher number, particularly in samples with high hydrocarbon-degrading bacteria.

Link Between *Flavobacteriales* and an Arctic Spring Bloom of *Phaeocystis pouchetii*

Flavobacteriales have been associated with the degradation of phytoplankton-derived polysaccharides [87, 96, 97]. In the N-ICE2015 data set, the two most abundant *Flavobacteriales* related-OTUs were as follows: "pOtu00007" assigned as *Polaribacter* and "pOtu00011" which did not possess enough similarity with any reference sequence in the SILVA database to be classified to genus level. *Flavobacteriales* related-OTUs were reported before for the Arctic Ocean; however, their contributions were considerably smaller than those reported here and elsewhere [89, 98], particularly for *Polaribacter* phylotypes [8, 76, 99]. Probably, the lower abundance observed in these previous works [8, 76, 99] was the consequence of prefiltration or size fractionation of seawater. This procedure led to the loss of those *Flavobacteria* tightly associated with larger phytoplankton, [100] especially, if the latter produces large colonies surrounded by external mucopolysaccharides, which is the case of *Phaeocystis pouchetii* [6]. As the genus name suggests, *Polaribacter* spp. were first retrieved from polar environments and have been commonly found in sea-ice [11–13, 100] and polar surface waters [98, 99]. Since it is consistently retrieved from sea-ice, it has been suggested that the melting sea-ice is a potential seed source of *Polaribacter* spp. for the under-ice water column [8, 89, 101]. With a distribution mostly limited to polar latitudes, [102] some *Polaribacter* strains are psychrophilic and endemic to polar regions [103, 104]. *Polaribacter* showed an increase in relative abundance from winter towards summer with the increase of PAR, accompanied by the melting of snow-covered sea ice. Probably, the sunlight, acted as the source of light to the energy-harvesting proteorhodopsins [105] or light stimulus to algal blooms, selected for the emergence of *Polaribacter*, at YP, explaining the abundant reads of this genus in the YP libraries. In agreement, PAR and Chl *a* are one to two orders of magnitude higher in ice-covered epipelagic waters sampled at YP, with a thinner snow layer, compared to ice-covered waters sampled at the other two sites, with a thicker snow layer under the sea-ice. Furthermore, *Bacteroidetes* and *Polaribacter* were positively correlated with *Prymnesiophyceae*, with the "pOtu00030" (NS9 marine group, *Flavobacteriales* order) directly connected with *Phaeocystis* ("eOtu00009"), and *Polaribacter* ("pOtu00007"), "pOtu00029" (*Ulvibacter*, *Flavobacteriales*), and "pOtu00031" (*Formosa*, *Flavobacteriales*) positively associated with the diatom *Thalassiosira* ("eOtu00015"). Thus, *Flavobacteriales* could be linked to the microbial degradation of an early *P. pouchetii* spring bloom in the Arctic Ocean, as it was reported for the Southern Ocean [106]. The occurrence of *Phaeocystis* in the YP sample at a 250-m depth was unexpected since the early under-ice spring bloom dominated by *P. pouchetii* (late May/early June) was observed in the upper 50 m [6]. However, this observation coincided with a deep export event of the *Phaeocystis* under-ice bloom reported from the same area [107]. Therefore, it is likely that we captured the deep export of *Phaeocystis* aggregates and possibly also diatoms in our YP sample from a 250-m depth.

The Seasonal Pattern of the Phylum *Thaumarchaeota*

In agreement with the literature, about 20% of the overall microbial community associated with the samples collected in the epipelagic waters above the Nansen Basin (NB_5 and NB_50) was affiliated with the phylum *Thaumarchaeota*. Irrespective of the limited temporal and spatial resolution of our study, we observed a seasonal pattern characteristic of polar thaumarchaeotes as reported in previous studies [15, 108–111]. Results showed a progressive migration from the surface waters in late winter to deeper mesopelagic waters in early spring. Indeed, recent findings [112, 113] point towards a protective behavior by thaumarchaeotes against the harmful effect of photochemically produced reactive oxygen species in surface waters. Thus, as expected, nitrifying activity in ice-covered surface waters is higher in winter, as it has been evidenced by higher abundance of *amoA* gene/transcripts as well as potential nitrification rates [114, 115]. The near absence of the *Thaumarchaeota* phylum from surface samples collected over the YP could be attributed to light inhibition as light levels increased towards early summer accompanied by

increased light transmission through the more transparent ice cover. However, *Thaumarchaeota* were nearly absent even at mesopelagic depths over the YP which could be attributed to water mass characteristics [16]. The YP mesopelagic sample was taken from the core of the AW inflow which probably had only recently subducted below the fresher and colder PSW and thus *Thaumarchaeota* have not had time to build up biomass. Thaumarchaeotes are thought to oxidize ammonia and fix carbon dioxide or bicarbonate as sole energy and carbon sources [116, 117]. However, recent studies highlight the use of alternative substrates, particularly urea [109, 118, 119]. Actually, we found a positive correlation between thaumarchaeotes and DOC, which can comprise organic nitrogen compounds such as urea, a source of carbon and energy for polar thaumarchaeotes [109, 118]. Interestingly, the distribution of thaumarchaeotes showed a positive correlation with the distribution of *Nitrospinae* (although not significant), with *Nitrospinae* OTUs, "pOtu000022" (an uncultured *Nitrospinaceae*), and "pOtu00025" (*Nitrospina*), in the proximity of the most abundant thaumarchaeal "pOtu00003" ("Ca. Nitrosopumilus") suggesting a syntrophic relationship. An increasing body of evidence points to a close relationship between nitrite-oxidizing bacteria, such as those belonging to *Nitrospinae*, and ammonia-oxidizing microorganisms, like *Thaumarchaeota* [120, 121]. The high nitrifying activity in wintertime of these cold-adapted archaea plays an important role for the N and C budgets in the Arctic Ocean [114, 122].

