

Arctic microbiome and N-functions during the winter-spring transition



ciimar

Centro Interdisciplinar
de Investigação
Marinha e Ambiental

António Gaspar Gonçalves de Sousa

Mestrado de Biologia Celular e Molecular

Departamento de Biologia da Faculdade de Ciências da Universidade do Porto

Dissertação de Mestrado

2016/2017

Orientador

Catarina Maria Pinto Mora Pinto de Magalhães,
Professora Auxiliar Convidada na Faculdade de Ciências da Universidade do Porto
e Investigadora no Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR)

Co-orientador

Pedro Manuel da Silva Duarte,
Investigador Científico no Instituto Polar Norueguês (NPI)

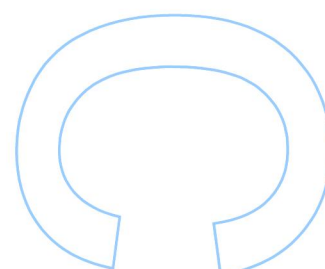
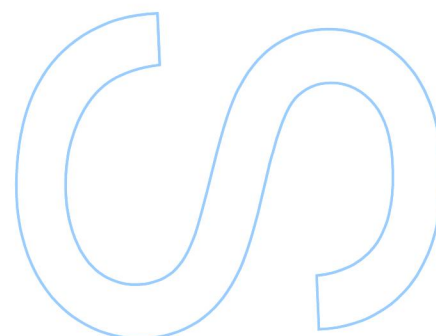
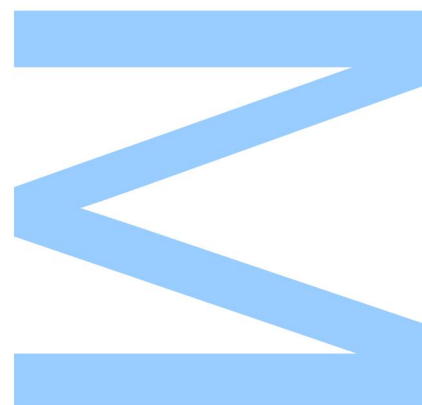
Co-orientador

Luís Fernando Rainho Alves Torgo,
Professor Associado na Faculdade de Ciências da Universidade do Porto
e Investigador Científico Instituto de Engenharia de Sistemas e Computadores,
Tecnologia e Ciência (INESC TEC)



FC

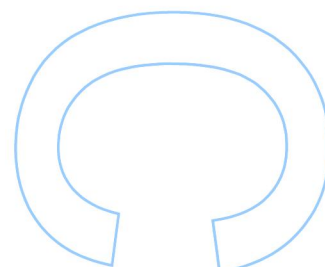
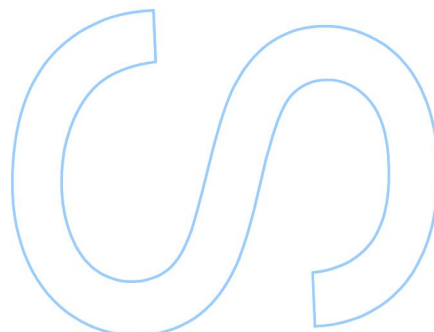
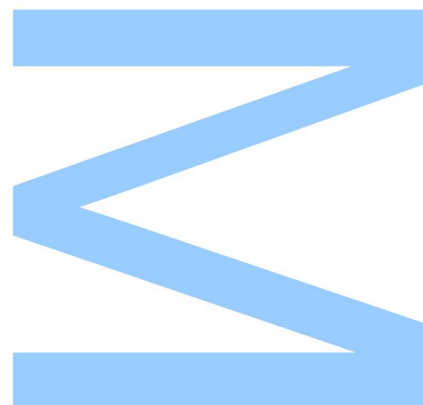
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO



Todas as correções
determinadas pelo júri, e só
essas, foram efetuadas.

O Presidente do Júri,

Porto, ____/____/____



Dissertação de candidatura ao grau de Mestre em Biologia Celular e Molecular submetida à Faculdade de Ciências da Universidade do Porto.

A presente tese foi desenvolvida no Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), sob a orientação científica da Doutora Catarina Magalhães, do Doutor Pedro Duarte, e do Doutor Luís Torgo.

Dissertation for applying to a Master's degree in Cell and Molecular Biology, submitted to the Faculty of Sciences of the University of Porto.

The present thesis was developed at the Interdisciplinary Center for Marine and Environmental Research (CIIMAR) under the scientific supervision of PhD Catarina Magalhães, PhD Pedro Duarte, and PhD Luís Torgo.

“Everything is everywhere, but the environment selects”

Baas-Becking (1934)

Agradecimentos

Antes de mais, gostaria de endereçar um agradecimento especial aos meus supervisores, Dra. Catarina Magalhães, Dr. Pedro Duarte e Dr. Luís Torgo, pela atenção, apoio e oportunidade que me foi dada para desenvolver esta dissertação.

Agradeço também a disponibilização de infraestruturas e equipamentos ao laboratório ‘Funcionamento de Ecossistemas e Biorremediação’ (EcoBioTec) assim como ao respetivo ‘Centro Interdisciplinar de Investigação Marinha e Ambiental’ (CIIMAR) que possibilitaram realizar parte das tarefas experimentais relacionadas com o presente trabalho. Aproveito a oportunidade para agradecer a excelente colaboração e receção por parte dos nossos colaboradores do Instituto Polar Norueguês (NPI), especialmente pela ajuda do Philipp Assmy, na interpretação dos dados de identificação de fitoplâncton.

O meu muito obrigado a uma extraordinária equipa multidisciplinar jovem e dedicada, que integra ou integrou o EcoBioTec, durante o meu período de dissertação e demonstrou um grande sentido crítico e companheirismo.

Particularmente àqueles que intervieram de forma proactiva e construtiva para esta tese. Um especial agradecimento à Maripà por todo o suporte, ajuda e esclarecimento nas análises bioinformáticas. Ao Hugo pela interpretação dos dados. Ao João, Paula, Jacinto e Adriana pelo esclarecimento das pequenas dúvidas intermináveis, e horas de companheirismo.

Também quero agradecer às instituições que financiaram este trabalho, nomeadamente ao *Research Council of Norway* (projeto *Boom or Bust* no. 244646) e ao NORTE2020, Fundo Europeu de Desenvolvimento Regional (FEDER) programas estruturados R&D&I CORAL - NORTE-01-0145-FEDER-000036 e MarInfo - NORTE-01-0145-FEDER-000031. Obrigado ao Programa Polar Português (PROPOLAR), financiado pela Fundação para a Ciência e a Tecnologia (FCT), pela bolsa de estudo de curto prazo para jovens investigadores 2016 - Fase 2 -, que permitiu a minha curta estadia no NPI.

Um último agradecimento a todos os colegas e professores do Mestrado de Biologia Celular e Molecular (M:BCM) dado durante 2015-2017, na FCUP.



Resumo

Uma das manifestações mais proeminentes das alterações climáticas é a mudança de regime do gelo marinho do Ártico, com uma redução na sua extensão durante o verão e uma substituição de grande parte do gelo perene e espesso por gelo com menos de um ano e de menor espessura. As comunidades microbianas são uma componente chave no momento de avaliar o impacto ecológico do regime de gelo em mudança do Ártico, pois constituem a base das redes tróficas marinhas do Ártico e dos ciclos biogeoquímicos.

Durante a *Norwegian young sea ICE expedition (N-ICE2015)*, que decorreu no gelo à deriva a norte de Svalbard entre Janeiro e Junho de 2015, foram recolhidas amostras de água, a 5, 20 ou 50 e 250 m de profundidade nos dias 9 de Março, 27 de Abril e 16 de Junho, juntamente com dados físicos e biogeoquímicos. Através da sequenciação massivamente paralela do *amplicon* da subunidade pequena ribossomal do ADN (*SSU rDNA*, sigla anglo-saxónica), bem como ADN ambiental (*i.e.*, metagenómica) obtivemos informação acerca da diversidade, da estrutura e das funções relacionadas com o ciclo do azoto (N) do microbioma do Ártico durante a transição inverno-primavera.

Os resultados mostram que, ao nível da composição, *Alpha-* (30.7%) e *Gammaproteobacteria* (28.6%) são os grupos mais abundantes na coleção procariótica N-ICE2015, e também os mais diversos filogeneticamente. As tendências de inverno para o início do verão são bastante evidentes com os thaumarchaeotas a representarem uma grande fração das comunidades na coluna de água abaixo do gelo na primavera e praticamente ausentes próximo ao verão. Além disso, o surgimento de *Flavobacteria* e do clado SAR92 no final da primavera pode estar associado à degradação de um *bloom* precoce de primavera dominado por *Phaeocystis*. Surpreendentemente, encontrou-se uma grande representatividade e elevada abundância relativa de bactérias hidrocarbonoclásticas, particularmente *Marinobacter* (6.3%) e *Alcanivorax* (54.3%). Estes filótipos suportam a evidência de uma biosfera rara do Ártico propensa a degradar hidrocarbonetos derivados de petróleo e provavelmente associada à infiltração de óleo natural. Além disso, as thaumarchaeotas oxidantes de amoníaco (*TAO*, sigla anglo-saxónica) não apenas têm uma elevada frequência de ocorrência nas águas cobertas de gelo durante o inverno e a primavera (5 e 50 m de profundidade) bem como as bactérias oxidantes de nitrito (*NOB*, sigla anglo-saxónica). No entanto, estes grupos estão quase

ausentes próximo ao verão, sugerindo atividade nitrificante ativa debaixo do gelo à deriva apenas no inverno e primavera. Os genes que codificam a urease e ammonia monooxygenase estão correlacionados positivamente com o N total dissolvido, que inclui a ureia, sugerindo que ambas as vias, a ureólise e a oxidação aeróbia do amoníaco estão acopladas. Um dos genes que codifica a urease aumenta ao longo da profundidade, sugerindo que populações de *TAO* distintas encontradas na coluna de água do Oceano Ártico têm um potencial genómico diferente para realizar a ureólise. Apesar do esforço de sequenciação feito, não foram encontradas evidências genómicas que suportem as vias de fixação, anammox e desnitrificação do azoto.

As bibliotecas microbianas e metagenómica da coleção N-ICE2015 analisadas neste estudo forneceram dados de sequenciação de alto rendimento ao longo da profundidade e da transição de inverno para primavera, que ajudam a melhorar o conhecimento atual sobre a microbiota assim como as comunidades e vias do ciclo-N no Oceano Ártico.

Palavras-chave: Oceano Ártico, Microbiota, Microbioma, SSU rDNA amplicon, Metagenómica Shotgun Ambiental, Diversidade, Estrutura, Procariotas, Protistas, Vias Biogeoquímicas-N

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. Microbial communities are a key component when evaluating the ecological impact of the Arctic's changing ice regime, as they constitute the basis of Arctic marine food webs and biogeochemical cycles.

During the Norwegian young sea ICE expedition (N-ICE2015), that took place in drifting pack ice north of Svalbard between January and June 2015, seawater was collected, at 5, 20 or 50 and 250 m depth in 9th March, 27th April and 16th June, together with physical and biogeochemical data. Through the massively parallel sequencing of small subunit ribosomal DNA (SSU rDNA) amplicon as well as environmental DNA (*i.e.*, metagenomics) we got a snapshot of the Arctic's microbiome diversity, structure and key N-cycling functions through the winter-spring transition.

Results shows that, at compositional level, *Alpha-* (30.7%) and *Gammaproteobacteria* (28.6%) are the most abundant across the prokaryotic N-ICE2015 collection, and also the most phylogenetically diverse. Winter to early summer trends are 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 late spring might be associated to the degradation of an early spring bloom of *Phaeocystis*. Surprisingly it was found a great representativeness and high relative abundance of hydrocarbonoclastic bacteria, particularly *Marinobacter* (6.3%) and *Alcanivorax* (54.3%). This phylotypes supports evidence of an Arctic's unexpected biosphere, prone to degrade petroleum-derived hydrocarbons and probably associated to natural oil seepage. In addition, not just thaumarchaeal ammonia oxidizers (TAO) have a high frequency of occurrence in the subsurface waters underneath the winter-spring pack ice (5 and 50 m depth), but also nitrite-oxidizing bacteria (NOB). However, they are nearly absent close to summer, suggesting active nitrifying activity underneath of winter-spring pack ice. Urease and ammonia monooxygenase encoding genes are positively correlated with total dissolved nitrogen (N), which includes urea, suggesting that both pathways, ureolysis and aerobic ammonia oxidation, are coupled. Urease encoding gene increases along depth suggesting

that distinct TAO populations found in the water column of the Arctic Ocean have different genomic potential to carry out ureolysis. In spite of the sequence effort made, it was not found genomic evidences to support nitrogen fixation, anammox and denitrification pathways.

The microbial and metagenomic libraries from N-ICE2015 collection analysed in the present study provides comprehensive new knowledge about the microbiota and N-cycling communities and pathways in the Arctic Ocean during the winter to spring transition.

Keywords: Arctic Ocean, Microbiota, Microbiome, SSU rDNA amplicon, Environmental Shotgun Metagenomics, Diversity, Structure, Prokaryotes, Protists, N-Biogeochemical Pathways

Table of contents

Agradecimientos.....	I
Resumo.....	III
Abstract.....	V
Table of contents.....	VII
List of figures.....	XI
List of tables.....	XIII
List of abbreviations.....	XV
Introduction.....	1
1. Assessing the Microbial Diversity and Functionality.....	2
1.1. Traditional Methods - DNA-independent.....	2
1.2. Modern Methods - DNA-dependent.....	2
1.2.1. Next-Generation Sequencing.....	4
1.3. Bioinformatics Pipelines and Databases.....	6
2. Objectives and Thesis Organization.....	9
1 st Chapter: Pelagic Microbial Communities from an Arctic Drift Ice during Winter to Spring Transition North of Svalbard.....	11
1. Background.....	12
1.1. Norwegian young sea Ice expedition 2015: motivation.....	12
1.2. Microbial Oceanography: polar oceans.....	12
1.3. Arctic's Microbial Communities: diversity and structure.....	13

1.4. Objectives.....	14
2. Methods.....	15
2.1. Sampling Sites.....	15
2.2. Water Column Sampling: environmental data and microbial collection.....	17
2.3. DNA Extraction, PCR, Library Preparation, and Sequencing of SSU rDNA amplicon.....	19
2.4. Bioinformatics Pipeline: SSU rDNA amplicon.....	19
2.4.1. Upstream Sequence Analysis: raw OTU table.....	19
2.4.2. Downstream Sequence Analysis: composition.....	20
2.4.3. Downstream Sequence Analysis: diversity and structure.....	21
2.4.4. Taxonomic Assignment of 18S rDNA Amplicon Libraries: Protist Ribosomal Reference database.....	21
2.5. Availability of Sequencing Data.....	22
3. Results and discussion.....	22
3.1. Metabarcoding the Arctic Microbiota through the Winter-Spring Transition: prokaryotic and protistan communities.....	22
3.1.1. The Dominant <i>Proteobacteria</i> and its Co-occurrence Patterns with Eukaryotes.....	23
3.1.2. Evidences of High Frequencies of Hydrocarbon-degrading Bacteria in the Arctic.....	27
3.1.3. Link Between <i>Flavobacteriales</i> and <i>Phaeocystis pouchetii</i> Arctic Spring Bloom.....	29
3.1.4. The Evident Seasonal Pattern of the phylum <i>Thaumarchaeota</i>	31
3.1.5. Composition and Distribution of Picoeukaryotes within the N-ICE2015 Collection.....	34

3.2. Seasonal and Depth Dependent Trends in Arctic's Microbial Diversity.....	35
4. Conclusions.....	39
2 nd Chapter: N-Cycling Microbial Communities and Pathways in the Arctic Ocean.....	41
1. Background.....	42
1.1. Primary Production in the Arctic Ocean: past, present and future.....	42
1.2. Inputs of Nitrogen in the Arctic Ocean: allochthonous and autochthonous sources.....	42
1.3. N-Biogeochemistry in the World's Oceans: communities and pathways.....	43
1.4. N-Biogeochemistry in the Arctic Ocean: communities and pathways.....	45
1.5. Objectives.....	47
2. Methods.....	48
2.1. Sampling Sites, Water Column Sampling, DNA Extraction, PCR, Library Preparation and Sequencing of 16S rDNA amplicon.....	48
2.2. Library Preparation and Sequencing of Metagenomes.....	48
2.3. EBI Metagenomics Pipeline: upstream analysis of metagenomic reads.....	48
2.4. EBI Metagenomics Pipeline: downstream analysis of metagenomic reads.....	49
2.5. ORCA Platform: Spearman correlations between IPRs related to N-biogeochemical pathways and environmental variables.....	50
3. Results and discussion.....	53
3.1. N-Cycling Microbial Communities and Pathways in the Arctic Ocean.....	53
3.1.1. Nitrifying Communities.....	53
3.1.2. Perspective on N-Cycling Microbial Communities in the Arctic Ocean.....	56

3.1.3. N-Biogeochemical Pathways.....	57
4. Conclusions.....	64
References.....	65
Additional files.....	84

List of figures

Figure 1 - Map highlighting the sampling sites (red dots) north of Svalbard.	16
Figure 2 - Taxonomic profile of prokaryotic N-ICE2015 collection.	24
Figure 3 - Taxonomic profile of eukaryotic N-ICE2015 collection.	26
Figure 4 - Heatmap highlighting the significant Spearman correlations between the phyla from the microbial N-ICE2015 collection.	32
Figure 5 - Heatmap highlighting the Spearman correlation matrix between microbial phyla and environmental variables collected during the N-ICE2015 cruise.	33
Figure 6 - Faith's Phylogenetic Diversity metric for sample groups snow-covered sea ice (red line, all samples collected in NB and TR) and sea-ice without snow (blue line, all samples collected in YP).	36
Figure 7 - PCoA of unweighted UniFrac distances across all samples.	38
Figure 8 - Distribution of nitrifying communities across the prokaryotic N-ICE2015 collection.	55
Figure 9 - Heatmap showing the significant Spearman correlations among the enzymes (i.e., family, domain) involved in the different N-biogeochemical pathways along the N- ICE2015 cruise.	61
Figure 10 - Heatmap showing the Spearman correlation matrix between enzymes (i.e., family, domain) involved in N-pathways and some environmental controls.	63

List of tables

Table 1 - Features of sampling conditions of microbial N-ICE2015 collection.	18
Table 2 - Nitrogen-biogeochemical cycle: pathways, genes, enzymes, and, the respective InterPro accession number (IPRs).	51
Table 3 - Absolute number of metagenomics reads that match nitrogen-cycle related InterPro accession number (IPRs) across samples collected during N-ICE2015 cruise.....	59

List of abbreviations

°C - degrees Celsius

µm – micrometre

16S rDNA - 16S ribosomal deoxyribonucleic acid

18S rDNA - 18S ribosomal deoxyribonucleic acid

aka - also known as

AMO - ammonia monooxygenase

APS - adenosine 5'-phosphosulfate

ATP - adenosine triphosphate

BAC - bacterial artificial chromosome

BLAST - Basic Local Alignment Search Tool

bp - base pair

C - carbon

Ca. - *Candidatus*

Chl *a* - Chlorophyll *a*

CO₂ - carbon dioxide

COG - Clusters of Orthologous Groups of proteins

CRT - cyclic reversible termination

DMSP - 3-dimethylsulphoniopropionate

DNA - deoxyribonucleic acid

dNTPs - deoxynucleotide triphosphates

DOC - dissolved organic carbon

DOM - dissolved organic matter

e.g. - *exempli gratia*

eDNA - environmental DNA

EMG - EBI Metagenomics

ENA - European Nucleotide Archive

FYI - first-year ice

Gb - gigabase pair

i.e. - *id est*

IAOOS - Ice-Atmosphere-Arctic Ocean Observing System

IMG/M - Integrated Microbial Genomes and Metagenomes

IPR - InterPro accession number

kb - kilobase pair

KEGG - Kyoto Encyclopedia of Genes and Genomes

L - litre

m – metre

MAW - modified Atlantic water

Mb - megabase pairs

MGI - Marine Group I

MG-RAST - MetaGenome Rapid Annotation using Subsystem Technology

MYI - multi-year ice

N - any base

N - nitrogen

N₂ - di-nitrogen

NB - Nansen Basin

NGS - next-generation sequencing

NH⁴⁺ - ammonium

N-ICE2015 - Norwegian young sea ICE expedition

NO²⁻ - nitrite

NO³⁻ - nitrate

NOB - nitrite-oxidizing bacteria

OMZ - oxygen minimum zones

OSD - Ocean Sampling Day

OTUs - operational taxonomic units

PAR - photosynthetically active radiation

PCoA - principal coordinate analysis

PCR - polymerase chain reaction

Pfam - Protein family database

pH - potential of hydrogen

PO₄³⁻ - phosphate

POC - particulate organic carbon

PON - particulate organic nitrogen

PPi – pyrophosphate

PR² - Protist Ribosomal reference database

PSW - polar surface water

PSW_w - warm polar surface waters

QIIME - Quantitative Insights into Microbial Ecology

RDP - Ribosomal Database Project

RV Lance - Research Vessel Lance

SBS - sequencing-by-synthesis

SiO₄⁴⁻ - silicate

SNA - single-nucleotide addition

SOP - standard operating procedure

sp. - species (singular)

spp. – species (plural)

TAO - thaumarchaeal ammonia oxidizers

TDN - total dissolved nitrogen

TR - Transition Region

v. - version

VBNC - viable but nonculturable

YP - Yermak Plateau

Introduction

1. Assessing the Microbial Diversity and Functionality

1.1. Traditional Methods - DNA-independent

Traditional microbial ecology methods rely on cultivation-dependent approaches which causes an enormous bottleneck in the composition assessment of microbial communities. Microbiologists refer to this phenomenon as “the great plate count anomaly” in which $\leq 1\%$ of microbial diversity is recovered from environmental soil samples (Staley and Konopka, 1985; Hugenholtz, 2002). The percentage of viable but nonculturable (VBNC) microbes is even lower to the much less explored vastness of the ocean. The incapacity to access the microbial “dark matter” (the VBNC fraction) arose from the inability of microbiologists to: replicate natural environments in artificial/synthetic media, in part due to the lack of knowledge of environmental/nutritional the remaining 99% to be discovered; and to look to microbes as highly complex, interactive individuals of multi-assemblages instead as tiny, independent parts of communities (Joint *et al.*, 2010). The complexity of the bacterial world is well mirrored in their ability to take concerted behaviours (phenotypes), such as virulence, competence and sporulation, in a population-density-dependent manner through cell-to-cell communication (Miller and Bassler, 2001; Waters and Bassler, 2005). Despite this quorum sensing activity has been studied for more than four decades since it was discovered (Nealson *et al.*, 1970) and their potential ecological impact in marine biogeochemistry cycles (Doberva *et al.*, 2015; Hmelo, 2016), our knowledge still remains scarce about it.

1.2. Modern Methods - DNA-dependent

Modern microbial ecology relies on advances in molecular biology during the second half of the last century that allowed the application of a broad-spectrum of molecular techniques in order to profile microbial communities' composition. The successful use of 16S rRNA (16S ribosomal ribonucleic acid) gene to assign single-species assemblages of a certain community to a unique taxon was a breakthrough in the field of microbial ecology (Pace, 1997). This kind of studies provided us an accurate and reliable way to access the assemble of species that comprise one environmental microbial community, but there is a lack of confidence about many of the functions predicted and

consequently the services delivered by the community to the ecosystem. Therefore, the low-cost and the high-throughput provided by next-generation sequencing (NGS) technologies, over the last decade until now, aided greatly to solve this problem with the rise of metagenomics. Metagenomics is the scientific field that study the collective genomes of a microbial community, whether archaea, bacteria, virus and eukaryotes (Handelsman *et al.*, 1998; Handelsman, 2004). The latter, to a lesser degree, due to the huge size of eukaryotic genomes with large portions of non-coding DNA (deoxyribonucleic acid) with repetitive sequences which makes difficult to sequence and assemble, respectively. The term metagenomics could also be erroneously applied to any study regarding the sequencing of multiple genomic regions or target genes, *e.g.*, 16S-based metagenomics, from environmental samples that allow to study one community (problem addressed in Esposito and Kirschberg, (2014)). Here, the term metagenomics will be employed to mention the random shotgun sequencing of all environmental DNA (aka whole genome shotgun metagenomics) in order to get the functional and taxonomic content of microbial communities circumventing the pure culture needs. Metagenomics is driven at two levels: functional and sequence (detailed in Handelsman, (2004)). Function-driven relies on the analysis of interesting traits/phenotypes that are heterologous expressed in one metagenomic library, like resistance to antibiotics. Clones displaying the interested trait are picked and sequenced (Handelsman, 2004; Riesenfeld *et al.*, 2004). Thus, function is easily associated to nucleotide sequence. Whereas, sequence-driven analysis focus on similarity searches against reference databases. All the clones or only those anchoring the taxonomic marker of interest are selected and sequenced (Handelsman, 2004; Riesenfeld *et al.*, 2004). The function of the genomic fragments sequenced can be predicted by means of similarity searches against reference databases (Handelsman, 2004; Riesenfeld *et al.*, 2004). The former lacks any phylogenetic information regarding the sequence origin, while the latter could have if included one taxonomic marker (Handelsman, 2004; Riesenfeld *et al.*, 2004). Both strategies were developed before the NGS advent relying on Sanger sequencing. Basically, the environmental DNA (eDNA) extracted is sheared in small pieces - shotgun -, which are inserted into a vector like BAC (bacterial artificial chromosome) and cloned in a host (usually *Escherichia coli*) in order to get a metagenomic library to be sequenced using the dye-termination biochemistry developed by Sanger (Handelsman *et al.*, 1998; Handelsman, 2004; Riesenfeld *et al.*, 2004). Since the sequence content, nucleotide and

their order, is determined by doing an electrophoresis in a 96 or 384-capillary-based gel, the throughput - number of reads sequenced by run - offered is very limited (Shendure and Ji, 2008). Although, it reaches a high read-length of ~1 kb (kilo base pairs) representing the “gold standard” of sequencing technologies with a base-call (the assignment of nucleotides to electropherogram peaks) accuracy of 99.999% (Shendure and Ji, 2008).

