Metabolism of microbiomes in a changing Arctic Ocean

Thomas Grevesse

A Thesis in the Department of Biology

Presented in Partial Fulfillment of the Requirements
For the Degree of
Doctor of Philosophy (Biology) at
Concordia University
Montreal, Quebec, Canada

January 2023

CONCORDIA UNIVERSITY

SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: Thomas Grevesse

Entitled: Metabolism of microbiomes in a changing Arctic Ocean

and submitted in partial fulfillment of the requirements for the degree of

Doctor Of Philosophy (Biology)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ Chair
Dr. Andrew Chapman

_____ External Examiner
Dr. Warwick Vincent

_____ Examiner
Dr. Pedro Peres-Neto

_____ Examiner
Dr. Mike Hallett

_____ Examiner
Dr. Brandon Findlay

_____ Thesis Supervisor
Dr. David Walsh

Approved by _____
Dr. Robert Weladji, Graduate Program Director

12/19/2022 _____
Dr. Pascale Sicotte, Dean
Faculty of Arts and Science

Abstract

Metabolism of microbiomes in a changing Arctic Ocean

Thomas Grevesse, Ph. D.

Concordia University, 2023.

The world's oceans are of utmost importance for us humans: they are a source of food and half of the oxygen we breathe, they act as climate regulators, trade routes, tourism attractions, and harbor an incredible diversity of life. The Arctic Ocean represents a particular ocean, with acute variations of temperatures, ice and solar radiation regimes throughout the year, and a strong terrestrial signature imparted by its immense watershed. But the oceans are now under threat of a changing climate. The polar oceans are especially susceptible to these changes with already dramatic visible consequences. The most visible consequence in the Arctic Ocean is a continuous loss of sea ice with impact on albedo, solar radiation regimes on the water surface, phytoplankton growth and primary productivity. The Arctic is also receiving increasing amounts of freshwater, leading to a freshening, disturbing the water column stratification, and increasing the load of organic matter from terrestrial origin. All these perturbations profoundly modify the sources and dynamics of organic and inorganic matter in the Arctic Ocean, perturbing the Arctic Ocean biogeochemical cycles. Given that microbial life is at the base of cycling this organic and inorganic matter, microbes play pivotal roles by controlling biogeochemical cycles and forming the base of the food web. Specifically, the diversity of metabolic processes carried out by microbes determines how they interact with and shape their environment. Despite the importance of understanding microbial metabolism in a rapidly changing Arctic Ocean, our knowledge of the microbial processes that distinguish the Arctic Ocean from the rest of the global oceans and how they are linked to the changing Arctic Ocean biogeochemical cycles is still very fragmented.

In this thesis, I undertook to address the lack of knowledge about the metabolism of the Arctic Ocean microbiomes by tackling two fundamental questions: *(i) What are the specificities and phylogenetic diversity of microbial metabolism in the Arctic Ocean compared to the other world oceans? (ii) What are the relationships between the Arctic Ocean microbial metabolic specificities and their biogeochemical environment?*

I first discovered that metabolic pathways for the degradation of aromatic compounds were enriched and expressed in the Canada Basin of the Arctic Ocean compared to the rest of the global ocean, in particular in the subsurface waters where organic matter of terrestrial origin accumulates. The capacity to degrade aromatic compound from terrestrial origin was phylogenetically concentrated in *Rhodspirillales*. These *Rhodospirillales* were enriched in aromatic compound degradation genes compared to close relatives from other oceans and their geographic distribution was restricted to the Arctic Ocean. These results suggest that the capacity to degrade aromatic compounds of terrestrial origin may be an adaptive trait of some Arctic Ocean microbial taxa. Furthermore, the aromatic-metabolizing bacteria may become more prominent

as organic matter inputs from land to ocean continue to rise with climate change, potentially impact the Arctic Ocean biogeochemical cycles.

In the second part of this thesis, I focused on the metabolism of neutral lipids, used to accumulate energy and carbon reserves. Within the global ocean, I discovered that the metabolism of neutral lipids was enriched in the microbial communities of the Arctic Ocean. In the photic zone, eukaryotic phototrophs dominated the synthesis of neutral lipids. I also discovered a large diversity of bacterial taxa able to degrade but not produce neutral lipids, suggesting that photosynthetic-based production of neutral lipids in eukaryotes may serve as an important carbon source for the heterotrophic bacterial community. Bacteria were the main producers in the aphotic zone and were equipped with a different set of enzymes targeting different compounds depending on their location within the water column. This study shows that the storage of neutral lipids may be a selective advantage for prokaryotes and picoeukaryotes in a context of extreme variations in energy and nutrients sources such as in the Arctic Ocean. In addition, I propose that, similarly to lipids from eukaryotic phototrophs sustaining the food web during the summer months, neutral lipids from prokaryotic origin may play an important role in sustaining the food web during the dark winter months.

Finally, I undertook a global ocean study to unravel the metabolic genes and pathways favored by the microbiomes of the Arctic Ocean. I confirmed the importance of aromatic compound degradation and neutral lipid metabolism. But I also uncovered a myriad of other metabolic processes favored by the microbiomes of the Arctic Ocean compared to other oceanic zones. In particular, in the photic zone of the Arctic Ocean, I discovered the prevalence of genes and pathways involved in the metabolism of glycans that might be involved in cold adaptation mechanisms. Importantly, I highlighted correspondences between the genes and pathways favored by the Arctic Ocean microbiomes and the composition and transformations of dissolved organic matter. Specifically, I found an enrichment in transformations involving sugars moieties in the photic zone and a strong aromaticity signature in the dissolved organic matter of the fluorescent dissolved organic matter maximum. These results show that the distinct metabolism of the Arctic Ocean microbiomes imprint the composition of the dissolved organic matter, uniquely influencing the Arctic Ocean biogeochemical cycles.

This thesis represents the first work to explore the metabolism of the Arctic Ocean microbiomes in such a comprehensive fashion. Not only does this thesis systematically uncover a multitude of metabolic processes of importance for the Arctic Ocean microbiomes, but it also brings new discoveries on their biogeography, ecological context, and phylogenetic diversity across prokaryotes and picoeukaryotes. Moreover, this thesis highlights the importance of these processes by linking them to the composition and transformation of dissolved organic matter, and hence biogeochemical cycles. As such, this thesis will serve as a base to guide experimental and field work that will quantify the role of microbiomes in the biogeochemical cycles of the Arctic Ocean. This will have important implications to understand and quantify how climate change perturbs Arctic Ocean ecosystems.

Acknowledgments

I would like to thank my supervisor David Walsh for hosting me in his lab and allowing me to get into and explore a totally new field of research for me. I would also like to thank my committee members Dr. Pedro Peres-Neto and Dr. Michael Hallett to keep me on the right track during this PhD.

I would like to thank all the members of the Walsh lab with whom I had fruitful scientific discussions, that helped me but that were also just a great bunch to work and hang out with. I also want to thank other people from the biology department that made my PhD a better experience, wink Rob and Alaa. Special thanks to Kevin and Simon for patiently correcting my laborious English.

Finally, work is only a part of life, and would mean nothing if there were no wonderful people outside of work to make life a truly beautiful and enjoyable journey. I could not be more grateful to my friends who, just by talking, laughing and hanging, can brush off any bad day like it was nothing. My family, that always supported me and without whom I would not be here today. And finally, cherry on the cake, Yasmine, my partner, who followed the whole process, was happy and celebrating in the highs and ever supporting in the lows.

Contribution of Authors

In all chapters, David Walsh designed the sampling strategy and Arthi Ramachandran collected metagenomics and metatranscriptomics samples. Thomas Grevesse performed the extraction of environmental DNA and RNA and analyzed the metagenomic and metatranscriptomic data. Sequencing, assembly and annotation of metagenomes and metatranscriptomes were performed by the Joint Genome Institute.

In chapter 2, Céline Guéguen (Sherbrooke University) collected and analyzed the FDOM samples. Vera Onana performed the genome annotation for publicly available genomes. Thomas Grevesse and David Walsh, with the help of Céline Guéguen wrote the manuscript.

In chapter 3, Thomas Grevesse and David Walsh designed the study. Susan McLatchie performed the genome reconstruction from metagenomes and Vera Onana performed the genomes annotation. Thomas Grevesse and David Walsh wrote the manuscript.

In chapter 4, Thomas Grevesse and David Walsh designed the study. Céline Guéguen collected the FTICR-MS samples. FTICR-MS spectra were obtained at the Environmental Molecular Sciences Laboratory. FTICR-MS spectra were analyzed by Roya Amini Tabrizi (Malak Tfaily lab, University of Arizona). Thomas Grevesse, Roya Amini Tabrizi, David Walsh and Malak Tfaily wrote the manuscript.

Table of Contents

List of figures

List of abbreviations

AAnBP	Aerobic anoxygenic photosynthetic bacteria
ATP	Adenosine triphosphate
C	Carbon
CAzyme	Carbohydrate active enzyme
cDNA	Complementary desoxyribonucleic acid
CDOM	Colored dissolved organic matter
DAG	Diacylglycerol
DGAT	Diacylglycerol O-acyltransferase
DIC	Dissolved inorganic carbon
DNA	Desoxyribonucleic acid
DOC	Dissolved organic carbon
DOM	Dissolved organic matter
EC	Enzyme commision
EEM	Emission/excitation matrix
FA	Fatty acid
FDOM	Fluorescent dissolved organic matter
FISH	Fluorescent in situ hybridization
FTICR	Fourier transform ion cyclotron resonance
GTDB	Genome taxonomy database
HS	Humic substances
KO	KEGG ortholog
MAG	Metagenome-assembled genome
mRNA	Messenger ribonucleic acid
MS	Mass spectrometry
N	Nitrogen
NGS	Next generation sequencing
NL	Neutral lipid
NMDS	Nonmetric multidimensional scaling
NMF	Non-negative matrix factorisation
NMR	Nuclear magnetic resonance
OM	Organic matter
P	Phosphorous
PARAFAC	Parallel factor analysis
PDAT	Phospholipid:diacylglycerol acyltransferase
PERMANOVA	Permutational analysis of variance
PHA	Polyhydroxyalkanoate
PHB	Polyhydroxybutanoate
PL	Phospholipid
PSU	Practical salinity unit
RDOM	Refractory dissolved organic matter

rRNA	ribosomal ribonucleic acid
SCM	Subsurface chlorophyll maximum
SE	Steryl ester
TAG	Triacylglycerol
tDOM	Terrestrial dissolved organic carbon
tOM	Terrestrial organic matter
WE	Wax ester

Chapter 1

Introduction

1.1 The marine microbiome

The ocean biome contains 95% of all the water on Earth and covers 70% of its surface. Microorganisms are ubiquitous and abundant in the ocean, with an estimated 5 x 10^5 cells/mL in the top 200 m and 5 x 10^4 cells/mL below 200 m[1]. The microbiome, defined as *"a characteristic microbial community occupying a reasonable well-defined habitat which has distinct physio-chemical properties"*[2], comprises 98% of the ocean biomass. The number of prokaryotes alone amounts to a total of 1.2 x 10^{29} cells in the world's oceans[1]. Considering an average of 20 fg of carbon per cell, the total carbon pool contained in the world's marine prokaryotes represents 2.2 x 10^{15} g[1]. In order to survive, grow or proliferate, all this biomass made up by marine prokaryotes and other microbial life (eukaryotes) create through primary productivity, use and transform organic and inorganic matter[3]. They therefore play essential roles in ocean biogeochemical cycles such as carbon, nitrogen, and oxygen cycles[4].

Microorganisms in the ocean are extremely diverse and have adapted to colonize every last corner of the ocean[5]. The diversification of their metabolic capacity and activity plays an important role in their ability to thrive in a wide range of physical, chemical and biological settings[6]. The diversity and interactions of metabolic processes in a microbiome will therefore dictate how this microbiome responds to its environment but also how it shapes its biogeochemical landscape[7]. The study of microbial metabolism is therefore key for understanding the biogeography, as well as the dynamics of microbiomes under changing environments[8]. As such the microbial metabolism is an important part of the big questions in marine microbiology: *What is the composition of microbiomes and how do they vary in space and time? What are the factors driving these variations? Do we find endemic microbial populations to specific environments? What adaptations drove these specializations? What are the metabolic processes used by microbiomes in specific environments and how do these metabolic processes spread through the microbiome? How do the metabolic processes of microbiomes impact the ocean biogeochemical cycles?*

From a wider perspective, answering these questions allows a better understanding of the influence that the microbiome and its environment have on each other. This is critical as global

atmospheric temperatures increase, leading to profound perturbations in oceanic ecosystems[9]. Answers to these fundamental questions provide an understanding on the influence of a changing environmental landscape on the composition, dynamics and metabolism of microbiomes. Consequentially, understanding the trajectories of microbiomes enables to predict how their metabolism will shape the environment they inhabit, linking microbiome metabolism to biogeochemical cycles in a changing ocean.

1.1.1 How to study microbiomes and their metabolism

In this thesis, we are interested in the capacity of microbiomes to carry out particular metabolic processes that have ecological importance in the Arctic Ocean. While studying environmental microbiomes, three fundamental interrogations underly all specific questions a microbiologist seeks to answer: *who is there?* (composition of the community), *what is it doing?* (metabolic transformations of organic and inorganic matter, physicochemical interactions with its environment, growth) and *why is it there?* (evolutionary history, adaptation).

Characteristics of macroorganisms used in traditional ecology to answer these questions, such as morphology and behavior of species do not apply to the microbial world. Microbiologists therefore have to use other ways to identify microbes and characterize their metabolic processes. The isolation and cultivation of specific microbial strains, with the study of the metabolites they transform was the original method to study environmental microbiomes. Isolation and cultivation techniques have made great progress, allowing the study of many microbial species and strains[10] and these techniques are still widely used today. The ability to study a microbial species cultured in isolation in the lab presents several advantages: (i) the phenotype (colony formation, physiology and biochemistry) of the strain can be characterized in detail (ii) effect of conditions (temperature, pH,etc.) on growth can be quantified as culture conditions can be easily manipulated (iii), and we can obtain complete genomes sequences, enabling us to explore the potential metabolism and evolution of the strain.

However, cultivation methods for marine microbiomes pose certain challenges. The biggest challenge is that the majority of marine microbes remain elusive to cultivation[11]. In addition, even in co-culture, cultivation is limited to few taxa, failing to recapitulate the complex metabolic interactions that are characteristic of environmental microbiomes. Finally, it is currently not possible to design artificial media capturing the complex dissolved organic matter (DOM) and trace elements composition of sea water[11]. Considering the limitations of cultivation-dependent techniques, meta-omics methods such as metagenomics, metatranscriptomics and metabolomics can help understand the complex metabolism of *in situ* environmental microbiomes.

16S ribosomal RNA gene analysis

When asking the *who is there?* question, the fundamental requirement is being able to

identify and differentiate taxonomic units from one another. It is impossible to differentiate prokaryotes from environmental microbiomes based on their morphology. In addition, the sheer number and complexity of chemical reactions carried on in environmental microbiomes prevent the delineation of taxonomic unit based on their metabolic activity (phenotype). The seminal work of Carl Woese using ribosomal RNA (rRNA) genes as a phylogenetic barcode for prokaryotes[12] provided an elegant way to overcome this issue and marks the beginning of the molecular methods era in the study of microbiomes. Using the sequencing of the 5S rRNA, Woese added a new branch to the tree of life with the Archaea as distinct from Bacteria. The 16S rRNA gene is a better and ideal candidate for phylogenetic reconstruction of prokaryotic communities as it is present in all prokaryotes, is not prone to lateral gene transfer and contains both highly conserved and highly variable regions. The sequencing of 16S rRNA genes is nowadays a widespread standard approach for the census of taxonomic diversity in environmental prokaryotic communities, and has been used in various environments, including the ocean[13]. However, the sequencing and analysis of whole microbiomes' ribosomal genes is limited to the identity and abundance of their taxa and does not allow access to the metabolic capacity of the microbiome.

Metagenomics

To answer the question: *what are microbes doing?* we need to access the functions of microbiomes, and hence, the ensemble of metabolic processes they are able to carry out. The ensemble of metabolic reactions a microbe can perform is encoded in its genes. The necessity to predict the metabolic potential of microbiomes to understand their role in various ecosystems led to the development of metagenomics. First coined by Jo Handelsman in 1998[14], metagenomics refers to the investigation of the collective genomes found within an environmental sample. Metagenomics was first achieved in 1991, by cloning a library of total DNA obtained from sea water in a λ phage vector[15]. The same method of cloning DNA retrieved from environmental samples in plasmid vectors followed by Sanger sequencing became more widely used[16–18]. The development of next generation sequencing (NGS) technologies allowed a substantial leap forward in studying microbiomes by enabling unparalleled sequencing in far shorter times and the scaling-up of the generation of nucleotide sequences by orders of magnitude at low cost[19]. In a typical metagenomics workflow, the total DNA of a microbiome is extracted, fragmented in small reads and sequenced. The reads are then assembled in longer sequences (scaffolds). The metagenomic assemblies obtained can be functionally annotated and provide a picture of the whole gene (protein-coding as well as house-keeping genes) pool of the sampled microbiome.

The ensemble of genes obtained from a metagenomics workflow serves as the base to answer fundamental questions in environmental microbiology related to the *what are microbes doing?* question. Metagenomic-based gene-centric analyses aim at producing a comprehensive view of the complete gene pool of a microbiome. Rather than representing a microbiome from the point of view of its constituent organisms (genomes), gene-centric analyses consider microbiomes as a collection of genes. This approach permits the evaluation of metabolic genes and pathways that are relevant to an ecosystem, and also enables to compare and contrast the occurrence

and frequencies of genes and pathways between different systems. The gene-centric approach has therefore been widely used and is still an approach of choice to facilitate the discovery of metabolic characteristic in various environments[20].

In addition, metagenome-assembled genomes (MAGs) can be reconstructed by grouping sequences based on genomic characteristics such as GC content, tetranucleotide frequencies and/or abundance in metagenomes (*i.e.* coverage)[21,22]. MAGs therefore enable us to uncover the taxonomic origin of genes of interest, establish the taxonomic composition of the community, reveal the biogeography of its members and also perform phylogenetic analysis of MAGs and their genes. MAGs have therefore been instrumental in answering questions about the evolutionary origin of taxa in specific environments and gene flow between these taxa[23,24].

Despite the great advances enabled by metagenomics, these methods can not differentiate active from inactive genes and genomes in a microbiome, but can only establish those that are present in the microbiome. As a result, metagenomics cannot delineate which genes and genomes are actively contributing to the ecosystem behavior (such as biogeochemical cycles) from the pool of genes and genomes present in the microbiome.

Metatranscriptomics

Metatranscriptomics focuses on studying the *in-situ* expression of genes. Before being sequenced, the whole microbiome RNA goes through a few extra steps compared to DNA. After the RNA fragmentation, the RNA fragments are used as a scaffold for the synthesis of a complementary DNA (cDNA) by a reverse transcriptase[25]. A second strand of cDNA, complementary to the first cDNA strand, is then synthesized by a DNA polymerase. This is the double-stranded cDNA that goes through the sequencing workflow. The nature of RNA poses additional challenges for sequencing. Contrary to DNA, RNA is notoriously unstable, which can compromise the integrity of the sample before sequencing. Extreme care is therefore necessary when processing RNA samples. In addition, microbiome environmental RNA samples are dominated by rRNA, which can significantly reduce the depth of coverage of messenger RNA (mRNA). mRNA can be enriched in samples either by separation using 16S and 23S probes and hybridization, or by degradation of rRNA with 5-exonucleases[25].

Metatranscriptomics provides a valuable complement to metagenomics by identifying which of the genes sequenced and annotated in the metagenome are transcribed and to what extent, enabling the study of gene expression of complex microbiomes at a given point in time and under a specific set of environmental conditions. Moreover, for certain genes, the abundance of transcripts show linear and proportional relationships with the activity level of the reaction they catalyze[26]. As the average half-time of a mRNA is $\sim$5 min, metatranscriptomics can reveal fast changes in microbiomes functional activities[27]. In that regard, metatranscriptomics provides a representation of the real-time functional activities of a microbiome and has better power to associate the microbiome metabolic processes to biogeochemical processes of their ecosystem.