Composition and Distribution of Picoeukaryotes

Picoeukaryotes are recognized to play an important ecological role in oligotrophic marine waters. As the Arctic Ocean becomes nutrient-depleted in early summer, a reaction by the picoplankton fraction that can account for most of photosynthetic biomass (as reported by Metfies et al. [17]) is expected. The dominant picoeukaryote within the eukaryotic N-ICE2015 collection was the OTU affiliated with the order *Chromulinales* ("eOtu00003", 30.8%), designated as *Spumella*-like which are phagotrophic nanoflagellates, mostly bacterivorous [123]. Therefore, the sudden appearance and dominance of this chrysophyte at 20 m depth over YP may be related with substrate availability. However, as we do not have any information on prey-specific relationships, it is impossible to point out which bacterial groups it fed on. Nevertheless, we cannot rule out seasonal patterns described previously [124] or the possibility of a mixotrophic lifestyle shown for many crysophytes [125] and dependent on light availability. Contrasting with this quite evident seasonal pattern, the picoeukaryotic green alga *Micromonas* (comprising mostly the *Mamiellales*) was spread across all depths and locations. One previous study [126] reported bacterivory for an Arctic *Micromonas* strain, *M. pusilla* CCMP20299, found in western Canadian Arctic [8, 127] and south of Svalbard [128]. The presence of *Micromonas* from late winter to early summer in our protist N-ICE2015 collection reinforces its omnipresence and mixotrophic lifestyle. However, distinct distributional patterns of *Micromonas* were observed at the OTU level. Whereas "eOtu00019" was more abundant at the two first sampling sites - NB and TR - and considerably less abundant over YP; the "eOtu00032" showed precisely the opposite pattern suggesting the presence of distinct ecotypes of *Micromonas* in the Arctic Ocean.

Main Conclusions

The N-ICE2015 prokaryote and eukaryote datasets offer a valuable resource to study the diversity and structure of the underexplored pelagic microbial communities from ice-covered waters north of Svalbard. The late winter to early summer transition was accompanied by the gradual melting of snow above the sea-ice with increasing light availability below the ice pack. The availability of light together with the water masses characteristics strongly impacted the prokaryotic and eukaryotic communities at the compositional level. Seasonal trends were quite evident for *Thaumarchaeota* and *Nitrospinae*, which were abundant in late winter and early spring but nearly absent in early summer. Both groups together are likely to contribute to the N and C budgets of the Arctic Ocean through nitrification and dark carbon fixation in the wintertime. During this period, prokaryotic microbial communities over the NB (at 250 m) and TR stations revealed high frequencies of two cosmopolitan alkane-degrading bacteria, *Alcanivorax* and *Marinobacter*, but the source of *n*-alkanes and/or other substrate is unknown. In early summer, *Flavobacteria* and SAR92 clade emerged with the presence of light over the YP, probably associated with the degradation of an under-ice bloom of *Phaeocystis*. Contrasting to these seasonal trends, SAR11 was apparently insensitive to environmental conditions inherent to this seasonal transition and abundant across epipelagic samples reflecting the high adaptability of this successful clade [75]. Moreover, picoeukaryotes are likely to play an important role in a future nutrient-depleted Arctic Ocean scenario. In agreement, the relatively high frequencies of *Spumella*-like phagotrophic nanoflagellate (which dominated the 20-m depth sample collected over YP with 31% of relative abundance) and the mixotrophic green alga *Micromonas* (1.7% in the entire N-ICE2015 dataset) emphasize a more important role of mixotrophy than those previously recognized by standard microscopic studies [129]. The late winter/early spring microbial communities showed higher phylogenetic diversity than early summer communities highlighting the need to continue with further studies to properly characterize the overlooked winter microbial communities under the pack-ice accounting for potential losses of microbial phylotypes with the gradually declining sea-ice cover.

Acknowledgments The author AGGS would like to address a special acknowledgement to the Master in Cell and Molecular Biology (M:BCM) held during 2015-17 at Faculty of Sciences, University of Porto (FCUP). We also would like to thank Dr. Anna Silyakova (Centre for Arctic Gas Hydrate, Environment and Climate, Tromsø) for critical reviewing the section "High frequencies of hydrocarbon-degrading bacteria in the Arctic."