1.2.1. Next-Generation Sequencing

Next-generation sequencing rely on massive parallel sequencing of thousands to millions of clonal clusters, the so-called polonies (=polymerase colonies; Shendure and Ji, 2008; Wooley *et al.*, 2010; Goodwin *et al.*, 2016). 454 (Roche), Ion Torrent (Thermo Fisher) and Illumina (sometimes referred as Solexa) are among the most widely used NGS platforms, because they are affordable for small laboratories and offer high-throughput. The workflow followed is quite similar to the traditional Sanger, but the clones and libraries are constructed *in vitro* instead of *in vivo*. Clusters (=clones) are amplified *in vitro* through PCR (polymerase chain reaction) cycles and the libraries (the set of clusters) sequenced using one of sequencing-by-synthesis (SBS) chemistries (Shendure and Ji, 2008; Metzker, 2010; Loman *et al.*, 2012; Goodwin *et al.*, 2016). SBS is the sequencing process catalysed by an enzyme, a polymerase or ligase, over a primed template (Goodwin *et al.*, 2016). It differs among the biochemistries whether it stops at each cycle or not: cyclic reversible termination (CRT) or single-nucleotide addition (SNA), respectively (Goodwin *et al.*, 2016). Quickly, 454 and Ion Torrent makes use of emulsion PCR to amplify the clusters linked to the beads (Loman *et al.*, 2012; Goodwin *et al.*, 2016). Then, SBS takes place at each cluster through the cyclic addition of each kind of dNTPs (deoxynucleotide triphosphates) - SNA (Loman *et al.*, 2012; Goodwin *et al.*, 2016). Since a different dNTP is added at each cycle, it is known which one dNTP is incorporated by the polymerase at each cycle. The detection of the signal released during the template extension differs between the two technologies. The former is based on pyrosequencing, in which the dNTPs incorporated release PPi (pyrophosphate) to yield ATP (adenosine triphosphate), in the presence of APS (adenosine 5'-phosphosulfate) and sulfurylase (Shendure and Ji, 2008; Metzker, 2010). ATP will be used by luciferase to catalyse a burst of light detected by a charge-coupled device camera (Loman *et al.*, 2012;

Goodwin *et al.*, 2016). The latter (Ion Torrent), at each time that a dNTP is incorporated, protons are released (Loman *et al.*, 2012; Goodwin *et al.*, 2016). The protons released change the pH (potential of hydrogen) which detected by a semi-conductor membrane (Loman *et al.*, 2012; Goodwin *et al.*, 2016). On the other hand, Illumina relies on bridge PCR to amplify the clusters linked to one of the two oligonucleotides tethered to the flow-cell - a glass slide with 8 lanes (Shendure and Ji, 2008; Metzker, 2010). This linkage is established by pairing the complementary sequence of adapters, introduced before. Also, the adapters possess the index (aka barcode) and sequencing primers regions to allow the multiplexing of several samples and the SBS process, respectively. Afterwards, the library is sequenced by using a polymerase to make the extension of the forward primed-template in a CRT way (Shendure and Ji, 2008; Metzker, 2010). CRT technology works by making used of reversible terminators. Reversible terminators are dNTPs with the 3'-hydroxyl (OH) group blocked by a cleavable moiety (Shendure and Ji, 2008; Metzker, 2010). This allows to stop the SBS reaction at each cycle in a reversible way. In addition, each one of the four 3'-blocked reversible terminators have a different cleavable fluorophore linked to the nucleobase (Shendure and Ji, 2008; Metzker, 2010). At each cycle that is incorporated one of the four reversible terminators added, the polymerase repairs the 3'-OH group and release the fluorophore (Shendure and Ji, 2008; Metzker, 2010). Ultimately, the wave-length of the excited fluorophore will be detected by a total internal reflection fluorescent device with two lasers (Shendure and Ji, 2008; Metzker, 2010). The same process can be repeated using the reverse sequence as template - paired-end sequencing. The same is valid to the other two platforms above-mentioned. Moreover, it ensures high quality reads in the overlapping regions between forward and reverse reads. The read-length is determined by the number of cycles and the base-call by the emission wave-length and signal intensity. It is difficult to compare the three sequencing technologies because they vary greatly according to the instrument model and the library preparation method used. Commonly, the Illumina offers the higher throughput, in the Gb (gigabase pairs - 10^9 bp) order, than 454 and Ion Torrent, that have similar throughputs, in the order of Mb (megabase pairs - 10^6 bp; Shokralla *et al.*, 2012; Loman *et al.*, 2012). The higher throughput provided by Illumina comes up against the higher run time, usually > 24 h, and smaller read-lengths (up to 150 bp), than the SNA technologies (< 24 h and > 150 bp; Shokralla *et al.*, 2012; Loman *et al.*, 2012). However, the paired-end sequencing overcomes this major drawback allowing to get longer read-

lengths, up to 2 x 300 bp, using V3 chemistry, with Illumina MiSeq instrument. Besides, it has the lowest error rate among the three (Laehnemann *et al.*, 2016). Altogether, the Illumina sequencing platforms are the preferable choice to large-scale projects such as Ocean Sampling Day (OSD; Kopf *et al.*, 2015) and Earth Microbiome Project (Gilbert *et al.*, 2014). Therefore, thanks to NGS, the metagenomics field underwent a great breakthrough. Now, microbial communities can be seen as basic units - superorganisms - delivering specific services assigned to their unique repertoire of genomes - metagenome - fulfilling the proposes of microbial ecology. Thus, beyond answer to “Who is there?”, microbial ecology has now the tools to address a more important question, “What are they doing (potentially)?”, to finally respond to “How they interact with each other and with the surrounding environment to do it?”.

1.3. Bioinformatics Pipelines and Databases

High-throughput DNA sequencing technologies generate large amounts of raw sequence data, or reads, that needs to be processed before being annotated against a non-redundant reference database. Processing reads into comprehensive biological information consists in decreasing the level of complexity from these huge datasets using high performance algorithms integrated in bioinformatics pipelines or workflows (Leipzig, 2017). Usually this requires the following steps, concerning the amplicon NGS data: demultiplex the raw amplicon libraries into samples; filter the reads based on quality parameters; align reads for further quality control; merge identical sequences - aka dereplication; detect and remove artificial sequences with multi parent origin produced during PCR amplification - aka chimeras; cluster unique sequences into operational taxonomic units (OTUs) based on user-defined threshold; assign OTUs to taxonomy using a non-redundant, reference database. Several bioinformatics pipelines have been developed to carry out the outlined steps. Within the most widely used are: mothur (Schloss *et al.*, 2009), Quantitative Insights into Microbial Ecology (QIIME; Caporaso *et al.*, 2010), and UPARSE (Edgar, 2013). Nonetheless, their utilization is limited by command-line and programming skills that are unusual among microbiologists. For this reason, emerged user-friendly options as well as automated pipelines in web servers, such as: MetaGenome Rapid Annotation using Subsystem Technology (MG-RAST; Meyer *et al.*, 2008), SILVAngs (Quast *et al.*, 2013) and RDPipeline (Cole *et al.*, 2014). Despite,

the latter being easier to deal with, the user has a standardized outcome and less flexibility and interactivity with their own data. Although, both allow to process amplicon sequencing data with more or less flexibility, they could be quite different in their “philosophical” approach focusing on particular step that they (developers) consider crucial and neglecting completely others. For instance, the team that manage the SILVAngs platform at Bremen does not include a step to identify and remove chimeric sequences because they do not trust in the current algorithms to identify them. However, chimeras could represent about 10% of the sequences in SSU rDNA (small subunit ribosomal DNA) libraries (Ashelford *et al.*, 2006; Schloss *et al.*, 2011; Porazinska *et al.*, 2012) being the most common source of error within reference databases (Hugenholtz and Huber, 2003; Ashelford *et al.*, 2005). By including chimeric sequences in the dataset, the SILVAngs pipeline incorporates those artificial sequences as novel taxa (=OTU) overestimating the sample richness. Meanwhile, highly efficient algorithms such as UCHIME (Edgar *et al.*, 2011) prove their worth identifying 99% (7% of the total dataset) of chimeric sequences within mock communities (a defined combination of DNA from known species that are used to mimic a natural microbial community; Schloss *et al.*, 2011). An OTU was the bioinformatics form found to discriminate a set of sequences that come from individuals of the same species (or other taxonomic level). The user-defined threshold to cluster similar sequences into OTU, at species level, is $\geq 97\%$ (Konstantinidis and Tiedje, 2005; Nguyen *et al.*, 2016). This threshold proved to be higher (Nguyen *et al.*, 2016), but at the same time the 3% dissimilarity can account for some intra-variability due to sequencing errors (Schloss and Wescott, 2011) and then considered to be a good threshold. Moreover, the SILVAngs pipeline does not discard the singletons (OTUs with just one sequence) from the taxonomy table provided, which could represent spurious OTUs or rare members of communities, with low coverage (Edgar, 2013; Aanderud *et al.*, 2015; Nguyen *et al.*, 2015). Nevertheless, the use of centroid-based clustering algorithms, such as UCLUST, by SILVAngs and QIIME (by default) produces high number of OTUs, that spurious or not, do not represent the true structure and diversity of microbial communities (Edgar, 2013). Still, these greedy algorithms, *e.g.*, UCLUST and CD-HIT, are widely used because they offer the best trade-off between accuracy and speed. This contrast with the robustness of average neighbor algorithm (Schloss and Wescott, 2011), a hierarchical-clustering based algorithm, supported by mothur, which can easily take 10 x more time than using QIIME with UCLUST to process the same kind

of data (Plummer *et al.*, 2015). An alternative is the greedy algorithm developed by Edgar (2013) which has been gained wide acceptance. Despite their greedy nature, it offers a better trade-off between robustness and speed than UCLUST with QIIME (Pylro *et al.*, 2014). Another important consideration is the representative sequence assigned to each OTU that goes to be used for classification. While SILVA picks the longest read as representative, QIIME considers the most abundant one. These differences will constraint the assignment at taxonomic level. Therefore, since these studies rely on sequence-based analysis, our ability to assign taxonomy to one OTU is by searching for the most similar sequence, usually through pairwise sequence alignments, within a known reference dataset of sequences whose origin was experimental determined.

Among the highly curated, chimera-free, non-redundant, reference databases of 16S rRNA gene sequences are: Greengenes (DeSantis *et al.*, 2006), SILVA (Quast *et al.*, 2013) and RDP (Ribosomal Database Project; Cole *et al.*, 2014). The most widely used non-redundant, reference database of 16S rRNA gene sequences, SILVA (version 123, released in September 2015), comprises 152 308 bacterial full-length sequences and 3 901 archaeal ones, an order of magnitude higher than RDP (version 14, released in May 2015), for both, *Bacteria* (10 244) and *Archaea* (435). While, Greengenes (the latest version, released August 2013) accounts for 202 421 of 16S rDNA sequences for both prokaryotic groups. Regarding the 18S rDNA reference databases, the SILVA (Quast *et al.*, 2013) and PR² (Protist Ribosomal reference database; Guillou *et al.*, 2013) are among the most widely used. Despite the fact that the PR² database has 135 110 sequences of 18S rDNA not equally distributed among the several groups of protists, just 29.4% ($\approx$ 39 722) are complete or almost complete, though 63.7% ($\approx$ 86 065) hold the V4 hypervariable region (Guillou *et al.*, 2013). Whereas, SILVA database (version 123, released in September 2015) comprises 16 209 eukaryotic sequences with full-length.

The multiple tiny pieces of DNA sequenced from a metagenomics library need to be put together like a puzzle. The reads are aligned in order to get larger consensus segments of contiguous stretches of DNA - contigs (Thomas *et al.*, 2012; Oulas *et al.*, 2015). Contigs could be assembled in order to yield scaffolds (Thomas *et al.*, 2012; Oulas *et al.*, 2015). Since there is not a reference to align against, the process has to be done by *de novo* (Thomas *et al.*, 2012; Oulas *et al.*, 2015). The last step in this process is bin and predict the taxonomy and function, respectively, of scaffolds. Depending on the complexity of the metagenome, the scaffolds could be taxonomically assigned or binned

based on composition, *e.g.*, GC-skew, codon usage, *k*-mer frequency, or by similarity means (Wooley *et al.*, 2010; Thomas *et al.*, 2012). Finally, BLASTing (perform a Basic Local Alignment Search Tool) the scaffolds against a curated, reference database could be the easiest path to predict the function and, at the same time, depend on the extension of conservation - identity -, the taxonomy (Wooley *et al.*, 2010; Scholz *et al.*, 2012). However, BLAST is computationally intensive which led researchers to deal with friendly options available online in order to get their metagenome annotated. MG-RAST, Integrated Microbial Genomes and Metagenomes (IMG/M; Markowitz *et al.*, 2014) and EBI metagenomics (Mitchell *et al.*, 2016) are currently available workbenches to annotate metagenomics sequencing data providing useful graphics visualization for publication. Some of the more used functional - protein - reference databases are KEGG (Kyoto Encyclopedia of Genes and Genomes; Kanehisa *et al.*, 2016) and COG (Clusters of Orthologous Groups of proteins; Tatusov *et al.*, 2000). Given the amount of raw sequence data that remain to be unassembled it's always very hard to find open-reading frames and other structural elements in metagenomics sequencing data. Nonetheless, searching for short highly conservative domains in reference databases such as Pfam (Protein family database; Finn *et al.*, 2016) and TIGRFAM (Haft *et al.*, 2003) is a good alternative to overcome this problem and annotate our data.

2. Objectives and Thesis Organization

The thesis presented herein - “Arctic microbiome and N-functions during the winter-spring transition” - makes part of an international collaboration with the Norwegian Polar Institute. The data used in this work was collected during the Norwegian young sea ICE expedition (N-ICE2015) that aimed to assess the ecological impact of changing sea-ice regime in the Arctic Ocean. The goals of this thesis are: (1) to describe, at compositional and structure levels, the Arctic's microbial communities, prokaryotes and protists, based on SSU rDNA amplicon sequencing data; and (2) identify important players in the nitrogen biogeochemical cycle and related pathways given their importance in this polar environment to primary production, through shotgun environmental metagenomics data. In order to give a comprehensive analysis and address properly both points, they were divided into two chapters: the first chapter - “Pelagic microbial communities from an Arctic drift ice during winter to spring transition north of Svalbard”

- focusing on (1); and a second chapter - “N-cycling microbial communities and pathways in the Arctic Ocean” - aiming at (2). Both chapters include “Background”, “Methods”, “Results and discussion”, and “Conclusions” sections.

1st Chapter:

Pelagic Microbial Communities from an Arctic Drift Ice
during Winter to Spring Transition North of Svalbard

1. Background

1.1. Norwegian young sea Ice expedition 2015: motivation

Over the last 30 years, the Arctic summer sea-ice extent and thickness have drastically decreased (Meier *et al.*, 2014; Lindsay and Schweiger, 2015). As a consequence, the ice pack became much younger, and the older and thicker multi-year ice (MYI) that survives summer melt, has largely disappeared and been replaced with first-year ice (FYI; Maslanik *et al.*, 2011; Polyakov *et al.* 2012; Parkinson and Comiso, 2013). The loss of the thicker (> 2 m) MYI pack intensifies positive feedback mechanisms, like the well-known ice-albedo feedback (Perovich and Richter-Menge, 2009), that leads to an increasing rate of decline of ice cover per decade (Serreze *et al.*, 2007). Arctic's thinner ice regime has been shaped by both atmospheric and oceanic forcing (Screen and Simmonds, 2010; Polyakov *et al.*, 2012; Polyakov *et al.*, 2017). Successive minimum records of sea-ice extent over the last years lead to the prediction of an Arctic summer completely ice-free at the end of the 21st century (Boé *et al.*, 2009).

The Norwegian young sea Ice expedition 2015 (Granskog *et al.*, 2016) that took place in the ice pack north of Svalbard, between January and June 2015, aimed at studying the new sea-ice regime combining atmospheric with oceanographic and sea-ice sampling.

1.2. Microbial Oceanography: polar oceans

The dark-light transition from polar night to polar day and from ice-covered into open waters impacts microbial metabolism through changes in light and nutrient availability. A metagenomics survey in the Southern Ocean showed a transition in dominance of chemolithoautotrophy towards phototrophy from winter to summer (Grzyski *et al.*, 2012). Bottom-up factors are particularly important in polar oceans controlling the growth and abundance of primary producers and heterotrophic bacteria (Kirchman *et al.*, 2009). At the end of winter light triggers not just the emergence of phytoplankton but also the dormant microbial life that relies on phototrophic metabolism, such as those carrying out anoxygenic photosynthesis and light-driven proton pumps - proteorhodopsins (originally found in one proteobacterium) and bacteriorhodopsin (*Halobacteria* class, *Euryarchaeota*; Bryant and Frigaard, 2006). While the remaining component of the microbial planktonic community relies directly on grazing on

phytoplankton or, indirectly, feeding on phytoplankton-derived dissolved organic matter (DOM) and “marine snow”. Although often referred as a vital constraint factor for microbial activity, temperature seems to play a minor role on bacterial growth and production when compared to light and nutrients, especially availability of labile DOM (although this could be correlated with temperature; Rivkin *et al.*, 1996; Deming, 2002; Kirchman *et al.*, 2009). Indeed, microbial activity was recorded at -20 °C in Arctic sea-ice cores (Junge *et al.*, 2004) and in one permafrost microcosm experiment (Tuorto *et al.*, 2014). In addition to a drastic seasonal pattern, all these factors show depth gradients that shape the microbial plankton at taxonomic and functional levels along the water column (DeLong *et al.*, 2006).

1.3. Arctic’s Microbial Communities: diversity and structure

The biogeochemical 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 the dietary basis of marine food webs and central players of biogeochemical cycles. With a more dynamic Arctic ice cover (Spreen *et al.*, 2011; Itkin *et al.*, 2017), the probability of lead and pressure ridge formation increases under the new Arctic sea-ice regime (Simmonds and Keay, 2009; Wadhams and Toberg, 2012; Willmes and Heinemann, 2016; Itkin *et al.*, 2017). Leads have been shown as an important source of light to initiate and sustain algal blooms dominated by the haptophyte *Phaeocystis pouchetii* under snow-covered sea ice (Assmy *et al.*, 2017). Early *P. pouchetii* spring blooms will deplete the nitrate surface inventory (Harrison and Cota, 1991; Comeau *et al.*, 2011) long before summer with possible negative effects on the magnitude of the diatom spring bloom (Assmy *et al.*, 2017). This trend could be further amplified by the observed long-term decline in silicic acid concentrations, a nutrient diatoms are critically dependent on, in the Norwegian and Barents Seas (Rey, 2012). 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 (Li *et al.*, 2009; Lasternas and Agustí, 2010; Lalande *et al.*, 2013; Nöthig *et al.*, 2015).

Previous studies about the structure and diversity of sea-ice microbial

assemblages concluded that sea-ice harbours microbial communities distinct from those in the underlying water column; however, comparable at diversity levels (Bowman *et al.*, 2012; Hatam *et al.*, 2014; Hatam *et al.*, 2016). Notably, the low abundance of the *Archaea* domain which is highly abundant and ecologically relevant in polar surface waters (Brown and Bowman, 2001; Bowman *et al.*, 2012). In addition, a transition from stable MYI to transient FYI bacterial communities is expected (Hatam *et al.*, 2016). Further, an increase in ciliates and a reduction in thaumarchaeotes was observed after the record sea-ice minimum in September 2007 (Comeau *et al.*, 2011).

Nevertheless, the Arctic Ocean is still poorly characterized at a genetic level, especially north of Svalbard (Pedrós-Alió *et al.*, 2015) which is well mirrored by the limited number of recently published works (Metfies *et al.*, 2016; Fernández-Méndez *et al.*, 2016; Meshram *et al.*, 2017). Metfies *et al.* (2016) and Meshram *et al.* (2017) reported the marine picoeukaryotes: as the main photosynthetic protists, mostly related to *Phaeocystis* sp. and *Micromonas* sp. (mixotrophic); and their distribution closely associated to water mass circulation, while Fernández-Méndez *et al.* (2016) found a high richness of the related nitrogen-fixing gene *nifH*, with few genes belonging to cyanobacteria (that are near absent in this region). Interestingly, a good part of it was affiliated to anaerobic diazotrophs (20%), from cluster III, that seem to thrive through this cold and highly oxygenated waters (Fernández-Méndez *et al.*, 2016). All these studies include seawater samples collected during summertime or close to it (May), north of Svalbard, in the Central Arctic Ocean (including Nansen Basin), when the melting of sea-ice is in an advanced state (at least for the south most stations). Therefore, the winter microbial communities (prokaryotic and protistan) underneath the Arctic ice pack remain overlooked. Also, those works focused on specific compartments of microbial communities based on their ecological role, *i.e.*, photosynthetic picoeukaryotes and diazotrophs, neglecting others, that cannot be dissociated.

1.4. Objectives

The Norwegian young sea Ice expedition 2015 (N-ICE2015) that took place in the ice pack north of Svalbard, between January and June 2015, provides high-throughput sequencing data along depth and from winter to spring seasons that will improve current knowledge about the microbiota of the Arctic Ocean. This chapter intends to describe the

Arctic's microbial communities, prokaryotes and protists: (1) at compositional and (2) phylogenetic (diversity) levels; and (3) how this diversity is structured during the winter-spring transition in order to get insight about how they are being shaped by the changing sea-ice regime.

2. Methods

2.1. 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 (Granskog *et al.*, 2016). This study uses data collected during the expedition and described in detail in other papers and datasets (Assmy *et al.*, 2016; Peterson *et al.*, 2017; Meyer *et al.*, 2017).

Three microbial sample sets ($n=9$) were collected along the floe 2, 3 and 4 drifts illustrated in Granskog *et al.* (2016), from winter (snow-covered sea ice) towards summer (sea-ice without snow). These three sampling sites are highlighted in Figure 1 and were chosen to collect seawater, at three distinct depths; hereafter designated as surface (5 m), middle (20 or 50 m) and bottom (250 m) samples. The north most site sampled is in the deep Nansen Basin (NB, Eurasian Basin) and the south most is in the shallow Yermak Plateau (YP). Between both sites, a sample was obtained in the marginal zone of YP, deeper than YP, but shallower than NB, and hereafter referred as Transition Region (TR). This area is involved by two arms (the Yermak and Svalbard Branches) of warmer and saltier Atlantic water that inflows through the Fram Strait towards the Arctic Ocean surrounding completely the Plateau (Meyer *et al.*, 2017). The advection of relatively warm Atlantic water masses has a profound effect on the Eastern European Arctic including the Svalbard Archipelago (Polyakov *et al.*, 2017). This advection of heat has a pronounced effect on the highly-stratified Arctic colder polar surface waters, resulting mostly in three distinct water masses, regarding temperature and salt properties, as characterized in Meyer *et al.* (2017): a layer of colder and less saline polar surface water (PSW); an intermediate layer of Atlantic water that suffers great influence of cooler polar waters presenting a sharp gradient of salt - halocline - (MAW, Modified Atlantic Water); lastly, a deeper layer of Atlantic water (AW). Additionally, the melting ice influence the

formation of warm Polar Surface Waters (PSWw) over the YP during springtime (Meyer *et al.*, 2017).

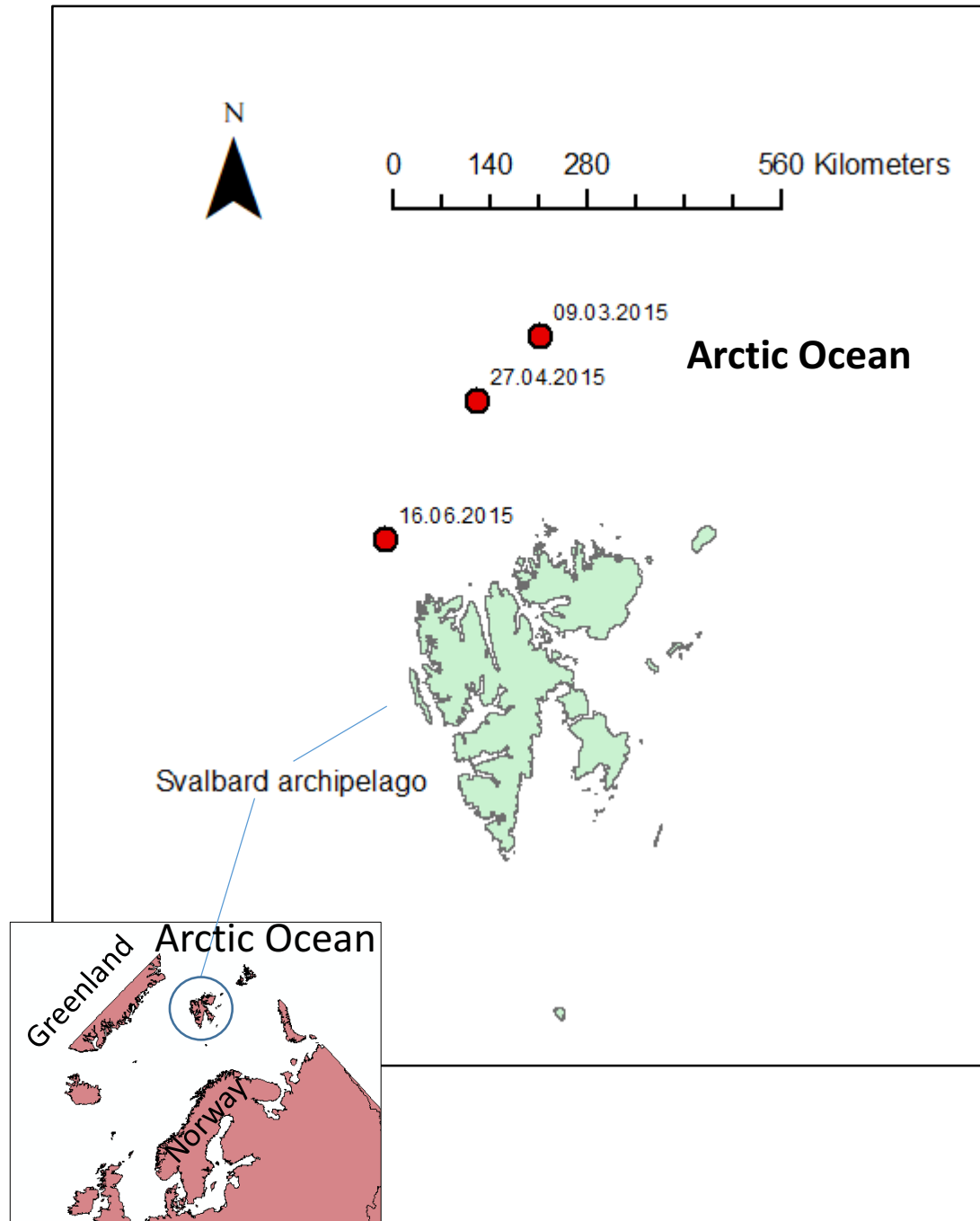


Figure 1 - Map highlighting the sampling sites (red dots) north of Svalbard. Samples collected during the N-ICE2015 cruise from the surface (S), middle (M) and bottom (B) at Nansen Basin (NB, 09.03.2015), Transition Region (TR, 27.04.2015) and Yermak Plateau (YP, 16.06.2015; see Table 1).

2.2. 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.* (2017). 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 holding. The former CTD included several sensors to measure, apart from salinity and temperature, fluorescence, chlorophyll and photosynthetically active radiation (PAR). 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^+), nitrite (NO_2^-) plus nitrate (NO_3^-), phosphate (PO_4^{3-}), silicate (SiO_4^{4-}), total dissolved nitrogen (TDN), particulate organic nitrogen (PON), dissolved organic carbon (DOC), particulate organic carbon (POC) and phaeopigments.

Water column physical and biogeochemical data mentioned in the previous paragraph and used here to describe the environmental context from where the microbial communities were collected are described and available in Peterson *et al.* (2016) and Assmy *et al.* (2016), respectively.

All the measurements made regarding the physical and biogeochemical data contextualizing the environment of microbial N-ICE2015 collection are given in Additional file 1: Table S1.

In order to get a reproducible, reliable and compatible results, we adopted the scheme from the Ocean Sampling Day (OSD) campaign, from sampling to sequencing the small subunit of ribosomal DNA (SSU rDNA) of microbial communities (Kopf *et al.*, 2015). 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 (SVG010RS, 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 attributed to each sample and description summarized below in Table 1.