Metatranscriptomics is most powerful if used in conjunction with metagenomics[28,29]. Indeed, metatranscriptomics alone cannot delineate if differences in transcripts abundance between samples originate from differences in organism abundances or from differences in gene expression. This is important information, specifically when studying metabolic changes in microbial communities over environmental gradients. Changes in gene expression within the studied gradient reflects the capacity of individual microbial taxa to withstand a change in environmental conditions through gene regulations mechanisms. In contrast, changes in microbial taxa abundance reveal that modifications of environmental conditions cause a turnover in community structure rather than a physiological adaptation. In conjunction with metagenomics, metatranscriptomics is a powerful method to provide information regarding what are the genes and genomes most active at any given conditions and time points in environmental microbiomes.

Metaproteomics is a useful way to explore the active part of metagenomes. However, transcripts need to be translated to proteins, and these proteins actively catalyzing enzymatic reactions. The levels of transcripts may therefore not mirror exactly the actual metabolic activity of a microbiome at any given condition and time point.

1.1.2 Composition and metabolism of the ocean microbiome

The global ocean biogeochemical cycles are modulated by microbiomes. Any changes in the structure and function of the ocean microbiome can have consequences at a large scale. It is therefore important to understand what mechanisms structure the composition and function of the ocean microbiome. The advent of metagenomics and metatranscriptomics enabled a great increase in knowledge of marine microbial systems. We have tremendously deepened our understanding of the ecology, biogeographic patterns and temporality of marine microbiomes. In addition, omics techniques revealed unsuspected metabolic capacities in marine microbes as well as their taxonomic origin.

Microbiology in the sunlit ocean

Since microbial primary producers are at the base of the food web, their phylogenetic diversity, the quality and quantity of OM they produce and how their ecology is affected by environmental gradients are of utmost importance for the ocean carbon cycle. Marine primary producers are responsible for approximately half of the global CO_2 fixation. Most of the ocean primary productivity takes place in the sunlit ocean[30]. The sunlit ocean (or euphotic zone, or epipelagic ocean) generally extends until light intensity is attenuated to 1% of the surface irradiance. The ocean primary producers are for the most part microbial and float freely in the water. These microbial phytoplankton are oxygenic photosynthetic organisms, include *Cyanobacteria* and single-celled eukaryotes, and are responsible for the bulk of ocean primary production[31]. Omics techniques have shown that the phototrophic oxygenic *Cyanobacteria*, including *Prochloroccocus* and *Synechoccocus*, are the most abundant primary producers in the

ocean and contribute to 25% of the ocean primary productivity[32]. *Prochloroccocus*, dominates the oxygenic phototrophic microbes of the epipelagic ocean between 40^0N and 40^0S and up to a depth of 150 m but its population size decreases beyond these latitudes[32]. *Synechoccocus* is not found so deep, but has a wider geographic distribution and is even present in polar regions[32]. Other strategies used by photoheterotrophs to harvest solar radiation[33] have also been uncovered in the ocean. Aerobic anoxygenic photosynthetic bacteria (AAnBP) use bacteriochlorophyll a to harvest energy from light and are found mostly in the *Alpha-* and *Gammaproteobacteria*[34]. At first considered rare and living in particular environments, we now know that they are widely distributed in the ocean[34], and can account for up to 24% of the communities[35]. The last strategy generates ATP using a light-driven proton-pump, proteorhodopsin[36]. Proteorhodopsins have been characterized in numerous phyla and many forms exist that can absorb different wavelengths[37].

Microbial heterotrophs strongly influence the carbon cycle of the ocean. Approximately half of the carbon fixed by primary producers is directly processed by bacteria[38] and ultimately, more than 70% of OM produced by phytoplankton is consumed by bacteria[39]. The ocean phytoplankton communities are characterized by local and transient increases of abundance, called blooms, when both nutrients (nitrogen and phosphorus), carbon (CO_2) and energy (light) are plentiful. For microbial heterotrophs, the epipelagic ocean is therefore generally oligotrophic with bursts of growth resources in the form of OM from phytoplankton exudates or dead phytoplankton. Some heterotrophs have adapted to efficiently degrade labile, low-molecular weight OM in nutrient-limited situations[40,41] and are cosmopolitan in the global ocean. For example, *SAR11* possesses a streamlined genome with a high level of enzyme multifunctionality (capable or catalyzing multiple reactions) and a large number of transporters to maximize its substrates uptake while minimizing its resource use[42]. *SAR11* is the most successful and abundant heterotroph in the euphotic zone and is represented by many ecotypes adapted to specific niches[40]. Similarly, various bacterial and archaeal taxa, such as *Rhodobacterales*, *Vibrio*, *Alteromonas*, *SAR 86* and the *MGII* clade of *Euryarcheaota* are ubiquitous due to a range of ecotypes adapted to particular niches[31]. Other bacterial taxa are adapted to grow rapidly by fastly consuming OM during phytoplankton blooms[43]. *Flavobacteria* possess many membrane-attached and extracellular hydrolytic enzymes to degrade high molecular weight compounds that cannot pass through the membrane. In addition, they also produce adhesins to facilitate their adhesion and gliding mobility on phytoplankton and proteases that may have algicidal properties[43]. *Flavobacteria*, with *Roseobacters* and members of *Gammaprotobacteria* are generally associated with phytoplankton blooms[43] and their abundance correlates with the bloom progression[44].

Microbiology in the dark ocean

The dark ocean is found below 200 m and is composed of the mesopelagic (200-1000 m), bathypelagic (1000-4000 m), abyssopelagic (4000-6000 m) and hadalpelagic (> 6000 m) zones. In terms of volume, the dark ocean is the largest biome on Earth. As such, it harbors most of the microbial biomass (75%) and half of the microbial production[45]. In terms of community composition, there is a strong shift from phototroph-dominated waters to heterotroph-dominated

waters between the sunlit and dark ocean[31]. The most remarkable feature is the vertical decrease in the bacteria to archaea ratio[46]. The relative abundance of archaea increases sharply in the mesopelagic waters and they dominate the communities throughout the mesopelagic and bathypelagic ocean. Archaea, particularly the *Crenarchaeota Group I/Thaumarchaeota* and the *Euryarchaeota Group II*, can make up to 50% of the community. Within the bacteria, *Alphaproteobacteria* and *Gammaproteobacteria* dominate most of the deep waters. However, various other taxa increase with depth: *Acidobacteria, Actinobacteria, Bacteroidetes, Firmicutes, Gemmatimonadetes, Lentisphaera, Nitrospirae, Planctomycetes, Verrucomicrobia, Deltaproteobacteria*[47]. Generally, the major groups of taxa found in the deep ocean are similar across different ocean basins[31].

The differences in community composition between the sunlit and the dark ocean is reflected in different metabolic processes[48]. The vertical increase in archaea relative abundance was reflected in inorganic carbon (DIC) fixation: it was reported that, in bathypelagic waters, 20% of archaeal cells fix inorganic carbon, while bacteria accounted for less than 2% of the C-fixing cells[49]. In addition, it was estimated that ∼83% of C in mesopelagic archaea is derived from DIC fixation[50]. The paradigm of the energy source for fuelling this DIC fixation in absence of light has been solved with the discovery of a marine archaeon able to oxidize ammonia[51]. Sinking of particulate organic matter is the main source of OM for the dark ocean microbial communities[52] and connects the epipelagic conditions to the dark ocean microbiomes[53]. The concentration of OM in particles can be four orders of magnitude higher than the background concentration[54]. Individual cells in the dark ocean therefore live in a resource desert interspersed by resource oases. Free-living and particle-attached microorganisms have therefore two different lifestyle reflected in functional differences: ammonia and CO_2 oxidation are enriched in the free-living fraction while dissimilatory nitrate reduction and H_2 oxidation are enriched in the particle-attached fraction[55]. In addition, some members of the dark ocean's microbiome, such as *SAR202* have the capacity to degrade recalcitrant organic matter, which makes up the bulk of DOM in the dark ocean[56]. The absence of light to fuel autotrophic C fixation, the predominance of recalcitrant DOM as well as the necessity to travel between OM-rich particles explain the shifts of energy metabolism, cell attachment, motility and host-viral interactions between the sunlit and the dark ocean[57].

Spatial and temporal patterns of the ocean microbiome

The ocean is a highly heterogeneous ecosystem and its physiochemistry is characterized by space and time gradients of biotic and abiotic factors but also by a multitude of different niches[58]. It is important to determine how these physicochemical gradients shape microbiome structure and function over spatial and temporal scales to understand how microbiomes impact their ecosystem, but also to predict the trajectories of microbiome structure and function with current and future changes in conditions and perturbations.

The heterogeneity of conditions in the ocean translates into a strong biogeography of microbial community structure and function[48]. Microbial communities are generally composed

of few abundant taxa and a wide diversity of rare (low abundant) taxa[59]. The most striking spatial pattern for microbial communities is the vertical segregation between the photic (epipelagic zone) and the aphotic (deep) ocean[57,60]. However, the microbial consortia that colonize phytoplankton-derived sinking particles in the epipelagic zone inoculate deep ocean communities, promoting the connectivity of the epipelagic and deep communities[61,62]. The composition of microbial communities also varies greatly within water masses across spatial gradients. Communities are characterized by higher diversity towards the equator and mid-latitudes[63,64], similarly to what has been observed for macroorganisms. Interestingly, despite their spatial separation, the microbial communities from the polar zones were more similar to each other than to communities from the equatorial and temperate oceans[65–67]. This latitudinal effect has been explained by temperature as the main predictor of the marine microbiome structure and function[48]. In addition, phosphorus availability has been shown as a selective pressure driving the divergence of some of the most abundant marine microbes in the epipelagic ocean (*Prochloroccocus, Pelagibacter*) populations[68].

Microbial communities must withstand the temporal variations of environmental conditions to survive and thrive in their ecosystem. It is essential to understand how microbial community structure and function change over time to gather important information such as their diversity patterns, what controls these patterns, their stability over changes in conditions, interactions within microbial communities or the niche of single taxa[31]. Time-series over decades revealed that the seasonal patterns of microbial community structure were stable and predictable both for the abundant[69–71], but also the rare taxa[72]. In addition, rare taxa can occasionally bloom and dominate the population if conditions are favorable[73,74]. The seasonal recurrence of microbial communities suggests that physical and chemical processes control the succession dynamics. However, biological interactions between bacteria, archaea, viruses and eukaryotes such as molecule exchanges, viral lysis or grazing have been shown to also contribute to observed abundance patterns[70,75]. The cohesion of clearly defined communities revealed by high resolution time-series confirms the importance of biological interactions in shaping the temporal dynamics of microbial communities[76].

We have acquired a lot of knowledge about the marine microbiome. From the biogeography, ecology and important metabolic processes of dominant microbial taxa, to the structure of microbiome in the global ocean or the temporal succession of microbial communities. However, we are still looking at the tip of the iceberg. We do not have a full grasp of the many metabolic processes used by microbial communities, nor of their ecology and phylogenetic diversity in a wealth of marine environments, specifically remote and under-sampled regions. In addition, we possess scarce information on the activity of important metabolic processes and how they are linked to the transformation of inorganic and organic matter. There is therefore a tremendous need to explore in detail the metabolic capacity and activity of microbial communities from remote environments, as well as how they impact the geochemical cycles through the myriad of biochemical transformations they carry out.

1.2 Dissolved organic matter in the ocean

1.2.1 Nature and diversity of marine dissolved organic matter

On average, a liter of sea water contains less than 1 mg of dissolved organic matter (DOM), but summed over the global ocean, the standing stock of DOM contains ~660 x 10^{15} g of C[77]. The pool of marine DOM therefore roughly contains the same amount of C as terrestrial biomass (~610 x 10^{15} g of C) or atmospheric CO_2 (750 x 10^{15} g of C) and 200 times the C inventory in marine biomass (~3 x 10^{15} g of C)[78]. DOM is also composed of large quantities of nitrogen, phosphorus, iron and other elements essential to life[79]. DOM therefore represents a central component in marine biogeochemical cycles[80]. As a consequence, microbes, through the transformation of DOM, operate an important fraction of marine geochemical cycles[81].

The rates at which microbial communities process DOM is central to understand the role of microbial processes in the budget of C and other essential elements. Microbes, through respiration, recycle up to 40 x 10^{15} g of primary produced C every year[82]. Most of the DOM released by phytoplankton is turned over by micro heterotrophs within hours to days[83]. This DOM therefore rapidly disappears from the water column and the bulk of detectable DOM in the water column has an average age of 16,000 years[84]. This old DOM therefore resists microbial degradation. Efforts trying to explain the long-term persistence of DOM in the ocean led to the classification of DOM based on its turnover time: labile (hours to days), semi-labile (weeks to months), semi-refractory (decades), refractory (millennia) and ultra-refractory (tens of millennia)[84]. The labile DOC (dissolved organic carbon) represents less than 1% of the marine DOC standing stocks (<0.2 x 10^{15} g) and refractory DOC makes up by far the largest fraction (630 $\pm$ 32 x 10^{15} g of C) while the rest of DOC is divided between semi-labile (6 $\pm$ 2 x 10^{15} g), semi-refractory (14 $\pm$ 2 x 10^{15} g) and ultra-refractory DOC (> 12 x 10^{15} g)[84].

Labile DOM supports most of heterotrophic microbial production in the euphotic zone and has the largest flux of all the DOM fractions[80]. Most of labile DOM in the euphotic zone is primary-produced, coming from phytoplankton exudates and lysates[85]. In the mesopelagic and bathypelagic, labile DOM can also come from the release of molecules from heterotrophs lysates[70], solubilization of sinking particles[86] or chemoautotrophy[87]. Labile DOM is therefore mostly composed of molecules from known building blocks such as polysaccharides, proteins, lipids and DNA or RNA[82]. Semi-labile DOM, due to its longer residence time, can accumulate in the euphotic zone and therefore be transported horizontally with surface currents to further support heterotrophic C demand but also vertically, being the most important source of C export to deeper waters[88]. Refractory DOM is the dominant form of DOM in deep waters (< 1000 m), where the depth DOC concentration stays constant. It is composed of a wide diversity of molecules at very low concentrations, among which carboxyl-rich alicyclic molecules are the most recognizable component[89].

The long-term persistence of recalcitrant DOM in the ocean could play an important role in long-term carbon storage by the ocean. However, the reasons behind the long-term residence

time of recalcitrant DOM are still not understood. Two theories have been proposed. For the intrinsic recalcitrance concept, a variety of compounds resist microbial degradation due to their chemical nature[79]. In this concept, the production of refractory DOM is due to microheterotrophs transforming labile molecules from primary producers into refractory molecules[90]. Other processes can also participate to the formation of recalcitrant DOM such as heat[91] or light-induced transformation of labile fractions[92]. In the intrinsic recalcitrance concept, refractory DOM cannot be removed by heterotrophs and its removal happens through abiotic processes such as adsorption to sinking particles[93], photodegradation[94] or entrainement in the hydrothermal circulation within Earth's crust[95]. In the framework of the emergent recalcitrance concept, all DOM molecules are continuously being transformed[79,96]. DOM constituents form complex ecological networks with biota from all trophic levels (phytoplankton, bacteria, viruses, grazers, etc.). The production, transformation and consumption of DOM constituents take place in these networks. The recalcitrance therefore emerges as a property at an ecosystem level. In the emergent recalcitrance framework, the lability of a single molecule is contingent with many factors, such as its concentration and chemical nature, but also the presence and concentration of other molecules, the composition of the microbial networks or the environmental conditions (presence/absence of light, temperature, salinity, etc). All compounds are therefore continuously produced and consumed and their concentration depends on the complex dynamic equilibrium of the ecological networks they are in[97].

Marine DOM represents an extremely complex mixture of molecules. A quick calculation estimated the diversity of marine DOM to be > 600,000 different compounds[79], but the diversity of DOM compounds could reach millions[98]. Since it is extremely hard to characterize and quantify the huge diversity of marine DOM molecules, simple questions such as what are the diversity and distributions of DOM molecules in various marine environment are questions that remain largely unanswered. As such, the exact composition of marine DOM and its transformations by microbes remain an active research topic[99].

1.2.2 Terrestrial organic matter in the ocean

If primary production (∼50 x 10^{15} g of C/year) is the main source of DOM in the ocean (3-20 x 10^{15} g of C/year)[80], terrestrial DOM (tDOM) also contributes significantly to the budget of marine DOM (0.5 x 10^{15} g of C/year)[100] and therefore represents an important source of nutrients and energy for marine microbiomes. The major pathway of tDOM input to the ocean is through riverine delivery (0.40 x 10^{15} g of C/year)[101], but ground water (0.23 x 10^{15} g of C/year)[102,103] and coastal erosion[104] are important sources of tDOM too. Most of tDOM comes from the decay of terrestrial plants[105]. As lignin, an aromatic polymer, can only be synthesized by vascular plants, lignin and its aromatic degradation products have been used to follow the fate of tDOM in the marine environment[106]. The ocean shelves and margins have been considered hotspots of tDOM degradation as open ocean waters are characterized by low or undetectable levels of traditional lignin markers such as lignin phenols[107]. However, results from recent methods have demonstrated the presence of tDOM-derived compounds far into the

open ocean[108].

The transformation of tDOM in the marine environment could contribute to long-term storage of recalcitrant DOM in the ocean. This could be explained by estimation showing that 90% of the reactive aromatic fraction of tDOM is photo-transformed into a more stable non-aromatic fraction in the shelf[104]. In addition, it has been proposed that microbial alteration of less labile fractions of tDOM by marine microheterotrophs may be facilitated by the consumption of labile primary-produced DOM, a process called priming effect[109]. In this scenario, the shelves, with their high primary production rates and tDOM inputs are prime environments for the priming effect and transformation of tDOM to refractory DOM.

1.2.3 Colored organic matter in the ocean

It is extremely challenging to describe marine DOM at the molecular level. However, bulk properties of DOM can be used to obtain valuable information on the origin and fates of DOM in the marine environment. Between 20% and 70% of marine DOM interacts with the UV and visible spectra and is coined colored DOM (CDOM)[110]. As the chemical nature of molecules dictates their optical properties, optical measurements of DOM have been widely used in the ocean to identify and follow different DOM fractions[111]. CDOM properties can be quantified by their capacity to absorb light in the UV-visible range[112]. CDOM in the ocean can be produced by primary producers or heterotrophs as well as results from the biological or photochemical alteration of other DOM[113]. CDOM presence and light absorption depends on its sources and sinks. It has been shown that CDOM can be produced both in the surface[114] by primary production and in deep ocean as a by-product of microbial degradation of sinking particulate organic matter[115]. In the global ocean, CDOM distribution is characterized by minima in the surface, in particular in the tropics, due to solar bleaching, but higher CDOM absorption in mid to high latitudes in the northern hemisphere[111]. Local surface maxima can also occur in regions of high primary production, upwelling, or major river inputs[116]. CDOM absorption values generally increase with depth to reach maxima in the Indian and Pacific but not the Atlantic oceans thermocline[112]. The global ocean relationships between DOC and CDOM suggest that the observable marine CDOM is a refractory component of DOM in the deep ocean[115].

A fraction of CDOM fluoresces and the capacity of excitation/emission spectra to discriminate between different fluorescent DOM (FDOM) fractions provides a convenient way to elucidate the patterns of sources and sinks of DOM[111]. FDOM is made up of two major types: amino acid-like compounds ($FDOM_A$) and humic-like compounds ($FDOM_H$)[117]. $FDOM_A$ displays narrow emission peaks typically below 400 nm and its fluorescence spectra are almost similar to those of pure tyrosine and tryptophan[118]. $FDOM_A$ is therefore thought to represent freshly produced proteins by phytoplankton and bacteria[118]. $FDOM_H$ is characterized by broad emission peaks generally above 400 nm. The chemical composition of various $FDOM_H$ fraction is more enigmatic but it has been linked to the oxidation and degradation of organic

material from terrestrial and marine sources[119]. The $FDOM_A$ fractions are usually higher in the surface (0-200 m), and decrease in the mesopelagic zone (200-1000 m) to reach constant low level in deep waters (>1000 m)[118]. In contrast the levels of $FDOM_H$ are low in the surface and increase with depth, a trend that has been attributed to high sensitivity to light but also the microbial transformation of DOM in the mesopelagic and bathypelagic waters[120]. Fluorescence emission/excitation matrices (EEM) of marine DOM, combined with statistical analysis (Parallel Factor Analysis, PARAFAC) enables the identification of different FDOM fractions[121,122]. As an example, the terrestrial origin of FDOM fractions in contrast to the marine origin could be obtained using these methods[123]. EEM and PARAFAC enabled to discriminate different fractions and elucidate general trends in the origin and transformations of marine DOM[124,125]. However, the key challenge now remains to better understand the diversity of chemical compounds responsible for the optical properties of FDOM fractions.