Authors' Contributions AGGS carried out the bioinformatics analysis of SSU rRNA amplicon datasets, analyzed all the data, and wrote the manuscript. MPT and CM supervised all the work. CM and PD designed the sampling campaign. PD and MFM collected and filtrated the samples aboard the R.V. Lance. AGGS and CM extracted the DNA to sequence. LT developed R scripts used in this study. JS, RBL, and JBPL performed phylogenetic analysis. MPT and HR critically reviewed the bioinformatics pipelines, particularly, the 18S rRNA amplicon analysis, and the marine hydrocarbon-degrading bacteria part, respectively. PA and CM funded the work. All authors improved, reviewed, and approved the final manuscript.

Funding This work was financially supported by the Centre for Ice, Climate and Ecosystems at the Norwegian Polar Institute and the Research Council of Norway (project Boom or Bust no. 244646) and Structured Program of R&D&I MarInfo-NORTE-01-0145-FEDER-000031, funded by the NORTE2020 through the European Regional Development Fund (ERDF). MF-M and PA were supported by Norwegian Ministries of Foreign Affairs and Climate and Environment through the program Arktis 2030 (project ID Arctic). This project was also funded by PROPOLAR through a grant to CM and a scholarship to AGGS (NITRONICE project) and by Portuguese Science and Technology Foundation (FCT) through a grant to CM (NITROLIMIT-PTDC 2017-PTDC/CTA-AMB/30997/2017).

Compliance with Ethical Standards

Competing Interests The authors declare that they have no competing interests.

Abbreviations *N-ICE2015*, Norwegian young sea ICE expedition 2015; *SSU rRNA*, small subunit ribosomal RNA; *FYI*, first-year ice; *MYI*, multiyear ice; *DOM*, dissolved organic matter; *CTD*, conductivity, temperature, and depth; *PAR*, photosynthetically active radiation; *SOP*, Standard Operating Procedure; *OTU*, operational taxonomic unit; *ML*, maximum likelihood; PR^2, Protist Ribosomal Reference database; *NB*, Nansen Basin; *TR*, Transition Region; *YP*, Yermak Plateau; *Chla*, chlorophyll alpha; *PSW*, polar surface water; *AW*, Atlantic water; *MAW*, modified Atlantic water; *DOC*, dissolved organic carbon

References

1. Lindsay R, Schweiger A (2015) Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. Cryosphere 9:269–283
2. Meier WN, Hovelsrud GK, van Oort BEH, Key JR, Kovacs KM, Michel C, Haas C, Granskog MA, Gerland S, Perovich DK, Makshtas A, Reist JD (2014) Arctic sea ice in transformation: a review of recent observed changes and impacts on biology and human activity. Rev Geophys 52:185–217
3. Maslanik J, Stroeve J, Fowler C, Emery W (2011) Distribution and trends in Arctic sea ice age through spring 2011. Geophys Res Lett 38:L13502. https://doi.org/10.1029/2011GL047735
4. Parkinson CL, Comiso JC (2013) On the 2012 record low Arctic sea ice cover: combined impact of preconditioning and an August storm. Geophys Res Lett 40:1356–1361
5. Polyakov IV, Walsh JE, Kwok R (2012) Recent changes of Arctic multiyear sea ice coverage and the likely causes. Bull Am Meteorol Soc 93:145–151
6. Assmy P, Fernández-Méndez M, Duarte P, Meyer A, Randelhoff A, Mundy CJ, Olsen LM, Kauko HM, Bailey A, Chierici M, Cohen L, Doulgeris AP, Ehn JK, Fransson A, Gerland S, Hop H, Hudson SR, Hughes N, Itkin P, Johnsen G, King JA, Koch BP, Koenig Z, Kwasniewski S, Laney SR, Nicolaus M, Pavlov AK, Polashenski CM, Provost C, Rösel A, Sandbu M, Spreen G, Smedsrud LH, Sundfjord A, Taskjelle T, Tatarek A, Wiktor J, Wagner PM, Wold A, Steen H, Granskog MA (2017) Leads in Arctic pack ice enable early phytoplankton blooms below snow-covered sea ice. Sci Rep 7:40850. https://doi.org/10.1038/srep40850
7. Arrigo KR, Perovich DK, Pickart RS, Brown ZW, van Dijken GL, Lowry KE, Mills MM, Palmer MA, Balch WM, Bahr F, Bates NR, Benitez-Nelson C, Bowler B, Brownlee E, Ehn JK, Frey KE, Garley R, Laney SR, Lubelczyk L, Mathis J, Matsuoka A, Mitchell BG, Moore GWK, Ortega-Retuerta E, Pal S, Polashenski CM, Reynolds RA, Schieber B, Sosik HM, Stephens M, Swift JH (2012) Massive phytoplankton blooms under arctic sea ice. Science 336:1408
8. Comeau AM, Li WKW, Tremblay JÉ, Carmack EC, Lovejoy C (2011) Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. PLoS One 6(11):e27492. https://doi.org/10.1371/journal.pone.0027492
9. Harrison WG, Cota GF (1991) Primary production in polar waters: relation to nutrient availability. Polar Res 10:87–104
10. Rey F (2012) Declining silicate concentrations in the Norwegian and Barents Seas. ICES J Mar Sci 69:208–212
11. 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. ISME J 6:11–20
12. Hatam I, Charchuk R, Lange B, Beckers J, Haas C, Lanoil B (2014) Distinct bacterial assemblages reside at different depths in Arctic multiyear sea ice. FEMS Microbiol Ecol 90:115–125
13. Hatam I, Lange B, Beckers J, Haas C, Lanoil B (2016) Bacterial communities from Arctic seasonal sea ice are more compositionally variable than those from multi-year sea ice. ISME J 10:2543–2552
14. Brown MV, Bowman JP (2001) A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). FEMS Microbiol Ecol 35:267–275
15. Wilson B, Müller O, Nordmann EL, Seuthe L, Bratbak G, Øvreås L (2017) Changes in marine prokaryote composition with season and depth over an Arctic polar year. Front Mar Sci 4:95. https://doi.org/10.3389/fmars.2017.00095
16. Müller O, Wilson B, Paulsen ML, Rumińska A, Armo HR, Bratbak G, Øvreås L (2018) Spatiotemporal dynamics of ammonia-oxidizing *Thaumarchaeota* in distinct Arctic water masses. Front Microbiol 9:24. https://doi.org/10.3389/fmicb.2018.00024
17. Metfies K, Von Appen WJ, Kilias E, Nicolaus A, Nöthig EM (2016) Biogeography and photosynthetic biomass of arctic marine pico-eukaroytes during summer of the record sea ice minimum 2012. PLoS One 11:e0148512. https://doi.org/10.1371/journal.pone.0148512
18. Meshram AR, Vader A, Kristiansen S, Gabrielsen TM (2017) Microbial eukaryotes in an arctic under-ice spring bloom north of Svalbard. Front Microbiol 8:1099. https://doi.org/10.3389/fmicb.2017.01099