Table 1 - Features of sampling conditions of microbial N-ICE2015 collection

sample ID	sampling date	floe	sampling site	latitude	longitude	snow- covered/sn ow-free	water mass	water depth (m)	collection filter pore (μ m)	volume filtrated (L)	filtration time (h)
NB_S	09/03/15	2	Nansen Basin	83° 10.002'N	22° 01.998'E	covered	PSW	5	0.22	5.7	22.0333
NB_M	09/03/15	2	Nansen Basin	83° 10.002'N	22° 01.998'E	covered	PSW	50	0.22	3.7	ND
NB_B	09/03/15	2	Nansen Basin	83° 10.002'N	22° 01.998'E	covered	MAW	250	0.22	4.5	ND
TR_S	27/04/15	3	Transition Region*	82° 23.195'N	15° 9.198'E	covered	PSW	5	0.22	11.0	≈13
TR_M	27/04/15	3	Transition Region*	82° 23.195'N	15° 9.198'E	covered	PSW	50	0.22	10.7	≈9
TR_B	27/04/15	3	Transition Region*	82° 23.195'N	15° 9.198'E	covered	MAW	250	0.22	9.2	≈13
YP_S	16/06/15	4	Yermak Plateau	80° 30.775'N	07° 52.428'E	free	PSW	5	0.22	3.0	ND
YP_M	16/06/15	4	Yermak Plateau	80° 30.775'N	07° 52.428'E	free	PSWw	20	0.22	3.3	ND
YP_B	16/06/15	4	Yermak Plateau	80° 30.775'N	07° 52.428'E	free	AW	250	0.22	4.0	ND

*Transition Region between the Eurasian Basin and Yermak Plateau. ND - Not Determined. PSW: Polar Surface Water; MAW; Modified Atlantic Water; PSWw: warm Polar Surface Water; AW: Atlantic Water (see description in “Methods”).

2.3. DNA Extraction, PCR, Library Preparation, and Sequencing of SSU rDNA amplicon

SterivexTM filters were thawed, at room temperature, and the plastic mold that surrounds the filter was cut off to remove the filter itself. Then, DNA extraction was proceeded using the PowerWater[®] DNA Isolation Kit protocol (MO BIO Laboratories, Inc., Portugal) 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 initially by Caporaso *et al.* (2011; 2012) and latter modified by Apprill *et al.* (2015) and Parada *et al.* (2016). The primer set has as the target the V4-V5 hypervariable regions of 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; Apprill *et al.*, 2015; Parada *et al.*, 2016). The 18S rRNA gene was amplified with the primer set described in Stoeck *et al.* (2010), TAREuk454FWD1 (5' - CCAGCASCYGC GGTAATTCC - 3') and TAREukREV3_modified (5' - ACTTTCGTTCTTGATYRATGA - 3'), with the exception of reverse primer (TAREukREV3_modified) which had an additional TGA triplet added at 3' end compared to the original introduced by Piredda *et al.* (2017). This primer set amplify the V4 region of 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 V3 Chemistry (Illumina). These steps were performed by LGC Genomics (LGC Genomics GmbH, Berlin, Germany) company and a detailed description is given in Ribeiro *et al.* (2017, submitted).

2.4. Bioinformatics Pipeline: SSU rDNA amplicon

2.4.1. Upstream Sequence Analysis: raw OTU table

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

<https://www.mothur.org>; Kozich *et al.*, 2013). Individually, the forward and reverse reads of each library were joined from raw Illumina fastq files (mothur v. 1.39.5; Schloss *et al.*, 2009). Merged reads with ambiguities (aka N) and out of the 360-380 bp and 365-385 bp range, for the 16S and 18S rDNA amplicon datasets, respectively, were excluded. The remaining sequences were dereplicated (based on 100% similarity) and aligned against the SILVA database (v. 1.2.8; Quast *et al.*, 2013) to screen the SSU rRNA gene region targeted in our study. Those that didn't align were excluded as well as the ones with homopolymers (n>8). After undergone a dereplication step, the unique sequences that differ within 3 base pairs similarity from a more abundant one were clustered together. Chimeric sequences were identified *de novo* and removed with UCHIME (Edgar *et al.*, 2011). Then, the unique reads were assigned against SILVA (v. 1.2.8) using the RDP naïve Bayesian Classifier (Wang *et al.*, 2007). Undesirable lineages, "Chloroplast", "Mitochondria", "unknown", "Eukaryota" were removed from the 16S dataset; while "Vertebrata", "Annelida", "Arthropoda", "Cnidaria", "Ctenophora", "Echinodermata", "Florideophycidae", "Mollusca", "Pav3", "D226", "FV18-2D11", "Tunicata" were excluded from 18S. Afterward, a distance matrix was built and the sequences clustered into OTUs (Operational Taxonomic Units) using 0.03 and 0.02 cutoff values for the 16S and 18S rDNA amplicon datasets, respectively, with OptiClust (Westcott and Schloss, 2017). Finally, the 16S and 18S rDNA amplicon-based OTU tables were built.

2.4.2. Downstream Sequence Analysis: composition

In order to fully exploit the downstream analysis, the OTU tables (one for each SSU rDNA library) produced through mothur pipeline were converted into biom format (mothur v. 1.39.5). The OTU tables in biom format were imported to QIIME (v. MacQIIME 1.9.1; Caporaso *et al.*, 2010) 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 rDNA amplicon-based OTU tables, respectively. The number of sequences filtered during each upstream step for the 16S and 18S rDNA amplicon datasets are summarized in Additional file 1: Tables S2 and S3, respectively. While, the distribution of prokaryotic taxa across N-ICE2015 collection at phylum, class, order, family, genus and OTU levels is provided in Additional file 2; the raw prokaryotic OTU table (without excluding any taxa, rare OTUs neither rarefying) in

Additional file 3; and, the distribution of eukaryotic taxa at different taxonomic ranks in Additional file 4. To clarify, all analysis described below and presented as figures in the text were conducted in duplicate for the 16S and 18S rDNA datasets based on information retrieved from Additional files 2 and 4, respectively.

2.4.3. Downstream Sequence Analysis: diversity and structure

A multiple sequence alignment with representative sequences of each OTU through 1 000 iterations was performed using MAFFT (v. 7.310; Katoh and Standley, 2013) to build a maximum likelihood tree under the GTRGAMMA model with 1 000 bootstrap replicates using RAxML (pthreads v. 8.0.26; Stamatakis, 2014). The OTU table and tree constructed were used to perform alpha (within samples) and beta (between samples) diversity metrics (using QIIME). Faith's Phylogenetic Diversity (Faith, 1992) was the alpha metrics estimated. While for beta-diversity the unweighted UniFrac metrics was estimated (Lozupone *et al.*, 2005; Lozupone *et al.*, 2007) subsampling the 16S and 18S datasets at 38 232 and 43 289 sequences, respectively, to produce a distance matrix that was visualized through the Principal Coordinate Analysis method (PCoA). Finally, heatmaps and Spearman correlations were generated using Hmisc, corrplot and ggplot2 packages in R (R Development Core Team, 2010; Wickham, 2009; Wei, 2013; Harrell and Harrell, 2015).

2.4.4. Taxonomic Assignment of 18S rDNA Amplicon Libraries: Protist Ribosomal Reference database

In order to improve the taxonomic assignment of 18S rDNA amplicon libraries classified against the SILVA database, the 18S rDNA amplicon dataset was additionally assigned against the Protist Ribosomal Reference database (PR², v. 4.5; Guillou *et al.*, 2013). Therefore, the 18S rDNA amplicon libraries were processed following the same steps aforementioned, with a few exceptions concerning the mothur MiSeq SOP pipeline. The alignment quality step was skipped because the PR² database is not aligned, and the sequences clustered using the VSEARCH algorithm (Rognes *et al.*, 2016). Since the PR² taxonomy is distinct of the SILVA database for some taxa; in addition to the lineages removed above, “Metazoa” was also removed. The rare clusters were excluded (<5

observations across samples) and rarefied at even sampling depth (43 647 sequences). The OTU table with taxonomy is presented in Additional file 5.

2.5. Availability of Sequencing Data

Raw Illumina fastq files, concerning the SSU rDNA amplicon data used in this study, were deposited in the European Nucleotide Archive (ENA) under the project accession number PRJEB21950.

3. Results and discussion

In the following paragraphs, results related with the abundant prokaryotic marine plankton (16S libraries) will be presented and discussed with greater detail than those related with the eukaryotic plankton (18S libraries). Despite the great effort to improve current metabarcoding protocols to reproduce eukaryotic diversity, the poor representation of many eukaryotic species in existing data bases and the lack of discrimination power from a unique universal marker remains a significant hurdle to uncover their true diversity (Stoeck *et al.*, 2010). Thus, in this case the results and discussion will focus on phylogenetic diversity and assignment at order/family level, where there is greater confidence. Also, those groups that fit into 0.2-3.0 μm range (picoeukaryotes) and that are not possible to classify through traditional taxonomic standards will also be discussed.

3.1. Metabarcoding the Arctic microbiota through the Winter-Spring Transition: prokaryotic and protistan communities

The prokaryotic marine plankton collected during the N-ICE2015 expedition, north of Svalbard, possesses a repertoire of 24 phyla (Additional file 2). The majority of them (15) represents < 1% of the sequence reads across the entire N-ICE2015 dataset (*Woesearchaeota* (DHVEG-6), *Acidobacteria*, unclassified *Bacteria*, *Chloroflexi*, *Cyanobacteria*, *Deinococcus-Thermus*, *Firmicutes*, *Gemmatimonadetes*, *Gracilibacteria*, *Lentisphaerae*, PAUC34f, SBR1093, *Spirochaetae*, TM6, *Tenericutes*). Among the most abundant are the bacterial phyla *Proteobacteria* (with 63.4% of relative

abundance in the entire dataset) and *Bacteroidetes* (14.1%) beyond the archaeal phylum *Thaumarchaeota* (10.4%) (Additional file 2 and Figure 2a). *Proteobacteria* dominates the majority (*i.e.*, > 50%) of the Arctic's prokaryotic libraries, with the exception of the mesopelagic sample from the Yermak Plateau, which is dominated by bacteroidetes (57.1%, comprises mostly *Flavobacteriales*; Figure 2a).

3.1.1. The Dominant *Proteobacteria* and its Co-occurrence Patterns with Eukaryotes

Contributing with almost equal parts to the high abundance of *Proteobacteria* are the *Alpha*- (30.6% in the prokaryotic N-ICE2015 dataset) and *Gammaproteobacteria* (28.6%) classes, which is in agreement with previous studies (Figure 2a; Kirchman *et al.*, 2010; Galand *et al.*, 2010; Comeau *et al.*, 2011; Ladau *et al.*, 2013). These two classes are 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* is mainly composed by one ribotype affiliated to “*Candidatus Pelagibacter*” from SAR11 clade (Figure 2b). SAR11 clade is the most abundant phylotype in the surface of the oceans representing up to one-third of the bacterioplankton (Morris *et al.*, 2002). In the N-ICE2015 dataset SAR11 represents on average $\approx 30.7\%$ of the epipelagic communities sampled and $\approx 9.2\%$ of the mesopelagic ones. Previous NGS-based studies (Kirchman *et al.*, 2010; Comeau *et al.*, 2011) reported a representativeness of this group on Arctic surface waters of $\approx 15\%$ on average. The difference to the results presented herein could be attributable to the primers used in our study that were designed to efficiently catch the diversity of marine and freshwater alphaproteobacteria from SAR11 clade (see “Methods” section; Apprill *et al.*, 2015; Parada *et al.*, 2016). Nevertheless, we cannot rule out the possibility that the differences observed merely reflect the different composition across distinct biogeographical regions of the Arctic Ocean, since the studies outlined before (Kirchman *et al.*, 2010; Comeau *et al.*, 2011) were conducted in the western Arctic Ocean, while the N-ICE2015 cruise occurred north of Svalbard.

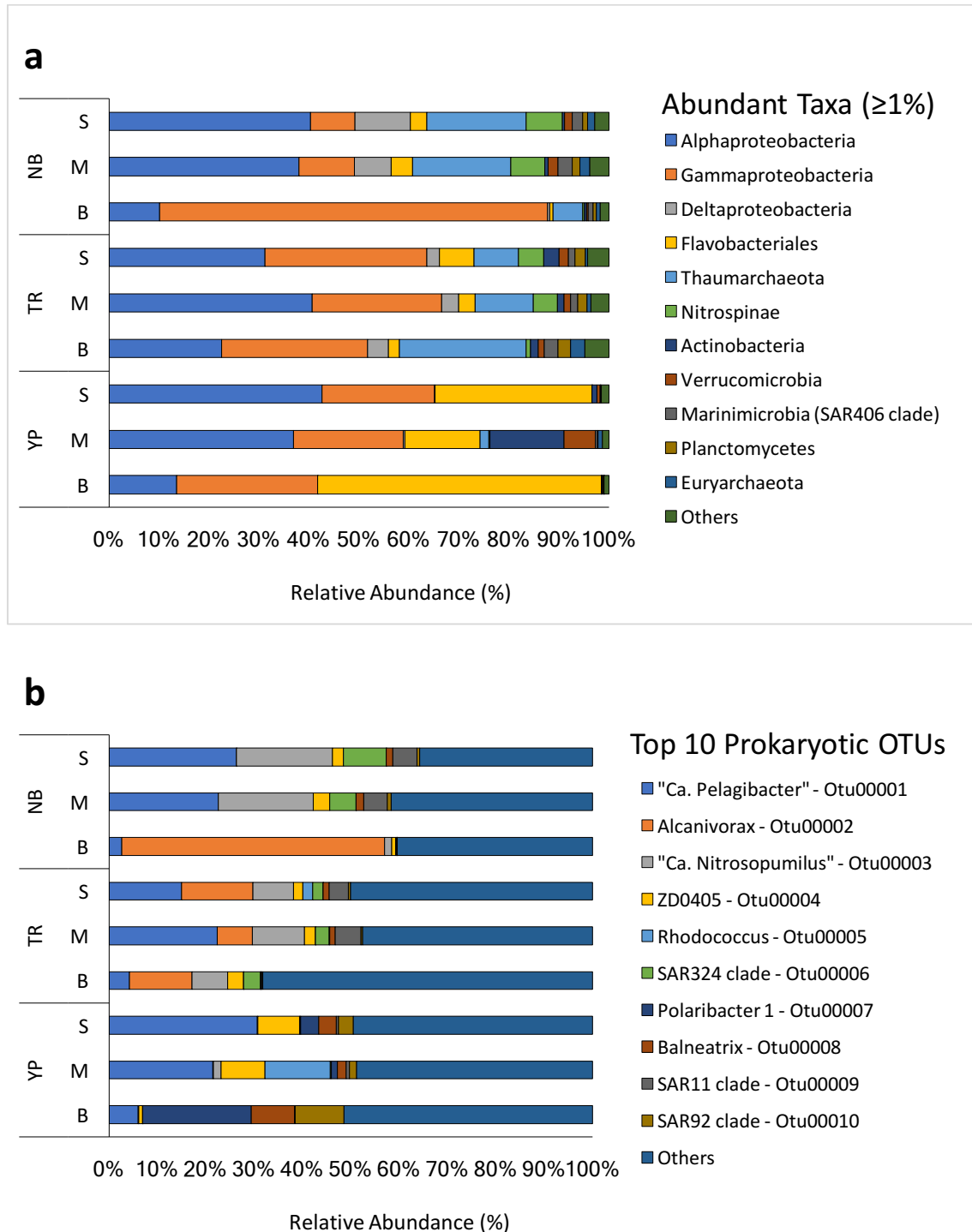


Figure 2 - Taxonomic profile of prokaryotic N-ICE2015 collection. N-ICE2015 collection possess bacterial and archaeal sequences retrieved from the surface (S), middle (M) and bottom (B) seawater at Nansen Basin (NB), Transition Region (TR) and Yermak Plateau (YP) (see Table 1). **a** Distribution of the abundant taxa ($\geq 1\%$) at higher taxonomic level, phylum (*Thaumarchaeota*, *Nitrospinae*, *Actinobacteria*, *Verrucomicrobia*, *Marinimicrobia* (SAR406 clade), *Planctomycetes*, *Euryarchaeota*), class (*Alpha*-, *Gamma*-, *Deltaproteobacteria*), and order (*Flavobacteriales*). **b** Distribution of the top 10 OTUs (Operational Taxonomic Units) 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%).

The genome from SAR11 alphaproteobacterial isolate “*Ca. Pelagibacter ubique*” unveils great metabolic ability to cope with the oligotrophic lifestyle (Giovannoni *et al.*, 2005). It is one of the smallest genome known for a free-living microorganism enriched in A+T bases which demands less nitrogen (Giovannoni *et al.*, 2005). Further physiological studies (*e.g.*, Steindler *et al.*, 2011) demonstrate an equal skill to grow heterotrophically under both regimes, darkness and light, with a specific requirement of reduced sulphur, like the phytoplankton-derived osmolyte 3-dimethylsulphoniopropionate (DMSP; Tripp *et al.*, 2008) being responsible for the consumption of 30% of assimilated DMSP in the North Atlantic (Malmstrom *et al.*, 2004). Several phylogenetically diverse algae have the ability to produce the osmoprotectant DMSP, such as diatoms, dinoflagellates, and prymnesiophytes (Liss *et al.*, 1994). Particularly, *P. pouchetii* produces high levels of DMSP and together with diatoms contribute to the DMSP pool in Arctic surface waters (Matrai and Vernet, 1997). While the DMSP production is light-dependent for the *P. pouchetii*, it is not for the diatoms, that tends to precede the *P. pouchetii* blooms (Matrai *et al.*, 1995) which could explain the presence of SAR11 clade under snow-covered sea ice. Nevertheless, diatoms and *P. pouchetii* are less represented in the eukaryotic libraries of N-ICE2015 collection for the corresponding surface (5 m) and middle (20 or 50 m) samples. Otherwise, dinoflagellates dominate our 18S libraries at those depths (Additional file 4 and Figure 3) and are much stronger producers of DMSP than diatoms (Levasseur, 2013). Therefore, this result suggests that the high relative abundance of SAR11 clade on surface Arctic Ocean under the pack-ice is sustained by DMSP production by dinoflagellates.

Gammaproteobacteria also represents a high quota of proteobacteria in the 16S libraries of the N-ICE2015 dataset along both seasons. However, the relative abundance of the frequency of occurrence of the most abundant gammaproteobacterial orders (*Cellvibrionales* (2.7%), *Alteromonadales* (3.9%), and *Oceanospirillales* (19.1%)) varied seasonally (Additional file 2). *Cellvibrionales* was mostly found in the Yermak Plateau, where the samples were collected in early summer (Table 1), the *Alteromonadales* was particularly abundant during springtime with a peak in the 250 m depth sample from Nansen Basin (20.1%, NB_B; Additional file 2), whereas *Oceanospirillales*, was highly abundant across all the samples, particularly at NB_B sample (55.7%) where it

predominates (Additional file 2).

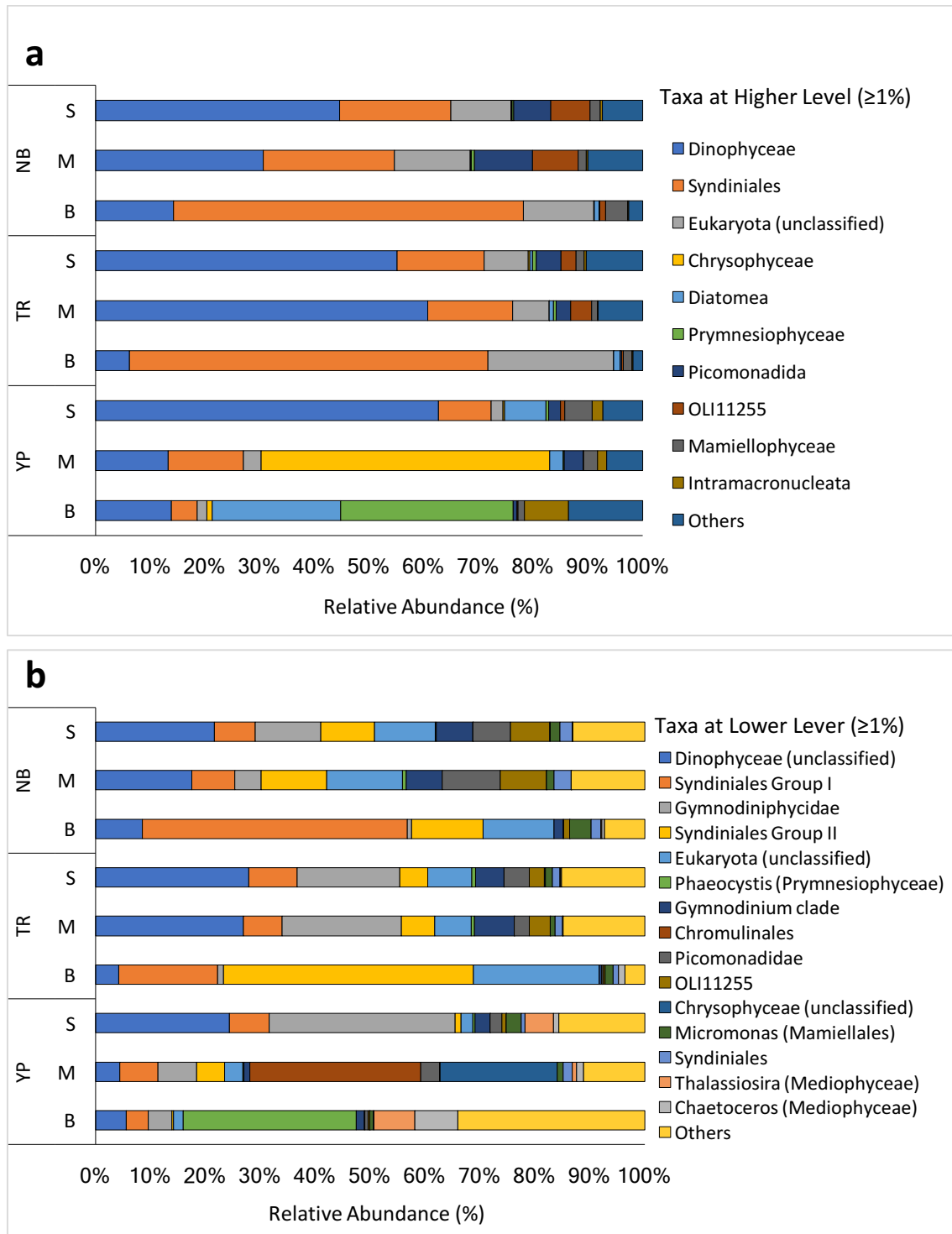


Figure 3 - Taxonomic profile of eukaryotic N-ICE2015 collection. N-ICE2015 collection possess protists sequences retrieved from the surface (S), middle (M) and bottom (B) seawater at Nansen Basin (NB), Transition Region (TR) and Yermak Plateau (YP) (see Table 1). **a** Distribution of the abundant taxa at higher level ($\geq 1\%$) retrieved from the OTU table assigned at class level (SILVA v. 1.2.8 classification). **b** Distribution of the abundant taxa at lower level ($\geq 1\%$) 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. “Others” represents the sum of the frequency of occurrence of the rare taxa (< 1%).

The greater representativeness of *Cellvibrionales* in the early summer waters of YP is confirmed by a proteorhodopsin-harboring SAR92 clade (*Porticoccaceae* family; Figure 2b), which demands for a photoheterotrophic metabolism (Stingl *et al.*, 2007). Previous reports successively observed an increased abundance and activity of this clade during phytoplankton spring blooms in the southern North Sea (Klindworth *et al.*, 2014; Wemheuer *et al.*, 2014; Teeling *et al.*, 2016). In agreement, the same closely related haptophytes (*Phaeocystis* spp.) and polar centric diatoms (*Thalassiosira* and *Chaetoceros* spp.) were found in our 18S libraries (Figure 3b). As expected, the highest percentage of SAR92 clade (10.4%; Figure 2b) appears together with the greatest fraction of diatoms (23.5%, *Diatomea* class, SILVA v. 1.2.8 classification; *Thalassiosira* (7.4%) and *Chaetoceros* (7.8%)) and, particularly, with a dominance of *Phaeocystis* phylotypes (31.5%; reported elsewhere; Assmy *et al.*, 2017), in YP_B sample (Figure 3a and b). Intriguingly, their highest frequency co-occurs at 250 m depth in the Yermak Plateau (Figure 3b), which might be the consequence of a vertical mixing event (see below). As far we know this is the first time that SAR92 clade is associated with the diatoms *Thalassiosira* and *Chaetoceros* as well the haptophyte *Phaeocystis* in the Arctic Ocean. This gammaproteobacterial clade was not reported in a recent study on the bacterial composition in first-year drift ice and under-ice seawater at Fram Strait, western of Svalbard (Eronen-Rasimus *et al.*, 2016). However, it was already detected on surface waters of the Canadian Arctic (Han *et al.*, 2014).

3.1.2. Evidences of High Frequencies of Hydrocarbon-degrading Bacteria in the Arctic

Regarding the two orders, *Alteromonadales* and *Oceanospirillales*, there is a clear dominance of three ribotypes, at NB_B sample, classified as: *Alteromonadales*, *Marinobacter* (“Otu00019”, 6.3%) and *Glaciecola* (“Otu00020”, 12.8%), and *Oceanospirillales*, *Alcanivorax* (“Otu00002”, 54.3% at NB_B and 9.9% across the entire

prokaryotic N-ICE2015 dataset; see Figure 2b and Additional file 2). Although all of these copiotrophs have been previously retrieved from the Arctic sea-ice (Bowman, 2015) or seawater (Bano and Hollibaugh, 2002), with the exception of the psychrophilic *Gaciecola* spp. usually found in Arctic and Antarctic sea-ice, polar seas and their coastal environments (von Scheibner *et al.*, 2017), *Alcanivorax* and *Marinobacter* are not expected to represent a significant part of the Arctic's microbial communities. In fact, *Marinobacter* and *Alcanivorax* are known as hydrocarbon degraders (McGenity *et al.*, 2012). Although hydrocarbon content was not measured, the three could be considered indicator species of hydrocarbon availability, since the dominance of hydrocarbonoclastic bacteria genera, such as *Marinobacter* and *Alcanivorax*, is often related with an oil spills in marine environments (Hara *et al.*, 2003; Schneiker *et al.*, 2006; Duran, 2010). Actually, two microcosms experiments amended with crude oil upon microbial communities from Arctic sea-ice selected the emergence of some oil-degrading bacteria, such as *Marinobacter* (Gerdes *et al.*, 2005; Brakstad *et al.*, 2008). Additionally, other studies (*e.g.*, Chronopoulou *et al.*, 2015) show that *Gaciecola* spp., isolated from the North Sea, can grow on most of the *n*-alkanes, and also became predominant in oil-contaminated ice cores experiments in the Arctic sea (Brakstad *et al.*, 2008).