1.2.4 How to elucidate the chemical composition of marine dissolved organic matter

In this thesis, we are interested in the metabolic processes of microbiomes and how they shape the geochemical cycles of the Arctic Ocean. Microbiomes control geochemical cycles through the chemical transformations of organic and inorganic matter from their environment. We therefore sought to characterize and quantify the molecular composition of organic matter in the water column and link it to the microbiome metabolic capacity. Metabolite profiling of microbial cultures' media cultures originally attempted to measure the diversity of biochemical compounds in microbes environment but quickly became used to differentiate the phenotype of closely related microbial strains or species as way for microbial identification[126,127]. These studies used various analytical methods to detect and quantify various metabolites. However, this is the pioneering work on mass-spectrometry based metabolomics for plants[128,129] that really allowed the non-targeted and systematic characterization of organic matter in microbiomes environment to take off.

Nowadays, nuclear magnetic resonance (NMR) and mass spectrometry (MS) coupled with separation techniques are widely used to study the chemical composition of organic matter in microbiomes environments[130]. NMR is highly reproducible and can quantify the identified chemical compounds. However, NMR has a poor sensitivity and separation capacity. Due to high sensitivity, quantitative and accurate mass determination in a high-throughput fashion, MS has become the method of choice for measuring and characterizing complex mixtures of organic matter[131]. MS detects and discriminates compounds by measuring their mass-to-charge ratios (m/z). MS techniques are usually separated into two stages: (i) the ionization of compounds, and (ii) measurement of their mass. To analyze complex samples such as marine organic matter samples, high throughput separation techniques (chromatograpy, capillary electrophoresis) are connected in series with MS to reduce the complexity of a given mass spectra[130]. Especially, liquid chromatography coupled to Fourier transform ion cyclotron resonance MS (FT ICR MS) fostered unique insights in the transformation of marine organic matter (OM) by microbiomes.

Mass spectrometry techniques, despite some ongoing developments, are very promising techniques to elucidate the molecular composition of marine organic matter. Dues to the immense complexity and diversity of compounds found in marine DOM, the grand challenge is to elucidate the links between DOM molecular composition and the metabolic capacity and activity of microbiomes.

1.2.5 Dissolved organic matter – microbe interactions

The diversity, quantities and molecular composition of marine DOM is important in shaping the microbiome structure and function. It was shown that OM from different phytoplankton species recruits and is degraded by different microbial consortia[132]. This could be explained by the fact that the interactions between DOM and microbiomes take place at a very high resolution, between species and individual molecules, but could not be detected at lower resolution (e.g. genus level and group of compounds level), revealing intricate interaction networks between DOM and microbiomes[133]. In addition, the progressive degradation of OM derived from a single phytoplankton species is associated with a succession in the microbial community composition, showing specific lineages involved in the transformation of specific compounds[134], further supporting close associations between the structure of DOM molecular composition and bacterial communities.

The microbial processing paths of OM dictates the fate of OM in the environment. The OM processed by microbes can either (i) be mineralized through respiration, (ii) used to build biomass that is grazed upon or (iii) released in the environment by viral lysis or other processes to feed the microbial loop and (iv) be transformed to metabolites that escape further processing and stays in the ocean for a long time, becoming refractory DOM[43]. Experimental work using metagenomics and molecular characterization of DOM demonstrated that only ∼0.4% of phytoplankton-derived OM is funneled to a DOM fraction that resembles refractory DOM from the deep ocean[135]. Despite its slow production rate, the long-term persistence of refractory DOM in the ocean allows an accumulation over time. The modes of production of RDOM is still an ongoing research area. Early studies demonstrated overall differences in OM composition between oceanic and estuarine waters or at different water depths[136,137], underlying different sources and microbial degradation processes. But FT ICR MS also showed that a large pool of marine DOM is indistinguishable in various and diverse samples[138]. This points to various modes of degradation and metabolic pathways leading to the formation of a universal marine DOM background, regardless of the OM source[136,138]. Studies proposed that the diversification of bacterial exometabolomes from single carbon sources and the resemblance to marine DOM they observed could be at the source of refractory DOM in the ocean[139].

Overall, marine DOM and microbial communities influence each other, and it is crucial to highlight these relationships to understand and predict the roles of marine microbiomes in shaping current and future biogeochemical cycles.

1.3 Microbiomes of the Arctic Ocean

1.3.1 Setting the stage: the Arctic Ocean under change

The role of the physicochemical environment is central in shaping microbiome composition and function. The ecological set-up of a microbiome is therefore of utmost importance to contextualize findings about its composition and function. Consequently, it is necessary to gain knowledge of factors shaping the physicochemical and ecological environment of the microbiomes we study. As this thesis focuses on microbiomes of the Arctic Ocean and what ecologically important metabolic processes differentiate them from other oceans, we will briefly introduce features that make the Arctic Ocean such a unique ocean.

Geography of the Arctic Ocean

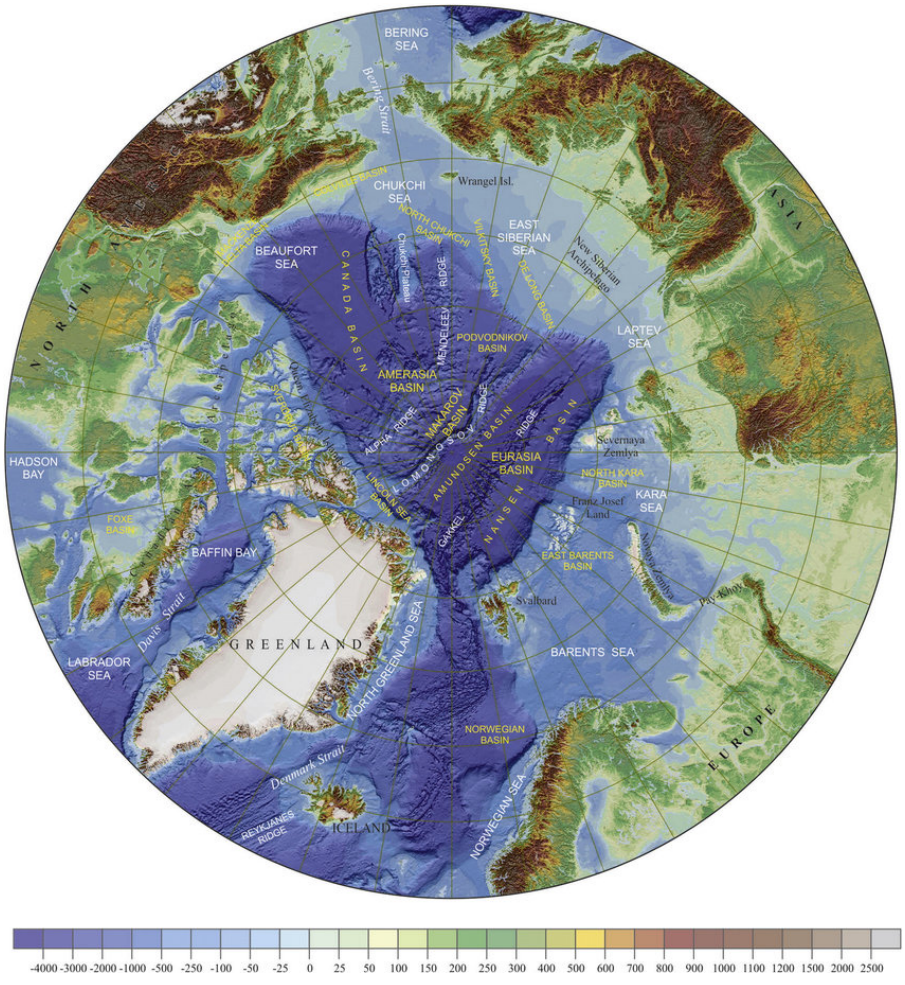

Figure 1.1: *Geography of the Arctic Ocean*

The Arctic Ocean is an enclosed ocean with small openings connecting to other oceans (Figure 1.1): the Bering Strait connects to the Pacific Ocean while channels of the Canadian archipelago and Fram Strait connect to the North Atlantic Ocean. The Arctic Ocean is constituted of large continental shelves surrounding four basins: the Canada Basin, the Makarov Basin, the Amundsen Basin and the Nansen basin. Together the continental shelves represent half of the total Arctic Ocean surface area and amount to a quarter of the World Ocean shelves while the Arctic Ocean only represents 1% of the total World Ocean volume[140]. In addition, with 11% of the world's river discharge, the Arctic Ocean collects the highest load of fresh water and associated terrestrial organic matter of any other oceans on a per volume basis[141].

Oceanography of the Arctic Ocean and the Canada Basin

The Canada Basin is the largest of the Arctic basins and is bordered by broad and shallow continental shelves (East Siberian and Chuckchi Sea shelves, Figure 1.1) to the west, narrow continental shelf to the south (Beaufort Shelf) and by the Canadian archipelago to the east. The Canada Basin is a highly stratified, oligotrophic ocean (Figure 1.2). The surface mixed layer water (0-∼50 m depth) is the freshest due to the influence of river runoff and ice melt and thaw cycles[142]. The stratification arises from a strong halocline (gradient of salinity where fresh water overlies more saline water, ∼50-250 m). The Canada Basin halocline is formed by an inflow of relatively fresh, nutrient-rich water from the Pacific through the Bering Strait that sinks under the fresher surface mixed layer. The salinity allows us to distinguish water from Pacific summer origin (32.3 PSU) and Pacific winter origin (33.1 PSU). Pacific waters mixing with underlying Atlantic water forms the bottom of the halocline (34-34.4 PSU). Below the halocline, we find the warmer and saline water of Atlantic origin, reaching the Canada Basin via two pathways. Atlantic water transported through Fram Strait (first pathway) creates a temperature maximum around 400 m and sits atop Atlantic water transported across and being modified in the Barents Sea shelf (second pathway)[143]. Saltier deep Arctic water extends below the Atlantic water up until the bottom.

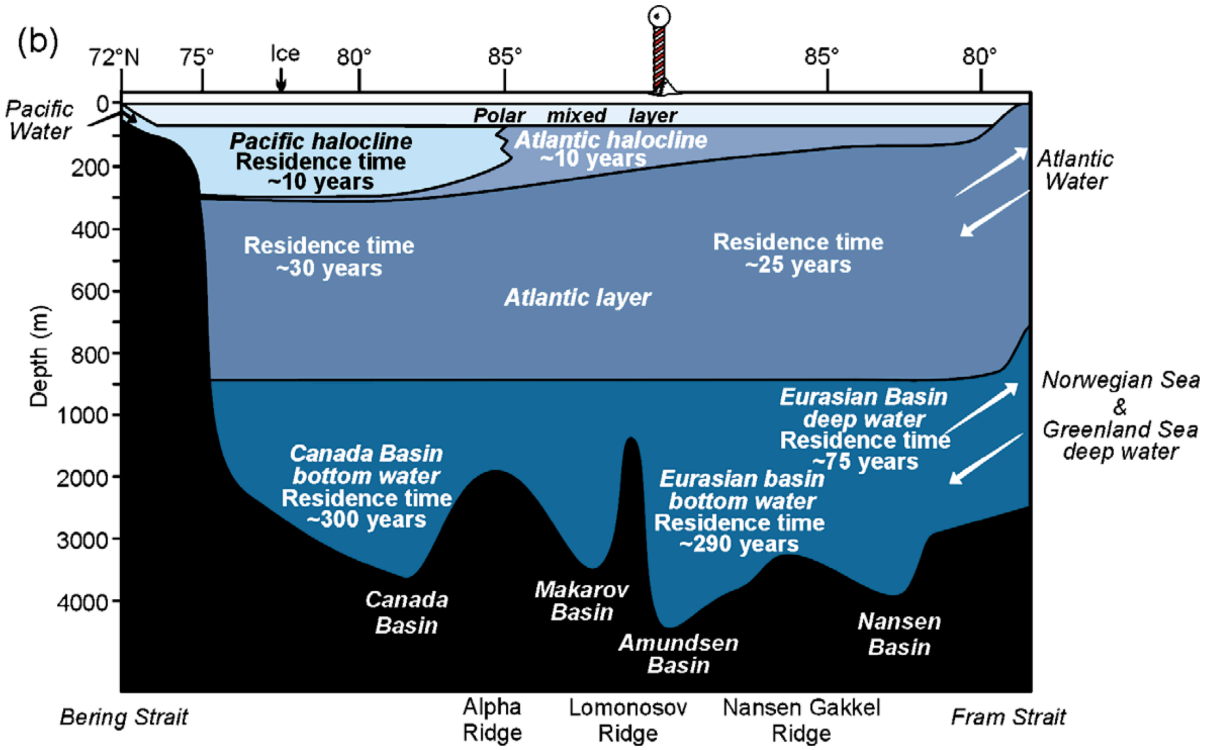

Figure 1.2: *Vertical startification of the Arctic Ocean. From MacDonald et al.*[144].

Seasonality in the Arctic Ocean

The Arctic Ocean is characterized by an intense seasonal forcing. In winter (September-March), it is cold and dark due to an ice cover blocking most of the very short period of daylight Figure 1.3. During spring and summer (March-September), the daylight period extends to cover most of the day. In parallel, the whole Arctic region warms up and a portion of the ocean ice cap melts away. This seasonal cycle has a profound effect on the life and carbon cycles in the Arctic Ocean. In the Canada Basin, the very low availability of light in winter prevents photosynthesis and hence primary production by phytoplankton[145]. During the winter period, the nutrients of the surface mixed layer is replenished through mixing with the deeper water masses[145]. However, the stronger stratification of the Canada Basin strongly limits this nutrient resupply[146]. Ice melt in the spring coupled with light availability trigger big phytoplankton blooms in the shelf areas[147]. In the Canada Basin, these blooms are very limited due to the scarcity of nutrients. The phytoplankton primary productivity in the Canada Basin is therefore more restricted to the subsurface water layer (subsurface chlorophyll maximum), where an optimum of light and nutrients is found[148]. The long winter period devoid of primary-produced organic matter imposes a big strain on the survival of microbiomes. However, the effect of the Arctic seasonality on the Arctic Ocean microbiomes is still poorly understood.

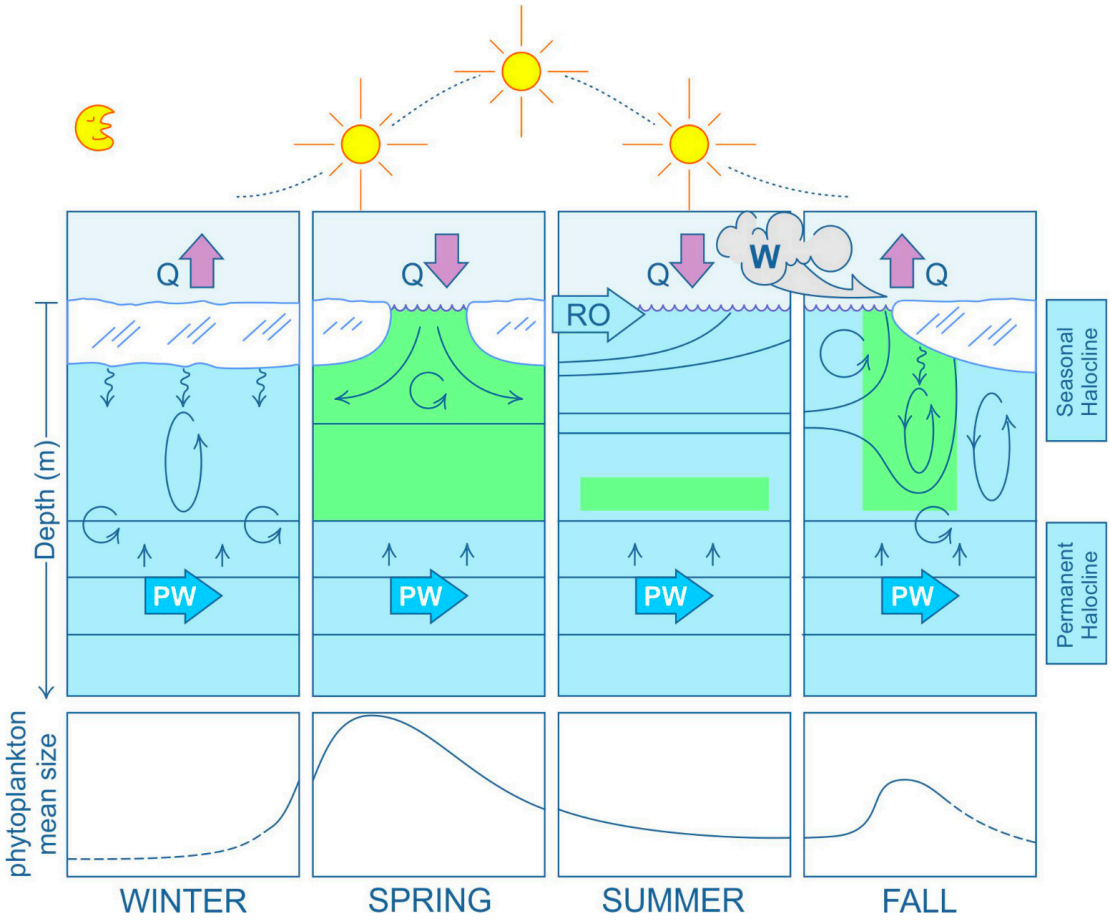

Figure 1.3: *Seasonal changes of environmental conditions in the Arctic Ocean. From Brown et al.*[149].

The seasonal cycle also dictates the timing of terrestrial organic matter input to the Arctic Ocean. As the drainage basin of the Arctic Ocean is twice its area, the spring freshet delivers substantial amount of terrestrial organic matter and nutrients to the surface waters of the continental shelves[150]. The labile fractions of this organic matter are modified in the surface during the spring and summer, leaving more recalcitrant fractions. In the fall, brine derived from ice formation, sinks and carry with it the recalcitrant fractions of terrestrial organic matter. Brine flows off the shelf, also carrying sinking primary produced particulate matter and exchanging organic matter with the shelf sediments and ends up in the halocline of the Arctic basins[151]. The halocline is therefore enriched with nutrients and more recalcitrant fractions of terrestrially, sediment and primary-produced organic matter[150,151]. Some studies have highlighted the capacity to use tDOM in taxa from the Arctic Halocline[23]. However, we still do notunderstand the extent of this capacity in the Arctic Ocean microbiomes, and how it compares with other regions with a weaker tDOM signature.

The Arctic Ocean under change

The Arctic is the region with the most prominent anthropogenic-fueled warming in the world[152]. In the Arctic Ocean, the most visible consequence is the loss of sea ice extent and thickness. During the last half century, the September sea ice extent has decreased by 0.8

million square kilometer per decade[153], with the largest sea ice minimum recorded in 2012. The loss of sea ice is the tip of the iceberg, as warming in the Arctic region has more profound consequences. Globally, the extended light period associated with the decline of sea ice extent has favored longer growth period for phytoplankton resulting in an increase of 30% in the primary productivity between 1998 and 2012[154]. However, the change in primary productivity is highly heterogeneous throughout the Arctic due to large differences in the local availability of nutrients. The highest increases were observed in the interior shelves (70-112% increase) as nutrients are replenished by deep water upwelling[154] and increasing freshwater inputs. In contrast, the Canada Basin has seen a 79% and 29% decrease in nitrates and phosphates respectively in the last three decades[146]. The increasing fresh water addition from sea-ice melt and riverine input, resulting in increased stratification, has led to the oligotrophication of the Canada Basin by preventing nutrient resupply from deep waters. As stratification is expected to increase in the Canada Basin[155], but not in other regions of the Arctic Ocean[156], the Canada Basin may be the only region of the Arctic Ocean showing a decreasing primary productivity with the ongoing climate change. Warming of the Arctic region also has a profound impact on the input of terrestrial organic matter to the Arctic Ocean. As temperatures rise, the volume of freshwater input to the Arctic Ocean is increasing, carrying larger amounts of terrestrial organic matter[150]. The thawing of permafrost further enhances the mobilization of long-stored organic matter that ends-up in the Arctic Ocean[141]. The yearly discharge of organic matter from terrestrial sources is therefore expected to increase significantly as the Arctic Ocean continues to warm up[141]. It is important to understand how these changes may affect the microbiomes structure and function to predict the evolution of their impact on Arctic geochemical cycles.