19. Li WKW, McLaughlin FA, Lovejoy C, Carmack EC (2009) Smallest algae thrive as the Arctic Ocean freshens. Science 326: 539
20. Lasternas S, Agustí S (2010) Phytoplankton community structure during the record Arctic ice-melting of summer 2007. Polar Biol 33:1709–1717
21. Lalande C, Bauerfeind E, Nöthig EM, Beszczynska-Möller A (2013) Impact of a warm anomaly on export fluxes of biogenic matter in the eastern Fram Strait. Prog Oceanogr 109:70–77
22. Nöthig EM, Bracher A, Engel A, Metfies K, Niehoff B, Peeken I, Bauerfeind E, Cherkasheva A, Gäbler-Schwarz S, Hardge K, Kilias E, Kraft A, Mebrahtom Kidane Y, Lalande C, Piontek J, Thomisch K, Wurst M (2015) Summertime plankton ecology in fram strait - a compilation of long-and short-term observations. Polar Res 34:23349. https://doi.org/10.3402/polar.v34.23349
23. Bowman JS (2015) The relationship between sea ice bacterial community structure and biogeochemistry: a synthesis of current knowledge and known unknowns. Elem Sci Anthr 3:000072. https://doi.org/10.12952/journal.elementa.000072
24. Fernández-Méndez M, Turk-Kubo KA, Buttigieg PL, Rapp JZ, Krumpen T, Zehr JP, Boetius A (2016) Diazotroph diversity in the sea ice, melt ponds, and surface waters of the Eurasian basin of the Central Arctic Ocean. Front Microbiol 7:1884. https://doi.org/10.3389/fmicb.2016.01884
25. Pedrós-Alió C, Potvin M, Lovejoy C (2015) Diversity of planktonic microorganisms in the Arctic Ocean. Prog Oceanogr 139: 233–243
26. Granskog MA et al (2016) Arctic research on thin ice: consequences of Arctic Sea Ice Loss. EOS Trans AGU 97(5):22–26
27. Peterson AK et al N-ICE2015 (2016) Ocean turbulent fluxes from under-ice turbulence cluster (TIC) [v1.0]. https://doi.org/10.21334/npolar.2016.ab29f1e2
28. Assmy P et al (2016) N-ICE2015 water column biogeochemistry [v1.0]. https://doi.org/10.21334/npolar.2016.3ebb7f64
29. Meyer A, Sundfjord A, Fer I, Provost C, Villacieros Robineau N, Koenig Z, Onarheim IH, Smedsrud LH, Duarte P, Dodd PA, Graham RM, Schmidtko S, Kauko HM (2017) Winter to summer oceanographic observations in the Arctic Ocean north of Svalbard. J Geophys Res Ocean 122:6218–6237
30. Bauer RK (2017) Oceanmap: a plotting toolbox for 2D oceanographic data
31. 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
32. Meyer et al (2016) N-ICE2015 ocean microstructure profiles (MSS90L). [Data set]. Norwegian Polar Institute. https://doi.org/10.21334/npolar.2016.774bf6ab
33. Kopf A, Bicak M, Kottmann R, Schnetzer J, Kostadinov I, Lehmann K, Fernandez-Guerra A, Jeanthon C, Rahav E, Ullrich M, Wichels A, Gerdts G, Polymenakou P, Kotoulas G, Siam R, Abdallah RZ, Sonnenschein EC, Cariou T, O'Gara F, Jackson S, Orlic S, Steinke M, Busch J, Duarte B, Caçador I, Canning-Clode J, Bobrova O, Marteinsson V, Reynisson E, Loureiro CM, Luna GM, Quero GM, Löscher CR, Kremp A, DeLorenzo ME, Øvreås L, Tolman J, LaRoche J, Penna A, Frischer M, Davis T, Katherine B, Meyer CP, Ramos S, Magalhães C, Jude-Lemeilleur F, Aguirre-Macedo ML, Wang S, Poulton N, Jones S, Collin R, Fuhrman JA, Conan P, Alonso C, Stambler N, Goodwin K, Yakimov MM, Baltar F, Bodrossy L, van de Kamp J, Frampton DMF, Ostrowski M, van Ruth P, Malthouse P, Claus S, Deneudt K, Mortelmans J, Pitois S, Wallom D, Salter I, Costa R, Schroeder DC, Kandil MM, Amaral V, Biancalana F, Santana R, Pedrotti ML, Yoshida T, Ogata H, Ingleton T, Munnik K, Rodriguez-Ezpeleta N, Berteaux-Lecellier V, Wecker P, Cancio I, Vaulot D, Bienhold C, Ghazal H, Chaouni B, Essayeh S, Ettamimi S, Zaid EH, Boukhatem N, Bouali A, Chahboune R, Barrijal S, Timinouni M, el Otmani F, Bennani M, Mea M, Todorova N, Karamfilov V, ten Hoopen P, Cochrane G, L'Haridon S, Bizsel KC, Vezzi A, Lauro FM, Martin P, Jensen RM, Hinks J, Gebbels S, Rosselli R, de Pascale F, Schiavon R, dos Santos A, Villar E, Pesant S, Cataletto B, Malfatti F, Edirisinghe R, Silveira JAH, Barbier M, Turk V, Tinta T, Fuller WJ, Salihoglu I, Serakinci N, Ergoren MC, Bresnan E, Iriberri J, Nyhus PAF, Bente E, Karlsen HE, Golyshin PN, Gasol JM, Moncheva S, Dzhembekova N, Johnson Z, Sinigalliano CD, Gidley ML, Zingone A, Danovaro R, Tsiamis G, Clark MS, Costa AC, el Bour M, Martins AM, Collins RE, Ducluzeau AL, Martinez J, Costello MJ, Amaral-Zettler LA, Gilbert JA, Davies N, Field D, Glöckner FO (2015) The ocean sampling day consortium. Gigascience 4:27. https://doi.org/10.1186/s13742-015-0066-5
34. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N, Knight R (2011) Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc Natl Acad Sci 108:4516–4522
35. Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G, Knight R (2012) Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624
36. Apprill A, Mcnally S, Parsons R, Weber L (2015) Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. Aquat Microb Ecol 75:129–137
37. Parada AE, Needham DM, Fuhrman JA (2016) Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. Environ Microbiol 18:1403–1414
38. Stoeck T et al (2010) Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. Mol Ecol 19:21–31
39. Piredda R et al (2017) Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. FEMS Microbiol Ecol 93:1–14
40. Ribeiro H, de Sousa T, Santos JP, Sousa AGG, Teixeira C, Monteiro MR, Salgado P, Mucha AP, Almeida CMR, Torgo L, Magalhães C (2018) Potential of dissimilatory nitrate reduction pathways in polycyclic aromatic hydrocarbon degradation. Chemosphere 199:54–67
41. Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the miseq illumina sequencing platform. Appl Environ Microbiol 79:5112–5120
42. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, van Horn DJ, Weber CF (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541
43. Quast C et al (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:590–596
44. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R (2011) UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27:2194–2200
45. Wang Q, Garrity GM, Tiedje JM, Cole JR (2007) Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol 73:5261–5267
46. Westcott SL, Schloss PD (2017) OptiClust, an improved method for assigning amplicon-based sequence data to operational