Besides that, *Alcanivorax* not just dominates the mesopelagic sample collected at Nansen Basin (54.4%), but also represents a significant part of the prokaryotic marine plankton in the Transition Region' samples (average 14.8%; Figure 2b). Several other hydrocarbon-degrading bacteria are also widespread across the prokaryotic N-ICE2015 dataset at TR (Additional file 2), such as the well-known *Pseudomonas* (1.7%, TR_S), *Shewanella* (1.9%, TR_S), *Rhodococcus* (2.1%, TR_S), *Flavobacterium* (3.8%, TR_S), *Hyphomonas* (1.1%, TR_B), *Marinobacter* (2.2%, TR_S; and 2.7%, TR_B), *Colwellia* (0.1%, TR_B), *Cycloclasticus* (0.3%, TR_B) genera (Giudice *et al.*, 2010). On the other hand, at Yermak Plateau, *Alcanivorax* genus was only found in the middle sample (YP_M) at a low (0.1%) relative abundance. Curiously, other known hydrocarbon-degrading bacteria genera were found at YP at low relative abundances in the middle sample (YP_M), such as *Rhodococcus* (13.5%, YP_M), *Sphingopyxis* (5.8%, YP_M), *Nocardioides* (0.8%, YP_M), *Marinobacter* (0.3%, YP_M), *Colwellia* (0.4%, YP_M; and 0.9% YP_B). Notwithstanding the evidences, we do not have results to demonstrate that north of Svalbard ice-covered waters have been contaminated with crude oil. None oil spill incident was reported recently; however, north of Svalbard and Nansen Basin region

are important targets for oil and gas exploration, although, to date, no information from drill core samples exists. Nevertheless, there is a strong possibility of hydrocarbon availability due to several naturally occurring seepages of crude oil in the Svalbard region (Hammer *et al.*, 2011; Roy *et al.*, 2014). Still there is no information available of recent seep activity in the Southern Nansen Basin and the Eastern Yermak Plateau. Another strong hypothesis is due to deep-water-oil plume from west Spitsbergen oil seepages, which represents a prominent methane/hydrocarbon-rich seep area (Gentz *et al.*, 2014; Mau *et al.*, 2017). The oil plume might be carried out by the West Spitzbergen Current, continuing as the Yermak Branch into the Nansen Basin, with an uncertain pathway through the Yermak Pass (Koenig *et al.*, 2017; Meyer *et al.*, 2017), passing through the studied sampling points. This could explain our results, particularly the ones that shows a lower abundance of *Alphaproteobacteria* and *Pelagibacteraceae*, and higher abundance of *Gammaproteobacteria* including *Oceanospirillaceae* in mesopelagic samples of the three sampling sites (Figure 2). This result is in accordance with other studies (*e.g.*, Hazen *et al.*, 2016), regarding the effect of deep-water–oil plume on microbial communities, and supports our hypothesis of hydrocarbon availability. Therefore, the 16S libraries of N-ICE2015 collection supports evidence of an Arctic’s rare biosphere harbouring several gammaproteobacteria that are prone to degrade petroleum-derived hydrocarbons.

3.1.3. Link Between *Flavobacteriales* and *Phaeocystis pouchetii* Arctic Spring Bloom

The Bacteroidetes phylum is abundant across all the marine prokaryotic plankton collected over Yermak Plateau. A detailed analysis at higher taxonomic resolution reveals *Flavobacteriales* order as the main responsible for the spread of bacteroidetes, ranging from 15.0 to 56.7% (Figure 2a). *Flavobacteriales* was associated with the degradation of phytoplankton-derived polysaccharides (Taylor *et al.*, 2014; Tully *et al.*, 2014), as well as hydrocarbons, as previously referred. An even more thorough analysis, pinpoint the two most abundant *Flavobacteriales* related-OTUs: “Otu00007” (1.3-22.4%, YP_M-YP_B) assigned as “Polaribacter_1” and “Otu00011” (1.3-12.8%, YP_M-YP_B) which didn’t possess enough similarity with any reference sequence in the SILVA database to be classified until genus level (Figure 2b and Additional file 2). The co-occurrence of two *Flavobacteriales* related-OTUs at similar levels of abundance was reported before for the Arctic Ocean (*Polaribacter* and AGG58 cluster (Malmstrom *et*

al., 2007); *Polaribacter* and unidentified *Flavobacteriaceae* (Comeau *et al.*, 2011)). However, the levels reported were considerably smaller compared to ours and elsewhere (Nikrad *et al.*, 2012; Eronen-Rasimus *et al.*, 2016), particularly for *Polaribacter* phylotypes (< 10% in Malmstrom *et al.* (2007) and Comeau *et al.* (2011); and ≈20% of *Flavobacteriia* class in Kirchman *et al.* (2010) compared to 31.4% in the shallower sample of YP (5 m) and the highest percentage, 56.7%, at 250 m depth; Additional file 2). Probably, the lower abundance observed in these previous works (Malmstrom *et al.*, 2007, Kirchman *et al.*, 2010; Comeau *et al.*, 2011) was the consequence of pre-filtration or size fractionation of seawater which led to the loss of those flavobacteria tightly associated with the phytoplankton (Abell and Bowman, 2005), especially, if this produces large colonies involved by external mucopolysaccharides, which seems to be the case (see Assmy *et al.* (2017)). As the genus name suggest, *Polaribacter* spp. were first retrieved from polar environments being commonly found in sea-ice (Gosink *et al.*, 1998; Bowman *et al.*, 2012; Hatam *et al.*, 2014; Hatam *et al.*, 2016) and abundant on polar surface waters (Malmstrom *et al.*, 2007; Nikrad *et al.*, 2012). With a distribution mostly limited to polar latitudes (Ladau *et al.*, 2013), some *Polaribacter* strains are psychrophilic and endemic to polar regions (Gosink *et al.*, 1998; Staley and Gosink, 1999). *Polaribacter* shows a clear shift from winter towards summer accompanied by the melting of snow-covered sea ice that starts at the end of May, beginnings of June. Since it is consistently retrieved from sea-ice, it has been suggested (Comeau *et al.*, 2011; Grzymiski *et al.*, 2012; Eronen-Rasimus *et al.*, 2016) that the melting sea-ice is a potential seed source of *Polaribacter* spp. for the under-ice water column. Probably, the sunlight, either as source of light to the energy-harvesting proteorhodopsins (see González *et al.* (2008)) or light stimulus to algal blooms, selects for the emergence of *Polaribacter*, at Yermak Plateau, explaining the abundant reads of this genus in the YP libraries. In agreement, PAR (Photosynthetically Active Radiation) and Chl *a* (Chlorophyll *a*) are one to two orders of magnitude higher on epipelagic waters sampled at YP compared to ice-covered waters sampled in the other two sites (Additional file 1: Table S1). Also, the greatest fraction of these *Flavobacteriales*-related OTUs at mesopelagic depths (YP_B) co-occurred with the highest presence of “Chloroplast” (see the raw 16S OTU table on Additional file 3). Corroborating this result, the 18S rDNA amplicon libraries show the greatest abundance of *Phaeocystis* (31.5%; *Prymnesiophyceae* class, *Haptophyta* division) in the YP_B sample (Figure 3b). During the N-ICE2015 cruise, an early under-ice spring bloom

dominated by the haptophyte algae *Phaeocystis pouchetii* (late May/early June) was observed in the upper 50 m (Assmy *et al.*, 2017). The same study hypothesizes that vertical mixing of Polar Surface Water (PSW) with Atlantic Water (AW) may have occurred in the shallower Yermak Plateau, with AW possibly supporting the seeding of the bloom (Assmy *et al.*, 2017), which can explain the results observed. Supporting this hypothesis, the PCoA analysis indicates that the YP sample from the surface (YP_S) and bottom (YP_B) share more prokaryotic and protist phylotypes between them than with the middle sample (YP_M, collected at 20 m depth; see below). Thus, *Flavobacteriales*, more specifically *Polaribacter*, 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 (Delmont *et al.*, 2014).

3.1.4. The Evident Seasonal Pattern of the phylum *Thaumarchaeota*

In agreement with literature, about 20% of the overall microbial community that compose the samples collected in the epipelagic waters above the Nansen Basin (NB_S and NB_M) are affiliated with *Thaumarchaeota* phylum. Regardless the scarce number of samples and sampled area, we observed the season pattern characteristic of polar thaumarchaeotes as reported in previous studies (DeLong *et al.*, 1994; Kirchman *et al.*, 2007; Alonso-Sáez *et al.*, 2008; Alonso-Sáez *et al.*, 2012). Results showed a clear progressive migration from the surface to deeper mesopelagic waters, from winter towards summer (see Figure 2a). This pattern is accompanied by the melting of sea-ice that starts at the end of May, beginnings of June. Indeed, recent findings (Kim *et al.*, 2016; Tolar *et al.*, 2016a) pinpoint to a protective behaviour by thaumarchaeotes against the harmful effect of reactive oxygen species levels, which are higher in open surface waters due to their photochemical production. Thus, as expected, nitrifying activity, in surface waters, is higher during sea-ice covered seasons, particular winter, as evidenced by higher abundance of *amoA* gene/transcripts as well as potential nitrification rates (Christman *et al.*, 2011; Pedneault *et al.*, 2014). The near absence of *Thaumarchaeota* phylum from samples collected at Yermak Plateau, even at mesopelagic depths, could result from the simultaneous absence of ice or it is just the effect of environmental constraint of the warmer and saltier Atlantic Water that inflows through this region. Thaumarchaeotes are thought to oxidize ammonia and fix carbon dioxide or bicarbonate

as sole energy and carbon sources (Berg *et al.*, 2007; Hatzenpichler, 2012). However, recent studies highlight the use of alternative organic compounds, particularly urea (Mußmann *et al.*, 2011; Alonso-Sáez *et al.*, 2012; Connelly *et al.*, 2014). Actually, we did not find any correlation between thaumarchaeotes and the nutrient ammonia (measured in the ionic form, ammonium), but instead a strong positive correlation with DOC (see Figure 5). The high nitrifying activity in wintertime of these cold-adapted archaea play a very important role for the N and C budgets in the Arctic Ocean (Christman *et al.*, 2011; Tolar *et al.*, 2016b).

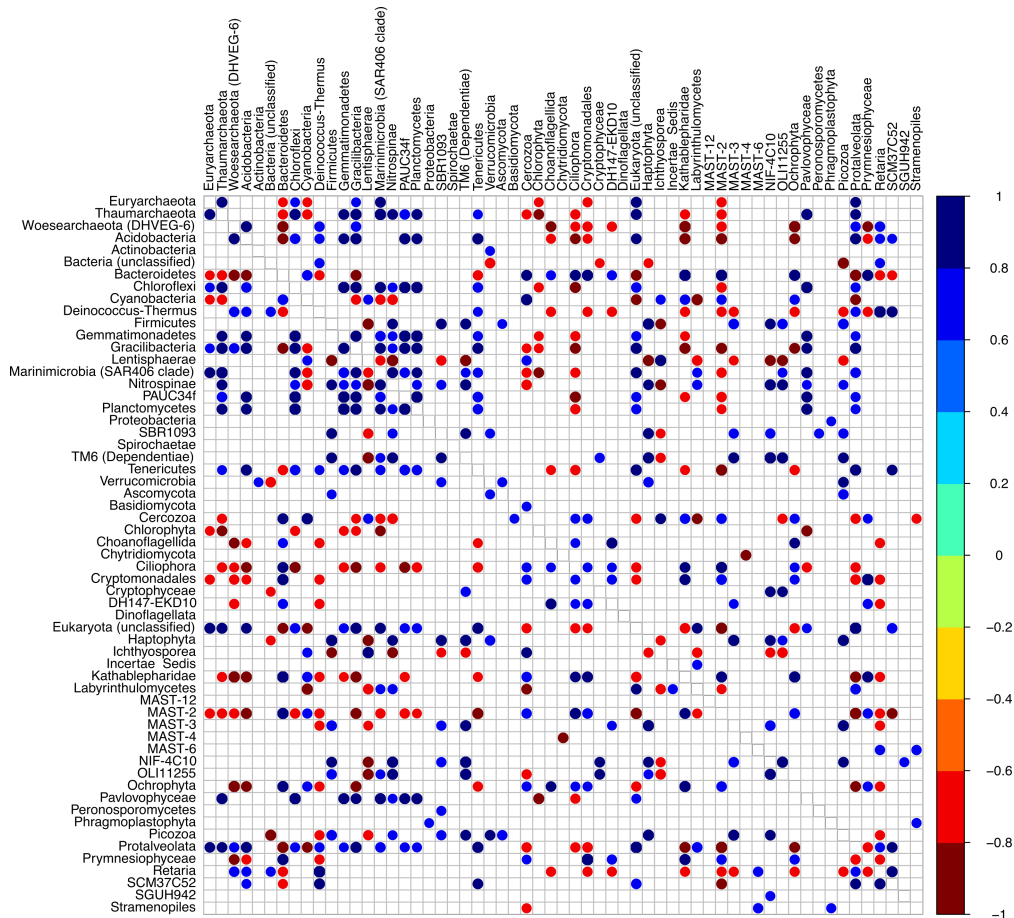


Figure 4 - Heatmap highlighting the significant Spearman correlations between the phyla from the microbial N-ICE2015 collection.

Concerning the reminiscent part of abundant prokaryotic phyla, *Euryarchaeota* (1.0%), *Planctomycetes* (1.1%), *Marinimicrobia* (1.3%), *Nitrospinae* (2.8%) are particularly plentiful in ice-cover waters sampled at Nansen Basin and Transition Region contrasting with *Verrucomicrobia* (1.7%) and *Actinobacteria* (2.6%), which have a peak at 20 m depth water sample of YP (6.3% and 14.8%, respectively; Figure 2a).

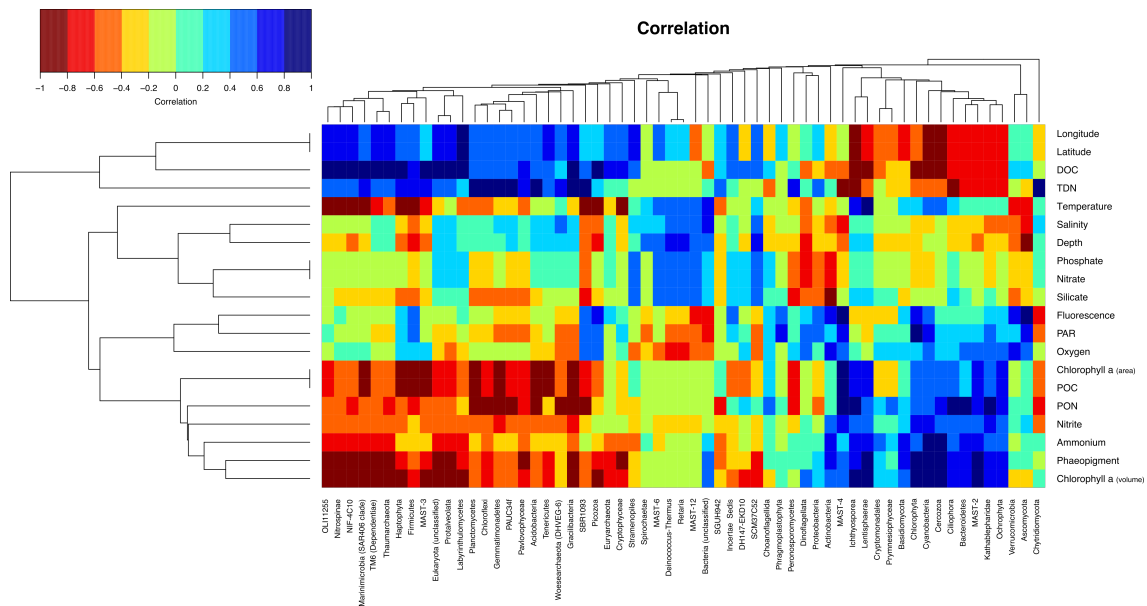


Figure 5 - Heatmap highlighting the Spearman correlation matrix between microbial phyla and environmental variables collected during the N-ICE2015 cruise. DOC stands for Dissolved Organic Carbon; TDN, Total Dissolved Nitrogen; PAR, Photosynthetically Active Radiation; Chlorophyll *a* (volume, mg.m⁻³); PON, Particulate Organic Nitrogen; Chlorophyll *a* (area, mg.m⁻²); POC - Particulate Organic Carbon (see in detail Additional file 1: Table S1).

Interestingly, the distribution of thaumarchaeotes is strong positively correlated with the distribution of *Euryarchaeota*, *Planctomycetes*, *Marinimicrobia* and *Nitrospinae* (Figure 4). 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* (Hawley *et al.*, 2014; Parada and Fuhrman, 2017). The remaining phyla are composed mostly by environmental sequences data, which makes it difficult to infer the ecophysiological role played in Arctic ice-covered waters. The *Planctomycetes* family, *Phycisphaeraceae* (0.4% across prokaryotic N-ICE2015; Additional file 2), and *Marinimicrobia* were implicated in the proteolysis of syntrophic amino acids through a multi-omics study upon a methanogenic reactor (Nobu *et al.*, 2015). Both phyla are usually observed in higher abundances in the dark ocean, associated to oxygen minimum zones (Yilmaz *et al.*, 2016). Indeed, both phyla show a slightly negative correlation with oxygen concentration (Figure 5). In addition, both seem to

benefit from local conditions, such as higher concentrations of DOC, TDN, and lower particulate organic matter as well as primary productivity (Chl *a* as indicator; Figure 5 and Additional file 1: Table S1). Interestingly, the *Euryarchaeota* phylotypes fell mainly within the marine group II, which belongs to *Thermoplasmata* order. This order is mostly composed of thermophilic archaea (Huber and Stetter, 2006). This suggest the existence of psychrophilic or psychrotolerant microorganisms (much lower than 15 °C; Huber and Stetter, 2006), that are phylogenetically related with *Thermoplasmata*. Although, they are somewhat negatively correlated with temperature (Figure 5). Opposite to these trends, *Verrucomicrobia* (mostly *Haloferula* ribotypes – 5.3%, YP_M) and *Actinobacteria* (mainly *Rhodococcus* – 13.5%, YP_M) have a maximum at YP_M sample which is strong positively correlated with oxygen and fluorescence levels (a proxy to chlorophyll *a* abundance; Additional file 2 and Figure 5).

3.1.5. Composition and Distribution of Picoeukaryotes within the N-ICE2015 Collection

Picoeukaryotes are recognized to play an important ecological role in oligotrophic marine waters. As the Arctic Ocean had become nutrient depleted in early summer, it's expected a reaction by these pico-plankton fraction, that can account for most of photosynthetic biomass (as reported by Metfies *et al.* (2016)). The dominant picoeukaryote within the eukaryotic N-ICE2015 collection is a phylotype affiliated with *Chromulinales* order (“Otu00003”, 30.8%), assigned as *Spumella* (and *S. elongata* with PR²; see Figure 3b, Additional file 4 and 5). However, *Spumella* spp. are mainly from freshwater and soil, and so this phylotype should be better designated *Spumella*-like (Grossmann *et al.*, 2016). *Spumella*-like ssp. are phagotrophic nanoflagellates, mostly bacterivorous (Grossmann *et al.*, 2016). Therefore, the suddenly appearance and dominance of this chrysophyte at 20 m depth in YP has to be related with substrate availability. Still, since there is not a prey-specific relationship it is impossible to point out which are the main bacterial groups on which it feeds. Contrasting with this previous quite evident seasonal pattern, the picoeukaryotic green alga *Micromonas* (comprising mostly the *Mamiellales*) is spread across all dataset (Figure 2b). One previous study (McKie-Krisberg and Sanders, 2014) reported bacterivory for an Arctic *Micromonas* strain, *M. pusilla* CCMP20299, found in western Canadian Arctic (Lovejoy *et al.*, 2007;

Comeau *et al.*, 2011) and south of Svalbard (Foulon *et al.*, 2008). The presence of *Micromonas* across the winter (snow-covered sea ice) towards summer (lighter period) in our protist N-ICE2015 collection reinforces the mixotrophic lifestyle mentioned before, although the distribution at OTU level showed a shift between the different *Micromonas* OTUs: whereas “Otu00019” has greater abundances in the two first sampling sites – NB and TR – and considerably less abundant in YP; the “Otu00032” showed precisely the opposite trend (Additional file 4). This result suggests the presence of distinct ecotypes of *Micromonas* in the Arctic Ocean. However, with the assignment against PR² the OTU differences are not so clear (which should be the result of the use of different clustering algorithms, OptiClust (with the SILVA database) vs. VSEARCH (PR²); see “Methods” section), the first two OTUs assigned to the green alga were: “Otu0017”, assigned as “*Micromonas* clade B arctic”, and “Otu0027”, as “*Micromonas* clade A, ABC, 1-2” (Additional file 5). A third OTU (“Otu0092”) assigned to “*Micromonas* pusilla clade C.D.5” was also found (Additional file 5).

3.2. Seasonal and Depth Dependent Trends in Arctic's Microbial Diversity

Understanding how microbial diversity shifts during the winter-spring transition is important to get insight about how it is being shaped by the changing sea-ice regime. Interestingly, the two north most sampling sites, that were sampled under the pack ice (Figure 1 and Table 1), harbour higher microbial diversity (Faith's Phylogenetic Diversity) than the waters sampled, close to the summer, under the YP (Figure 6a and b). A deeper look at OTU level, shows an opposite pattern between prokaryotic and protistan communities along the water column. While the deepest protistan communities sampled showed less number of OTUs than the shallower communities (Additional file 1: Table S3); the prokaryotic samples displayed the opposite trend, with the exception of YP' samples that harbour higher protistan richness (at OTU level) in surface waters (until 20 m depth; Additional file 1: Table S2). Influencing positively the higher OTU richness found on prokaryotic libraries collected on MAW – NB_B and TR_B – could be the higher temperatures registered in these waters (average temperatures of MAW $\approx$ 1.76 °C compared to above PSW = -1.84 °C). This result is in agreement to previous studies on latitudinal diversity gradients controlling marine bacterial communities (Fuhrman *et al.*, 2008; Ghiglione *et al.*, 2012). Although, the conclusions of these works points to opposite

directions. Whereas Fuhrman *et al.* (2008) found a strong positive correlation between temperature and richness; Ghiglione *et al.* (2012) not, pointing instead to the distinct sequence evolution rates between surface and deep pelagic compartments.

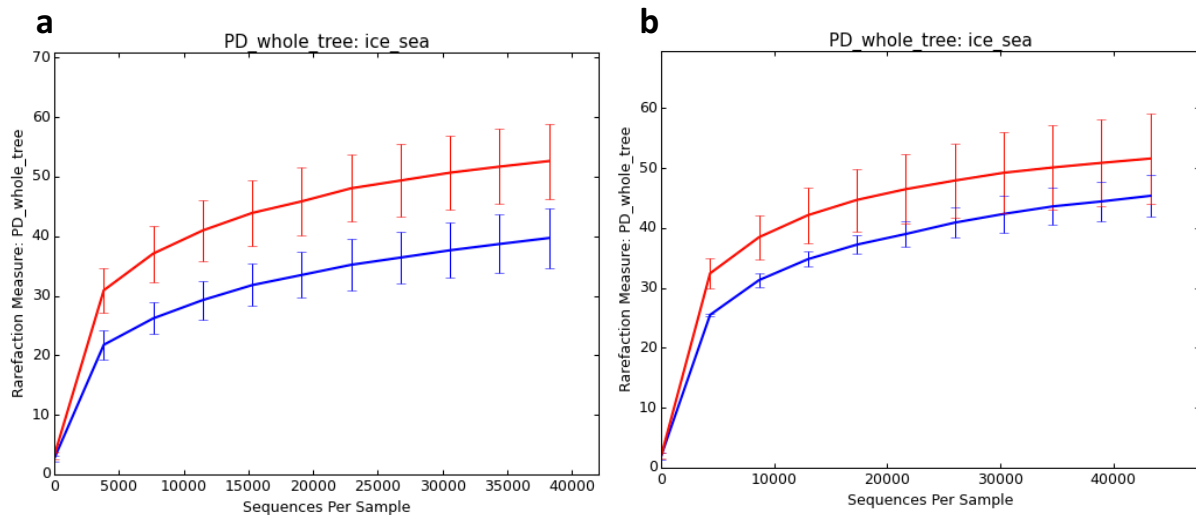


Figure 6 - Faith's Phylogenetic Diversity metric for sample groups snow-covered sea ice (red line, all samples collected in NB and TR) and se-ice without snow (blue line, all samples collected in YP). Error bars show the standard deviation of each sample set (see Table 1). **a** Prokaryotic samples. **b** Eukaryotic samples.

The eukaryotic pattern described here is precisely the opposite observed for two stations (sampled during May 2010) under the Yermak Plateau by one of the few reports conducted north of Svalbard (Meshram *et al.*, 2017). Despite the strategy applied be close enough to allow comparisons, we cannot rule out time and/or space gaps as well as differences with experimental procedures. Unfortunately, since the aim of studying microbial eukaryotes is often related with primary production, *i.e.*, photosynthetic protists, the target water layer is the sunlit ocean, which keep the deeper communities overlooked. Therefore, the lack of depth-dependent studies upon eukaryotic diversity makes difficult to compare our results with others. Nevertheless, Xu *et al.* (2017) described the same trend as herein for eukaryotic communities from the South China Sea screened during the summertime. A plausible explanation for the inverse trend observed of protist-prokaryotic OTU richness can be related with top-down effects. OTUs affiliated to well-known bacterivorous taxa, such as *Picomonadidae* (family) and *Micromonas* (genus), have a higher representativeness on epipelagic waters than mesopelagic ones

(Figure 3b), where it was found greater prokaryotic richness (with the exception of YP). Therefore, grazing on bacteria by picoeukaryotes could reduce the prokaryotic richness in surface Arctic Ocean under winter pack-ice.

The N-ICE2015' microbial plankton was collected along drift ice during the winter to spring transition north of Svalbard which corresponds to the gradual snow melt and an increase in the light availability below the ice (see PAR values in the Additional file 1: Table S1; PAR values were below detection limit for the NB site). Therefore, our study allows us to question if the epipelagic communities are more similar to the mesopelagic ones in the dark period – snow-covered sea ice - than in the lighter period – sea-ice without snow. Unexpectedly, the PCoA, based on unweighted UniFrac metric, showed that samples collected at YP (lighter) are more similar between them than the epipelagic communities under the pack ice, *i.e.*, NB and TR, with the respective mesopelagic ones (Figure 7a). This tendency is even more evident for the eukaryotic communities (Figure 7b). Additionally, the samples from the two north most sampling sites (NB and TR) formed one tight cluster of protistan communities; whereas the prokaryotic plankton is more transient between sites. Nevertheless, the epipelagic microbial communities are always more similar than the mesopelagic ones for the snow-covered sea ice period (Figure 7a and b). Therefore, this observation should 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.* (2017)) during the snow-covered sea ice period, which guarantees a homogenous environment to the upper (epipelagic) microbial plankton inhabiting the PSW rather than the imposition of a dark phase. Opposite, YP was sampled close to the summer, which coincides with the sea-ice without snow, and therefore, the lighter period. The thinner ice (in advanced melting state) imposes a much shallower mixed layer depth (≈ 5 m) (Meyer *et al.*, 2017) and, therefore, we expected a sharper shift among the communities collected. Contrary to anticipated, the most distant communities, *i.e.*, from the surface and from mesopelagic depths, are closer between them than with the middle one (Figure 7a and b). This result together with the composition of microbial communities described above as well as literature (Assmy *et al.*, 2017), led us to hypothesize the occurrence of a vertical mixing event upon the YP waters before the sampling took place.