Dissolved organic matter in the Arctic Ocean

The transformation of tDOM to refractory DOM could be strongly enhanced in the Arctic Ocean due to its unique situation: the Arctic receives 11% of the total tDOM input but represents only 1% of the total world ocean volume. In addition, half of the Arctic Ocean is constituted of continental shelves[157]. The concentrations of lignin phenols were estimated to be 10 times higher in the surface waters of the Arctic Ocean (360 ng/L) compared to the Atlantic (36 ng/L) and Pacific oceans (25 ng/L) but similar in deep waters (60 ng/L in the Arctic and Atlantic, 28 ng/L in the Pacific)[158], suggesting an important transformation of lignin phenols in the Arctic Ocean. In addition, it has recently been shown that nutrients from riverine input and erosion sustain a third of the Arctic primary productivity[159]. The resulting synchronizing of tDOM input and primary productivity may therefore enhance the priming effect to transform tDOM into refractory DOM in the Arctic Ocean. In the Canada Basin, the use of EEM and PARAFAC demonstrated a disproportionately large FDOM maximum in the halocline compared to other oceans, associated with organic matter from terrestrial origin[122,160]. In addition, the importance of terrestrial FDOM fractions increased from 2007 to 2017, concurrently with sea-ice loss in the Canada Basin[125]. However, the details of tDOM transformation by microbes in the Arctic Ocean remain vastly unexplored. It would be valuable to determine if the Arctic Ocean microbiomes remineralize tDOM or transform it into refractory DOM.

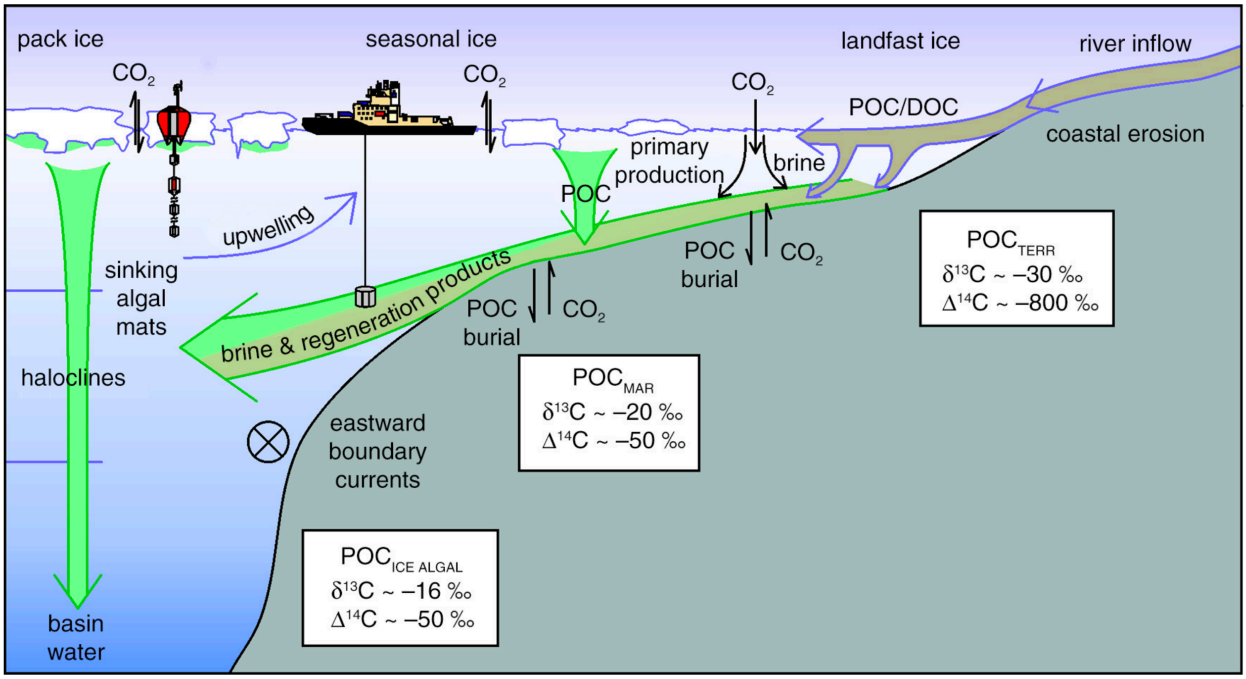

Figure 1.4: *Schematics of the sources and sinks of carbon in the Arctic Ocean. From Anderson et al.*[150].

1.3.2 The Arctic Ocean microbiome

Microbial communities seem to be adapted to Arctic conditions and show significant differences in assemblage compared to lower latitudes oceans. Several studies have highlighted marked differences in the composition of microbial communities between the Arctic Ocean and other oceanic zones[66,67]. Pedrós-Alió *et al.*[161] explored the International Census of Marine Microbes (ICoMM). On average, in temperate waters, for the most abundant groups they found 50% of *Alphaproteobacteria*, 30% of *Gammapreotebacteria*, 10% of *Bacteroidetes* and 4% of *Cyanobacteria*. In the Arctic Ocean, *Alphaproteobacteria* were considerably less present with 20%, on average, the same percentage as *Gammaproteobacteria*. These proportions were however highly variable depending on the sampling site. This could be attributed to the fact that *Alphaproteobacteria* are in general better adapted to the open ocean with more stable conditions throughout the year while *Gammaproteobacteria* do better in seasonally ice-covered ocean with strong variation in conditions. *Betaproteobacteria* were significantly more abundant in the Arctic Ocean, reaching 8% for the inshore communities[162]. *Betaproteobacteria* are one of the most prevalent group in freshwater system. The high amount of fresh water input to the Arctic and the persistence of a surface layer with low salinity are probably responsible for the persistence of *Betaproteobacteria* in the Arctic Ocean[162]. This is supported by the disap-

pearance of *Betaproteobacteria* with salinity increase from deltas to open ocean in temperate systems[163]. *Bacteroidetes* are also more prevalent in the Arctic Ocean (12 to 50%) compared with global average in temperate waters (4%). Higher percentage of *Bacteroidetes* in Southern oceans (Antarctica) suggests this observation stems from adaptation to cold conditions but the driving force behind it is still unknown.

Adaptations to conditions of the Arctic have also been shown in individual taxa. A study identified taxa found only in the Arctic Ocean[66]. Our lab showed the existence of *SAR11* ecotypes whose global distribution was restricted to the Arctic Ocean[164]. Another study attributed the adaptation of some *Chloroflexi* taxa restricted to the Arctic Ocean to their capacity to degrade terrestrial OM thanks to their unusually large number of aromatic compound degradation genes[165]. The relatively fresh surface waters of the Arctic Ocean also explained the presence of a previously undescribed *Methylophilaceae* clade that has undergone a freshwater to marine transition[24]. Rapid adaptation mechanisms to stressor in the microbiomes of the Arctic Ocean may also be different than in other oceans. Using metagenomes and metatranscriptomes, one study of the global ocean microbiome showed that the relative contribution of gene expression changes to the metatranscriptome differences along environmental gradients was lower in polar water than non-polar waters[29]. This suggests that in polar zones, changes in the community activity will be driven mainly by changes in community composition rather than by gene regulation mechanisms. However, experiments also showed that Arctic Ocean microbiomes prioritize functional restructuring over taxonomy following a rapid perturbations in the composition of OM simulating a phytoplankton bloom[166].

1.3.3 Seasonality of the Arctic Ocean microbiome

Despite strong seasonal changes, microbial communities must somehow maintain their composition and function over the years to maintain the biogeochemical cycles in the Arctic. It is therefore important to understand how microbial communities change through the seasons in the Arctic. Practically, the complexity of sampling the Arctic Ocean water in winter (harsh conditions, ice-covered water, etc.) has prevented extended investigations. However, a few studies have attempted to highlight seasonal changes in the Arctic Ocean microbiomes structure and function. Kirchman *et al.*[167] surprisingly did not notice significant differences of community structure between the sites sampled in winter and those in summer (Beaufort sea, Chucki sea and Franklin bay) using 16S rRNA sequencing. On the phylum levels, winter and summer communities were highly similar, with the exception of higher abundance of non-*Flavobacteria Bacteroidetes* in winter. At lower phylogenetic levels, they observed some differences between summer and winter. Such as for the *SAR11* clade for which only 25% of the ribotypes were common to winter and summer samples, but ribotypes unique to each season were not abundant. Using fluorescence in situ hybridization (FISH), Alonso-Sáez *et al.*[168] showed significant differences in composition during summer and winter. Archaea, in particularly *Crenarchaea*, were more abundant in winter (up to 12% compared to 1-2% in high summer), while bacteria overly dominated the communities in summer with all major phyla (*Alphaproteobacteria, Gammaproteobacteria,*

Bacteroidetes) increasing in abundance. Similarly, autonomous sampling throughout the year revealed a marked annual cycle in the microbial communities composition[169]. Interestingly a single ribotype of *Betaproteobacteria* (*Janthinobacterium*) was reported to bloom in winter to reach 20% of the total community in the Admunsen gulf[74]. This is even more surprising considering the fact that *Betaproteobacteria* usually show low abundances in Arctic waters. Han *et al.*[170] elucidated microbial community composition (using 16S tag sequencing) during the ice melt in the Western Arctic. This allowed them to collect the surface water communities along a gradient of ice coverage (or the equivalent of different stages of ice melting), and consequently a gradient of salinity. They showed that *Alphaproteobacteria* abundances correlated negatively while *Flavobacteria* correlated positively with salinity. They associated the change of salinity with ice melt (salinity decreasing with decreasing ice cover/increasing ice melt), and therefore ice melt as the driver of community composition change. Similarly, the presence of sea ice dictates the composition of phytoplankton communities as well as free-living and particle-attached microbial communities[171]. The unique community composition of the Arctic Ocean microbiomes suggests differences in the repertoire of functional genes and metabolic processes of the Arctic Ocean microbiome compared to temperate oceans. However, despite the importance of the microbiome metabolism in Arctic geochemical cycles, we only possess very scarce information on the metabolism of the Arctic Ocean microbiome, its phylogenetic diversity and how it compares to the rest of the global ocean.

1.4 Objectives and structure of the thesis

This thesis work revolves around understanding the suite of metabolic processes that distinguish the Arctic Ocean microbiome from the microbiomes of the rest of the global ocean and how they are linked to the composition of organic matter and the unique biogeochemical situation of the Arctic Ocean. The few studies touching on the metabolism of the Arctic Ocean microbiome have so far been limited to very particular processes and the eco-evolutionary history of the few taxa carrying them on. There is a lack of knowledge regarding the variety and importance of the Arctic Ocean microbial metabolic processes. In addition, we poorly understand how the metabolism of the Arctic Ocean microbiome distinguishes itself from the global ocean and the potential consequences on the transformation and composition of the organic matter in biogeochemical cycles.

This thesis aims at filling this knowledge gap by systematically exploring the metabolism of the Arctic Ocean microbiomes and contrasting it with the global ocean to highlight what distinguishes the Arctic Ocean within the global ocean. This work also seeks to link the distinguishable metabolic features of the Arctic Ocean microbiomes to the transformation and composition of the dissolved organic matter. Finally, this thesis aims to uncover the ecological importance of the microbial actors undertaking the metabolic processes distinguishing the Arctic Ocean within the global ocean.

In chapter 2, we demonstrated that metabolic pathways involved in the degradation of aromatic compounds are enriched in a subsurface layer characterized by the maximum of fluorescent dissolved organic matter (FDOMmax). This FDOMmax corresponds to the accumulation of terrestrial organic matter that has washed off to the Arctic Ocean. We confirmed that the capacity to degrade aromatic compounds in the Arctic Ocean microbiomes reveals their capacity to process terrestrial organic matter by showing that the most abundant aromatic compound degradation pathways targeted a variety of aromatic compounds typical of the local riverine discharge. Finally, by reconstructing a set of 664 metagenome-assembled genomes (MAGs), we discovered that the capacity to process aromatic compounds and terrestrial organic matter was phylogenetically concentrated in *Rhodospirillales*, an *Alphaproteobacteria* group. We showed that the distribution of most of these MAGs was restricted to the Arctic Ocean, and that they were enriched in aromatic compound degradation genes compared to their most closely related sisters from other oceans. These findings suggest a role for aromatic compound degradation for the adaptation of *Rhodospirillales* in the Arctic Ocean.

In chapter 3, we investigated the metabolism of neutral lipids (NLs) in the Arctic Ocean microbiomes. The accumulation of NLs is a ubiquitous way to store carbon in living organisms. We demonstrated that the Arctic Ocean microbiomes invest a higher fraction of their gene pool in the synthesis of NLs compared to other oceans. We highlighted the important contribution of eukaryotic phytoplankton in the synthesis of NLs within the photic zone, but also uncovered an unexpected diversity of bacterial taxa with the genomic capacity to synthesize NLs throughout the water column. Within these taxa, the capacity to synthesize different NLs was phylogenetically segregated and co-occurred with the metabolism of different carbon sources. In addition, we showed that a significant part of the microbiomes can use NLs as a growth substrate. These findings highlighted the importance of NLs in the Arctic Ocean and led us to postulate that NLs of bacterial origin may support the food web during the winter polar night.

In chapter 4, we extended our study to the global ocean to systematically highlight all the metabolic processes that distinguished the Arctic from other oceans. We confirmed and strengthened findings of chapter 2 and 3. However, the chief discovery was the prevalence of a suite of metabolic processes involved in cold adaptations mechanisms in the polar oceans' photic zone, among which the biosynthesis of glycans (polysaccharides) was dominant. Elucidation of the chemical composition of organic matter in the Arctic Ocean revealed a signature of sugars and their transformations in the photic zone, as well as aromatic compounds in the FDOMmax. Altogether, these results highlight the importance of terrestrial organic matter as a growth resource for the microbiomes and geochemical cycles of the Arctic Ocean, and reveal an unsuspected importance of glycans for the cold adaption of Arctic microbiomes.

Bibliography

[1] W.B. Whitman, D.C. Coleman, and W.J. Wiebe. *Prokaryotes: The unseen majority*. **Proc. Natl. Acad. Sci. U. S. A.**, 95:6578–83, 1998.

[2] G. Berg, D. Rybakova, D. Fischer, et al. *Microbiome definition re-visited: old concepts and new challenges*. **Microbiome**, 8:103, 2020.

[3] E.K. Hall, E.S. Bernhardt, R.L. Bier, et al. *Understanding how microbiomes influence the systems they inhabit*. **Nat. Microbiol**, 3:977–82, 2018.

[4] P.G. Falkowski, T. Fenchel, and E.F. Delong. *The Microbial Engines That Drive Earth 's Biogeochemical Cycles*. **Science**, 320:1034–9, 2008.

[5] L.R. Thompson, J.G. Sanders, D. McDonald, et al. *A communal catalogue reveals Earth's multiscale microbial diversity*. **Nature**, 551:457–63, 2017.

[6] D.M. Downs. *Understanding microbial metabolism*. **Annu. Rev. Microbiol.**, 60:533–59, 2006.

[7] E.B. Kujawinski. *The Impact of Microbial Metabolism on Marine Dissolved Organic Matter*. **Annu. Rev. Mar. Sci.**, 3:567–99, 2011.

[8] B. Bajic and A. Sanchez. *The ecology and evolution of microbial metabolic strategies*. **Curr. Opin. Biotechnol.**, 62:123–8, 2020.

[9] E.H. Allison and H.R. Bassett. *Climate change in the oceans: Human impacts and responses*. **Science**, 350(6262):778–82, 2015.

[10] O. Prakash, Y. Shouche, K. Jangid, and J.E. Kostka. *Microbial cultivation and the role of microbial resource centers in the omics era*. **Appl. Microbiol. Biotechnol.**, 97(1):51–62, 2013.

[11] I. Joint, M. Mühling, and J. Querellou. *Culturing marine bacteria - an essential prerequisite for biodiscovery*. **Microb. Biotechnol.**, 3(5):564–75, 2010.

[12] C.R. Woese and G.E. Fox. *Phylogenetic structure of the prokaryotic domain: The primary kingdoms*. **Proc. Natl. Acad. Sci. U. S. A.**, 74(11):5088–90, 1977.

[13] T.B. Britschgi and S.J. Giovannoni. *Phylogenetic analysis of a natural marine bacterioplankton population by rRNA gene cloning and sequencing*. **Appl. Environ. Microbiol.**, 57(6):1707–13, 1991.

[14] J. Handelsman, M.R. Rondon, S.F. Brady, et al. *Molecular biological access to the chemistry of unknown soil microbes: A new frontier for natural products*. **Chem. Biol.**, 5:R245–49, 1998.

[15] T.M. Schmidt, E.F. DeLong, and N.R. Pace. *Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing*. **J. Bacteriol.**, 173(14):4371–8, 1991.

[16] J.L. Stein, T.L. Marsh, K.Y. Wu, et al. *Characterization of uncultivated prokaryotes: Isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon*. **J. Bacteriol.**, 178(3):591–9, 1996.

[17] G.W. Tyson, J. Chapman, P. Hugenholtz, et al. *Community structure and metabolism through reconstruction of microbial genomes from the environment*. **Nature**, 428:37–43, 2004.

[18] C.J. Venter, K. Remington, J.F. Heidelberg, et al. *Environmental Genome Shotgun Sequencing of the Sargasso Sea*. **Science**, 304:66–74, 2004.

[19] S. Goodwin, J.D. McPherson, and W.R. McCombie. *Coming of age: ten years of next-generation sequencing technologies*. **Nat. Rev. Genet.**, 17:333–51, 2016.

[20] H-P. Grossart, R. Massana, K.D. McMahon, and D.A. Walsh. *Linking metagenomics to aquatic*

microbial ecology and biogeochemical cycles. **Limnol. Oceanogr.**, 65:S2–S20, 2020.

[21] T.J. Sharpton. *An introduction to the analysis of shotgun metagenomic data*. **Front. Plant Sci.**, 5: 209, 2014.

[22] T.J. Sharpton. *A review of computational tools for generating metagenome-assembled genomes from metagenomic sequencing data*. **Comput. Struct. Biotechnol. J.**, 19:6301–14, 2021.

[23] D. Colatriano, P.Q Tran, C. Gueguen, W.J. Williams, C. Lovejoy, et al. *Genomic evidence for the degradation of terrestrial organic matter by pelagic Arctic Ocean Chloroflexi bacteria*. **Commun. Biol.**, 1(1)):90, 2018.

[24] A. Ramachandran, S. McLatchie, and D.A. Walsh. *A Novel Freshwater to Marine Evolutionary Transition Revealed within Methylophilaceae Bacteria from the Arctic Ocean*. **MBio**, 12(3):e0130621, 2021.

[25] M. Peimbert and L.D. Alcaraz. *Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing, Chapter: A Hitchhiker's Guide to Metatranscriptomics*. Springer International Publishingr, 2016.

[26] D.E. Helbling, M. Ackermann, K. Fenner, et al. *The activity level of a microbial community function can be predicted from its metatranscriptome*. **ISME J.**, 6:902–4, 2012.

[27] M.A. Moran, B. Satinsky, S.M. Gifford, et al. *Sizing up metatranscriptomics*. **ISME J.**, 7:237–43, 2013.

[28] Y. Shi, G.W. Tyson, J.M. Eppley, and E.F. DeLong. *Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean*. **ISME J.**, 5:999–1013, 2011.

[29] G. Salazar, L. Paoli, A. Alberti, et al. *Gene Expression Changes and Community Turnover Differentially Shape the Global Ocean Metatranscriptome*. **Cell**, 179:1068–83, 2019.

[30] F. Baltar and G.J. Herndl. *Ideas and perspectives: Is dark carbon fixation relevant for oceanic primary production estimates?* **Biogeosciences**, 16:3793–9, 2019.

[31] L.J. Stal and M.S. Cretoiu. *The marine microbiome*. Springer, 2022.

[32] P. Flombaum, J.L. Gallegos, R.A. Gordillo, et al. *Present and future global distributions of the marine Cyanobacteria Prochlorococcus and Synechococcus*. **Proc. Natl. Acad. Sci. U. S. A.**, 110(24): 9824–9, 2013.

[33] M. Koblizek. *Ecology of aerobic anoxygenic phototrophs in aquatic environments*. **FEMS Microbiol. Rev.**, 39:854–70, 2015.