taxonomic units. mSphere 2:e00073–e00017. https://doi.org/10.1128/mSphereDirect.00073-17
47. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R (2010) QIIME allows analysis of high-throughput community sequencing data. Nat Methods 7:335–336
48. Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780
49. Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313
50. Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423
51. Faith DP (1992) Conservation evaluation and phylogentic diversity. Biol Conserv 61:1–10
52. Faith DP, Baker AM (2006) Phylogenetic diversity (PD) and biodiversity conservation: some bioinformatics challenges. Evol Bioinforma 2:121–128
53. Pielou EC (1966) The measurement of diversity in different types of biological collections. J Theor Biol 13:131–144
54. Lozupone C, Knight R (2005) UniFrac : a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–8235
55. Lozupone CA, Hamady M, Kelley ST, Knight R (2007) Quantitative and qualitative β diversity measures lead to different insights into factors that structure microbial communities. Appl Environ Microbiol 73:1576–1585
56. McMurdie PJ, Holmes S (2013) Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8(4):e61217. https://doi.org/10.1371/journal.pone.0061217
57. Harrell JFE (2018) Hmisc: Harrell Miscellaneous. https://CRAN.R-project.org/package=Hmisc. Accessed March 2017
58. Wei T & Simko V (2017) R package 'corrplot': visualization of a correlation matrix (version 0.84). https://github.com/taiyun/corrplot. Accessed March 2017
59. R Core Team (2018) R: a language and environment for statistical computing. R Found Stat Comput, Vienna, Austria
60. Wickham H. ggplot2: elegant graphics for data analysis. (Springer-Verlag New York, 2009)
61. Kurtz ZD, Müller CL, Miraldi ER, Littman DR, Blaser MJ, Bonneau RA (2015) Sparse and compositionally robust inference of microbial ecological networks. PLoS Comput Biol 11:1–25. https://doi.org/10.1371/journal.pcbi.1004226
62. Guillou L et al (2013) The Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res 41:597–604
63. Rognes T, Flouri T, Nichols B, Quince C, Mahé F (2016) VSEARCH: a versatile open source tool for metagenomics. Peer J 4:e2584. https://doi.org/10.7717/peerj.2584
64. Mitchell A, Bucchini F, Cochrane G, Denise H, Hoopen P, Fraser M, Pesseat S, Potter S, Scheremetjew M, Sterk P, Finn RD (2016) EBI metagenomics in 2016 - an expanding and evolving resource for the analysis and archiving of metagenomic data. Nucleic Acids Res 44:D595–D603. https://doi.org/10.1093/nar/gkv1195
65. DeSantis TZ et al (2006) Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol 72:5069–5072
66. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410
67. Chen IMA et al (2017) IMG/M: integrated genome and metagenome comparative data analysis system. Nucleic Acids Res 45:D507–D516. https://doi.org/10.1093/nar/gkr975
68. Becker RA, Wilks AR, Brownrigg R, Minka TP & Deckmyn A (2017) Maps: draw geographical maps
69. Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL, Brown JH (2008) A latitudinal diversity gradient in planktonic marine bacteria. Proc Natl Acad Sci 105:7774–7778
70. Ghiglione J-F, Galand PE, Pommier T, Pedros-Alio C, Maas EW, Bakker K, Bertilson S, Kirchman DL, Lovejoy C, Yager PL, Murray AE (2012) Pole-to-pole biogeography of surface and deep marine bacterial communities. Proc Natl Acad Sci 109:17633–17638
71. Countway PD, Gast RJ, Dennett MR, Savai P, Rose JM, Caron DA (2007) Distinct protistan assemblages characterize the euphotic zone and deep sea (2500 m) of the western North Atlantic (Sargasso Sea and Gulf Stream). Environ Microbiol 9:1219–1232
72. Xu D, Jiao N, Ren R, Warren A (2017) Distribution and diversity of micriobial eukaryotes in bathypelagic waters of the south China Sea. J Eukaryot Microbiol 64:370–382
73. Xu D, Li R, Hu C, Sun P, Jiao N, Warren A (2017) Microbial eukaryote diversity and activity in the water column of the South China sea based on DNA and RNA high throughput sequencing. Front Microbiol 8:1121. https://doi.org/10.3389/fmicb.2017.01121
74. Galand PE, Potvin M, Casamayor EO, Lovejoy C (2010) Hydrography shapes bacterial biogeography of the deep Arctic Ocean. ISME J 4:564–576
75. Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, Giovannoni SJ (2002) SAR11 clade dominates ocean surface bacterioplankton communities. Nature 420:806–810
76. Kirchman DL, Cottrell MT, Lovejoy C (2010) The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. Environ Microbiol 12:1132–1143
77. Giovannoni SJ et al (2005) Genome streamlining in a cosmopolitan oceanic bacterium. Science 309:1242–1245
78. Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ (2011) Energy starved candidatus pelagibacter ubique substitutes light-mediated ATP production for endogenous carbon respiration. PLoS One 6:1–10. https://doi.org/10.1371/journal.pone.0019725
79. Tripp HJ, Kitner JB, Schwalbach MS, Dacey JWH, Wilhelm LJ, Giovannoni SJ (2008) SAR11 marine bacteria require exogenous reduced sulphur for growth. Nature 452:741–744
80. Malmstrom RR, Kiene RP, Cottrell MT, Kirchman DL (2004) Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the North Atlantic Ocean. Appl Environ Microbiol 70:4129–4135
81. Liss P, Malin G, Turner S, Holligan P (1994) Dimethyl sulphide and *Phaeocystis*: a review. J Mar Syst 5:41–53. https://doi.org/10.1016/0924-7963(94)90015-9
82. Matrai PA, Vernet M (1997) Dynamics of the vernal bloom in the marginal ice zone of the Barents Sea: dimethyl sulfide and dimethylsulfoniopropionate budgets. J Geophys Res 102:22965–22979
83. Levasseur M (2013) Impact of Arctic meltdown on the microbial cycling of sulphur. Nat Geosci 6:691–700
84. Dacey JWD, Wakeham SG (1986) Oceanic dimethylsulfide: production during zooplankton grazing on phytoplankton. Science 233:1314–1316
85. Stingl U, Desiderio RA, Cho JC, Vergin KL, Giovannoni SJ (2007) The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. Appl Environ Microbiol 73:2290–2296