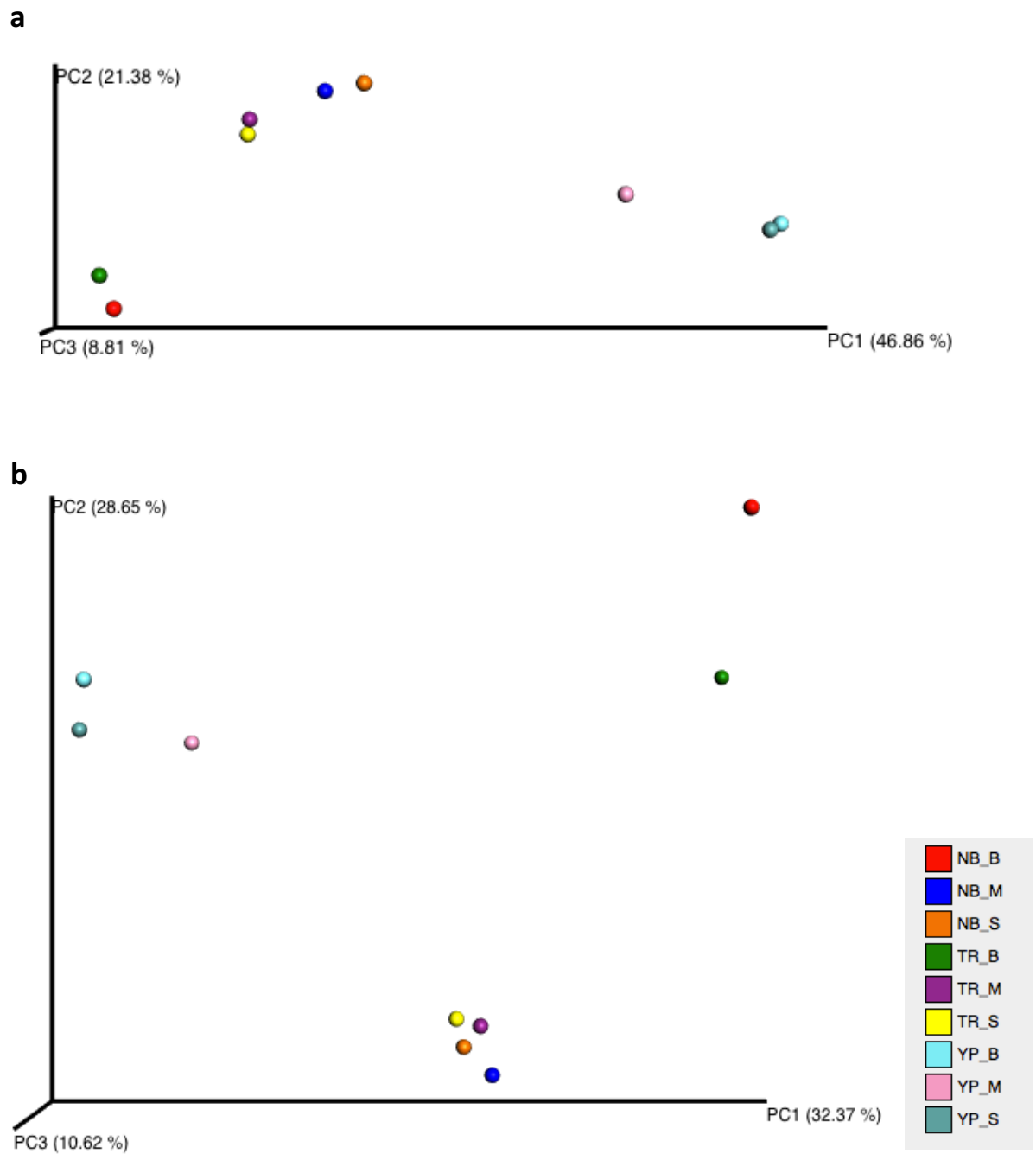


Figure 7 - PCoA of unweighted UniFrac distances across all samples. **a** Prokaryotic samples. **b** Eukaryotic samples. Samples retrieved from the surface (S), middle (M) and bottom (B) seawater at Nansen Basin (NB), Transition Region (TR) and Yermak Plateau (YP) (see Table 1). The coordinates are scaled by the percent explained.

Ultimately, the congruent correspondence pattern regarding the clustering of prokaryotic communities with the protistan ones, is according to previous works (Galand *et al.*, 2010; Ghiglione *et al.*, 2012; Metfies *et al.*, 2016), linking the biogeographical pattern of microbial communities to distinct water masses.

4. Conclusions

The microbial N-ICE2015 collection offers a valuable resource to study the diversity and structure from the relatively underexplored pelagic microbial communities north of Svalbard. Unexpectedly, the winter snow-covered sea ice waters harbour higher diversity than the lighter waters sampled during spring. Much of this diversity fell within *Alpha*- (30.6%) and *Gammaproteobacteria* (28.6%) classes, which are not just the most abundant across the prokaryotic N-ICE2015 collection but also the most phylogenetically diverse. Interestingly, diversity shows opposite patterns along depth for prokaryotic and protistan communities. Seasonal trends are quite evident with thaumarchaeotes abundant under sea-ice waters in springtime and nearly absent towards summer; and the emergence of *Flavobacteria* and SAR92 clade close to summer, probably associated to the degradation of an early spring bloom of *Phaeocystis*. Unexpectedly, the high relative abundance and representativeness of several hydrocarbonoclastic bacteria (phylotypes) expected to be rare members of Arctic's microbial communities, supports evidence of an Arctic's rare biosphere harbouring several gammaproteobacteria that are prone to degrade petroleum-derived hydrocarbons probably influenced by natural hydrocarbon-rich seeps.

2nd Chapter: N-Cycling Microbial Communities and Pathways in the Arctic Ocean

1. Background

1.1. Primary Production in the Arctic Ocean: past, present and future

Phytoplankton, unicellular algae and cyanobacteria, drives roughly half of photosynthetic carbon dioxide (CO₂) fixation of the world's ocean (Field *et al.*, 1998; Kirchman, 2008). Intrinsically dependent on light and nutrients supply (Sigman and Hain, 2012), the ocean productivity is dictated by environmental variables and the fitness of photosynthetic microbes to cope with them. Since light penetration is limited to the first tens of meters' depth (≈ 100 m) - euphotic zone -, the primary production relies on nutrient inputs (known as 'new production') and/or remineralization of existing organic matter ('regenerated production') within the sunlit ocean (Sigman and Hain, 2012). As virtually everything denser than seawater that falls into the ocean will sink into the seabed, photosynthetic biomass production relies on the constant nutrient supply, such as nitrogen (N), phosphorous, iron, and silicate (Moore *et al.*, 2013).

Since the end of the last century, it has been reported an increase of 30% of annual net primary production in the Arctic Ocean (Arrigo *et al.*, 2015). As the Arctic sea-ice shrinks (Serreze *et al.*, 2007), primary production increases in open water (Arrigo *et al.*, 2008; Arrigo *et al.*, 2011; Arrigo *et al.*, 2015). This is due to a combined effect of less and thinner ice and delayed freeze-up (Arrigo *et al.*, 2008; Arrigo *et al.*, 2011; Arrigo *et al.*, 2015). Phytoplankton blooms are not restricted to the broader open ocean areas (due to sea-ice retreat), but also have been reported beneath ponded sea-ice (Arrigo *et al.*, 2012; Arrigo *et al.*, 2014a) or even underneath snow-cover pack ice (Assmy *et al.*, 2017), as a consequence of the Arctic's thinner ice regime. These evidences suggest that this upward trend will continue; however, as Arrigo *et al.* (2011) pointed out, their sustainability will come up against the ability of the Arctic Ocean to provide and replenish the inventory of nutrients.

1.2. Inputs of Nitrogen to the Arctic Ocean: allochthonous and autochthonous sources

Nitrate, as the first inorganic nutrient being depleted during Arctic's phytoplankton blooms (Taylor *et al.*, 2013; Assmy *et al.*, 2017), plays a pivotal role on polar surface waters co-limiting (together with light) the primary production of the Arctic

Ocean (Harrison and Cota, 1991; Tremblay and Gagnon, 2009). Nitrate concentrations in Arctic surface waters vary seasonally and spatially (Codispoti *et al.*, 2013). In nutrient-rich coastal and shelf areas the nutrient provision, inclusive nitrate, is greatly influenced by allochthonous sources, such as river run-off (Dittmar and Kattner, 2003) and mostly by Atlantic and Pacific waters inflows through Fram Strait/Barents Sea and Bering Strait, respectively (Wheeler *et al.*, 1997; Codispoti *et al.*, 2013). In the oligotrophic open ocean, *i.e.*, in the upper layers of the deep basins of the Arctic Ocean, the influence exerted by these two allochthonous sources is much lesser. The nitrate surface inventory is maximum in late winter/early spring, with approximately half of the ice-covered continental shelves with $>10 \mu\text{M}$ (Arrigo *et al.* (2012) and references herein). Whereas, the minimum nitrate records ($\sim 0 \mu\text{M}$) are reached in summer, in the Arctic Basin (Codispoti *et al.*, 2013). Nevertheless, the concentrations during summer under the Eurasian Basin can be as high as $\sim 5 \mu\text{M}$ in the upper 20 m, due to Atlantic water inflow (Codispoti *et al.*, 2013).

Besides the allochthonous origins mentioned before, local events of wind mixing, upwelling, eddies and diffusion can make nitrate available in surface waters from deeper nutrient-rich waters (Voss *et al.* (2013), Codispoti *et al.* (2013) and references herein). This new input of nitrate will contribute to what is referred as new production (Codispoti *et al.* (2013) and references herein). In addition, the recycling of organic matter into ammonium and/or urea provides an additional inorganic and organic N sources, respectively, which will fuel phytoplankton production (Codispoti *et al.* (2013) and references herein). Urea uptake by Arctic phytoplankton may be as high as ammonium (even though usually less than nitrate; Fouilland *et al.*, 2007). This reflects the diversity of N assimilation pathways found in flagellates, that are able to grow with both N sources, nitrate and urea (Terrado *et al.*, 2015). These latter works (Fouilland *et al.*, 2007; Terrado *et al.*, 2015) highlight the advantages of Arctic phytoplankton using urea in a nitrate-limiting ecosystem as the Arctic Ocean, and over heterotrophic bacteria that out compete phytoplankton in nitrate uptake.

1.3. N-Biogeochemistry in the World's Oceans: communities and pathways

As the N-biogeochemical cycle is microbial ruled, the ocean fertility is also intrinsically dependent of biologically driven processes (Voss *et al.*, 2013; Fowler *et al.*, 2013). Depending on prevailing environmental conditions and metabolic potential,

microbes can transform nitrogen in different chemical forms that can lead to the gain and/or loss of N that can be assimilated by the ecosystem. Therefore, the balance of N budget in the Arctic Ocean is influenced by complex microbial interactions, which bear the nitrogen fixation, assimilation, nitrification, anammox and denitrification pathways (Voss *et al.*, 2013; Fowler *et al.*, 2013).

The activity of nitrifiers and dinitrogen (N₂) fixing prokaryotes are important sources to the availability of the limiting nutrient nitrogen in open ocean waters (Capone *et al.*, 1997; Zehr *et al.*, 2001; Montoya *et al.*, 2004; Wuchter *et al.*, 2006; Yool *et al.*, 2007). For a long time, N₂-fixation in the oligotrophic open ocean was thought to be driven primarily by *Trichodesmium* (Capone *et al.*, 1997), a filamentous diazotrophic cyanobacteria, and by a symbiotic (of several diatoms) heterocystous cyanobacterium, from the genus *Richelia* (Carpenter *et al.*, 1999). This paradigm kept untouched until the discover of two widely distributed unicellular cyanobacteria, “*Candidatus Atelocyanobacterium thalassa*” (UCYN-A) and *Crocosphaera watsonii* (UCYN- B) (Zehr *et al.*, 2001; Montoya *et al.*, 2004; Moisander *et al.*, 2010). Whereas the latter has a culture representative; the photoheterotrophic cyanobacterium, “*Ca. A. thalassa*” (UCYN-A), that lives symbiotically with the *Braarudosphaera bigelowii* prymnesiophyte, does not have any representative in culture (Zehr *et al.*, 2001; Montoya *et al.*, 2004; Moisander *et al.*, 2010; and references herein). It is now becoming evident that marine heterotrophic diazotrophs other rather than cyanobacteria have been overlooked (Farnelid *et al.*, 2001; Bombar *et al.*, 2016; and references herein).

Apart from biological N fixation, the nitrate made available in the surface ocean by nitrifiers, through a two-step oxidation reactions called nitrification, can account for half of the nitrate consumed by phytoplankton at the global scale (Yool *et al.*, 2007). Since nitrification is a two-step reaction performed in tandem by two functional groups of prokaryotes, the aerobic ammonia oxidizers and nitrite oxidizers, it demands the co-occurrence of both in order to fully oxidize ammonia to nitrate. The exception to this canonical process is the complete ammonia oxidation - comammox - to nitrate by a single bacteria, enriched from fresh water biofilms, that is closely related to the genus *Nitrospira* (a canonical nitrite oxidizer; reviewed in Daims *et al.* (2016)). The current view is that mesophilic ammonia-oxidizing *Crenarchaeota*, renamed *Thaumarchaeota* (phylum) are the main drivers of aerobic ammonia oxidation in the twilight ocean (disphotic) zones (transition region between euphotic-aphotic zones; Hutchins and Fu, 2017). While the co-

occurrence of nitrite-oxidizing bacteria, such as *Nitrospina* and *Nitrospira*, near disphotic zones, completes the nitrification process, yielding nitrate to the upper layers of the ocean (Hutchins and Fu, 2017).

Potential nitrogen losses from marine environments are mainly recognized to microorganisms that carry out anammox or denitrification pathways (Ulloa *et al.*, 2012; Voss *et al.*, 2013; Hutchins and Fu, 2017). Nevertheless, these losses are quite distinct in quantitative and qualitative terms. Whereas, denitrification, the dissimilatory reduction of nitrate to N₂ gas, by fungi and several non-related prokaryotes (polyphyletic) was estimated to represent a removal of 71% of fixed N; the anaerobic ammonia oxidation - anammox -, which together with nitrite, yield N₂, carry out by some planctomycetes, (bacterial phylum *Planctomycetes*) is less than half, *i.e.*, 29% (Voss *et al.*, 2013; Hutchins and Fu, 2017). Both processes are negatively correlated with oxygen concentrations, reason why, zones depleted in oxygen, such as marine sediments and Oxygen Minimum Zones (OMZs) are the most probable places to encounter them (Ulloa *et al.*, 2012; Voss *et al.*, 2013; Hutchins and Fu, 2017).

The balance of N budget in the oceans is dependent on the stability of factors that control the above-mentioned N-biogeochemical processes. As the ocean warms, OMZs expand (Keeling *et al.*, 2010), which presumably will increase the rates of anammox and denitrification. In addition to these losses, the acidification of the oceans in the next decades is predicted to decrease the nitrification rates by 3-44% (Beman *et al.*, 2011). Therefore, there is a continued need to assess the microbial communities that drives the N-biogeochemical cycle in order to identify the effects of current environmental changes.

1.4. N-Biogeochemistry in the Arctic Ocean: communities and pathways

The Arctic Ocean has experienced rapid changes with increased primary production that possibly can be reflected upon microbial communities that have an important influence on the N-biogeochemical cycle. Since the Arctic is a highly-oxygenated ocean, anammox and denitrification are mainly restricted to marine sediments, with denitrification accounting roughly for two thirds of the N lost (Rysgaard *et al.*, 2004; Gihring *et al.*, 2010), in shelf sediments from Greenland (Rysgaard *et al.*, 2004) and Svalbard (Gihring *et al.*, 2010). However, the latter study, from Gihring *et al.*

(2010), which addressed the same N-pathways discussed herein on two Arctic fjord sediments, found that the nitrate gained through nitrification doubles that lost through denitrification, with a significant fraction moving upward into the water column. In addition, other work points the abundance of nitrate found in the subglacial meltwaters of a high Arctic glacier (in Svalbard) partially due to the nitrifying communities that were using as substrate the snowpack ammonium and mineralized organic nitrogen (Wynn *et al.*, 2007). Nitrification is expected to have a significant contribution to the nitrate surface inventory in the Arctic Ocean. However, a transect from the equatorial Pacific to the Arctic Ocean showed decreased contribution of nitrification to the nitrate assimilation over the euphotic zone, with an average 55.6% for the subtropical zone and a minimum of 0-4.74% for the three north most zones - Bering and Chukchi sea shelves (Shiozaki *et al.*, 2016). Nonetheless, as this study shows and others focusing on the same subject (*e.g.*, Kalanetra *et al.*, 2009), ammonia oxidizers are negatively correlated with light, explaining the insignificant contribution of nitrification over the shallow stations (bottom ≤ 67 m) sampled in the Canadian Arctic during summer (Shiozaki *et al.*, 2016). Moreover, ammonia oxidation activity in the same Arctic sector was reported to be higher in wintertime during a two-year period (Christman *et al.*, 2011).

Thaumarchaeal ammonia oxidizers (TAO) are particularly abundant during winter in the subsurface of polar oceans, where they are the most dominant phylotype recovered, reaching around 20-30% of the overall prokaryotic community (DeLong *et al.*, 1994; Alonso-Sáez *et al.*, 2008; Kalanetra *et al.*, 2009; Grzymiski *et al.*, 2012; Alonso-Sáez *et al.*, 2012). This huge fraction mirrors the dominance of their chemoautotrophic metabolism during the dark period (wintertime) in the Southern Ocean (Grzymiski *et al.*, 2012). Unexpectedly, one study conducted on polar thaumarchaeotes found that neither the uptake of bicarbonate (indicative of chemoautotrophism) or leucine (heterotrophism) occurred in significant portions, suggesting alternative metabolic pathways concerning C sources (Alonso-Sáez *et al.*, 2012). In addition, high copy numbers of *amoA* and *ureC* genes were found through a metagenomics survey, implying an active coupling of ammonia-oxidation and ureolysis pathways (Alonso-Sáez *et al.*, 2012). Therefore, ureolysis, which yields ammonia and CO₂, seems to play an essential role providing the energy - ammonia - and carbon source - CO₂ - to polar TAO (Alonso-Sáez *et al.*, 2012; Connelly *et al.*, 2014) that helps to support the greater nitrifying activity observed in wintertime (Christman *et al.*, 2011; Tolar *et al.* 2016b). Despite that ammonia-oxidizing

bacteria had been retrieved from the Arctic Ocean, their presence should be minor and attributed to the influence of coastal and shelf waters, due to their requirement of high ammonium levels (Bano and Hollibaugh, 2000; Hollibaugh *et al.*, 2002; Christman *et al.*, 2011).

Contrasting to this wealth of studies, much less is known about the N₂-fixing communities as well as their contribution to the N fixed in the Arctic Ocean. Despite, the high richness of the related nitrogen-fixing gene *nifH* found in the Central Arctic Ocean, few belong to cyanobacteria (Fernández-Méndez *et al.*, 2016), that are nearly absent in this region. Instead, anaerobic diazotrophs, from cluster III, seem to thrive through this highly cold and oxygenated waters (Farnelid *et al.*, 2001; Blais *et al.*, 2012; Fernández-Méndez *et al.*, 2016). Also, N₂-fixation rates in the Canadian Arctic were highest under the influence of the Mackenzie River plume (Blais *et al.*, 2012). These observations points to a minor contribution of N₂-fixation to the Arctic Ocean.

1.5. Objectives

Sea ice covers 4.1-6.1% of the world's ocean surface (Arrigo, 2014b). Still, the N-biogeochemistry associated to it and to the pelagic compartments underneath is largely unknown. Particularly, the metabolic potential of microbial communities that influence the N cycle remains undersampled in the Arctic Ocean. The only metagenomic study conducted so far in the Arctic Ocean by Alonso-Sáez *et al.* (2012) helped to elucidate how polar TAO are so successful in a N-limiting ocean like the Arctic, and understand the seasonal variation among Arctic prasinophytes (see Joli *et al.*, 2017).

The first chapter of this thesis is a starting discussion about the pelagic microbiota collected during an Arctic expedition in the pack ice, during the winter to spring transition, north of Svalbard and it is primarily a microbiota census data. Here, in the second chapter, the focus will be the prokaryotic communities with a putative role on the N-biogeochemical cycle based on 16S rDNA next-generation sequencing (V4-V5 regions) data. Additionally, the functional potential of these communities will be addressed interrogating the respective metagenomes with related N-biogeochemical pathways.

2. Methods

2.1. Sampling Sites, Water Column Sampling, DNA Extraction, PCR, Library Preparation and Sequencing of 16S rDNA amplicon

A detailed description of the sampling sites, water column sampling, DNA extraction, PCR, library preparation and sequencing of 16S rDNA amplicon is given in the “Methods” section, in the first chapter of this thesis. Specifically, the sampling stations are described in “Sampling Sites”; sampling procedures in “Water Column Sampling: environmental data and microbial collection”; and information about the DNA extraction, PCR, library preparation and sequencing of 16S rDNA amplicon is given in “DNA Extraction, PCR, Library Preparation and Sequencing of SSU rDNA amplicon”.

2.2. Library Preparation and Sequencing of Metagenomes

The source of environmental DNA (eDNA) used for the library preparation and sequencing of the nine 16S rDNA amplicon samples was the same used for the library preparation and sequencing of nine metagenomes described below.

The preparation of the nine libraries for shotgun metagenomics sequencing followed the following steps. First, the quality of eDNA was checked in an agarose gel. Then, eDNA was sheared in small pieces (< 600 bp) using a Covaris sonicator and the fragments of selected size purified using AMPure XP beads (Agencourt). About 200 ng per sample of fragmented DNA was picked to construct Illumina libraries using the Ovation Rapid DR Multiplex System 1-96 (NuGEN). 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) company.

2.3. EBI Metagenomics Pipeline: upstream analysis of metagenomic reads

The forward and reverse raw fastq files ($n=18$) were submitted and archived to the European Nucleotide Archive (ENA) under the study accession PRJEB15043 - *Shotgun Metagenomic Sequencing of the Arctic Ocean during Winter-Spring Transition*

(made private until 30th April 2018) in order to run the Illumina metagenomic reads in the EBI Metagenomics (EMG) pipeline (v.3.0; <https://www.ebi.ac.uk/metagenomics/>; Mitchell *et al.*, 2015). 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.* (2015). First, Illumina forward reads were merged with overlapping reverse ones with SeqPrep (v.1.1; <https://github.com/jstjohn/SeqPrep>). The low quality paired-end reads were trimmed at ends and sequences with > 10% undetermined nucleotides excluded using Trimmomatic (v.0.35; Bolger *et al.*, 2014). Additionally, sequences < 100 bp in length were discarded using biopython (v.1.65; <http://biopython.org>) tools. A detailed description of the number of reads filtered during the upstream analysis by the EMG pipeline as well as the average read length and GC-content (for a subsampling size $\approx 2\,000$ k reads) for the good-quality reads annotated, is given in Additional file 7. Then, the good-quality paired-end reads undergone an analysis using profile hidden Markov models in order to identify and mask non-coding RNAs (ncRNAs) through HMMER (v.3.1b1; <http://hmmer.org>).

2.4. EBI Metagenomics Pipeline: downstream analysis of metagenomic reads

During the identification of ncRNAs, there is a bifurcation in the EMG pipeline in order to assign taxonomy to the 16S rDNA Operational Taxonomic Units (OTUs), and to annotate the predicted coding DNA sequences (pCDS). At taxonomic level, the ncRNAs genes, *i.e.* rRNA, inclusive tRNA, are identified and extracted from metagenomic datasets using the HMMER (v.3.1b1; <http://hmmer.org>). With QIIME (v.1.9.1; Caporaso *et al.*, 2010) the 16S rRNA genes selected are clustered into OTUs using the closed-reference OTU picking strategy. Then, OTUs are classified against the Greengenes reference database (v.13.8; DeSantis *et al.*, 2006). At functional level, the pCDS > 60 bp in length are predicted with FragGeneScan (v.1.20; Rho *et al.*, 2010), a tool that makes use of hidden Markov model to detect partial/fragmented genes in short reads based on codon usage bias, sequencing error models and start/stop codon patterns. Finally, the pCDS are annotated against InterPro signature database (release 58.0; Finn *et al.*, 2017) using protein signature recognition methods with InterProScan (v.5.19-58.0). Basically, the application queries the pCDS (*i.e.*, translated pCDS) that through predictive models matches against one InterPro entry, *i.e.*, one protein signature, that makes part of the InterPro consortium (a sub set of member databases that includes Pfam, TIGRFAM,

PRINTS, PROSITE patterns, Gene3d). Therefore, our pCDS (*i.e.*, translated pCDS) are classified into protein families and conserved domains and sites predicted. Concerning the functional analysis, in addition to InterPro (IPR) matches, *i.e.*, a table with IPRs abundance across samples (see Additional file 6), the Gene Ontology (GO) terms derived from IPRs it is provided at the same time as well as a reduced list of GO terms designated GO slim. Regarding the taxonomic analysis, the EMG pipeline provides a taxonomic table classified at phylum and species level given in Additional file 8. The IPR (general), GO and GO-slim tables are given in Additional file 6.

2.5. ORCA platform: Spearman correlations between IPRs related to N-biogeochemical pathways and environmental variables

Since this chapter focus on the prokaryotic communities that have the genomic potential to carry out pathways involved in the N-cycle; the IPR table ('IPR table (general)', Additional file 6) was interrogated for those IPRs that are potentially related with N pathways. Such search was made using the IPRs (InterPro accession number), their related genes, enzymes (family, domain) and pathways highlighted in Table 2. The result was the table designated 'N cycle related IPR (specific)' which is given in Additional file 6. The IPR table ('IPR table (general)', Additional file 6) coupled with environmental data (Additional file 1: Table S1) were imported to the ORCA platform (Leite, 2016). Then, the IPR table was automatically normalized with the DeSEQ package (Anders and Huber, 2010), and filtered for IPRs of interest (those within the 'N cycle related IPR (specific)', Additional file 6). The ORCA platform is a R-based scripts workbench with a graphical user interface that allows to explore meaningful biological interrelationships from high-throughput sequencing data (Leite, 2016). Here, the ORCA was used to study strong Spearman correlations among the IPRs of interest, *i.e.*, those present in the N-ICE2015 metagenomic dataset with a putative role on N-cycle ('N cycle related IPR (specific)', Additional file 6), and between these and environmental variables. In order to get the results presented herein, *i.e.*, heatmaps and Spearman correlations, the ORCA platform made used of Hmisc, corrplot and ggplot2 packages in R (R Development Core Team, 2008; Wickham, 2009; Wei, 2013; Harrell and Harrell, 2015).

Table 2 - Nitrogen-biogeochemical cycle: pathways, genes, enzymes, and, the respective InterPro accession number (IPRs).