[34] A-C. Lehours, F. Enault, D. Boeuf, and C. Jeanthon. *Biogeographic patterns of aerobic anoxygenic phototrophic bacteria reveal an ecological consistency of phylogenetic clades in different oceanic biomes*. **Sci. Rep.**, 8:4105, 2018.

[35] R. Lami, M.T. Cottrell, J. Ras, et al. *High abundances of aerobic anoxygenic photosynthetic bacteria in the South Pacific Ocean*. **Appl. Environ. Microbiol.**, 73(13):4198–205, 2007.

[36] O. Béjà, L. Aravind, E.V. Koonin, et al. *Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea*. **Science**, 289:1902–6, 2000.

[37] J.A. Fuhrman, M.S. Schwalbach, and U. Stingl. *Proteorhodopsins: An array of physiological roles?* **Nat. Rev. Microbiol.**, 6:488–94, 2008.

[38] H.W. Ducklow, D.L. Kirchman, H.L. Quinby, et al. *Stocks and dynamics of bacterioplankton carbon during the spring bloom in the eastern North Atlantic Ocean*. **Deep. Res. Part II Top. Stud. Oceanogr.**, 40(1/2):245–63, 1993.

[39] M.J. Church. *Microbial Ecology of the Oceans, Chapter 10: Resource Control of Bacterial Dynamics in the Sea*. John Wiley & Sons, Ltd, 2008.

[40] S.J. Giovannoni. *SAR11 Bacteria: The Most Abundant Plankton in the Oceans*. **Ann. Rev. Mar. Sci.**, 9:12.1–12.25, 2017.

[41] S.J. Giovannoni, H.J. Tripp, S. Givan, et al. *Genome streamlining in a cosmopolitan oceanic bacterium*. **Science**, 309:1242–5, 2005.

[42] S.E. Noell, G.E. Barrell, C. Suffridge, et al. *SAR11 cells rely on enzyme multifunctionality to metabolize a range of polyamine compounds*. **MBio**, 12:e01091–21, 2021.

[43] A. Buchan, G.R. LeCleir, C.A. Gulvik, and J.M. González. *Master recyclers: features and functions of bacteria associated with phytoplankton blooms*. **Nat. Rev. Microbiol.**, 12:686–98, 2014.

[44] D.M. Needham, R. Sachdeva, and J.A. Fuhrman. *Ecological dynamics and co-occurrence among marine phytoplankton, bacteria and myoviruses shows microdiversity matters*. **ISME J.**, 11:1614–29, 2017.

[45] J. Arístegui, J.M. Gasol, C.M. Duarte, and G.J. Herndl. *Microbial oceanography of the dark ocean's pelagic realm*. **Limnol. Oceanogr.**, 54(5):1501–29, 2009.

[46] M.B. Karner, E.F. DeLong, and D.M. Karl. *Archaeal dominance in the mesopelagic zone of the Pacific Ocean*. **Nature**, 409:507–10, 2001.

[47] F.M. Lauro and D.H. Bartlett. *Archaeal dominance in the mesopelagic zone of the Pacific Ocean*. **Nature**, 12:15–25, 2008.

[48] S. Sunagawa, L.P. Coelho, S. Chaffron, et al. *Structure and function of the global ocean*. **Science**, 348(6237):1261359, 2015.

[49] G.J. Herndl, T. Reinthaler, E. Teira, et al. *Contribution of Archaea to total prokaryotic production in the deep atlantic ocean*. **Appl. Environ. Microbiol.**, 71(5):2303–9, 2005.

[50] H. Agogué, M. Brink, J. Dinasquet, and G.J. Herndl. *Major gradients in putatively nitrifying and non-nitrifying Archaea in the deep North Atlantic*. **Nature**, 456:788–92, 2008.

[51] M. Könneke, A.E. Bernhard, J.R. de la Torre, et al. *Isolation of an autotrophic ammonia-oxidizing marine archaeon*. **Nature**, 437:543–6, 2005.

[52] P.W. Boyd, H. Claustre, M. Levy, et al. *Multi-faceted particle pumps drive carbon sequestration in the ocean*. **Nature**, 568:327–35, 2019.

[53] J.A. Cram, L.C. Xia, D.M. Needham, et al. *Cross-depth analysis of marine bacterial networks suggests downward propagation of temporal changes*. **ISME J**, 9:2573–86, 2015.

[54] R. Stocker. *Marine Microbes See a Sea of Gradients*. **Science**, 338:628–33, 2012.

[55] S. Acinas, P. Sánchez, G. Salazar, et al. *Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities*. **Commun. Biol.**, 4:604, 2021.

[56] Z. Landry, B.K. Swan, G.J. Herndl, et al. *SAR202 genomes from the dark ocean predict pathways for the oxidation of recalcitrant dissolved organic matter*. **MBio**, 8(2):e00413–17, 2017.

[57] E.F. DeLong, C.M. Preston, T. Mincer, et al. *Community genomics among stratified microbial assemblages in the ocean's interior*. **Science**, 311:496–503, 2006.

[58] B. Balino, M. Fasham, and M. Bowles. *Ocean biogeochemistry and global change: JGOFS research highlights 1988-2000*. **IGBP Science**, 2, 2011.

[59] M.L. Sogin, H.G. Morrison, J.A. Huber, and others. *Microbial diversity in the deep sea and the underexplored "rare biosphere"*. **Proc. Natl. Acad. Sci. U. S. A.**, 103(32):12115–20, 2006.

[60] L. Amaral-Zettler, L.F. Artigas, J. Baross, et al. *Life in the World's oceans: Diversity, Distribution, and Abundance, Chapter 12: A global census of marine microbes*. Blackwell Publishing, 2010.

[61] J.A. Cram, C-E. T. Chow, R. Sachdeva, et al. *Seasonal and interannual variability of the marine bacterioplankton community throughout the water column over ten years*. **ISME J.**, 9:563–80, 2015.

[62] M. Mestre, C. Ruiz-Gonzalez, R. Logares, et al. *Sinking particles promote vertical connectivity in the ocean microbiome*. **Proc. Natl. Acad. Sci. U.S.A.**, 2018.

[63] F.M. Ibarbalz, N. Henry, M.C. Brandao, et al. *Global Trends in Marine Plankton Diversity across Kingdoms of Life*. **Cell**, 179:1084–97, 2019.

[64] A.S. Amend, T.A. Oliver, L.A. Amaral-Zettler, et al. *Macroecological patterns of marine bacteria on a global scale*. **J. Biogeogr.**, 40:800–11, 2013.

[65] J-F. Ghiglione, P.E. Galand, T. Pommier, et al. *Pole-to-pole biogeography of surface and deep marine bacterial communities*. **Proc. Natl. Acad. Sci. U.S.A.**, 109(43):17633–8, 2012.

[66] M. Royo-Llonch, P. Sanchez, C. Ruiz-Gonzalez, et al. *Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean*. **Nat. Microbiol.**, 2021.

[67] S. Cao, W. Zhang, W. Ding, et al. *Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics*. **Microbiome**, 8:47, 2020.

[68] M.L. Coleman and S.W. Chisholm. *Ecosystem-specific selection pressures revealed through comparative population genomics*. **Proc. Natl. Acad. Sci. U. S. A.**, 107(43):18634–9, 2010.

[69] J.A. Fuhrman, I. Hewson, M.S. Schwalbach, et al. *Annually reoccurring bacterial communities are predictable from ocean conditions*. **Proc. Natl. Acad. Sci. U. S. A.**, 103(35):13104–9, 2006.

[70] C. Bunse and J. Pinhassi. *Marine Bacterioplankton Seasonal Succession Dynamics*. **Trends Microbiol.**, 25(6):494–505, 2017.

[71] S. Lambert, M. Tragin, J-C. Lozano, et al. *Rhythmicity of coastal marine picoeukaryotes, bacteria and archaea despite irregular environmental perturbations*. **ISME J.**, 13:388–401, 2019.

[72] L. Alonso-Saez, L. Diaz-Perez, and X.A.G. Moran. *The hidden seasonality of the rare biosphere in coastal marine bacterioplankton*. **Environ. Microbiol.**, 17(10):3766–80, 2015.

[73] J.A. Gilbert, J.A. Steele, J.G. Caporaso, et al. *Defining seasonal marine microbial community dynamics*. **ISME J.**, 6:298–308, 2012.

[74] L. Alonso-Saez, M. Zeder, T. Harding, et al. *Winter bloom of a rare betaproteobacterium in the Arctic Ocean*. **Front. Microbiol.**, 5:425, 2014.

[75] D.M. Needham and J.A. Fuhrmans. *Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom*. **Nat. Microbiol.**, 1:16005, 2016.

[76] A.M. Martin-Platero, B. Clreary, K. Kauffman, et al. *High resolution time series reveals cohesive but short-lived communities in coastal plankton*. **Nat. Commun.**, 9:266, 2018.

[77] H. Ogawa and E. Tanoue. *Dissolved Organic Matter in Oceanic Waters*. **J. Oceanogr.**, 59:129–47, 2003.

[78] D.A. Hansell, C.A. Carlson, D.J. Repeta, and R. Schlitzer. *Dissolved organic matter in the ocean: a controversy stimulates new insights*. **Oceanography**, 22(4):202–11, 2009.

[79] T. Dittmar, S.T. Lennartz, H. Buck-Wiese, et al. *Enigmatic persistence of dissolved organic matter in the ocean*. **Nat. Rev. Earth Environ.**, 2:570–83, 2021.

[80] S. Wagner, F. Schubotz, K. Kaiser, et al. *Soothsaying DOM: A Current Perspective on the Future of Oceanic Dissolved Organic Carbon*. **Front. Mar. Sci.**, 7:341, 2020.

[81] F. Azam and F. Malfatti. *Microbial structuring of marine ecosystems*. **Nat. Rev. Microbiol.**, 5: 782–91, 2007.

[82] T. Dittmar and A. Stubbins. *Treatise on Geochemistry: Second Edition, 12.6: Dissolved Organic Matter in Aquatic Systems*. Elsevier Ltd., 2013.

[83] B. Pontiller, S. Martinez-Garcia, D. Lundin, and J. Pinhassi. *Labile Dissolved Organic Matter Compound Characteristics Select for Divergence in Marine Bacterial Activity and Transcription*. **Front. Microbiol.**, 11:588778, 2020.

[84] D. Hansell. *Recalcitrant Dissolved Organic Carbon Fractions*. **Ann. Rev. Mar. Sci.**, 5:421–45, 2013.

[85] B. Kieft, Z. Li, S. Bryson, et al. *Phytoplankton exudates and lysates support distinct microbial consortia with specialized metabolic and ecophysiological traits*. **Proc. Natl. Acad. Sci.**, 118(41): e2101178118, 2021.

[86] F. Azam and R.A. Long. *Sea snow microcosms*. **Nature**, 1414:495–8, 2001.

[87] M. Hügler and S.M. Sievert. *Beyond the Calvin cycle: Autotrophic carbon fixation in the ocean*. **Ann. Rev. Mar. Sci.**, 3:261–89, 2011.

[88] D.A. Hansell, C.A. Carlson, and R. Schlitzer. *Net removal of major marine dissolved organic carbon fractions in the subsurface ocean*. **Global Biogeochem.**, 26:GB1016, 2012.

[89] N. Hertkorn, R. Benner, M. Frommberger, et al. *Characterization of a major refractory component of marine dissolved organic matter*. **Geochim. Cosmochim. Acta**, 70:2990–3010, 2006.

[90] P.F. Hach, H.K. Marchant, A. Krupke, et al. *Rapid microbial diversification of dissolved organic matter in oceanic surface waters leads to carbon sequestration*. **Sci. Rep.**, 10:13025, 2020.

[91] T. Dittmar and J. Paeng. *A heat-induced molecular signature in marine dissolved organic matter*. **Nat. Geosci.**, 2:175–9, 2009.

[92] R. Benner and B. Biddanda. *Photochemical transformations of surface and deep marine dissolved organic matter: Effects on bacterial growth*. **Limnol. Oceanogr.**, 46(3):1373–8, 1998.

[93] E.R.M. Druffel, P.M. Williams, J.E. Bauer, and J.R. Ertel. *Cycling of dissolved and particulate organic matter in the open ocean*. **J. Geophys. Res.**, 97(C10):639–59, 1992.

[94] A.M. Hansen, T.E.C. Kraus, B.A. Pellerin, et al. *Optical properties of dissolved organic matter (DOM): Effects of biological and photolytic degradation*. **Limnol. Oceanogr.**, 61:1015–32, 2016.

[95] J.A. Hawkes, P.E. Rossel, A. Stubbins, et al. *Efficient removal of recalcitrant deep-ocean dissolved organic matter during hydrothermal circulation*. **Nat. Geosci.**, 8:856–60, 2015.

[96] A. Mentges, C. Feenders, C. Deutsch, et al. *Long-term stability of marine dissolved organic carbon emerges from a neutral network of compounds and microbes*. **Sci. Rep.**, 9:17780, 2019.

[97] E.J. Zakem, B.B. Cael, and N.M. Levine. *A unified theory for organic matter accumulation*. **Proc. Natl. Acad. Sci. U. S. A.**, 118(6):e2016896118, 2021.

[98] N. Hertkorn, C. Ruecker, M. Meringer, et al. *High-precision frequency measurements: Indispensable tools at the core of the molecular-level analysis of complex systems*. **Anal. Bioanal. Chem.**, 389: 1311–27, 2007.

[99] T.S. Catalá, S. Shorte, and T. Dittmar. *Marine dissolved organic matter: a vast and unexplored molecular space*. **Appl. Microbiol. Biotechnol.**, 105:7225–39, 2021.

[100] R. Benner. *What happens to terrestrial organic matter in the ocean?* **Mar. Chem.**, 92:307–10, 2004.

[101] X. Cao, G.R. Aiken, K.D. Butker, et al. *Evidence for major input of riverine organic matter into the*

ocean. **Org. Geochem.**, 116:62–76, 2017.

[102] W.S. Moore. *The effect of submarine groundwater discharge on the ocean*. **Ann. Rev. Mar. Sci.**, 2:59–88, 2010.

[103] J. Zhang and A.K.. Mandal. *Linkages between submarine groundwater systems and the environment*. **Curr. Opin. Environ. Sustain.**, 4:219–226, 2012.

[104] J.E. Bauer, W-J. Cai, P.A. Raymond, et al. *The changing carbon cycle of the coastal ocean*. **Nature**, 504:61–70, 2013.

[105] N.E. Blair and R.C. Aller. *The Fate of Terrestrial Organic Carbon in the Marine Environment*. **Ann. Rev. Mar. Sci.**, 4:401–23, 2012.

[106] S.M. Cragg, D.A. Friess, L.G. Gillis, et al. *Vascular Plants Are Globally Significant Contributors to Marine Carbon Fluxes and Sinks*. **Ann. Rev. Mar. Sci.**, 12:469–97, 2020.

[107] T.R. Anderson, E.C. Rowe, L. Polimene, et al. *Unified concepts for understanding and modelling turnover of dissolved organic matter from freshwaters to the ocean: the UniDOM model*. **Biogeochemistry**, 146:105–23, 2019.

[108] A. Vorobev, S. Sharma, M. Yu, et al. *Identifying labile DOM components in a coastal ocean through depleted bacterial transcripts and chemical signals*. **Environ. Microbiol.**, 20(8):3012–30, 2018.

[109] T.S. Bianchi. *The role of terrestrially derived organic carbon in the coastal ocean: A changing paradigm and the priming effect*. **Proc. Natl. Acad. Sci. U. S. A.**, 108(49):19473–81, 2011.

[110] R.W.P.M. Laane and L. Koole. *The relation between fluorescence and dissolved organic carbon in the Ems-Dollart Estuary and the western Wadden Sea*. **Netherlands J. Sea Res.**, 15(2):217–27, 1982.

[111] N.B. Nelson and J.M. Gauglitz. *Optical signatures of dissolved organic matter transformation in the global ocean*. **Front. Mar. Sci.**, 2:118, 2016.

[112] N.B. Nelson and D.A. Siegel. *The global distribution and dynamics of chromophoric dissolved organic matter*. **Ann. Rev. Mar. Sci.**, 5:447–76, 2013.

[113] J.D. Kinsey, G. Corradino, K. Ziergovel, et al. *Formation of chromophoric dissolved organic matter by bacterial degradation of phytoplankton-derived aggregates*. **Front. Mar. Sci.**, 4:430, 2018.

[114] N.B. Nelson, D.A. Siegel, and A.F. Michaels. *Seasonal dynamics of colored dissolved material in the Sargasso Sea*. **Deep. Res. Part I Oceanogr. Res. Pap.**, 45:931–57, 1998.

[115] N.B. Nelson, D.A. Siegel, C.A. Carlson, and C.M. Swan. *Tracing global biogeochemical cycles and meridional overturning circulation using chromophoric dissolved organic matter*. **Geophys. Res. Lett.**, 37:L03610, 2010.

[116] D.A. Siegel, S. Maritonera, N.B. Nelson, et al. *Global distribution and dynamics of colored dissolved and detrital organic materials*. **J. Geophys. Res. Ocean.**, 107(C12):21, 2002.

[117] P.G. Coble. *Marine optical biogeochemistry: The chemistry of ocean color*. **Chem. Rev.**, 107: 402–18, 2007.

[118] Y. Yamashita and E. Tanoue. *Chemical characterization of protein-like fluorophores in DOM in relation to aromatic amino acids*. **Mar. Chem.**, 82:255–71, 2003.

[119] T.S. Catalá, I. Reche, A. Fuentes-Lema, et al. *Turnover time of fluorescent dissolved organic matter in the dark global ocean*. **Nat. Commun.**, 6:5986, 2015.

[120] L. Jorgensen, C.A. Stedmon, T. Kragh, et al. *Global trends in the fluorescence characteristics and distribution of marine dissolved organic matter*. **Mar. Chem.**, 126:139–48, 2011.

[121] K.R. Murphy, C.A. Stedmon, T.D. Waite, and G.M. Ruiz. *Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy*. **Mar. Chem.**, 108:40–58, 2008.

[122] C. Guéguen and P. Kowalczuk. *Colored Dissolved Organic Matter in Frontal Zones. In: Chemical Oceanography of Coastal Zones*. Springer-Verlag Berlin Heidelberg, 2013.

[123] R. Goncalves-Araujo, M.A. Granskog, A. Bracher, et al. *Using fluorescent dissolved organic matter to trace and distinguish the origin of Arctic surface waters*. **Sci. Rep.**, 6:33978, 2016.

[124] Z. Gao and C. Gueguen. *Distribution of thiol, humic substances and colored dissolved organic matter during the 2015 Canadian Arctic GEOTRACES cruises*. **Mar. Chem.**, 203:1–9, 2018.

[125] C. DeFrancesco and C. Guéguen. *Tracing riverine dissolved organic carbon and its transport to the halocline layer in the Chukchi Sea (western Arctic Ocean) using humic-like fluorescence fingerprinting*. **J. Geophys. Res. Ocean**, 126(2):e2020JC016578, 2021.

[126] C. Collar. *Review: Biochemical and technological assessment of the metabolism of pure and mixed cultures of yeast and lactic acid bacteria in breadmaking applications*. **Food Sci. Technol. Int.**, 2: 349–67, 1996.

[127] J. Smedsgaard and J.C. Frisvard. *Using direct electrospray mass spectrometry in taxonomy and secondary metabolite profiling of crude fungal extracts*. **J. Microbiol. Methods**, 25:5–17, 1996.

[128] O. Fiehn, J. Kopka, P. Dörmann, et al. *Metabolite profiling for plant functional genomics*. **Nat. Biotechnol.**, 18:1157–61, 2000.

[129] O. Fiehn. *Metabolomics - The link between genotypes and phenotypes*. **Plant Mol. Biol.**, 48: 155–71, 2002.

[130] D. Ye, X. Li, J. Shen, and X. Xia. *Microbial metabolomics: From novel technologies to diversified applications*. **Trends Anal. Chem.**, 148:116540, 2022.

[131] E.E.K. Baidoo. *Microbial metabolomics, Chap 1: Microbial metabolomics: a general overview*. Humana Press, 2019.

[132] J-B. Raina, B.S. Lambert, D.H. Parks, et al. *Chemotaxis shapes the microscale organization of the ocean ' s microbiome*. **Nature**, 605:132–8, 2022.

[133] H. Osterholz, G. Singer, B. Wemheuer, et al. *Deciphering associations between dissolved organic molecules and bacterial communities in a pelagic marine system*. **ISME J.**, 10:1717–30, 2016.