86. Klindworth A, Mann AJ, Huang S, Wichels A, Quast C, Waldmann J, Teeling H, Glöckner FO (2014) Diversity and activity of marine bacterioplankton during a diatom bloom in the North Sea assessed by total RNA and pyrotag sequencing. Mar Genomics 18:185–192
87. Teeling H et al (2016) Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. Elife 5:1–31. https://doi.org/10.7554/eLife.11888.001
88. Wemheuer B, Güllert S, Billerbeck S, Giebel HA, Voget S, Simon M, Daniel R (2014) Impact of a phytoplankton bloom on the diversity of the active bacterial community in the southern North Sea as revealed by metatranscriptomic approaches. FEMS Microbiol Ecol 87:378–389
89. Eronen-Rasimus E, Piiparinen J, Karkman A, Lyra C, Gerland S, Kaartokallio H (2016) Bacterial communities in Arctic first-year drift ice during the winter/spring transition. Environ Microbiol Rep 8:527–535
90. Han D, Kang I, Ha HK, Kim HC, Kim OS, Lee BY, Cho JC, Hur HG, Lee YK (2014) Bacterial communities of surface mixed layer in the pacific sector of the western Arctic Ocean during sea-ice melting. PLoS One 9:e86887. https://doi.org/10.1371/journal.pone.0086887
91. Bano N, Hollibaugh JT (2002) Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. Appl Environ Microbiol 68:505–518
92. Yakimov MM, Timmis KN, Golyshin PN (2007) Obligate oil-degrading marine bacteria. Curr Opin Biotechnol 18:257–266
93. Giudice AL, Bruni V, de Domenico M & Michaud L (2010) Psychrophiles-cold-adapted hydrocarbon-degrading microorganisms. Springer Berlin Heidelberg.
94. Green DH, Llewellyn LE, Negri AP, Blackburn SI, Bolch CJS (2004) Phylogenetic and functional diversity of the cultivable bacterial community associated with the paralytic shellfish poisoning dinoflagellate Gymnodinium catenatum. FEMS Microbiol Ecol 47:345–357
95. Lea-Smith DJ, Biller SJ, Davey MP, Cotton CAR, Perez Sepulveda BM, Turchyn AV, Scanlan DJ, Smith AG, Chisholm SW, Howe CJ (2015) Contribution of cyanobacterial alkane production to the ocean hydrocarbon cycle. Proc Natl Acad Sci 112:13591–13596
96. Taylor JD, Cottingham SD, Billinge J, Cunliffe M (2014) Seasonal microbial community dynamics correlate with phytoplankton-derived polysaccharides in surface coastal waters. ISME J 8:245–248
97. Tully BJ, Tully BJ, Sachdeva R, Heidelberg KB, Heidelberg JF (2014) Comparative genomics of planktonic *Flavobacteriaceae* from the Gulf of Maine using metagenomic data. Microbiome 2:1–14. https://doi.org/10.1186/2049-2618-2-34
98. Nikrad MP, Cottrell MT, Kirchman DL (2012) Abundance and single-cell activity of heterotrophic bacterial groups in the western Arctic Ocean in summer and winter. Appl Environ Microbiol 78:2402–2409
99. Malmstrom R, Straza T, Cottrell M, Kirchman D (2007) Diversity, abundance, and biomass production of bacterial groups in the western Arctic Ocean. Aquat Microb Ecol 47:45–55
100. Abell GCJ, Bowman JP (2005) Ecological and biogeographic relationships of class *Flavobacteria* in the Southern Ocean. FEMS Microbiol Ecol 51:265–277
101. Grzymski JJ, Riesenfeld CS, Williams TJ, Dussaq AM, Ducklow H, Erickson M, Cavicchioli R, Murray AE (2012) A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. ISME J 6:1901–1915
102. Ladau J, Sharpton TJ, Finucane MM, Jospin G, Kembel SW, O'Dwyer J, Koeppel AF, Green JL, Pollard KS (2013) Global marine bacterial diversity peaks at high latitudes in winter. ISME J 7:1669–1677