N-Pathway	Gene	Enzyme (family, domain)	IPRs
aerobic ammonia oxidation to nitrite (nitrification)	bacterial <i>amoA</i>	ammonia monooxygenase/particulate methane monooxygenase, subunit A	IPR003393
	archaeal <i>amoA</i>	ammonia monooxygenase, subunit A, archaeal	IPR024656
	<i>amoB</i>	ammonia monooxygenase/particulate methane monooxygenase, subunit B	IPR006833
	<i>amoC</i>	ammonia monooxygenase/particulate methane monooxygenase, subunit C	IPR006980
	<i>hao</i>	hydroxylamine oxidase	IPR012138
nitrite oxidation to nitrate (nitrification)	<i>nxrA</i>	nitrite oxidoreductase alpha subunit	-
	<i>nxrB</i>	nitrite oxidoreductase, beta subunit	-
nitrate reduction to nitrite (assimilatory or dissimilatory metabolism)	<i>narI</i>	nitrate reductase, gamma subunit	IPR003816
	<i>narH</i>	nitrate reductase, beta subunit	IPR006547
	<i>narG</i>	nitrate reductase, alpha subunit	IPR006468
	<i>napA</i>	periplasmic nitrate reductase, large subunit	IPR010051
	<i>napB</i>	nitrate reductase cytochrome c-type subunit NapB	IPR005591
	<i>napC</i>	periplasmic nitrate reductase c-type cytochrome, NapC/NirT	IPR011885
	<i>napC</i>	denitrification system component NapC/NirT/NrfH	IPR024717
	<i>napH</i>	ferredoxin-type protein, NapH/MauN family	IPR011886
	<i>nasA</i>	assimilatory nitrate reductase (NADH) alpha subunit apoprotein	-
	<i>nasB</i>	assimilatory nitrate reductase electron transfer subunit	-
nitrite reduction to ammonium (assimilation)	<i>nrfA</i>	cytochrome c552	IPR003321
	<i>nrfA</i>	formate-dependent cytochrome c nitrite reductase, c552 subunit	IPR017570
	<i>nirB</i>	nitrite reductase [NAD(P)H] large subunit, NirB	IPR012744
	<i>nirB</i>	nitrite reductase [NAD(P)H], large subunit	IPR017121
	<i>NiRs</i>	nitrite/sulfite reductase ferredoxin-like domain	IPR005117
nitrite reduction to nitric oxide	<i>nirK</i>	nitrite reductase, copper-type	IPR001287

	<i>nirS</i>	cytochrome cd1-nitrite reductase-like, haem d1 domain	IPR011048
	<i>nnrS</i>	NnrS	IPR010266
nitric oxide to nitrous oxide (incomplete denitrification)	<i>norB</i>	nitric oxide reductase subunit B	-
	<i>norC</i>	nitric oxide reductase subunit C	-
nitrous oxide reduction to dinitrogen (denitrification)	<i>nosZ</i>	nitrous-oxide reductase	IPR023644
	<i>nosZ</i>	nitrous-oxide reductase, Sec-dependent	IPR026468
	<i>nosD</i>	nitrous oxide reductase family maturation protein NosD	IPR026464
	<i>nosL</i>	nitrous oxide reductase accessory protein NosL	IPR008719
	<i>nosR</i>	nitrous oxide reductase expression regulator NosR	IPR011399
nitrate transport/nitrite excretion	<i>narK</i>	nitrate transporter	IPR004737
	<i>ntrB</i>	nitrate transport permease	IPR005889
	<i>ntrC/D</i>	nitrate transport ATP-binding subunit C/D	IPR005890
dinitrogen fixation	<i>nifN</i>	nitrogenase molybdenum-iron cofactor biosynthesis protein	IPR005975
	<i>nifH</i>	nitrogenase iron protein NifH	IPR005977
	<i>nifK</i>	nitrogenase molybdenum-iron protein beta chain	IPR005976
	<i>vnf/anfD</i>	nitrogenase alpha chain	IPR005974
	<i>nifE</i>	nitrogenase MoFe cofactor biosynthesis protein NifE	IPR005973
	<i>nifD</i>	nitrogenase molybdenum-iron protein alpha chain	IPR005972
	<i>nifX</i>	nitrogen fixation protein NifX	IPR013480
	<i>nifH/frxC</i>	<i>nifH/frxC</i> family	IPR000392
anammox	<i>hzoB</i>	hydrazine-oxidizing enzyme	-
	<i>hzoA</i>	hydrazine oxidoreductase HzoA	-
ureolysis	<i>ureAB</i>	urease, gamma/gamma-beta subunit	IPR002026
urea transporter	-	urea transporter	IPR004937

3. Results and discussion

In the next paragraphs, two distinct types of high-throughput sequencing data will be discussed: (1) 16S rDNA amplicon (V4-V5 regions), and (2) environmental shotgun metagenomic. (2) offers an opportunity to assess the composition and structure of microbial communities independently of primer affinity and PCR bias, and, then, an alternative to (1); the number of 16S rDNA metagenomic reads is considerably much lower (from a minimum of 2 716 - TR_B - to a maximum of 6 381 - YP_B - reads/sample; Additional file 8) than the 16S rDNA amplicon (V4-V5 regions; >35 000 reads/sample; Additional file 1). In addition, on average, approximately less than one third of these 16S rDNA metagenomics reads are not taxonomically classified (*i.e.*, “unassigned”; Additional file 8), which is, at least partially, the consequence of a lower average read length of 302.32 bp (Additional file 7; the amplicon read length is 360-380 bp, Additional file 1). Therefore, (1) will be used in order to identify potential N-cycling microbial communities within the prokaryotic N-ICE2015 collection; while (2) will support the search for protein families/domains with a putative role on N-biogeochemical pathways in the Arctic Ocean. The composition of Arctic’s pelagic microbial communities during the winter-spring transition and the environmental variables that shape its distribution was discussed in the first chapter of this thesis. The second chapter tries to give a deeper insight into the functional capabilities of these communities in order to elucidate their role in the Arctic N-cycling.

3.1. N-Cycling Microbial Communities and Pathways in the Arctic Ocean

3.1.1. Nitrifying Communities

Nitrifying communities represent between 6.3-27.4% of the prokaryotic libraries from Nansen Basin - NB - and Transition Region - TR -, but are practically absent from Yermak Plateau - YP ($\leq 2.0\%$; Figure 8). Likely inhibited by the increasing intensity of sunlight, as the winter pack ice was melting along the N-ICE2015 cruise as well as the snow, the nitrifying communities let to dominate the epipelagic waters of NB to be more abundant in the mesopelagic sample of TR (TR_B; see the “Results and discussion” section in the first chapter about this matter).

Regarding the two first sampling sites, the distribution of phylotypes affiliated

to thaumarchaeal ammonia oxidizers (TAO) along the water column is highly distinct (Figure 8). Phylotypes assigned to the genus “*Candidatus Nitrosopumilus*” are the dominant TAO in the upper layers (5 and 50 m depth) of NB and TR; while the phylotypes related to the thaumarchaeal Marine Group I (MGI), are the predominant TAO within the mesopelagic waters sampled (at 250 m depth; Figure 8). MGI *Thaumarchaeota* is $\leq 1\%$ across the epipelagic water samples, but increases > 2 -fold than “*Ca. Nitrosopumilus*” at mesopelagic depths, for the NB_B and TR_B samples.

In spite of the great sampling variability, in time and space, the differential distribution of the two distinct TAO genera presented here suggests different niche occupation and specialization. The previous identification of distinct archaeal *amoA* clades in the Arctic Ocean (Pedneault *et al.*, 2014) supports our results. Gene sequences retrieved from shallower waters tended to fall into clade A; whereas the deeper ones fall into clade B (Pedneault *et al.*, 2014). Intriguingly, the same study (Pedneault *et al.*, 2014) found a stoichiometry < 1 between *ureC* gene copy numbers and thaumarchaeal 16S rRNA gene. However, they obtained a close stoichiometry when considered only the TAO populations from the bottom, suggesting that just the most representative clade from the deeper communities have the ability to carry out ureolysis (Pedneault *et al.*, 2014). Actually, *Nitrosopumilus maritimus*, the most abundant thaumarchaeal ribotype in the upper layers of the Arctic Ocean (Galand *et al.*, 2009; Kalanetra *et al.*, 2009), lacks urease encoding genes, like *ureC* (Walker *et al.*, 2010). However, these genes were identified in single amplified genomes of MGI thaumarchaeotes from subtropical gyres (Swan *et al.*, 2014), and expressed in one *Nitrosopelagicus* strain (Carini *et al.*, 2017). Hence, the intricate result found by Pedneault *et al.* (2014) should reflect the metabolic ability of distinct TAO populations in the water column of the Arctic Ocean to carry out ureolysis that in turn will be responsible by niche differentiation.

Despite the majority of polar ammonia-oxidizers be restricted to thaumarchaeotes, it was also found one betaproteobacterial genus, *Nitrosomonas*, which is known to have the ability to oxidize ammonia to nitrite (Figure 8). Although, with insignificant representativeness among epipelagic microbial communities sampled at NB and TR (*i.e.*, $< 0.5\%$ *per sample*), being almost absent from the deeper and YP samples. Presenting just low levels in surface waters that were under snow-covered sea ice at time of collection could suggest the sea-ice as potential seed source; however, *Nitrosomonas* ribotypes were never identified in sea-ice (Hatam *et al.*, 2014; Bowman *et al.*, 2012;

Bowman *et al.*, 2015; Hatam *et al.*, 2016). Yet *Nitrosomonas*- and *Nitrosospira*-like sequences were already detected in polar oceans (Hollibaugh *et al.*, 2002). The former was not detected in the Southern Ocean and their presence in the Arctic Ocean waters was attributed to the influence of waters from eutrophic Chukchi Shelf (Bano and Hollibaugh, 2000; Hollibaugh *et al.*, 2002). Therefore, our results suggested that betaproteobacterial ammonia-oxidizers should play a minor role in the biogeochemical cycle of N in the Arctic Ocean.

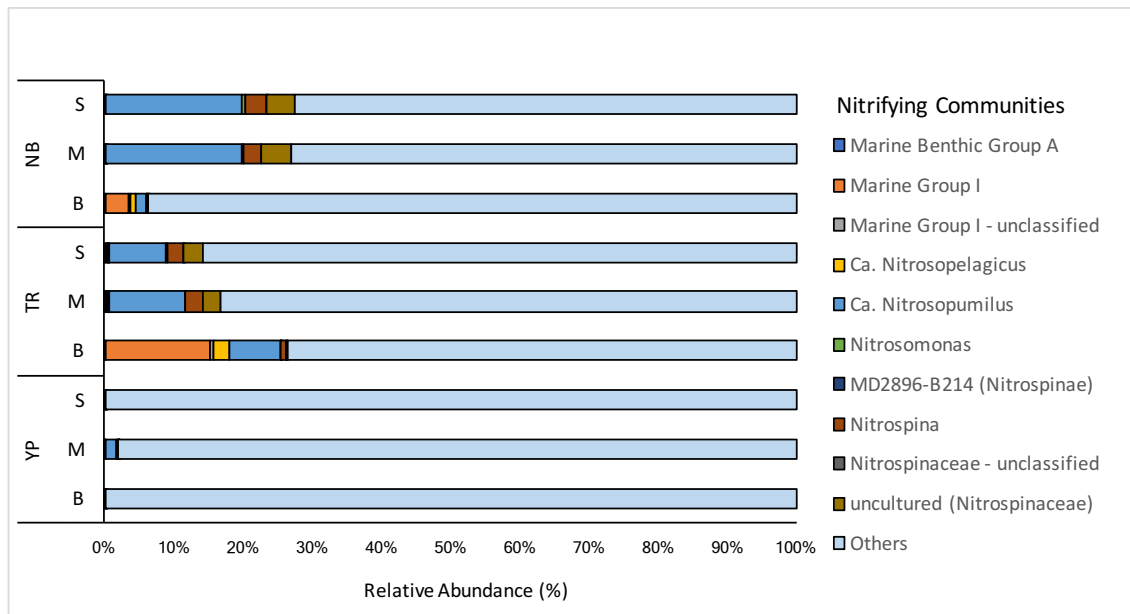


Figure 8 - Distribution of nitrifying communities across the prokaryotic N-ICE2015 collection. The N-ICE2015 collection possess bacterial and archaeal sequences retrieved from the surface (S), middle (M) and bottom (B) seawater at Nansen Basin (NB), Transition Region (TR) and Yermak Plateau (YP) (see Table 1). The taxonomic level given is the lowest given at genus level in the Additional file 2 derived from 16S rDNA amplicon data analysis. Nitrifying communities includes ammonia-oxidizing archaea (*i.e.*, “marine benthic group A”, “marine group I”, “marine group I - unclassified”, “*Ca. Nitrosopelagicus*”, “*Ca. Nitrosopumilus*”), ammonia-oxidizing bacteria (*i.e.*, “*Nitrosomonas*”), and nitrite-oxidizing bacteria genera (*i.e.*, “MD2896-B214 (*Nitrospinae*)”, “*Nitrospina*”, “*Nitrospinaceae* - unclassified”, “uncultured (*Nitrospinaceae*)”). “*Ca.*” stands for *candidatus* genera. From the top of the legend, the first three classifications are representative of environmental sequences. The environmental clade “MD2896-B214” belongs to the phylum *Nitrospinae* as well as the unclassified and uncultured sequences affiliated to the family *Nitrospinaceae*. These nomenclatures were given according to the SILVA reference database (v. 1.2.8). “Others” represents the sum of the frequency of occurrence of non-nitrifying taxa.

Nitrite-oxidizing bacteria (NOB) within the N-ICE2015 collection falls within the *Nitrospinae* phylum, and displays the same pattern of TAO (Figure 8). But unlike TAO, NOB do not increase with depth in NB and TR samples. Less abundant than TAO, NOB range, on average, from 5.9%, in the upper layers of the two north most sampling sites, to < 1%, in the remaining samples (Figure 8). NOB belonging to the phylum *Nitrospinae* are recognized as the most important nitrite oxidizers in marine environments (Daims *et al.*, 2016) and found close to their counterparts, TAO (Hawley *et al.*, 2014; Parada and Fuhrman, 2017). This is the first study reporting such high levels of NOB in the Arctic Ocean, reflecting the importance of sampling the microbial communities under the winter pack ice, where they are expected to play an active role in the nitrite cycling. Curiously, the most dominant NOB taxon was assigned to an uncultured representative within the family *Nitrospinaceae* (with $\approx 1.6\%$ within the prokaryotic N-ICE2015 collection, Additional file 2; Figure 8). The second most abundant belongs to the genus *Nitrospina* ($\approx 1.2\%$ within the prokaryotic N-ICE2015 collection, Additional file 2; Figure 8). Both have similar percentages within the upper 50 m depth of the two north most sampling sites, but are almost absent from the respective deeper samples (at 250 m depth). This observation suggests that nitrification, the full aerobic ammonia oxidation to nitrate, in the Arctic Ocean, probably happens in the subsurface waters underneath of winter pack ice, but not in the upper mesopelagic zones. In agreement to other studies (Church *et al.*, 2010; Newell *et al.*, 2011; Ulloa *et al.*, 2012; Hawley *et al.*, 2014), this result reinforces the need to re-assess the role of putative ammonia oxidizers at mesopelagic depths.

3.1.2. Perspective on N-Cycling Microbial Communities in the Arctic Ocean

As introduced in the “Background” section, the N budget in the Arctic Ocean is influenced by complex microbial interactions, which bear the nitrogen fixation, assimilation, nitrification, anammox and denitrification pathways. The polyphyletic nature of assimilation and denitrification pathways, that are spread among several unrelated lineages of prokaryotes (not only), suggests horizontal gene transfer or gene duplication events (*e.g.*, Jones *et al.*, 2008), making it difficult to identify which members could carry out these pathways. Hence, the study of these two pathways is concerned to the functional related protein families/domains retrieved from the metagenomes. Regarding the distribution of prokaryotic communities among our samples that could bear

the remaining N-biogeochemical pathways, the results observed vary widely. As expected, nitrifying communities are abundant and spread in the N-ICE2015 collection. While, despite the identification of *Planctomycetes* (1.1%; Additional file 2), the phylum that includes all known anammox, the taxa identified does not include the *Scalindua* genus, the anammox bacteria retrieved from marine environments, neither other known anammox genera (Kuenen, 2008). Ultimately, most of the knowledge about the N₂-fixing prokaryotes in the oceans is limited to diazotrophic cyanobacteria, that are near absent in the Arctic as evidenced by the prokaryotic N-ICE2015 collection (0.1%, *Cyanobacteria*; Additional file 2). Likely, in the Arctic Ocean, the N-fixation is performed mostly by anaerobic diazotrophs, from cluster III (Farnelid *et al.*, 2001; Blais *et al.*, 2012; Fernández-Méndez *et al.*, 2016), which are affiliated to sulfate-reducing δ -*Proteobacteria*, *Spirochetes*, *Firmicutes*, and methanogens (Bombar *et al.*, 2016). Methanogens were not found; whereas sulfate-reducing δ -*Proteobacteria* (*Desulfurellaceae* family), *Spirochetes* (aka *Spirochaetae*) and *Firmicutes* represent each $\approx$ 0% within the prokaryotic N-ICE2015 collection.

3.1.3. N-Biogeochemical Pathways

In order to link putative N-cycling microbial communities to the correspondent N-biogeochemical pathways, the nine metagenomes from the nine prokaryotic communities were screened for enzyme families/domains involved in these pathways (highlighted in Table 2). The results presented in Table 3 highlight the presence of at least one of the N-cycle related IPR for all N-pathways screened. Among those IPRs interrogated, enzyme families/domains associated to ammonia oxidation are not just present but also among the most abundant. Particularly, the enzyme ammonia monooxygenase (AMO, encoded by *amoCAB* operon), which performs the ammonia oxidation to hydroxylamine (Table 3). Although, none read matched the hydroxylamine oxidase (HAO) IPR, which carry out the detoxification of hydroxylamine yielding nitrite (Table 3). However, this result needs to be carefully interpreted. There are several reasons to explain the result observed: (1) a bad annotation of HAO-encoding genes; (2) an underrepresentation/absence of HAO-encoding genes in the nine metagenomes; or (3) a fragmentary nature of predicted coding DNA sequences that made impossible to annotate the HAO-encoding genes. Of course, these reasons are applied for all IPRs searched.

Unfortunately, the nitrite oxidoreductase enzyme is not distinguished by any IPR, what makes impossible to search for it. Actually, a InterProScan search using the protein subunit encoded by *norA* - nitrite oxidoreductase alpha subunit - as query, against the InterPro reference database, will match the nitrate reductase alpha subunit, which is associated to denitrifying processes and not nitrifying ones. Although this subunit was not found either among the metagenomic libraries of N-ICE2015 (Table 3).

	NnrS	IPR010266	11	21	42	31	57	41	9	10	19
nitrous oxide reduction to dinitrogen	nitrous-oxide reductase	IPR023644	-	-	-	-	-	-	-	-	-
	nitrous-oxide reductase, Sec-dependent	IPR026468	-	-	-	-	-	-	-	-	-
	nitrous oxide reductase family maturation protein NosD	IPR026464	-	-	-	-	-	-	-	-	-
	nitrous oxide reductase accessory protein NosL	IPR008719	2	2	7	1	0	7	0	0	2
	nitrous oxide reductase expression regulator NosR	IPR011399	-	-	-	-	-	-	-	-	-
nitrate transport/nitrite excretion	nitrate transporter	IPR004737	-	-	-	-	-	-	-	-	-
	nitrate transport permease	IPR005889	0	0	0	1	0	0	0	0	0
	nitrate transport ATP-binding subunit C/D	IPR005890	-	-	-	-	-	-	-	-	-
dinitrogen fixation	nitrogenase molybdenum-iron cofactor biosynthesis protein	IPR005975	-	-	-	-	-	-	-	-	-
	nitrogenase iron protein NifH	IPR005977	-	-	-	-	-	-	-	-	-
	nitrogenase molybdenum-iron protein beta chain	IPR005976	-	-	-	-	-	-	-	-	-
	nitrogenase alpha chain	IPR005974	-	-	-	-	-	-	-	-	-
	nitrogenase MoFe cofactor biosynthesis protein NifE	IPR005973	-	-	-	-	-	-	-	-	-
	nitrogenase molybdenum-iron protein alpha chain	IPR005972	-	-	-	-	-	-	-	-	-
	nitrogen fixation protein NifX	IPR013480	0	0	0	0	0	1	0	0	1
	NifH/frxC family	IPR000392	24	40	15	25	17	13	16	17	26
ureolysis	urease, gamma/gamma-beta subunit	IPR002026	35	55	89	83	74	108	43	60	52
urea transporter	urea transporter	IPR004937	0	4	0	0	2	5	6	2	14

As expected, a greater number of reads annotate as urease gamma/gamma-beta subunit are found at 250 m depth of the two north most sampling sites (NB_B and TR_B samples, Table 3). This observation is according to Alonso-Sáez *et al.* (2012) as well as Pedneault *et al.* (2014), and mirror, partially, the distinct genomic potential of different TAO populations along the water column to carry out ureolysis. This conclusion is also supported by the absence of any correlation between urease and AMO subunits (Figure 9). Of course, other prokaryotes rather than TAO have also the genomic potential to possess urease encoding genes (Solomon *et al.*, 2010) and should contribute with a significant fraction. Nevertheless, both enzyme subunits, urease and AMO, are positively correlated with total dissolved nitrogen, which includes urea, suggesting that both pathways, ureolysis and aerobic ammonia oxidation, are coupled (Figure 10).

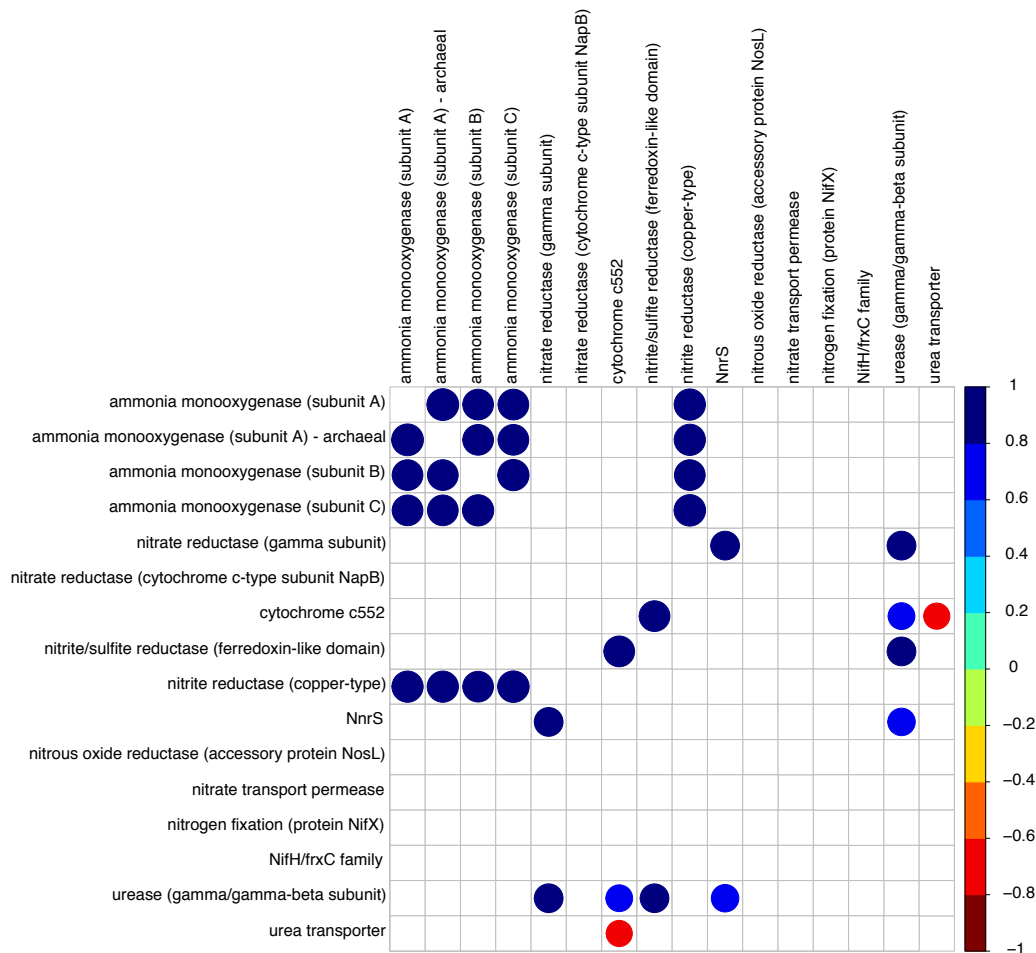


Figure 9 - Heatmap showing the significant Spearman correlations among the enzymes (*i.e.*, family, domain) involved in the different N-biogeochemical pathways along the N-ICE2015 cruise.

Few metagenomic reads matched the nitrate reductase gamma subunit and cytochrome c-type NapB, both related with denitrification processes, explicitly the nitrate reduction to nitrite. In addition, the remaining reduction steps of denitrification includes the reduction of nitrite to nitric oxide, nitric oxide to nitrous oxide, and, finally, nitrous oxide reduction to di-nitrogen gas. Several reads across the metagenomes samples were annotated to the family of nitrite reductase copper-type (3-104 reads), involved in the reduction of nitrite to nitric oxide (Table 3). Also, NnrS, a protein involved in the nitric oxide metabolism, was identified across the samples (9-57 reads; Table 3). Unfortunately, the step reduction of nitric oxide to nitrous oxide was not possible to check for the same reason of nitrite oxidoreductase, mentioned before, *i.e.*, the nitric oxide reductase subunit B, one of the enzyme subunits involved in the reduction of nitric oxide to nitrous oxide, shares homology with the family cytochrome c oxidase subunit I (van der Oost *et al.* 1994). Although the family cytochrome c oxidase subunit I (IPR000883) is highly abundant (1 664-2 946 reads, Additional file 6), it is involved in aerobic respiration, and therefore is common and widespread among aerobic microorganisms rather than denitrifying ones (van der Oost *et al.*, 1994). The last reduction step is very important because the incomplete denitrification without the reduction of nitrous oxide to N₂ will promote the release of nitrous oxide to the atmosphere. This compound is a greenhouse gas much stronger than CO₂, with an atmospheric lifetime $\approx$ 100 years (Fowler *et al.*, 2013; Voss *et al.*, 2013). Finally, few reads matched IPRs associated to the N pathway of nitrous oxide reduction to N₂ (0-7 reads, Table 3). As anticipated in the “Background” section, the highly-saturated oxygen waters of the Arctic Ocean (average 8.1 mL.L⁻¹, Additional file 1: Table S1) make this an unsuitable environment to denitrification and anammox (indistinguishable pathway using the annotation method herein) processes, usually found in OMZ (Ulloa *et al.*, 2012; Voss *et al.*, 2013; Hutchins and Fu, 2017). Indeed, some IPRs (nitrate reductase cytochrome c-type subunit NapB, NnrS, and nitrous oxide reductase, accessory protein NosL) are negatively correlated with oxygen (Figure 10). In the Arctic Ocean, the N loss throughout such processes should be limited to shelf sediments (Rysgaard *et al.*, 2004; Gihring *et al.*, 2010).