[134] Y. Wang, R. Xie, Y. Shen, et al. *Linking Microbial Population Succession and DOM Molecular Changes in Synechococcus-Derived Organic Matter Addition Incubation*. **Microbiol. Spectr.**, 2022.

[135] H. Osterholz, J. Niggemann, H-A. Giebel, et al. *Inefficient microbial production of refractory dissolved organic matter in the ocean*. **Nat. Commun.**, 6:7422, 2015.

[136] B.P. Koch, M. Witt, R. Engbrodt, et al. *Molecular formulae of marine and terrigenous dissolved organic matter detected by electrospray ionization Fourier transform ion cyclotron resonance mass spectrometry*. **Geochim. Cosmochim. Acta**, 69(13):3299–308, 2005.

[137] R.L. Sleighter, Z. Liu, J. Xue, and P.G. Hatcher. *Multivariate statistical approaches for the characterization of dissolved organic matter analyzed by ultrahigh resolution mass spectrometry*. **Environ. Sci. Technol.**, 44:7576–82, 2010.

[138] M. Zark and T. Dittmar. *Universal molecular structures in natural dissolved organic matter*. **Nat. Commun.**, 9:3178, 2018.

[139] B.E. Noriega-Ortega, G. Wienhausen, A. Mentges, et al. *Does the chemodiversity of bacterial exometabolomes sustain the chemodiversity of marine dissolved organic matter?* **Front. Microbiol.**,

10:215, 2019.

[140] C. Michel, J. Hamilton, E. Hansen, et al. *Arctic Ocean outflow shelves in the changing Arctic: A review and perspectives*. **Prog. Oceanogr.**, 139:66–88, 2015.

[141] F-J.W. Parmentier, T.R. Christensen, S. Rysgaard, et al. *A synthesis of the arctic terrestrial and marine carbon cycles under pressure from a dwindling cryosphere*. **Ambio**, 46(S1):S53–S69, 2017.

[142] F. McLaughlin, E. Carmack, A. Proshutinsky, et al. *The rapid response of the Canada basin to Climate forcing*. **Oceanography**, 24(3):146–159, 2011.

[143] B.A. Bluhm, K.N. Kosobokova, and E.C. Carmack. *A tale of two basins: An integrated physical and biological perspective of the deep Arctic Ocean*. **Prog. Oceanogr.**, 139:89–121, 2015.

[144] R.W. Macdonald, Harner. T., and J. Fyfe. *Recent climate change in the Arctic and its impact on contaminant pathways and interpretation of temporal trend data*. **Sci. Total Environ.**, 342:5–86, 2005.

[145] J. Berge, P.E. Renaud, G. Darnis, et al. *In the dark: A review of ecosystem processes during the Arctic polar night*. **Prog. Oceanogr.**, 139:258–71, 2015.

[146] Y. Zhuang, H. Jin, W-J. Cai, et al. *Freshening leads to a three-decade trend of declining nutrients in the western Arctic Ocean*. **Environ. Res. Lett.**, 16:054047, 2021.

[147] M Ardyna and K.R. Arrigo. *Phytoplankton dynamics in a changing Arctic Ocean*. **Nat. Clim. Chang.**, 10:892–903, 2020.

[148] J-E. Tremblay, L.G. Anderson, P. Matrai, et al. *Global and regional drivers of nutrient supply, primary production and CO_2 drawdown in the changing Arctic Ocean*. **Prog. Oceanogr.**, 139:171–96, 2015.

[149] K.A. Brown, J.M. Holding, and E.C. Carmack. *Understanding Regional and Seasonal Variability Is Key to Gaining a Pan-Arctic Perspective on Arctic Ocean Freshening*. **Front. Mar. Sci.**, 7:606, 2020.

[150] L.G. Anderson and R.W. Macdonald. *Observing the Arctic Ocean carbon cycle in a changing environment*. **Polar Res.**, 34(1):26891, 2015.

[151] J. Jung, J.E. Son, Y.K Lee, K.H Cho, Y. Lee, et al. *Tracing riverine dissolved organic carbon and its transport to the halocline layer in the Chukchi Sea (western Arctic Ocean) using humic-like fluorescence fingerprinting*. **Sci. Total Environ.**, 772:145542, 2021.

[152] Z. Li, Q. Ding, M. Steele, and A. Schweiger. *Recent upper Arctic Ocean warming expedited by summertime atmospheric processes*. **Nat. Commun.**, 13:363, 2022.

[153] J.E. Box, W.T. Colgan, T.R. Christensen, et al. *Key indicators of Arctic climate change: 1971-2017*. **Environ. Res. Lett.**, 14:045010, 2019.

[154] K.R. Arrigo and G.L. van Dijken. *Continued increases in Arctic Ocean primary production*. **Prog. Oceanogr.**, 135:60–70, 2015.

[155] J. Toole, M. Timmermans, D. Perovich, et al. *Journal of Geophysical Research: Oceans*. **Prog. Oceanogr.**, 115, 2010.

[156] I.V. Polyakov, A.V. Pnyushkov, M.B. Alkire, et al. *Greater role for Atlantic inflows on sea-ice loss in the Eurasian Basin of the Arctic Ocean*. **Science**, 356:285–91, 2017.

[157] L.E. Kipp, M.A. Charette, W.S. Moore, et al. *Increased fluxes of shelf-derived materials to the central arctic ocean*. **Sci. Adv.**, 4:eaao1302, 2018.

[158] P.J. Hernes and R. Benner. *Terrigenous organic matter sources and reactivity in the North Atlantic Ocean and a comparison to the Arctic and Pacific oceans*. **Mar. Chem.**, 100:66–79, 2006.

[159] J. Terhaar, R. Lauerwald, P. Regnier, et al. *Around one third of current Arctic Ocean primary production sustained by rivers and coastal erosion*. **Nat. Commun.**, 120:169, 2021.

[160] J. Jung, J.E. Son, Y.K. Lee, et al. *Tracing riverine dissolved organic carbon and its transport to the halocline layer in the Chukchi Sea (western Arctic Ocean) using humic-like fluorescence fingerprinting*. **Sci. Total Environ.**, 772:145542, 2021.

[161] C. Pedrós-Alio, M. Potvin, and C. Lovejoy. *Diversity of planktonic microorganisms in the Arctic Ocean*. **Prog. Oceanogr.**, 139:233–43, 2015.

[162] P. Galand, C. Lovejoy, J. Pouliot, et al. *Microbial community diversity and heterotrophic production in a coastal Arctic ecosystem: A stamukhi lake and its source waters*. **Limnol.and Oceanograph.**, 53(2):813–23, 2008.

[163] M-E. Garneau, W.F. Vincent, L. Alonso-Saéz, et al. *Prokaryotic community structure and heterotrophic production in a river-influenced coastal arctic ecosystem*. **Aquat. Microb. Ecol.**, 42: 27–40, 2006.

[164] S. Kraemer, A. Ramachandran, D. Colatriano, et al. *Diversity and biogeography of SAR11 bacteria from the Arctic Ocean*. **ISME J**, 14(1):79–90, 2020.

[165] D. Colatriano, P.Q. Tran, C. Guéguen, et al. *Genomic evidence for the degradation of terrestrial organic matter by pelagic Arctic Ocean Chloroflexi bacteria*. **Commun. Biol.**, 1:90, 2018.

[166] M. Mikan, H.R. Harvey, E. Timmins-Schiffman, et al. *Metaproteomics reveal that rapid perturbations in organic matter prioritize functional restructuring over taxonomy in western Arctic Ocean microbiomes*. **ISME J.**, 14:39–52, 2020.

[167] D.L. Kirchman, M.T. Cottrell, and C. Lovejoy. *The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes*. **Environ. Microbiol.**, 12:1132–43, 2010.

[168] L. Alonso-Saéz, O. Sánchez, J.M. Gasol, et al. *Winter-to-summer changes in the composition and single-cell activity of near-surface Arctic prokaryotes*. **Environ. Microbiol.**, 10(9):2444–54, 2008.

[169] M. Wietz, C. Bienhold, K. Metfies, et al. *The polar night shift: seasonal dynamics and drivers of Arctic Ocean microbiomes revealed by autonomous sampling*. **ISME Commun.**, 1, 2021.

[170] D. Han, I. Kang, H.K. Ha, et al. *Bacterial communities of surface mixed layer in the pacific sector of the western Arctic Ocean during sea-ice melting*. **PLoS One**, 9(1):e86887, 2014.

[171] E. Fadeev, A. Rogge, S. Ramondenc, et al. *Sea ice presence is linked to higher carbon export and vertical microbial connectivity in the Eurasian Arctic Ocean*. **Commun. Biol.**, 4:1255, 2021.

Chapter 2

Degradation pathways for aromatic compounds of terrestrial origin are widespread and expressed in Arctic Ocean microbiomes

2.1 Abstract

Background: The Arctic Ocean receives massive freshwater input and a correspondingly large amount of humic-rich organic matter of terrestrial origin. Global warming, permafrost melt, and a changing hydrological cycle will contribute to an intensification of terrestrial organic matter release to the Arctic Ocean. Although considered recalcitrant to degradation due to complex aromatic structures, humic substances can serve as substrate for microbial growth in terrestrial environments. However, the capacity of marine microbiomes to process aromatic-rich humic substances, and how this processing may contribute to carbon and nutrient cycling in a changing Arctic Ocean, is relatively unexplored. Here, we used a combination of metagenomics and metatranscriptomics to assess the prevalence and diversity of metabolic pathways and bacterial taxa involved in aromatic compound degradation in the salinity-stratified summer waters of the Canada Basin in the western Arctic Ocean.

Results: Community-scale meta-omics profiling revealed that 22 complete pathways for processing aromatic compounds were present and expressed in the Canada Basin, including those for aromatic ring fission and upstream funnelling pathways to access diverse aromatic compounds of terrestrial origin. A phylogenetically diverse set of functional marker genes and transcripts were associated with fluorescent dissolved organic matter, a component of which is of terrestrial origin. Pathways were common throughout global ocean microbiomes, but were more abundant in the Canada Basin. Genome-resolved analyses identified 12 clades of Alphaproteobacteria, including Rhodospirillales, as central contributors to aromatic compound processing. These genomes were mostly restricted in their biogeographical distribution to the Arctic Ocean, and were enriched in aromatic compound processing genes compared to their closest relatives from other oceans.

Conclusion: Overall, the detection of a phylogenetically diverse set of genes and transcripts implicated in aromatic compound processing supports the view that Arctic Ocean microbiomes

have the capacity to metabolize humic substances of terrestrial origin. In addition, the demonstration that bacterial genomes replete with aromatic compound degradation genes exhibit a limited distribution outside of the Arctic Ocean suggests that processing humic substances is an adaptive trait of the Arctic Ocean microbiome. Future increases in terrestrial organic matter input to the Arctic Ocean may increase the prominence of aromatic compound processing bacteria and their contribution to Arctic carbon and nutrient cycles.

2.2 Introduction

Humic substances (HS) are a heterogeneous mixture of organic compounds resulting from biochemical transformations of dead plants and microbes. HS are ubiquitous in both terrestrial and aquatic systems and constitute the largest fraction of organic matter (OM) in terrestrial ecosystems[1,2], reaching 60-80% in soils[3] and 50-80% in freshwaters[4]. The fraction of HS is relatively high (20-60%) in shelves, coastal zones and estuaries[5] due to the input of terrestrial OM (tOM) with freshwater runoff and exchange with sediments. HS constitute a smaller fraction of dissolved OM (DOM) in the open ocean (0.7-2.4%)[6]. The lesser amount of HS in the DOM of open oceans indicates that HS is removed by ocean microbiomes and additional non-biological processes[7,8].

The Arctic Ocean receives a disproportionately high input of freshwater (10% of total global freshwater input for 1.3% of total ocean volume), and a correspondingly high tOM input (10% of ocean total tOM input)[9]. Rivers annually discharge 25-36 Tg of dissolved organic carbon and 12 Tg of particulate organic carbon to the Arctic Ocean[10,11]. Climate change is strongly influencing the Arctic region, which in turn is influencing Arctic hydrology and organic matter dynamics[12,13]. More specifically, permafrost thawing[14], combined with intensifying river runoff[15], coastal erosion[16] and groundwater input[17], is driving an increase in the amount of humic-rich DOM input into the Arctic Ocean. The humic-rich DOM consequently contributes significantly to the carbon pool of the Arctic Ocean DOM compared to other oceans[18], and potentially represents a significant and increasing growth resource for the Arctic Ocean microbiome.

The origins and distributions of tOM in the Canada Basin in the western Arctic Ocean has been extensively studied, making it a useful system for investigating interactions between tOM and ocean microbiomes. In spring and summer, humic-rich OM is transported by riverine inputs to the surface mixed layer of the Arctic Ocean shelves[19,20]. In shelf waters, tOM is partially photodegraded[21], while some flocculates upon mixing with salt water and sinks to the sediments along with particulate OM[15]. In fall and winter, the tOM remaining in the surface layer sinks with the dense brine expelled during ice-formation. This brine flows along Chukchi Sea and Beaufort Sea shelves, exchanging organic matter with bottom sediments, ultimately accumulating in the deeper and more saline water of Pacific Ocean origin[16,22]. The interactions with shelf sediments and pore waters constitute a substantial source of tOM which

may have been reprocessed by sediment microbiomes[23]. It has been estimated that 11-44% of Arctic Ocean sediment OM is of terrestrial origin[24]. As a consequence of the OM dynamics, the Canada Basin is characterized by a strong and distinctive signal of humic-rich DOM that extends from the subsurface water to a depth of 300 m[25,26].

HS are heterogeneous supramolecular assemblies formed by microbial and physico-chemical transformations[27] of organic matter. In terrestrial systems, HS originate from vascular plant residues (lignin and other biopolymers) and other organic detritus[28], giving rise to HS rich in aromatic moieties. In contrast, HS produced in marine environments have a strong aliphatic and branched structure[29]. In the Arctic Ocean however, HS are aromatic-rich due to their terrestrial and sediment origin[30]. HS usually show a high degree of recalcitrance that is dependent on their physicochemical interactions with the environment[31]. In soils, sorption of HS to mineral particles drives a physical separation of HS from microbes and their enzymes, preventing fast degradation of HS[32,33]. Numerous studies have therefore demonstrated that HS freed from their soil environment can be used to support microbial growth[7,34,35].

The capacity of microbiomes to couple HS transformation to growth relies on the ability to degrade aromatic compounds from HS. The degradation of aromatic compounds follows two main steps. Funneling pathways transform (e.g. via oxidation, decarboxylation, and/or demethylation) larger and more substituted aromatic compounds to a small set of key aromatic compounds (e.g. gentisate, catechol, protocatechuate), which then undergo an aromatic ring-fission step followed by further processing to generate central carbon metabolism intermediates. In humic-rich environments such as soils, microbiomes use a wide variety of funneling pathways to access the diverse set of lignin-derived aromatic compounds (e.g. vanillate, syringate, benzoate and their derivatives) that have been incorporated in HS[36].

In soils, fungi degrade most of the humic substances[37]. In the ocean, bacteria are considered the main actors in OM degradation[38], even if recent studies have highlighted an important role for fungi[39,40], for example in processing OM in marine snow[41] or by parasitizing phytoplankton[42]. Certain bacteria inhabiting humic-rich environments can grow on HS as sole carbon and energy sources, and are therefore are able to access aromatic compounds within HS[5,8,34]. Transcriptomic analysis from the humic acid-degrading bacterium *Pseudomonas sp.* isolated from sub-arctic tundra soils showed that genes involved in the funneling and ring-opening steps of aromatic compound degradation pathways were up-regulated when fed with humic acids compared to glucose[43]. Recently it was shown that *Chloroflexi* genomes from the Canada Basin encoded a diverse set of genes associated with aromatic compound degradation[44]. The *Chloroflexi* populations appeared to be endemic to the Arctic Ocean and were associated with the humic-rich fluorescence DOM maximum (FDOMmax). These observations suggest the disproportionately high fraction and diversity of aromatic-rich HS in the Arctic Ocean DOM compared to other oceans may select for a diverse HS-degrading microbiome.

The genomic diversity and metabolic pathways in the Arctic Ocean microbiomes can provide important insights regarding the fate of HS and its impact on Arctic Ocean biogeochemical

cycles. However, outside of perhaps the *Chloroflexi*, we know very little about how phylogenetically widespread HS degradation is in the Arctic Ocean microbiomes, nor the diversity of metabolic pathways employed by the Arctic Ocean microbiomes to process HS. We hypothesized that the capacity for aromatic compound degradation was linked to the distribution of humic-rich tOM and enhanced in the Arctic Ocean compared to other oceans with a lesser amount of HS. Finally, we hypothesized that the vast amount of HS in the Arctic Ocean may have played a role as an ecological pressure for the adaptive evolution of the taxa most implicated in aromatic compound degradation.

2.3 Results

2.3.1 Environmental context

We surveyed the microbiomes along a latitudinal transect (73-81^{0}N) of salinity-stratified waters of the Canada Basin using a combination of metagenomics and metatranscriptomics (Figure 2.1a-b). The sampling design targeted distinct water column features, including the relatively fresh surface mixed layer (surface; 5 m and 20 m), the subsurface chlorophyll maximum (SCM; 55-95 m), the FDOMmax associated with colder Pacific-origin water (32.3 and 33.1 PSU; 90-250 m), the warmer Atlantic-origin water (Tmax and AW; 360-1000 m depth), and Arctic bottom water (∼3800 m). The warmer Atlantic-origin water and Arctic bottom water are herein collectively referred to as deep waters.

We sought to determine the distribution and composition of OM in the Canada Basin, with a focus on the distribution of tOM. Optical properties of the OM, such as fluorescence, have previously been used to assess the composition of OM in the ocean, and differentiate between terrestrial and marine OM sources[45,46]. We used excitation emission matrix (EEC) fluorescence spectroscopy combined with parallel factor analysis (PARAFAC) to determine the distribution of fluorescent DOM components. In the Canada Basin, seven components (C1-C7) were identified, as previously defined in DeFrancesco *et al.*[25]. These components corresponded to terrestrially derived humic-like DOM (C1 and C4), amino acid or protein material (C3 and C6), or microbially-derived humic-like DOM (C2, C5 and C7) (Figure 2.1c). The aromatic-rich C1 was the most abundant component within the FDOMmax samples (25-27 %), but also in the whole water column below the surface (20-22% in the SCM and 21-23% in the deep), verifying that a significant fraction of OM is of terrestrial origin. Of the terrestrial components, C4 was the dominant component in the surface (19-30 %). The reduced contribution of C1 in the surface is because C1 is more red-shifted than C4 indicating a stronger aromatic character and thus enhanced photosensitivity. Overall, these results indicate a strong contribution of a photostable fraction from terrestrial origin in the FDOM of the surface and an aromatic-rich fraction from terrestrial origin in the FDOM of the whole water column below the surface.