103. Gosink JJ, Woese CR, Staley JT (1998) Polaribacter gen. nov., with three new species, P. irgensii sp. nov., P. franzmannii sp. nov. and P. filamentus sp. nov., gas vacuolate polar marine bacteria of the Cytophaga-Flavobacterium-Bacteroides group and reclassification of 'Flectobacillus glomera'. Int J Syst Bacteriol 48:223–235
104. Staley JT, Gosink JJ (1999) Poles apart: biodiversity and biogeography of sea ice bacteria. Annu Rev Microbiol 53:189–215
105. Gonzalez JM, Fernandez-Gomez B, Fernandez-Guerra A, Gomez-Consarnau L, Sanchez O, Coll-Llado M, del Campo J, Escudero L, Rodriguez-Martinez R, Alonso-Saez L, Latasa M, Paulsen I, Nedashkovskaya O, Lekunberri I, Pinhassi J, Pedros-Alio C (2008) Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (Flavobacteria). Proc Natl Acad Sci 105:8724–8729
106. Delmont TO, Hammar KM, Ducklow HW, Yager PL, Post AF (2014) *Phaeocystis antarctica* blooms strongly influence bacterial community structures in the Amundsen Sea polynya. Front Microbiol 5:1–13. https://doi.org/10.3389/fmicb.2014.00646
107. Wollenburg JE, Katlein C, Nehrke G, Nöthig EM, Matthiessen J, Wolf- Gladrow DA, Nikolopoulos A, Gázquez-Sanchez F, Rossmann L, Assmy P, Babin M, Bruyant F, Beaulieu M, Dybwad C, Peeken I (2018) Ballasting by cryogenic gypsum enhances carbon export in a *Phaeocystis* under-ice bloom. Sci Rep 8:1–9. https://doi.org/10.1038/s41598-018-26016-0
108. Alonso-Sáez L, Sánchez O, Gasol JM, Balagué V, Pedrós-Alio C (2008) Winter-to-summer changes in the composition and single-cell activity of near-surface Arctic prokaryotes. Environ Microbiol 10:2444–2454
109. Alonso-Saez L, Waller AS, Mende DR, Bakker K, Farnelid H, Yager PL, Lovejoy C, Tremblay JE, Potvin M, Heinrich F, Estrada M, Riemann L, Bork P, Pedros-Alio C, Bertilsson S (2012) Role for urea in nitrification by polar marine Archaea. Proc Natl Acad Sci 109:17989–17994
110. DeLong EF, Wu KY, Prézelin BB, Jovine RVM (1994) High abundance of *Archaea* in Antarctic marine picoplankton. Nature 371:695–697
111. Kirchman DL, Elifantz H, Dittel AI, Malmstrom RR, Cottrell MT (2007) Standing stocks and activity of *Archaea* and *Bacteria* in the western Arctic Ocean. Limnol Oceanogr 52:495–507
112. Kim J-G, Park SJ, Sinninghe Damsté JS, Schouten S, Rijpstra WIC, Jung MY, Kim SJ, Gwak JH, Hong H, Si OJ, Lee SH, Madsen EL, Rhee SK (2016) Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. Proc Natl Acad Sci 113:7888–7893
113. Tolar BB, Powers LC, Miller WL, Wallsgrove NJ, Popp BN, Hollibaugh JT (2016) Ammonia oxidation in the ocean can be inhibited by nanomolar concentrations of hydrogen peroxide. Front Mar Sci 3:1–16. https://doi.org/10.3389/fmars.2016.00237
114. Christman GD, Cottrell MT, Popp BN, Gier E, Kirchman DL (2011) Abundance, diversity, and activity of ammonia-oxidizing prokaryotes in the coastal arctic ocean in summer and winter. Appl Environ Microbiol 77:2026–2034
115. Pedneault E, Galand PE, Potvin M, Tremblay JÉ, Lovejoy C (2014) Archaeal *amoA* and *ureC* genes and their transcriptional activity in the Arctic Ocean. Sci Rep 4:4661. https://doi.org/10.1038/srep04661
116. Berg IA, Kockelkorn D, Buckel W, Fuchs G (2007) A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in archaea. Science 318:1782–1786
117. Hatzenpichler R (2012) Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea. Appl Environ Microbiol 78:7501–7510
118. Connelly TL, Baer SE, Cooper JT, Bronk DA (2014) Urea uptake and carbon fixation by marine pelagic bacteria and archaea during