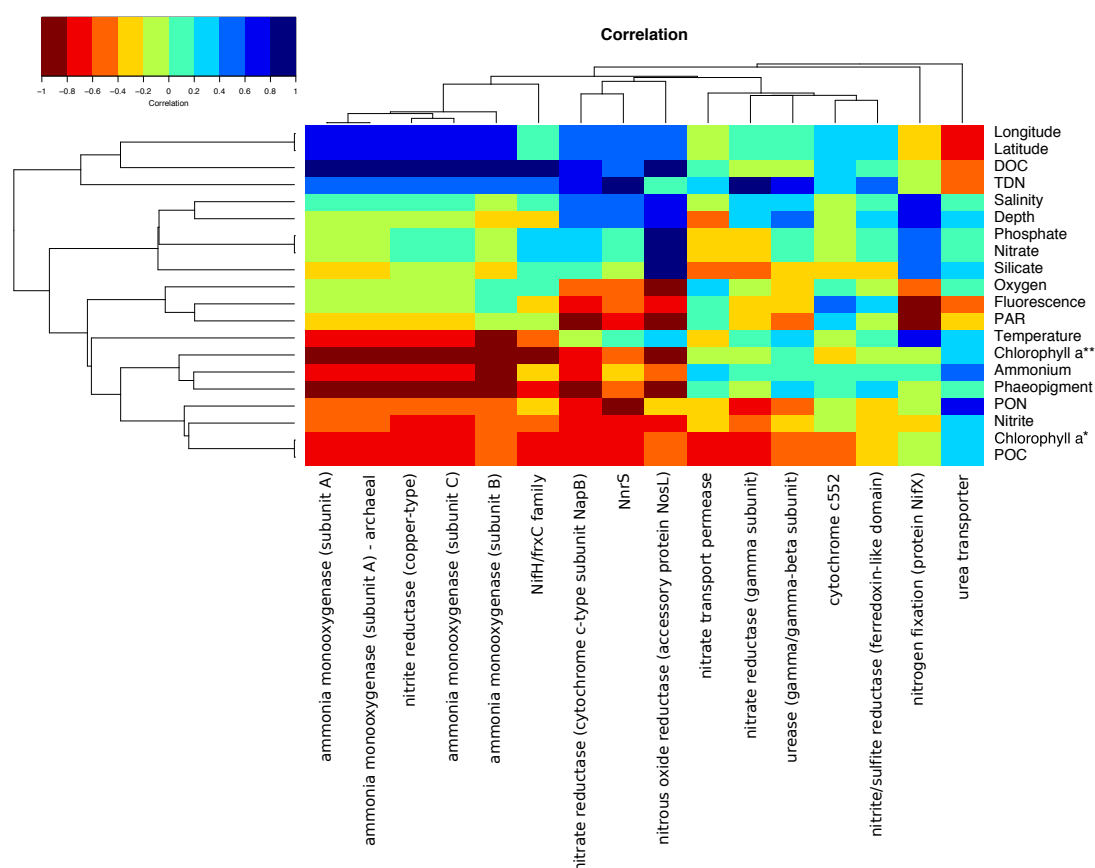


Figure 10 - Heatmap showing the Spearman correlation matrix between enzymes (*i.e.*, family, domain) involved in N-pathways and some environmental controls. DOC stands for Dissolved Organic Carbon; TDN, Total Dissolved Nitrogen; PAR, Photosynthetically Active Radiation; Chlorophyll *a*** (Chlorophyll *a* in mg.m^{-3}); PON, Particulate Organic Nitrogen; Chlorophyll *a** (Chlorophyll *a* in mg.m^{-2}); POC - Particulate Organic Carbon (see in detail Additional file 1: Table S1).

Since metagenomics represents the collective genomes of a microbial community (Handelsman *et al.*, 1998), it offers a powerful way to search/screen the genomic/functional potential of a community to carry out one process/pathway, but doesn't ensure their expression even less their activity. Therefore, the presence of several denitrifying related IPRs, particularly those involved in the step of nitrite reduction to nitric oxide, does not imply that the pathway is being active. Actually, the heterotrophic denitrification is an alternative metabolic pathway to the respiration of oxygen, in suboxic/anoxic conditions (Zhu *et al.*, 2013). Additionally, denitrification can be also put in practise by nitrifiers when oxygen is depleted - nitrifier denitrification (Zhu *et al.*, 2013). Hence, both heterotrophs and nitrifiers were found within the prokaryotic N-

ICE2015 collection (see the “Results and discussion” section, in the first chapter), with potentially some of them to carry out denitrifying related genes, what could explain the identification of several denitrifying related IPRs.

From the several N-fixation related IPRs screened, only two were found, nitrogen fixation protein NifX and NifH/frxC family, but just the latter has representativeness across all the samples (Table 3). However, this result needs to be carefully interpreted, since the NifH/frxC family shares homology with the nitrogenase but also with the light-independent protochlorophyllide oxidoreductase (Shi and Shi, 2006). The latter enzyme is involved in chlorophyll biosynthesis (Shi and Shi, 2006). Based on this observation, the absence of clear diazotrophic taxa, and previous works (Farnelid *et al.*, 2001; Blais *et al.*, 2012; Fernández-Méndez *et al.*, 2016), the contribution of N₂-fixation to the input of N in the Arctic Ocean should be minor or none.

4. Conclusions

Sea ice covers 4.1-6.1% of the world’s ocean surface (Arrigo, 2014b) and the biogeochemistry of N cycle associated to it and to the pelagic compartments underneath is largely unknown. The prokaryotic and metagenomic libraries from N-ICE2015 cruise allows to assess the community and the functional potential of nitrogen fixation, assimilation, nitrification, anammox and denitrification pathways in the Arctic Ocean. Nitrifiers, both TAO and NOB, have a high frequency of occurrence in the subsurface waters underneath the winter-spring pack ice (5 and 50 m depth), but they are nearly absent close to summer. Urease and AMO encoding genes, are positively correlated with total dissolved nitrogen, which includes urea, suggesting that both pathways, ureolysis and aerobic ammonia oxidation, are coupled. The number of reads matching the urease encoding gene increases along depth suggesting that distinct TAO populations found in the water column of the Arctic Ocean have different genomic potential to carry out ureolysis. In spite of the effort made, it was not found genomic evidences to support nitrogen fixation, anammox and denitrification pathways. Therefore, future studies focusing on the expression and quantification of these physiological processes are needed.

References

- Aanderud ZT, Jones SE, Fierer N, Lennon JT. Resuscitation of the rare biosphere contributes to pulses of ecosystem activity. *Front Microbiol.* 2015;6:24.
- Abell GC, Bowman JP. Ecological and biogeographic relationships of class *Flavobacteria* in the Southern Ocean. *FEMS Microbiol Ecol.* 2005;51:265-77.
- Alonso-Sáez L, Sánchez O, Gasol JM, Balagué V, Pedrós-Alió C. 2008. Winter-to-summer changes in the composition and single-cell activity of near-surface Arctic prokaryotes. *Environ Microbiol.* 2008;10:2444-54.
- Alonso-Sáez L, Waller AS, Mende DR, Bakker K, Farnelid H, Yager PL, et al. Role for urea in nitrification by polar marine Archaea. *Proc Natl Acad Sci U S A.* 2012;109:17989-94.
- Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol.* 2010;11:R106.
- Apprill A, McNally S, Parsons R, Weber L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquat Microb Ecol.* 2015;75:129-37.
- Arrigo KR, Perovich DK, Pickart RS, Brown ZW, van Dijken GL, Lowry KE. Massive phytoplankton blooms under Arctic sea ice. *Science.* 2012;336:6087.
- Arrigo KR, Perovich DK, Pickart RS, Brown ZW, van Dijken GL, Lowry KE. Phytoplankton blooms beneath the sea ice in the Chukchi sea. *Deep Sea Res Part II Top Stud Oceanogr.* 2014a;105:1-14.
- Arrigo KR, van Dijken G, Pabi S. Impact of a shrinking Arctic ice cover on marine primary production. *Geophys Res Lett.* 2008;35:L19603.
- Arrigo KR, van Dijken GL. Continued increases in Arctic Ocean primary production. *Progr Oceanogr.* 2015;136:60-70.
- Arrigo KR, van Dijken GL. Secular trends in Arctic Ocean net primary production. *J Geophys Res Oceans.* 2011;116: C09011.
- Arrigo KR. Sea ice ecosystems. *Ann Rev Mar Sci.* 2014b;6:439-67.
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol.* 2005;71:7724-36.

Ashelford KE1, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. New screening software shows that most recent large 16S rRNA gene clone libraries contain chimeras. *Appl. Environ. Microbiol.* 2006;72:5734-41.

Assmy P, Duarte P, Dujardin J, Fernández-Méndez M, Fransson A, Hodgson R, et al. N-ICE2015 water column biogeochemistry. Norwegian Polar Institute. 2016. <https://doi.org/10.21334/npolar.2016.3ebb7f64>.

Assmy P, Fernández-Méndez M, Duarte P, Meyer A, Randelhoff A, Mundy CJ. Leads in Arctic pack ice enable early phytoplankton blooms below snow-covered sea ice. *Sci Rep.* 2017;7:40850.

Bano N, Hollibaugh JT. Diversity and distribution of DNA sequences with affinity to ammonia-oxidizing bacteria of the beta subdivision of the class Proteobacteria in the Arctic Ocean. *Appl Environ Microbiol.* 2000;66:1960-9.

Bano N, Hollibaugh JT. Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. *Appl Environ Microbiol.* 2002;68:505-18.

Beman JM, Chow C-E, King AL, Feng Y, Fuhrman JA, Andersson A, et al. Global declines in oceanic nitrification rates as a consequence of ocean acidification. *Proc Natl Acad Sci U S A.* 2011;108:208-13.

Berg IA, Kockelkorn D, Buckel W, Fuchs G. A 3-hydroxypropionate/4-hydroxybutyrate autotrophic carbon dioxide assimilation pathway in Archaea. *Science.* 2007;318:1782.

Biopython. Python tools for computational molecular biology. <http://biopython.org> (2017) Accessed 2nd Sep 2017.

Blais M, Tremblay J-É, Jungblut AD, Gagnon J, Martin J, Thaler M, et al. Nitrogen fixation and identification of potential diazotrophs in the Canadian Arctic. *Glob Biogeochem Cycles.* 2012;26:GB3022.

Boé J, Hall A, Qu X. September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nat Geosci.* 2009;2:341-3.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* 2014;30:2114-20.

Bombar D, Paerl RW, Riemann L. Marine non-cyanobacterial diazotrophs: moving beyond molecular detection. *Trends Microbiol.* 2016;24:916-27.

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

- Bowman JS. The relationship between sea ice bacterial community structure and biogeochemistry: A synthesis of current knowledge and known unknowns. *Elem Sci Anth*. 2015;3:72.
- Brakstad OG, Nonstad I, Faksness LG, Brandvik PJ. Responses of microbial communities in Arctic sea ice after contamination by crude petroleum oil. *Microb Ecol*. 2008;55:540-52.
- Brown MV, Bowman JP. A molecular phylogenetic survey of sea-ice microbial communities (SIMCO). *FEMS Microbiol Ecol*. 2001;35:267-75.
- Bryant DA, Frigaard NU. Prokaryotic photosynthesis and phototrophy illuminated. *Trends Microbiol*. 2006;14:488-96.
- Capone DG, Zehr JP, Paerl HW, Bergman B, Carpenter EJ. *Trichodesmium*, a globally significant marine cyanobacterium. *Science*. 1997;276:1221-9.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. *Nat Methods*. 2010;7:335-6.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, et al. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J*. 2012;6:1621-4.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, et al. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc Natl Acad Sci U S A*. 2011;108:4516-22.
- Carini P, Dupont C, Santoro AE. Correlated expression of archaeal ammonia oxidation machinery across disparate environmental and culture conditions. *bioRxiv*. 2017;DOI:<https://doi.org/10.1101/175141>.
- Carpenter EJ, Montoya JP, Burns J, Mulholland MR, Subramaniam A, Capone DG. Extensive bloom of a N₂-fixing diatom/cyanobacterial association in the tropical Atlantic Ocean. *Mar Ecol Prog Ser*. 1999;185:273-83.
- Christman GD, Cottrell MT, Popp BN, Gier E, Kirchman DL. Abundance, diversity, and activity of ammonia-oxidizing prokaryotes in the coastal Arctic Ocean in summer and winter. *Appl Environ Microb*. 2011;77:2026-34.
- Chronopoulou PM, Sanni GO, Silas-Olu DI, van der Meer JR, Timmis KN, Brussaard CP, et al. Generalist hydrocarbon-degrading bacterial communities in the oil-polluted water column of the North Sea. *Microb Biotechnol*. 2015;8:434-47.
- Church MJ, Wai B, Karl DM, DeLong EF. Abundances of crenarchaeal *amoA* genes and transcripts in the Pacific Ocean. *Environ Microbiol*. 2010;12:679-88.

- Codispoti LA, Kelly V, Thessen A, Matrai P, Suttles S, Hill V, et al. Synthesis of primary production in the Arctic Ocean: III. Nitrate and phosphate based estimates of net community production. *Prog Oceanogr.* 2013;110:126-50.
- Cole JR, Wang Q, Fish JA, Chai B, McGarrell DM, Sun Y, et al. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* 2014;42:633-42.
- Comeau AM, Li WKW, Tremblay JÉ, Carmack EC, Lovejoy C. Arctic Ocean microbial community structure before and after the 2007 record sea ice minimum. *PLoS One.* 2011;6:e27492.
- Connelly TL, Baer SE, Cooper JT, Bronk DA, Wawrik B. Urea uptake and carbon fixation by marine pelagic bacteria and archaea during the Arctic summer and winter seasons. *Appl Environ Microbiol.* 2014;80:6013-22.
- Daims H, Lückner S, Wagner M. A new perspective on microbes formerly known as nitrite-oxidizing bacteria. *Trends Microbiol.* 2016;24:699-712.
- Delmont TO, Hammar KM, Ducklow HW, Yager PL, Post AF. *Phaeocystis antarctica* blooms strongly influence bacterial community structures in the Amundsen Sea polynya. *Front Microbiol.* 2014; 5:646.
- DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU. Community genomics among stratified microbial assemblages in the ocean's interior. *Science.* 2006;311:496-503.
- DeLong EF, Wu KY, Prézelin BB, Jovine RVM. High abundance of Archaea in Antarctic marine picoplankton. *Nature.* 1994;371:695-7.
- Deming JW. Psychrophiles and polar regions. *Curr Opin Microbiol.* 2002;5:301-9.
- DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, et al. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol.* 2006;72:5069-72.
- Dittmar T, Kattner G. The biogeochemistry of the river and shelf ecosystem of the Arctic Ocean: a review. *Mar Chem.* 2003;83:103-20.
- Doberva M, Sanchez-Ferandin S, Toulza E, Lebaron P, Lami R. Diversity of quorum sensing autoinducer synthases in the Global Ocean Sampling metagenomic database. *Aquat Microb Ecol.* 2015;74:107-19.
- Duran R. Marinobacter. In: Timmis KN, editor. *Handbook of hydrocarbon and lipid microbiology.* Springer, Berlin Heidelberg;2010. p. 1725-35.

EBI Metagenomics. Pipeline version 3.0. <https://www.ebi.ac.uk/metagenomics/pipelines/3.0> (2017) Accessed 2nd Sep 2017.

EBI Metagenomics. Submit, analyse, visualize and compare your data. <https://www.ebi.ac.uk/metagenomics/> (2017) Accessed 2nd Sep 2017.

Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*. 2011;27:2194-200.

Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods*. 2013;10:996-8.

Eronen-Rasimus E, Piiparinen J, Karkman A, Lyra C, Gerland S, Kaartokallio H. Bacterial communities in Arctic first-year drift ice during the winter/spring transition. *Environ Microbiol Rep*. 2016;8:527-35.

Esposito A, Kirschberg M. How many 16S-based studies should be included in a metagenomic conference? It may be a matter of etymology. *FEMS Microbiol Lett*. 2014;351:145-6.

Faith DP. Conservation evaluation and phylogenetic diversity. *Biol Conserv*. 1992;61:1-10.

Farnelid H, Andersson AF, Bertilsson S, Al-Soud WA, Hansen LH, Sørensen S, et al. Nitrogenase gene amplicons from global marine surface waters are dominated by genes of non-cyanobacteria. *PLoS One*. 2001;6:e19223.

Fernández-Méndez M, Turk-Kubo KA, Buttigieg PL, Rapp JZ, Krumpen T, Zehr JP, et al. Diazotroph diversity in the sea ice, melt ponds, and surface waters of the Eurasian Basin of the Central Arctic Ocean. *Front Microbiol*. 2016;7:1884.

Field CB, Behrenfeld MJ, Randerson JT, Falkowski P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science*. 1998;281:237-40.

Finn RD, Attwood TK, Babbitt PC, Bateman A, Bork P, Bridge AJ, et al. InterPro in 2017 - beyond protein family and domain annotations. *Nucleic Acids Res*. 2017;45:D190-9.

Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, et al. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44:D279-85.

Fouilland E, Gosselin M, Rivkin RB, Vasseur C, Mostajir B. Nitrogen uptake by heterotrophic bacteria and phytoplankton in Arctic surface waters. *J Plankton Res*. 2007;29:369–76.

Foulon E, Not F, Jalabert F, Cariou T, Massana R, Simon N. Ecological niche partitioning in the picoplanktonic green alga *Micromonas pusilla*: evidence from environmental surveys using phylogenetic probes. *Environ Microbiol*. 2008;10:2433-43.

Fowler D, Coyle M, Skiba U, Sutton MA, Cape JN, Reis S, et al. The global nitrogen cycle in the twenty-first century. *Philos Trans R Soc Lond B Biol Sci.* 2013;368: 20130164.

Fuhrman JA, Steele JA, Hewson I, Schwalbach MS, Brown MV, Green JL, et al. A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci U S A.* 2008;105:7774-8.

Galand PE, Casamayor EO, Kirchman DL, Potvin M, Lovejoy C. Unique archaeal assemblages in the Arctic Ocean unveiled by massively parallel tag sequencing. *ISME J.* 2009;3:860-9.

Galand PE, Potvin M, Casamayor EO, Lovejoy C. Hydrography shapes bacterial biogeography of the deep Arctic Ocean. *ISME J.* 2010;4:564-76.

Gentz T, Damm E, von Deimling JS, Mau S, McGinnis DF, Schlüter M. A water column study of methane around gas flares located at the West Spitsbergen continental margin. *Cont Shelf Res.* 2014; 72:107-18.

Gerdes B, Brinkmeyer R, Dieckmann G, Helmke E. Influence of crude oil on changes of bacterial communities in Arctic sea-ice. *FEMS Microbiol Ecol.* 2005;53:129-39.

Ghiglione JF, Galand PE, Pommier T, Pedrós-Alió C, Maas EW, Bakker K, et al. Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proc Natl Acad Sci U S A.* 2012;109:17633-8.

Gihring TM, Lavik G, Kuypers MMM, Kostka JE. Direct determination of nitrogen cycling rates and pathways in Arctic fjord sediments (Svalbard, Norway). *Limnol Oceanogr.* 2010;55:740-52.

Gilbert JA, Jansson JK, Knight R. The Earth Microbiome project successes and aspirations. *BMC Biol.* 2014;12:69.

Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome streamlining in a cosmopolitan oceanic bacterium. *Science.* 2005;309:1242-5.

Giudice AL, Bruni V, de Domenico M, Michaud L. Psychrophiles-cold-adapted hydrocarbon-degrading microorganisms. In: Timmis KN, editor. *Handbook of hydrocarbon and lipid microbiology.* Springer, Berlin Heidelberg;2010. p. 1897-1921.

González JM, Fernández-Gómez B, Fernández-Guerra A, Gómez-Consarnau L, Sánchez O, Coll-Lladó M, et al. Genome analysis of the proteorhodopsin-containing marine bacterium *Polaribacter* sp. MED152 (*Flavobacteriia*). *Proc Natl Acad Sci U S A.* 2008;105:8724-9.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17:333-51.

Gosink JJ, Woese CR, Staley JT. *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 glomeratus*' as *Polaribacter glomeratus* comb. nov. Int J Syst Bacteriol. 1998;48:223-35.

Granskog MA, Assmy P, Gerland S, Spreen G, Steen H, Smedsrud LH. Arctic research on thin ice: consequences of Arctic sea ice loss. EOS Trans AGU. 2016;97:22-6.

Grossmann L, Bock C, Schweikert M, Boenigk J. Small but manifold - hidden diversity in "*Spumella*-like flagellates". J Eukaryot Microbiol. 2016;63:419-39.

Grzymski JJ, Riesenfeld CS, Williams TJ, Dussaq AM, Ducklow H, Erickson M, et al. A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. ISME J. 2012;6:1901-15.

Guillou L, Bachar D, Audic S, Bass D, Berney C, Bittner L, et al. The Protist Ribosomal Reference database (PR²): a catalog of unicellular eukaryote small sub-unit rRNA sequences with curated taxonomy. Nucleic Acids Res. 2013;41:D597-604.

Haft DH, Selengut JD, White O. The TIGRFAMs database of protein families. Nucleic Acids Res. 2003;31:371-3.

Hammer Ø, Nakrem HA, Little CTS, Hryniewicz K, Sandy MR, Hurum JH, et al. Hydrocarbon seeps from close to the Jurassic–Cretaceous boundary, Svalbard. Palaeogeogr Palaeoclimatol Palaeoecol. 2011;306:15-26.

Han D, Kang I, Ha HK, Kim HC, Kim OS, Lee BY, et al. Bacterial communities of surface mixed layer in the Pacific sector of the western Arctic Ocean during sea-ice melting. PLoS One. 2014;9:e86887.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM. Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. Chem Biol. 1998;5:245-9.

Handelsman J. Metagenomics: application of genomics to uncultured microorganisms. Microbiol Mol Biol Rev. 2004;68:669-85.

Hara A, Sytsubo K, Harayama S. Alcanivorax which prevails in oil-contaminated seawater exhibits broad substrate specificity for alkane degradation. Environ Microbiol. 2003;5:746-53.

Harrell JrFE, Harrell JrMF. Package 'Hmisc'. 2015.

Harrison WG, Cota GF. Primary production in polar waters: relation to nutrient availability. Polar Res. 1991;10:87-104.

Hatam I, Charchuk R, Lange B, Beckers J, Haas C, Lanoil B. Distinct bacterial assemblages reside at different depths in Arctic multiyear sea ice. FEMS Microbiol Ecol. 2014;90:115-25.

- Hatam I, Lange B, Beckers J, Haas C, Lanoil B. Bacterial communities from Arctic seasonal sea ice are more compositionally variable than those from multi-year sea ice. *ISME J.* 2016;10:2543-52.
- Hatzenpichler R. Diversity, physiology and niche differentiation of ammonia-oxidizing archaea. *Appl Environ Microbiol.* 2012;78:7501-10.
- Hawley AK, Brewer HM, Norbeck AD, Paša-Tolić L, Hallam SJ. Metaproteomics reveals differential modes of metabolic coupling among ubiquitous oxygen minimum zone microbes. *Proc Natl Acad Sci U S A.* 2014;111:11395-400.
- Hazen TC, Prince RC, Mahmoudi N. Marine oil biodegradation. *Environ Sci Technol.* 2016;50:2121-29.
- Hmelo LR. Quorum sensing in marine microbial environments. *Ann Rev Mar Sci.* 2016;9:i-493.
- HMMER. HMMER: biosequence analysis using profile hidden Markov models <http://hmmer.org> (2017) Accessed 2nd Sep 2017.
- Hollibaugh JT, Bano N, Ducklow HW. Widespread distribution in polar oceans of a 16S rRNA gene sequence with affinity to *Nitrosospira*-like ammonia-oxidizing bacteria. *Appl Environ Microbiol.* 2002;68:1478-84.
- Huber H, Stetter KO. Thermoplasmatales. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. *The Prokaryotes*, vol. 3. New York: Springer Verlag; 2006. p. 101-12.
- Hugenholtz P, Huber T. Chimeric 16S rDNA sequences of diverse origin are accumulating in the public databases. *Int J Syst Evol Microbiol.* 2003;53:289-93.
- Hugenholtz P. Exploring prokaryotic diversity in the genomic era. *Genome Biol.* 2002;3:REVIEWS0003.
- Hutchins DA, Fu F. Microorganisms and ocean global change. *Nat Microbiol.* 2017;2:17058.
- Itkin P, Spreen G, Cheng B, Doble M, Girard-Arduin F, Haapala J, et al. Thin ice and storms: sea ice deformation from buoy arrays deployed during N-ICE2015. *J Geophys Res Oceans.* 2017;122:4661-74.
- John JS. SeqPrep. <https://github.com/jstjohn/SeqPrep> (2016) Accessed 2nd Sep 2017.
- Joint I, Mühling M, Querellou J. Culturing marine bacteria - an essential prerequisite for biodiversity. *Microb Biotechnol.* 2010;3:564-75.
- Joli N, Monier A, Logares R, Lovejoy C. Seasonal patterns in Arctic prasinophytes and inferred ecology of *Bathycoccus* unveiled in an Arctic winter metagenome. *ISME J.* 2017;11:1372-85.

- Jones CM, Stres B, Rosenquist M, Hallin S. Phylogenetic analysis of nitrite, nitric oxide, and nitrous oxide respiratory enzymes reveal a complex evolutionary history for denitrification. *Mol Biol Evol.* 2008;25:1955-66.
- Junge K, Eicken H, Deming JW. Bacterial activity at -2 to -20 °C in Arctic wintertime sea ice. *Appl Environ Microbiol.* 2004;70:550-7.
- Kalanetra KM, Bano N, Hollibaugh JT. Ammonia-oxidizing Archaea in the Arctic Ocean and Antarctic coastal waters. *Environ Microbiol.* 2009;11(9):2434-45.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* 2016;44:D457-62.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30:772-80.
- Keeling RE, Körtzinger A, Gruber N. Ocean deoxygenation in a warming world. *Ann Rev Mar Sci.* 2010;2:199-229.
- Kim JG, Park SJ, Damsté JSS, Schouten S, Rijpstra WI, Jung MY, et al. Hydrogen peroxide detoxification is a key mechanism for growth of ammonia-oxidizing archaea. *Proc Natl Acad Sci U S A.* 2016;113:7888-93.
- Kirchman DL, Cottrell MT, Lovejoy C. The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ Microbiol.* 2010;12:1132-43.
- Kirchman DL, Elifantz H, Dittel AI, Malmstrom RR, Cottrell MT. Standing stocks and activity of Archaea and Bacteria in the western Arctic Ocean. *Limnol Oceanogr.* 2007;52:495-507.
- Kirchman DL, Morán XA, Ducklow H. Microbial growth in the polar oceans - role of temperature and potential impact of climate change. *Nat Rev Microbiol.* 2009;7:451-9.
- Kirchman DL. Introduction and overview. In: Kirchman DL editor. *Microbial Ecology of the Oceans*. Second edition. Wiley-Liss, 2008. p. 1-23.
- Klindworth A, Mann AJ, Huang S, Wichels A, Quast C, Waldmann J, et al. Diversity and activity of marine bacterioplankton during a diatom bloom in the North Sea assessed by total RNA and pyrotag sequencing. *Mar Genomics.* 2014;18:185-92.
- Koenig Z, Provost C, Villacieros-Robineau N, Sennechael N, Meyer A, Lellouche J-M, et al. Atlantic Waters inflow north of Svalbard: Insights from IAOOS observations and Mercator Ocean global operational system during N-ICE2015. *J Geophys Res Oceans.* 2017; 122:1254-73.