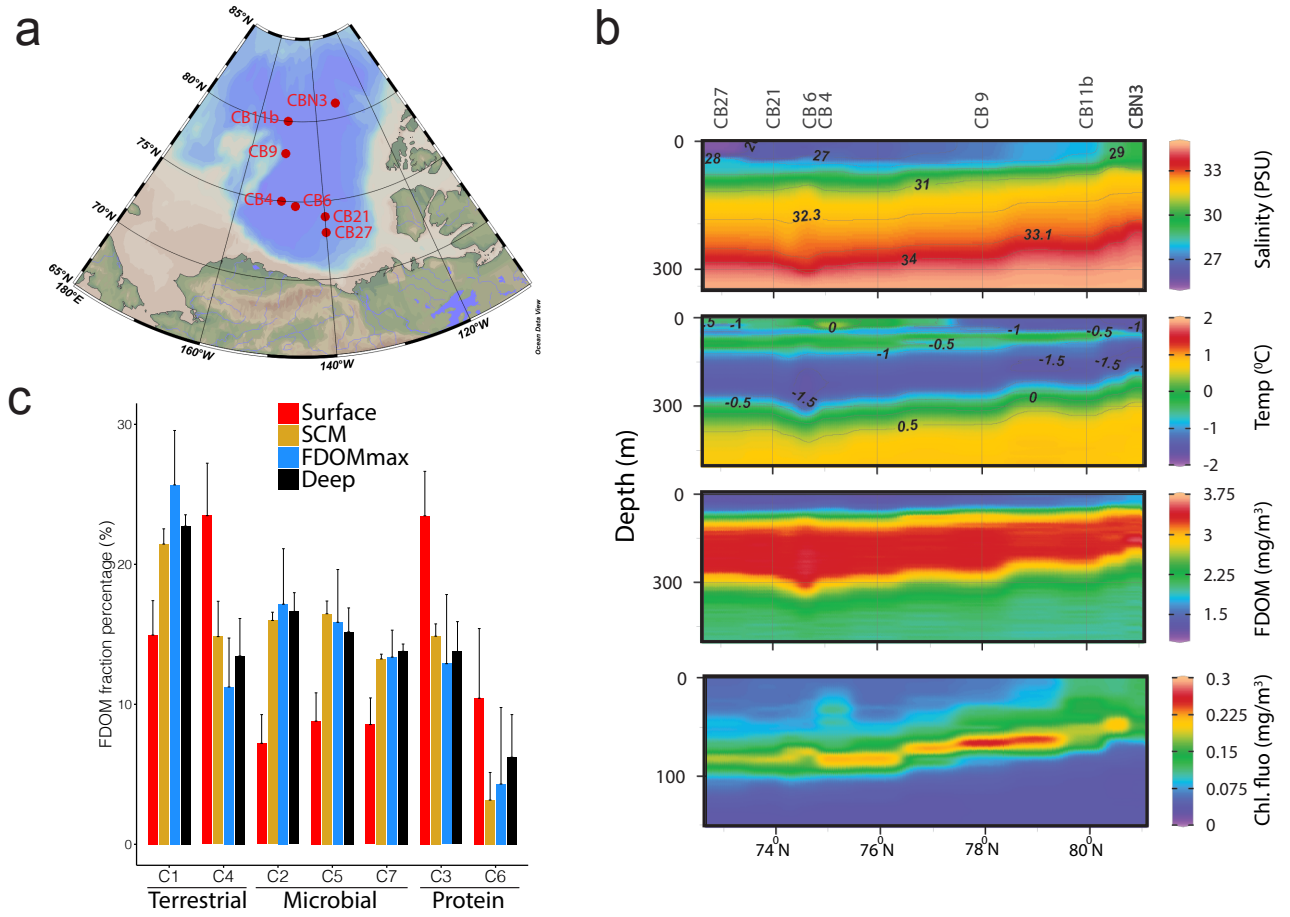

Figure 2.1: *Spatial biogeochemistry of the Canada Basin. a) Map of the 8 stations sampled in this study. b) depth profile of Salinity (PSU), temperature (^0C), fluorescent dissolved organic matter (FDOM, mg/m^3), and chlorophyll fluorescence (mg/m^3) at the 8 stations sampled in this study. c) Percentage of the 7 FDOM fractions identified using excitation emission matrix fluorescence spectroscopy combined with parallel factor analysis. Samples are grouped in 4 samples water features: Surface, subsurface chlorophyll maximum (SCM) fluorescent dissolved organic matter maximum (FDOMmax) and deep waters.*

2.3.2 Vertical-partitioning of metabolic features in metagenomes and metatranscriptomes

We investigated the abundance and distribution of aromatic compound degradation pathways in the Canada Basin microbiomes in relation to tOM availability. To investigate how the metabolic system of microbiomes was distributed across the Canada Basin, we first performed nonmetric multidimensional scaling (NMDS) analysis on the abundance of enzyme-encoding genes (genes assigned to enzyme commission (EC) numbers) annotated from metagenome or metatranscriptome assemblies. NMDS ordination showed that metagenomes (stress=0.11) were partitioned into four clusters consisting of samples collected from either the surface, the SCM, FDOMmax, or deep water (Figure 2.2a). A similar pattern was observed in the NMDS ordination of metatranscriptomes (stress=0.0503), although the variation between samples from within the same water column feature was higher than observed in metagenomes (Figure 2.2b). In addition, there was less separation between the samples from the FDOMmax and from deeper Atlantic-origin waters in the ordination based on metatranscriptomes compared to metagenomes.

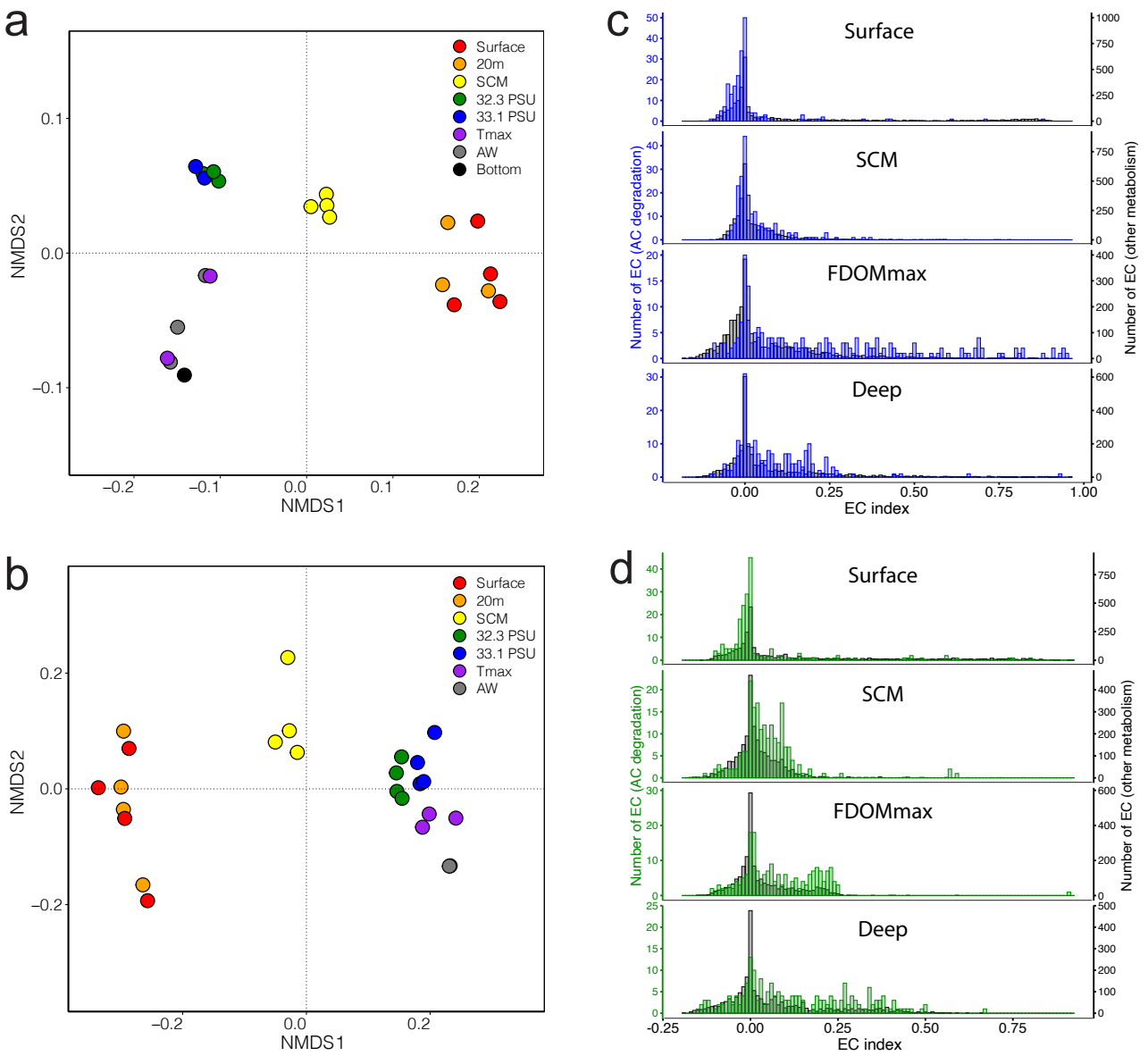

Figure 2.2: *Identification of the microbial enzymatic reactions associated with the 4 water column features. Non metric multidimensional scaling of the matrices of enzymatic reactions abundances from a) metagenomes and b) metatranscriptomes. c,d) Distribution of gene indices calculated for the enzymatic reactions from c) metagenomes and d) metatranscriptomes in the 4 water features. Blue/green: distribution for enzymatic reactions involved in the degradation of aromatic compounds from metagenomes (blue) and metatranscriptomes (green); grey: distribution for all enzymatic reactions not involved in the degradation of aromatic compounds.*

We next determined which enzymatic reactions differentiated the metagenomes across the stratified water column using non-negative matrix factorization (NMF), which is a tool for extracting meaningful features from high dimensional data[47]. In our analyses, NMF decomposes the matrix of EC number abundances into two matrices. Matrix 1 presents a reduced number of elements that describe the overall similarities of the metagenomes based on EC number composition, while matrix 2 presents the weighted contribution of individual EC numbers on each of the elements in matrix 1. We determined that a decomposition with four elements best represented the overall enzyme composition of metagenomes (Figure 2.3). The four elements, herein referred to as sub-metagenomes (Figure 2.4), represented the same patterns as observed in the NMDS ordination, corresponding to the surface, SCM, FDOMmax, and deep waters (Figure 2.2a).

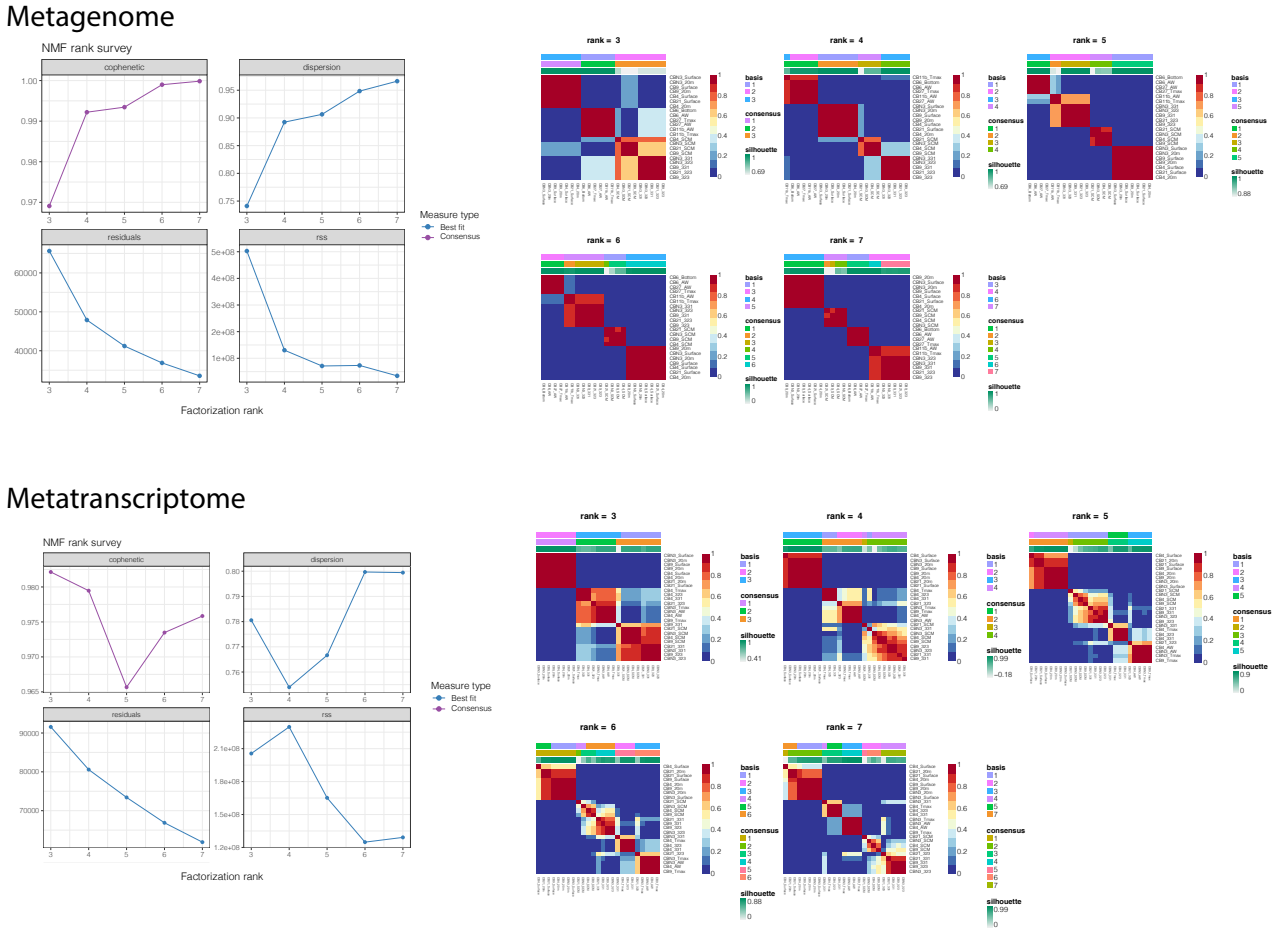

Figure 2.3: *Estimation of the rank for the NMF analysis of the EC abundance matrices annotated from metagenomes (top panels) and metatranscriptomes (bottom panels). Left panels: evolution of various parameters as a function of the rank used in the NMF analysis. Cophenetic correlation represents the correlation between the sample distances from the consensus matrix and the cophenetic distance between these samples when they are clustered. The rss is the residual sum of squares between the original EC abundance matrix and its estimate using the NMF algorithm. The dispersion is defined as 1-rss/*$\Sigma_{i,j}(V_{i,j})^2$ *(Vi,j are the entries of the EC abundance matrix) and estimates the fraction of variance of the EC abundance matrix explained by the NMF results. Residuals is the sum of residuals between the original EC abundance matrix and the matrix estimated using the NMF. Right panels: consensus matrices based on clustering the coefficient matrices at each of the 100 runs of the NMF analysis. The heatmap represents the fraction of times 2 samples fall in the same clusters out of 100 runs.*

We then assessed which EC numbers were strongly associated with each of the four sub-metagenomes by calculating an EC index value. This EC index value quantifies the tendency of an EC number to be specific to a single sub-metagenome (EC index values range between -1 and 1). The distribution of EC indices was plotted for each of the four sub-metagenomes. Overall, the means of the EC indices associated with aromatic compound degradation and other metabolic pathways in the four water column features were significantly different (PERMANOVA, F=89.8, p<0.0001). Each sub-metagenome has a collection of EC numbers with relatively high indices (>0.5) (Figure 2.2c). However, the most striking observation was that EC numbers involved in aromatic compound degradation were predominantly associated with the FDOMmax sub-metagenome, as demonstrated by the higher index values for EC numbers from aromatic compound degradation pathways than from other metabolic pathways in the FDOM-

max (Student t-test, t=13.26, p<0.0001). The EC indices for aromatic compound degradation genes were smaller than the EC indices associated with other metabolic processes in the surface (Student t-test, t=8.89, p<0.0001) and not significantly different for the SCM (Student t-test, t=0.369, p=0.414) and the deep (Student t-test, t=0.56, p=0.545) sub-metagenomes.

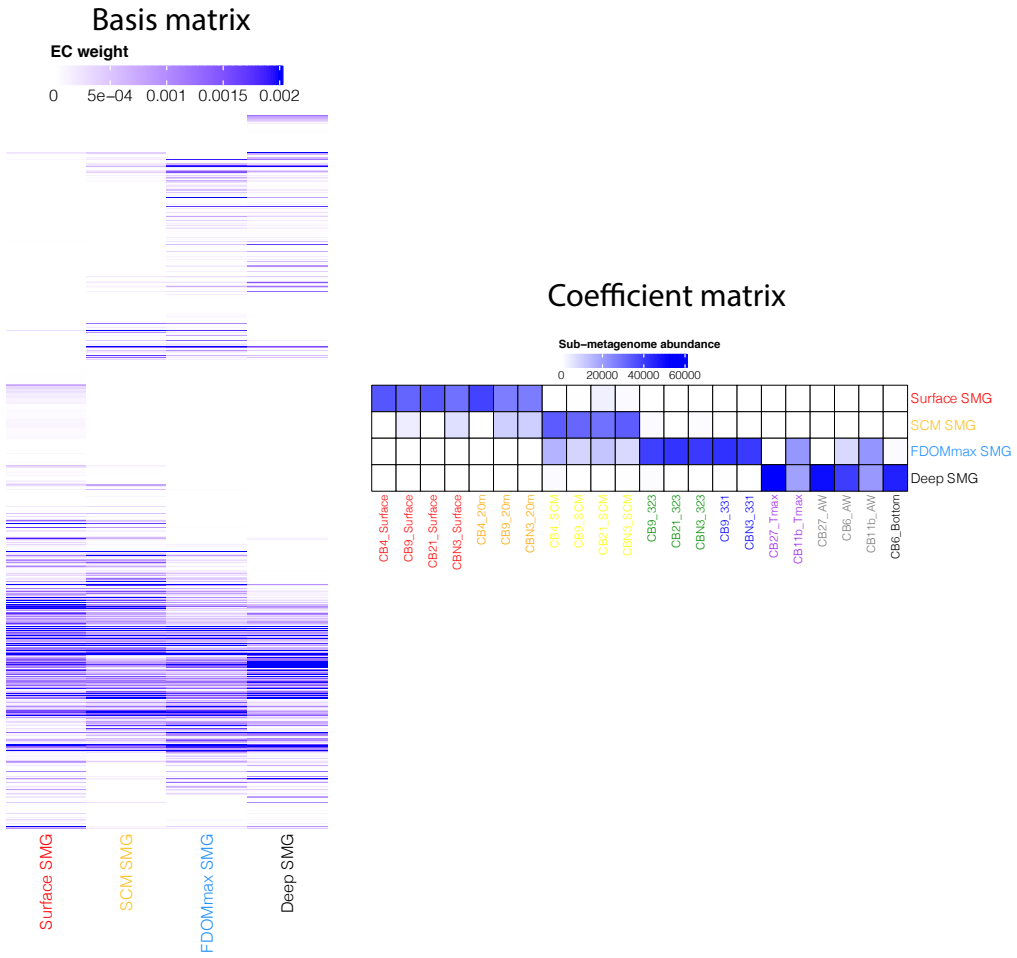

Figure 2.4: *Heatmaps of the basis matrix (left) and coefficient matrix (right) obtained after running an NMF analysis on the EC abundance matrix annotated from metagenomes and using a rank value of 4.*

We performed a similar NMF analysis on EC numbers in the metatranscriptomes (Figure 2.3). Similar with the NMF analysis of metagenomes, decomposition resulted in four elements, herein referred to as sub-metatranscriptomes (Figure 2.3), corresponding to the surface, SCM, FDOMmax, and deep waters (Figure 2.5). For the sub-metatranscriptomes, the means of the EC indices from aromatic compound degradation and from other metabolic pathways in the four water column features were significantly different (PERMANOVA, F=121, p<0.0001). For the sub-metatranscriptomes, however, we observed higher indices for the EC numbers from aromatic compound degradation than for EC numbers from other metabolic processes in the SCM (Student t-test, t=0.0218, p<0.0001), the FDOMmax (Student t-test, t=0.0444, p<0.0001) and the deep waters (Student t-test, t=0.0611, p<0.0001) (Figure 2.2d).