the Arctic summer and winter seasons. Appl Environ Microbiol 80:6013–6022

119. Mußmann M et al (2011) Thaumarchaeotes abundant in refinery nitrifying sludges express *amoA* but are not obligate autotrophic ammonia oxidizers. Proc Natl Acad Sci 108:16771–16776
120. Hawley AK, Brewer HM, Norbeck AD, Pa a-Toli L, Hallam SJ (2014) Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. Proc Natl Acad Sci 111:11395–11400
121. Needham DM, Sachdeva R, Fuhrman JA (2017) Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters. ISME J 11:1614–1629
122. Tolar BB, Wallsgrove NJ, Popp BN, Hollibaugh JT (2017) Oxidation of urea-derived nitrogen by thaumarchaeota-dominated marine nitrifying communities. Environ Microbiol 19:4838–4850
123. Grossmann L, Bock C, Schweikert M, Boenigk J (2016) Small but Manifold-Hidden Diversity in 'Spumella-like Flagellates'. J Eukaryot Microbiol 63:419–439
124. Nolte V et al (2010) Contrasting seasonal niche separation between rare and abundant taxa conceals the extent of protist diversity. Mol Ecol 19:2908–2915
125. Rottberger J, Gruber A, Boenigk J, Kroth PG (2013) Influence of nutrients and light on autotrophic, mixotrophic and heterotrophic freshwater chrysophytes. Aquat Microb Ecol 71:179–191
126. McKie-Krisberg ZM, Sanders RW (2014) Phagotrophy by the picoeukaryotic green alga Micromonas: implications for Arctic Oceans. ISME J 8:1953–1961
127. Lovejoy C, Vincent WF, Bonilla S, Roy S, Martineau MJ, Terrado R, Potvin M, Massana R, Pedrós-Alió C (2007) Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in arctic seas. J Phycol 43:78–89
128. Foulon E, Not F, Jalabert F, Cariou T, Massana R, Simon N (2008) Ecological niche partitioning in the picoplanktonic green alga *Micromonas pusilla*: evidence from environmental surveys using phylogenetic probes. Environ Microbiol 10:2433–2443
129. Kilias E, Kattner G, Wolf C, Frickenhaus S, Metfies K (2014) A molecular survey of protist diversity through the central Arctic Ocean. Polar Biol 37:1271–1287