- Konstantinidis KT, Tiedje JM. Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci U S A*. 2005;102:2567-72.
- Kopf A, Bicak M, Kottmann R, Schnetzer J, Kostadinov I, Lehmann K, et al. The ocean sampling day consortium. *Gigascience*. 2015;4:27.
- Kozich JJ, Westcott SL, Baxter NT, Highlander SK, Schloss PD. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl Environ Microbiol*. 2013;79:5112-20.
- Kuenen JG. Anammox bacteria: from discovery to application. *Nat Rev Microbiol*. 2008;6:320-6.
- Ladau J, Sharpton TJ, Finucane MM, Jospin G, Kembel SW, O'Dwyer J, et al. Global marine bacterial diversity peaks at high latitudes in winter. *ISME J*. 2013;7:1669-77.
- Laehnemann D, Borkhardt A, McHardy AC. Denoising DNA deep sequencing data - high-throughput sequencing errors and their correction. *Brief Bioinform*. 2016;17:154-79.
- Lalande C, Bauerfeind E, Nöthig EM, Beszczynska-Moller A. Impact of a warm anomaly on export fluxes of biogenic matter in the eastern Fram Strait. *Progr Oceanogr*. 2013;109:70-7.
- Lasternas S, Agustí S. Phytoplankton community structure during the record Arctic ice-melting of summer 2007. *Polar Biol*. 2010;33:1709-17.
- Leipzig J. A review of bioinformatic pipeline frameworks. *Brief. Bioinform*. *Brief Bioinform*. 2017;18:530-6.
- Leite C: "Domain Oriented Biclustering". Master Degree in Computer Sciences from FCUP, University of Porto, Portugal - concluded in 9th December 2016.
- Levasseur M. Impact of Arctic meltdown on the microbial cycling of sulphur. *Nature Geosciences*. 2013;6:691-700.
- Li WK, McLaughlin FA, Lovejoy C, Carmack EC. Smallest algae thrive as the Arctic Ocean freshens. *Science*. 2009;326:539.
- Lindsay R, Schweiger A. Arctic sea ice thickness loss determined using subsurface, aircraft, and satellite observations. *Cryosphere*. 2015;9:269-83.
- Liss PS, Malin G, Turner SM, Holligan PM. Dimethyl sulphide and *Phaeocystis*: a review. *J Mar Syst*. 1994;5:41-53.
- Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Loman, N. J. et al. Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol*. 2012;30:434-9.

Lovejoy C, Vincent WF, Bonilla S, Roy S, Martineau MJ, Terrado R, et al. Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in arctic seas. *J Phycol.* 2007;43:78–89.

Lozupone C, Knight R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol.* 2005;71:8228-35.

Lozupone CA, Hamady M, Kelley ST, Knight R. Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl Environ Microbiol.* 2007;73:1576-85.

Malmstrom RR, Kiene RP, Cottrell MT, Kirchman DL. Contribution of SAR11 bacteria to dissolved dimethylsulfoniopropionate and amino acid uptake in the north Atlantic Ocean. *Appl Environ Microbiol.* 2004;70:4129-35.

Malmstrom RR, Straza TRA, Cottrell MT, Kirchman DL. Diversity, abundance, and biomass production of bacterial groups in the western Arctic Ocean. *Aquat Microb Ecol.* 2007;47:45-55.

Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Pillay M, et al. IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucleic Acids Res.* 2014;42:D568-73.

Maslanik J, Stroeve J, Fowler C, Emery W. Distribution and trends in Arctic sea ice age through spring 2011. *Geophys Res Lett.* 2011;38: L13502.

Matrai PA, Vernet M, Hood R, Jennings A, Brody E, Saemundsdóttir. Light-dependence of carbon and sulfur production by polar clones of the genus *Phaeocystis*. *Mar Biol.* 1995;124:157-67.

Matrai PA, Vernet M. Dynamics of the vernal bloom in the marginal ice zone of the Barents Sea: dimethyl sulfide and dimethylsulfoniopropionate budgets. *J Geophys Res Oceans.* 1997;102:22965-79.

Mau S, Römer M, Torres ME, Bussmann I, Pape T, Damm E, et al. Widespread methane seepage along the continental margin off Svalbard - from Bjørnøya to Kongsfjorden. *Sci Rep.* 2017;7:42997.

McGenity TJ, Folwell BD, McKew BA, Sanni GO. Marine crude-oil biodegradation: a central role for interspecies interactions. *Aquat Biosyst.* 2012;8:10.

McKie-Krisberg ZM, Sanders RW. Phagotrophy by the picoeukaryotic green alga *Micromonas*: implications for Arctic Oceans. *ISME J.* 2014;8:1953-61.

Meier WN, Hovelsrud GK, van Oort BEH, Key JR, Kovacs KM, Michel C, et al. Arctic sea ice in transformation: A review of recent observed changes and impacts on biology and human activity. *Rev Geophys.* 2014;51:185-217.

- Meshram AR, Vader A, Kristiansen S, Gabrielsen TM. Microbial Eukaryotes in an Arctic Under-Ice Spring Bloom North of Svalbard. *Front Microbiol.* 2017;8:1099.
- Metfies K, von Appen WJ, Kiliyas E, Nicolaus A, Nöthig EM. Biogeography and photosynthetic biomass of Arctic marine pico-eukaryotes during summer of the record sea ice minimum 2012. *PLoS One.* 2016;11:e0148512.
- Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet.* 2010;11:31-46.
- Meyer A, Sundfjord A, Fer I, Provost C, Robineau NV, Koenig Z, et al. Winter to summer oceanographic observations in the Arctic Ocean north of Svalbard. *J Geophys Res Oceans.* 2017;122:6218-37.
- Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M, et al. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics.* 2008;9:386.
- Miller MB, Bassler BL. Quorum sensing in Bacteria. *Annu Rev Microbiol.* 2001;55:165-99.
- Mitchell A, Bucchini F, Cochrane G, Denise H, ten Hoopen P, Fraser M, et al. EBI metagenomics in 2016 - an expanding and evolving resource for the analysis and archiving of metagenomic data. *Nucleic Acids Res.* 2016;44:D595-D603.
- Moisander PH, Beinart RA, Hewson I, White AE, Johnson KS, Carlson CA, et al. Unicellular cyanobacterial distributions broaden the oceanic N₂ fixation domain. *Science.* 2010;327:1512-4.
- Montoya JP, Holl CM, Zehr JP, Hansen A, Villareal TA, Capone DG. High rates of N₂ fixation by unicellular diazotrophs in the oligotrophic Pacific Ocean. *Nature.* 2004;430:1027-31.
- Moore CM, Mills MM, Arrigo KR, Berman-Frank I, Bopp L, Boyd PW, et al. Processes and patterns of oceanic nutrient limitation. *Nat Geosci.* 2013;6:701-10.
- Morris RM, Rappé MS, Connon SA, Vergin KL, Siebold WA, Carlson CA, et al. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature.* 2002;420:806-10.
- mothur. MiSeq SOP. <https://www.mothur.org> (2017) Accessed 20th Jul 2017.
- Mußmann M, Brito I, Pitcher A, Damsté JSS, Hatzenpichler R, Richter A, et al. Thaumarchaeotes abundant in refinery nitrifying sludges express amoA but are not obligate autotrophic ammonia oxidizers. *Proc Natl Acad Sci U S A.* 2011;108:16771-6.
- Nealson KH, Platt T, Hastings JW. Cellular control of the synthesis and activity of the bacterial luminescent system. *J Bacteriol.* 1970;104:313-22.
- Newell SE, Babbitt AR, Jayakumar A, Ward BB. Ammonia oxidation rates and nitrification in the Arabian Sea. *Glob Biogeochem Cycles.* 2011;25: GB4016.

- Nguyen NH, Smith D, Peay K, Kennedy P. Parsing ecological signal from noise in next generation amplicon sequencing. *New Phytol.* 2015;205:1389-93.
- Nguyen N-P, Warnow T, Pop M, White B. A perspective on 16S rRNA operational taxonomic unit clustering using sequence similarity. *NPJ Biofilms Microbiomes.* 2016;2:16004.
- Nikrad MP, Cottrell MT, Kirchman DL. Abundance and single-cell activity of heterotrophic bacterial groups in the western Arctic Ocean in summer and winter. *Appl Environ Microbiol.* 2012;78:2402-9.
- Nobu MK, Narihiro T, Rinke C, Kamagata Y, Tringe SG, Woyke T, et al. Microbial dark matter ecogenomics reveals complex synergistic networks in a methanogenic bioreactor. *ISME J.* 2015;9:1710-22.
- Nöthig EM, Bracher A, Engel A, Metfies K, Niehoff B, Peeken I, et al. Summertime plankton ecology in Fram Strait - a compilation of long- and short-term observations. *Polar Res.* 2015;34:23349.
- Oulas A, Pavludi C, Polymenakou P, Pavlopoulos GA, Papanikolaou N, Kotoulas G, et al. Metagenomics: tools and insights for analyzing next-generation sequencing data derived from biodiversity studies. *Bioinform Biol Insights.* 2015;9:75-88.
- Pace NR. A molecular view of microbial diversity and the biosphere. *Science.* 1997;276:734-40.
- Parada AE, Fuhrman JA. Marine archaeal dynamics and interactions with the microbial community over 5 years from surface to seafloor. *ISME J.* 2017; doi:10.1038/ismej.2017.104.
- Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol.* 2016;18:1403-14.
- Parkinson CL, Comiso JC. On the 2012 record low Arctic sea ice cover: combined impact of preconditioning and an August storm. *Geophys Res Lett.* 2013;40:1356-61.
- Pedneault E, Galand PE, Potvin M, Tremblay JÉ, Lovejoy C. Archaeal amoA and ureC genes and their transcriptional activity in the Arctic Ocean. *Sci Rep.* 2014;4:4661.
- Pedrós-Alió C, Potvin M, Lovejoy C. Diversity of planktonic microorganisms in the Arctic Ocean. *Prog Oceanogr* 2015;139:233-43.
- Perovich DK, Richter-Menge JA. Loss of sea ice in the Arctic. *Annu Rev Mar Sci.* 2009;1:1417-41.
- Peterson AK, Fer I, McPhee MG, Randelhoff A. Turbulent heat and momentum fluxes in the upper ocean under Arctic sea ice. *J Geophys Res Oceans.* 2017;122:1439-56.

Peterson AK, Fer I, Randelhoff A, Meyer A, Håvik L, Smedsrud LH, et al. N-ICE2015 Ocean turbulent fluxes from under-ice turbulence cluster (TIC). Norwegian Polar Institute. 2016. <https://doi.org/10.21334/npolar.2016.ab29f1e2>.

Piredda R, Tomasino MP, D'Erchia AM, Manzari C, Pesole G, Montresor M, et al. Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. *FEMS Microbiol Ecol*. 2017;93:fiw200.

Plummer E, Twin J, Bulach DM, Garland SM, Tabrizi SN. A comparison of three bioinformatics pipelines for the analysis of preterm gut microbiota using 16S rRNA gene sequencing data. *J Proteomics Bioinform*. 2015;8:283-91.

Polyakov IV, Pnyushkov AV, Alkire MB, Ashik IM, Baumann TM, Carmack EC, et al. Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean. *Science*. 2017;356:285-91.

Polyakov IV, Walsh JE, Kwok R. Recent changes of Arctic multiyear sea ice coverage and the likely causes. *Bull Am Meteorol Soc*. 2012;93:145-51.

Porazinska DL, Giblin-Davis RM, Sung W, Thomas WK. The nature and frequency of chimeras in eukaryotic metagenetic samples. *J Nematol*. 2012;44:18-25.

Pyro VS, Roesch LF, Morais DK, Clark IM, Hirsch PR, Tótola MR. Data analysis for 16S microbial profiling from different benchtop sequencing platforms. *J Microbiol Methods*. 2014;107:30-7.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, et al. The SILVA ribosomal RNA gene database project improved data processing and web-based tools. *Nucleic Acids Res*. 2013;41:D590-6.

R Development Core Team: R: a language and environment for statistical computing. 2008, [<http://cran.r-project.org>].

Rey F. Declining silicate concentrations in the Norwegian and Barents Sea. *ICES J Mar Sci*. 2012;69:208-12.

Rho M, Tang H, Ye Y. FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res*. 2010;38:e191.

Ribeiro H, De Sousa T, Santos J, Sousa GGS, Teixeira C, Monteiro MR, et al. Potential of dissimilatory nitrate reduction pathways in polycyclic aromatic hydrocarbon degradation. *Chemosphere*. 2017;submitted.

Riesenfeld CS, Schloss PD, Handelsman J. Metagenomics: genomic analysis of microbial communities. *Annu Rev Genet*. 2004;38:525-52.

- Rivkin RB, Anderson MR, Lajzerowicz C. Microbial processes in cold oceans. I. Relationship between temperature and bacterial growth rate. *Aquat Microb Ecol.* 1996;10:243-54.
- Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ.* 2016;4:e2584.
- Roy S, Senger K, Braathen A, Noormets R, Hovland M, Olaussen S. Fluid migration pathways to seafloor seepage in inner Isfjorden and Adventfjorden, Svalbard. *Norw J Geol.* 2014; 94:99-119.
- Rysgaard S, Glud RN, Risgaard-Petersen N, Dalsgaard T. Denitrification and anammox activity in Arctic marine sediments. *Limnol Oceanogr.* 2004;49:1493-502
- Schloss PD, Gevers D, Westcott SL. Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS One.* 2011;6:e27310.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, et al. Introducing mothur: open-source, platform independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol.* 2009;75:7537-41.
- Schloss PD, Westcott SL. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Appl Environ Microbiol.* 2011;77:3219-26.
- Schneiker S, Martins dos Santos VA, Bartels D, Bekel T, Brecht M, Buhrmester J, et al. Genome sequence of the ubiquitous hydrocarbon-degrading marine bacterium *Alcanivorax borkumensis*. *Nat Biotechnol.* 2006;24:997-1004.
- Scholz MB, Lo CC, Chain PS. Next generation sequencing and bioinformatic bottlenecks: the current state of metagenomic data analysis. *Curr Opin Biotechnol.* 2012;23:9-15.
- Screen JA, Simmonds I. The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature.* 2010;464:1334-7.
- Serreze MC, Holland MM, Stroeve J. Perspectives on the Arctic's shrinking sea-ice cover. *Science.* 2007;315:1533-36.
- Shendure J, Ji H. Next-generation DNA sequencing. *Nat Biotechnol.* 2008;26:1135-45.
- Shi C, Shi X. Characterization of three genes encoding the subunits of light-independent protochlorophyllide reductase in *Chlorella protothecoides* CS-41. *Biotechnol Prog.* 2006;22:1050-5.

- Shiozaki T, Ijichi M, Isobe K, Hashihama F, Nakamura K, Ehama M, et al. Nitrification and its influence on biogeochemical cycles from the equatorial Pacific to the Arctic Ocean. *ISME J.* 2016;10:2184-97.
- Shokralla S, Spall JL, Gibson JF, Hajibabaei M. Next-generation sequencing technologies for environmental DNA research. *Mol Ecol.* 2012;21:1794-805.
- Sigman DM, Hain MP. The biological productivity of the ocean. *Nature Education Knowledge.* 2012;3:1-16.
- Simmonds I, Keay K. Extraordinary September Arctic sea ice reductions and their relationships with storm behavior over 1979-2008. *Geophys Res Lett.* 2009;36:L19715.
- Solomon CM, Collier JL, Berg GM, Glibert PM. Role of urea in microbial metabolism in aquatic systems: a biochemical and molecular review. *Aquat Microb Ecol.* 2010;59:67-88.
- Spren G, Kwok R, Menemenlis D. Trends in Arctic sea ice drift and role of wind forcing: 1992–2009. *Geophys Res Lett.* 2011;38:L19501.
- Staley JT, Gosink JJ. Poles apart: biodiversity and biogeography of sea ice bacteria. *Annu Rev Microbiol.* 1999;53:189-215.
- Staley JT, Konopka A. Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. *Annu Rev Microbiol.* 1985;39:321-46.
- Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30:1312–13.
- Steindler L, Schwalbach MS, Smith DP, Chan F, Giovannoni SJ. Energy starved *Candidatus Pelagibacter* ubique substitutes light-mediated ATP production for endogenous carbon respiration. *PLoS One.* 2011;6:e19725.
- Stingl U, Desiderio RA, Cho JC, Vergin KL, Giovannoni SJ. The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol.* 2007;73:2290-6.
- Stoeck T, Bass D, Nebel M, Christen R, Jones MDM, Breiner HW, et al. Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol Ecol.* 2010;19:21-31.
- Swan BK, Chaffin MD, Martinez-Garcia M, Morrison HG, Field EK, Poulton NJ, et al. Genomic and metabolic diversity of Marine Group I Thaumarchaeota in the mesopelagic of two subtropical gyres. *PLoS One.* 2014;9:e95380.

- Tatusov RL, Galperin MY, Natale DA, Koonin EV. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 2000;28:33-6.
- Taylor JD, Cottingham SD, Billinge J, Cunliffe M. Seasonal microbial community dynamics correlate with phytoplankton-derived polysaccharides in surface coastal waters. *ISME J.* 2014;8:245-8.
- Taylor RL, Semeniuk DM, Payne CD, Zhou J, Tremblay J-E, Cullen JT, et al. Colimitation by light, nitrate, and iron in the Beaufort Sea in late summer. *J Geophys Res.* 2013;118:3260-77.
- Teeling H, Fuchs BM, Bennke CM, Krüger K, Chafee M, Kappelmann L, et al. Recurring patterns in bacterioplankton dynamics during coastal spring algae blooms. *Elife.* 2016;5:e11888.
- Terrado R, Monier A, Edgar R, Lovejoy C. Diversity of nitrogen assimilation pathways among microbial photosynthetic eukaryotes. *J Phycol.* 2015;51:490-506.
- Thomas T, Gilbert J, Meyer F. Metagenomics - a guide from sampling to data analysis. *Microb Inform Exp.* 2012;2:3.
- Tolar BB, Powers LC, Miller WL, Wallsgrove, Popp BN, Hollibaugh JT. Ammonia oxidation in the ocean can be inhibited by nanomolar concentrations of hydrogen peroxide. *Front Mar Sci.* 2016a;3:237.
- Tolar BB, Wallsgrove NJ, Popp BN, Hollibaugh JT. Oxidation of urea-derived nitrogen by thaumarchaeota-dominated marine nitrifying communities. *Environ Microbiol.* 2016b; doi:10.1111/1462-2920.13457.
- Tremblay J-E, Gagnon J. The effects of irradiance and nutrient supply on the productivity of Arctic waters: a perspective on climate change. In: Nihoul JCJ, Kostianoy AG editors. *Influence of climate change on the changing Arctic and Sub-Arctic conditions.* Springer, Dordrecht, Netherlands, 2009. p. 73-93.
- Tripp HJ, Kitner JB, Schwalbach MS, Dacey JW, Wilhelm LJ, Giovannoni SJ. SAR11 marine bacteria require exogenous reduced sulphur for growth. *Nature.* 2008;452:741-4.
- Tully BJ, Sachdeva R, Heidelberg KB, Heidelberg JF. Comparative genomics of planktonic Flavobacteriaceae from the Gulf of Maine using metagenomic data. *Microbiome.* 2014;2:34.
- Tuorto SJ, Darias P, McGuinness LR, Panikov N, Zhang T, Häggblom MM, Kerkhof LJ. Bacterial genome replication at subzero temperatures in permafrost. *ISME J.* 2014;8:139-49.
- Ulloa O, Canfield DE, DeLong EF, Letelier RM, Stewart FJ. Microbial oceanography of anoxic oxygen minimum zones. *Proc Natl Acad Sci U S A.* 2012;109:15996-6003.

van der Oost J, de Boer AP, de Gier JW, Zumft WG, Stouthamer AH, van Spanning RJ. The heme-copper oxidase family consists of three distinct types of terminal oxidases and is related to nitric oxide reductase. *FEMS Microbiol Lett.* 1994;121:1-9.

von Scheibner M, Sommer U, Jürgens K. Tight coupling of *Glaciecola* spp. and diatoms during cold-water phytoplankton spring blooms. *Front Microbiol.* 2017;8:27.

Voss M, Bange HW, Dippner JW, Middelburg JJ, Montoya JP, Ward B. The marine nitrogen cycle: recent discoveries, uncertainties and the potential relevance of climate change. *Philos Trans R Soc Lond B Biol Sci.* 2013;368:20130121.

Wadhams P, Toberg N. Changing characteristics of arctic pressure ridges. *Polar Sci.* 2012;6:71-7.

Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 2007;73:5261-7.

Waters CM, Bassler BL. Quorum sensing: communication in Bacteria. *Annu Rev Cell Dev Biol.* 2005;21:319-46.

Wei T. Corrplot: visualization of a correlation matrix, Version 0.71. *Compr R Arch Network*, Vienna, Austria. 2013.

Wemheuer B, Güllert S, Billerbeck S, Giebel HA, Voget S, Simon M, et al. 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.* 2014;87:378-89.

Westcott SL, Schloss PD. OptiClust, an improved method for assigning amplicon-based sequence data to operational taxonomic units. *mSphere.* 2017;2:e00073-17.

Wheeler PA, Watkins JM, Hansing RL. Nutrients, organic carbon and organic nitrogen in the upper water column of the Arctic Ocean: implications for the sources of dissolved organic carbon. *Deep Sea Res Part II Top Stud Oceanogr.* 1997;44:1571-92.

Wickham H. *ggplot2: elegant graphics for data analysis.* Springer Science and Business Media. 2009.

Willmes S, Heinemann G. Sea-Ice wintertime lead frequencies and regional characteristics in the Arctic, 2003–2015. *Remote Sens.* 2016;8:4.

Wooley JC, Godzik A, Friedberg I. A Primer on metagenomics. *PLOS Comput Biol.* 2010;6:e1000667.

Wuchter C, Abbas B, Coolen MJL, Herfort L, van Bleijswijk J, Timmers P, et al. Archaeal nitrification in the ocean. *Proc Natl Acad Sci U S A.* 2006;103(33):12317-22.

Wynn PM, Hodson AJ, Heaton THE, Chenery SR. Nitrate production beneath a High Arctic glacier, Svalbard. *Chem Geol.* 2007;244:88-102.

Xu D, Li R, Hu C, Sun P, Jiao N, Warren A. Microbial eukaryote diversity and activity in the water column of the South China Sea based on DNA and RNA high throughput sequencing. *Front Microbiol.* 2017;8:1121.

Yilmaz P, Yarza P, Rapp JZ, Glöckner FO. Expanding the world of marine bacterial and archaeal clades. *Front Microbiol.* 2016;6:1524.

Yool A, Martin AP, Fernández C, Clark DR. The significance of nitrification for oceanic new production. *Nature.* 2007;447:999-1002.

Zehr JP, Waterbury JB, Turner PJ, Montoya JP, Omoregie E, Steward GF, et al. Unicellular cyanobacteria fix N₂ in the subtropical North Pacific Ocean. *Nature.* 2001;412:635-8.

Zhu X, Burger M, Doane TA, Horwath WR. Ammonia oxidation pathways and nitrifier denitrification are significant sources of N₂O and NO under low oxygen availability. *Proc Natl Acad Sci U S A.* 2013;110:6328-33.

Additional files

Note: Additional files are the same as supplementary information. They are given as ‘figshare private links’ to be accessed online through the figshare web service. Instead of divide and subset the data due to their size, *e.g.*, some tables have thousands of rows, all the tables that were generated during this work, and support the main results/findings, are given in full, to ensure transparency and reproducibility. They can be displayed online; however, the relative abundance of taxa cannot be displayed in percentage. These specific results can be displayed online as relative abundance, ranging between 0-1; or the files can be downloaded, and the values observed in percentage. The following eight ‘Additional files’ are given as ‘figshare private links’ (follow the link):

Additional file 1: Tables. Additional file 1 contains three supplementary tables: (1) Additional file 1: Table S1. Physical and biogeochemical conditions contextualizing the environment of microbial N-ICE2015 collection; (2) Additional file 1: Table S2. Description of 16S rDNA libraries from N-ICE2015 project; (3) Additional file 1: Table S3. Description of 18S rDNA libraries from N-ICE2015 project. (figshare private link: <https://figshare.com/s/a1bb3b0d53375e98ecf3>)

Additional file 2 – Distribution of prokaryotic taxa across N-ICE2015 collection at phylum, class, order, family, genus and OTU levels. Percentage of taxa at given taxonomic level, inclusive the absolute number of reads assigned to each OTU (97% sequence similarity threshold), within the samples and across the prokaryotic N-ICE2015 collection. The OTU table includes just prokaryotic taxa (see the lineages removed in the “Methods” section) classified against SILVA reference database (v. 1.2.8) after excluding the rare clusters (<5 observations across samples) and rarefying at even sampling depth (38 232 sequences). (figshare private link: <https://figshare.com/s/bc1eda6e0eda9e6439c0>)

Additional file 3 – Raw prokaryotic OTU table from the 16S rDNA libraries of N-ICE2015 collection. The OTU table includes all the taxa (inclusive those lineages that were removed from the main OTU table used in this manuscript, see the “Methods” section) classified against SILVA reference database (v. 1.2.8) without excluding the rare

clusters neither rarefying (the no of sequences goes from 59 627, at TR_S, to 212 529, at YP_S). (figshare private link: <https://figshare.com/s/7b0eccb3ebb825ab7869>)

Additional file 4 – Distribution of eukaryotic taxa across N-ICE2015 collection at phylum, class, order, family, genus and OTU levels. Percentage of taxa at given taxonomic level, inclusive the absolute number of reads assigned to each OTU (98% sequence similarity threshold), within the samples and across the eukaryotic N-ICE2015 collection. The OTU table includes just interesting protist taxa (see the lineages removed in the “Methods” section) classified against SILVA reference database (v. 1.2.8) after excluding the rare clusters (<5 observations across samples) and rarefying at even sampling depth (43 289 sequences). (figshare private link: <https://figshare.com/s/01ed266bd5ab58f600fb>)

Additional file 5 – Eukaryotic OTU table from the 18S rDNA libraries of N-ICE2015 collection assigned against the Protist Ribosomal Reference database (PR²). Percentage and absolute number of reads assigned to each OTU (98% sequence similarity threshold), within the samples and across the eukaryotic N-ICE2015 collection. The OTU table includes protist taxa (see the lineages removed in the “Methods” section) classified against PR² (v. 4.5) after excluding the rare clusters (<5 observations across samples) and rarefying at even sampling depth (43 647 sequences). (figshare private link: <https://figshare.com/s/a575656635661ab3996d>)

Additional file 6 – Metabolic profile of N-ICE2015 microbial collection. Absolute number of metagenomic reads that match an InterPro signature (IPR accession number) from the InterPro protein reference database (release 58.0), the derived Gene Ontology (GO) terms, the GO slim and the N cycle related IPRs across the samples collected during the N-ICE2015 campaign. Samples retrieved from the surface (S), middle (M) and bottom (B) seawater at Nansen Basin (NB), Transition Region (TR) and Yermak Plateau (YP) (see Table 1). (figshare private link: <https://figshare.com/s/e2a7803e27218402984d>)

Additional file 7: Table S1 - Description of shotgun metagenomic libraries from N-ICE2015 project. (figshare private link: <https://figshare.com/s/3a6b2573ab4acb051a2a>)

Additional file 8 - Taxonomic assignment of 16S rDNA reads extracted from the shotgun metagenomic libraries of N-ICE2015 collection. Absolute number of 16S rDNA reads extracted from the shotgun metagenomic libraries of N-ICE2015 collection assigned at phylum and species level. (figshare private link: <https://figshare.com/s/3635a05861fe77947232>)