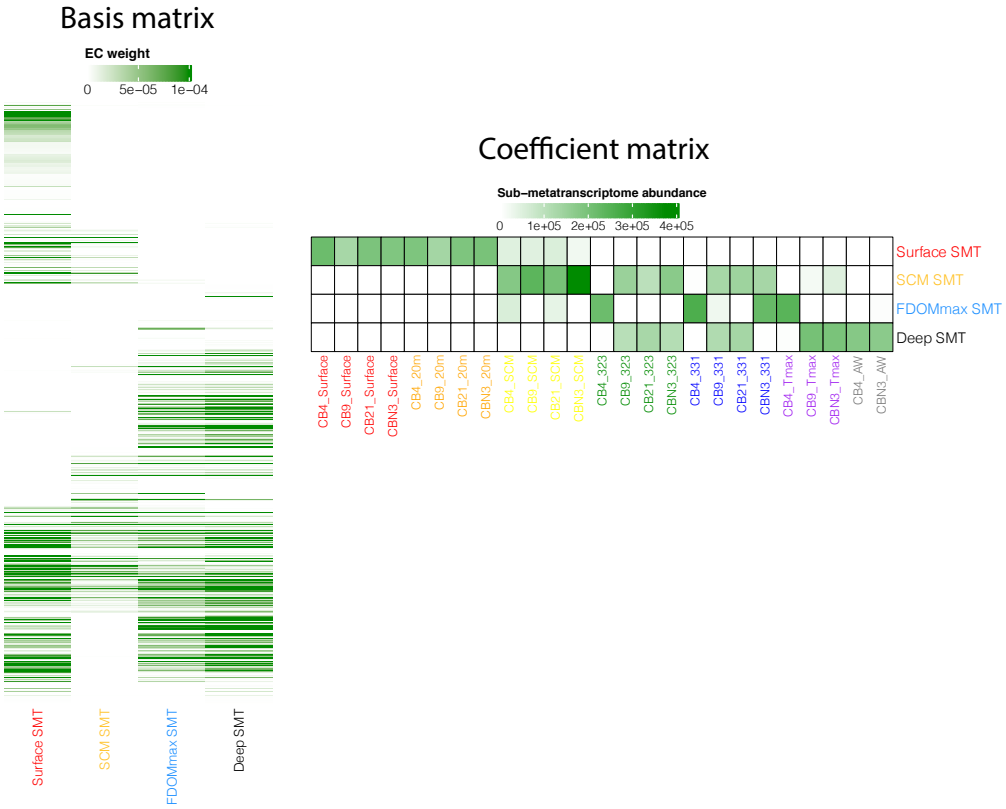

Figure 2.5: *Heatmaps of the basis matrix (left) and coefficient matrix (right) obtained after running an NMF analysis on the EC abundance matrix annotated from metatranscriptomes and using a rank value of 4.*

2.3.3 Aromatic compound degradation genes in global ocean metagenomes

As the humic-rich OM input to the Arctic Ocean is disproportionately high compared to other oceans, we investigated if genes associated with aromatic compound degradation were more abundant in the Canada Basin metagenomes compared to other oceanic metagenomes. As terrestrial OM is a significant contributor of HS to the Arctic Ocean, we restricted our analysis to genes involved in processing aromatic compounds of terrestrial origin. We focused the analysis on a set of 46 pathways previously implicated in degrading aromatic compounds from lignin (Figure 2.6a). We compared the relative abundance of aromatic compound degradation genes between metagenomes of the Canada Basin water features (surface, SCM, FDOMmax, deep) and metagenomes from both the surface and subsurface waters (SCM, mesopelagic, bathypelagic) of the Atlantic, Pacific, Indian and Southern Oceans as well as the Mediterranean Sea and Red Sea (Figure 2.6b). The Canada Basin FDOMmax metagenomes contained the highest fraction of aromatic compound degradation genes (3.4%). Aromatic compound degradation genes were identified in other oceanic metagenomes (1.5-2.5 % of total protein-coding genes) and the relative abundance of aromatic compound degradation genes increased with water depth. Overall, the mean percentage of aromatic compound degradation genes in the water column features of the Canada Basin were significantly different than in other oceans (PERMANOVA, $F=27.8$, $p<0.0001$). Specifically, the relative abundance was consistently higher (1.3-1.7 fold) in the microbiomes of the Canada Basin upper water column features compared to microbiomes from other oceanic zones (Student t-test: surface/surface, $t=0.58$,

p=0.0013; SCM/SCM, t=0.92, p=0.0001; FDOMmax/mesopelagic, t=0.81, p=0.0001) (Figure 2.6c). However, we did not observe significant differences between the percentage of aromatic compound degradation genes of Arctic deep-water microbiomes and the microbiomes of other oceans deep waters (Student t-test, t=0.20, p=0.490) (Figure 2.6c).

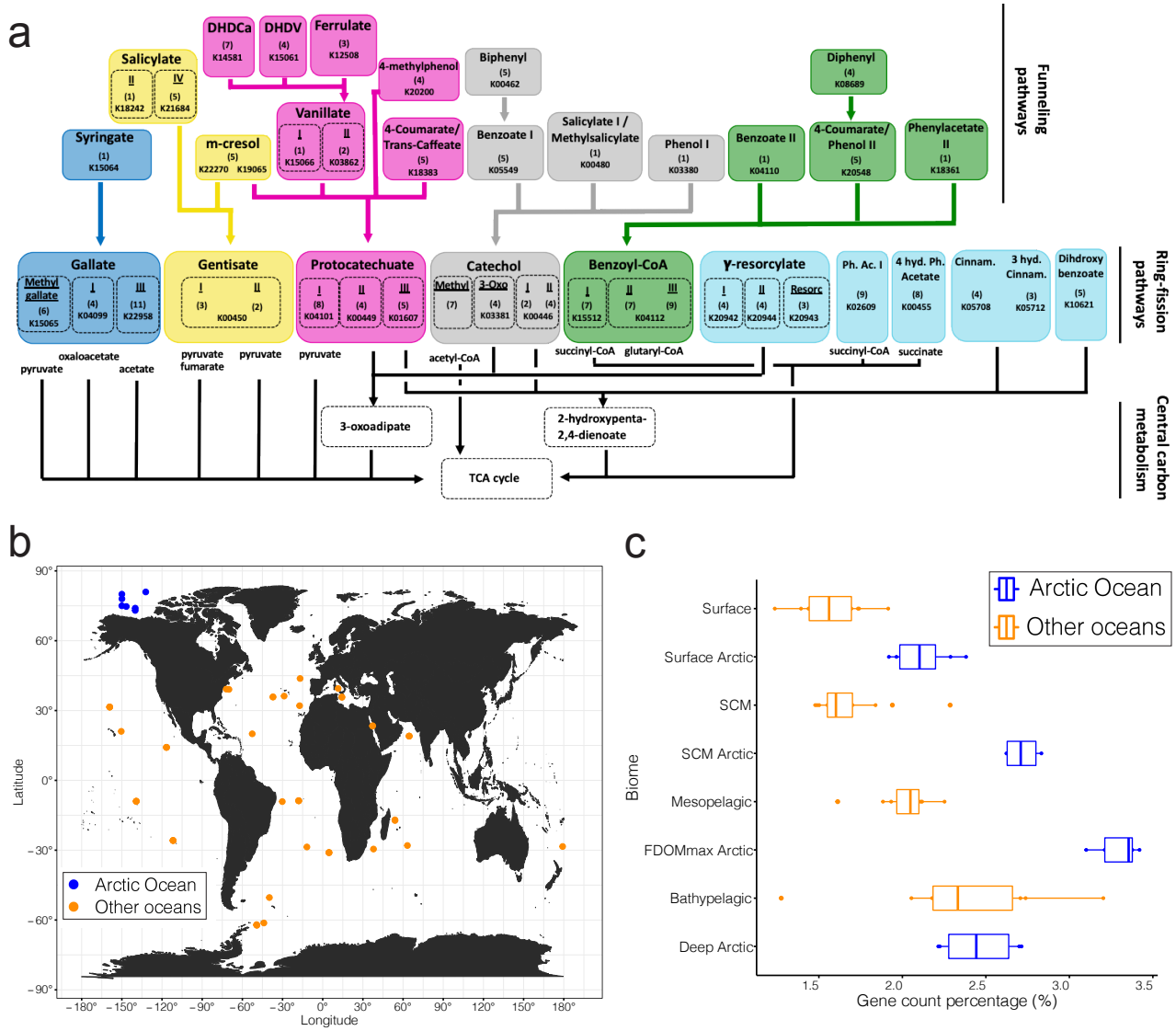

Figure 2.6: *a) Diagram of the 46 funneling and ring-opening pathways involved in the degradation of lignin-derived aromatic compounds. b) Map of the metagenomic samples used to compare the capacity to degrade aromatic compounds in the Arctic Ocean (blue) compared to other seas and oceans (orange). c) Fraction of metabolic genes involved in the degradation of aromatic compounds in the Arctic Ocean (blue) and other seas and oceans (orange). The line in the box represents the median. The left and right hinges of the box represent the 25th and 75th percentiles. Whiskers extend to the smallest and largest values (no further than 1.5 the inter-quartile range).*

2.3.4 Distribution of aromatic compound degradation genes and pathways in metagenomes and metatranscriptomes

To elucidate the diversity of aromatic compounds that the Arctic Ocean microbiomes can access as growth substrates, we assessed the diversity and the completeness of aromatic compound degradation pathways in Canada Basin metagenomes. We found evidence for the presence of 44 of the 46 aromatic compound degradation pathways in the metagenomes (Figure 2.7). A

complete set of genes were identified for over half of these pathways in the metagenomes, irrespective of the water column feature (Figure 2.7). Evidence for the 44 pathways was also identified in the metatranscriptomes, including expression of the full complement of genes for 22 pathways (Figure 2.7).

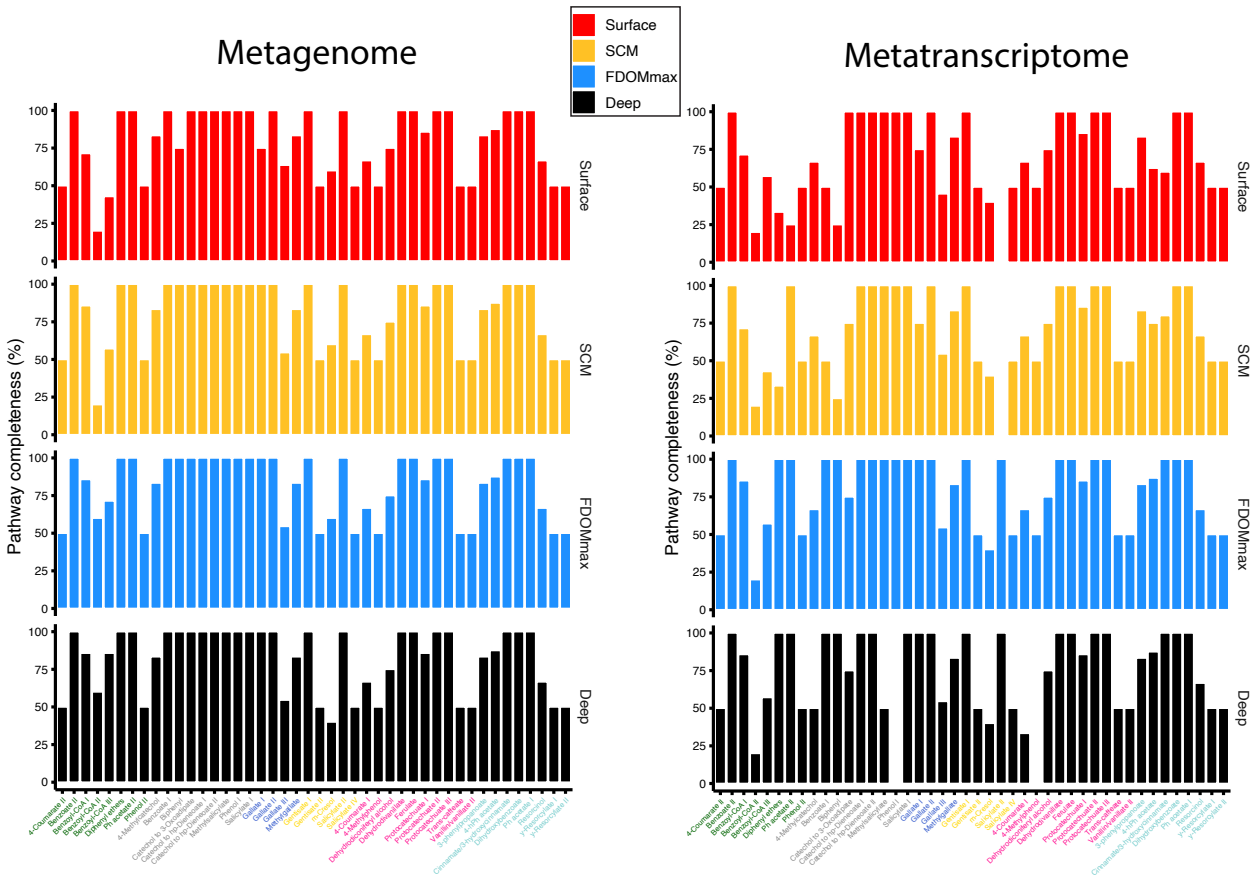

Figure 2.7: *Lignin-derived aromatic compound degradation pathways completeness in the 4 water features (surface, SCM, FDOMmax, deep) for metagenomes (left) and metatranscriptomes (right).*

To measure the distribution of the aromatic compound degradation pathways through the water column, we used a selection of 39 unique marker genes for the 46 aromatic compound degradation pathways. To provide a measure of pathway abundance and expression, we summed the depth of coverage of each marker gene or transcript and corrected for differences in metagenome sequencing effort (Figure 2.8). Out of the 39 unique marker genes, 32 were detected in the Canada Basin metagenomes (Figure 2.8a, Figure 2.9). Generally, the most abundant genes were also most abundant in the metatranscriptomes (Figure 2.8a-b, Figure 2.9, Figure 2.10). Most of the marker genes were most abundant in the FDOMmax, yet were most highly expressed in deep waters. (Figure 2.8a-b).

Vanillate monooxygenase (K03862) was the most abundant marker gene within all water column features (20 copies/10^6 reads in the FDOMmax) for pathways degrading aromatic compounds from terrestrial sources, while 3-O-methylgallate 3,4-dioxygenase (K15065) showed a lower abundance (8 copies/10^6 reads in the FDOMmax). While vanillate monooxygenase was more abundant in the metatranscriptomes of the FDOMmax, 3-O-methylgallate

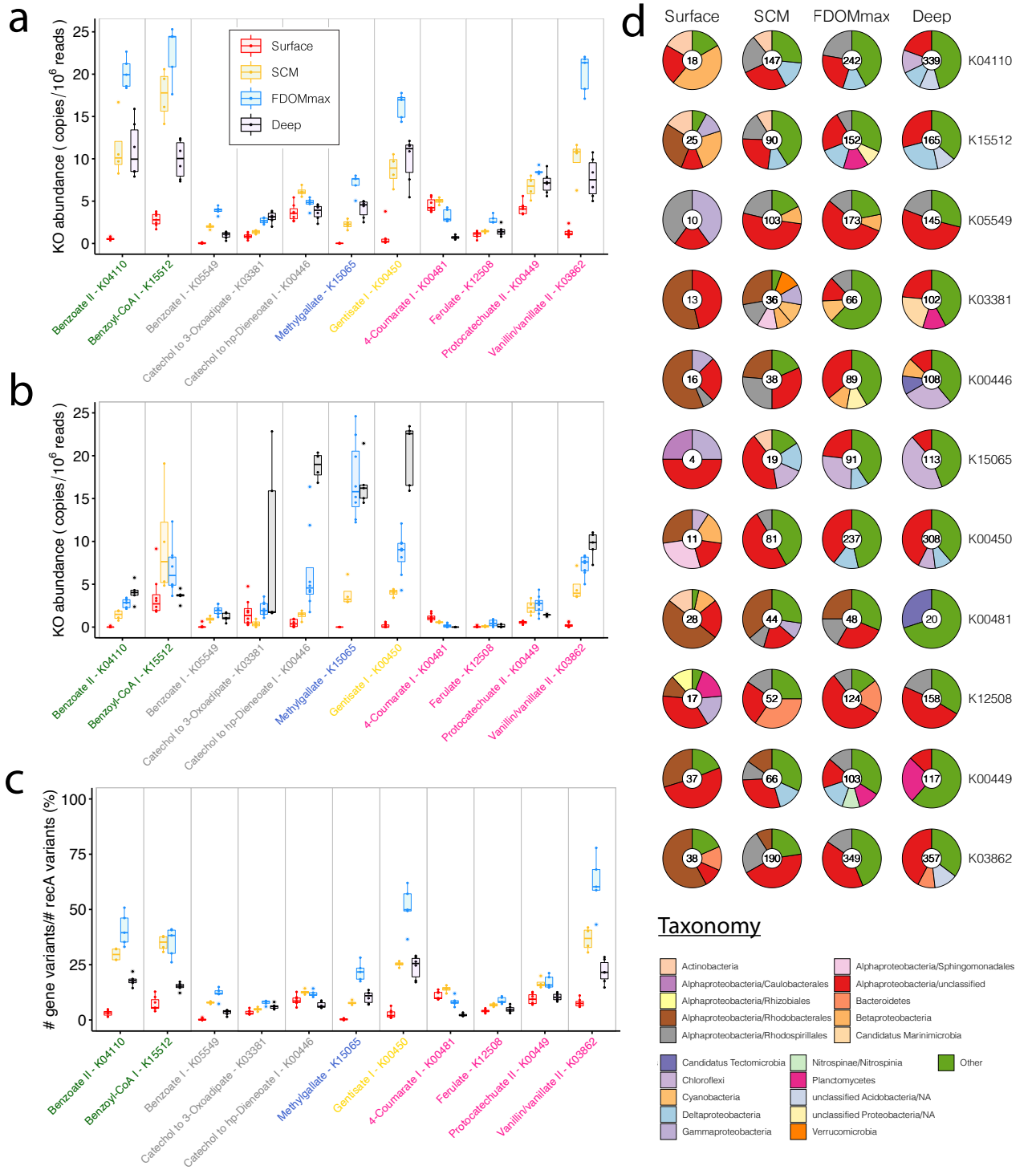

Figure 2.8: *Distribution and taxonomy of marker genes involved in the degradation of aromatic compounds in the Arctic Ocean. a) normalized abundance of genes annotated with KEGG orthology numbers (KOs) marker of aromatic degradation pathways in the 4 different water features. b) normalized abundance of transcripts annotated with KEGG orthology numbers marker of aromatic degradation pathways in the 4 different water features. c) estimated fraction of the microbiomes harboring genes annotated with KOs marker of aromatic degradation pathways in the 4 different water features. The line in boxes represents the medians. The bottom and upper hinges of the box represent the 25^{th} and 75^{th} percentiles. Whiskers extend to the smallest and largest values (no further than 1.5 x the inter-quartile range). d) taxonomy of genes annotated with KOs marker of aromatic degradation pathways in the 4 different water features. Numbers within pie charts represent the number of gene clusters (clustered at 95% identity) identified in each water feature. Red: surface samples; yellow: SCM samples; blue: FDOMmax samples; black: deep samples.*

3,4-dioxygenase was more abundant in the metatranscriptomes of the deep layers. Both ring fission protocatechuate dioxygenases (K00449 and K04101) were abundant in metagenomes (8 and 6 copies/10^6 reads in the FDOMmax) and metatranscriptomes (2.5 and 7.5 copies/10^6 reads in the FDOMmax). Overall, these results show that the Canada Basin microbiomes can fully transform aromatic compounds from terrestrial sources into central carbon metabolism intermediates, with an enhanced capacity in the FDOMmax.

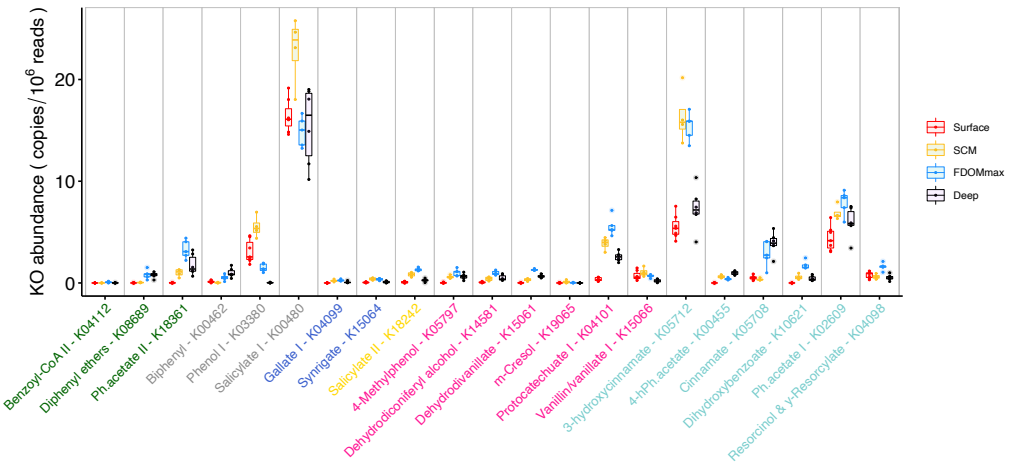

Figure 2.9: *Normalized abundance per water column feature of KO numbers markers of aromatic compounds degradation pathways annotated from metagenomes and b) metatranscriptomes. c) Estimated fraction of the microbiome harboring genes annotated with KO marker of aromatic compounds degradation patwhays.*

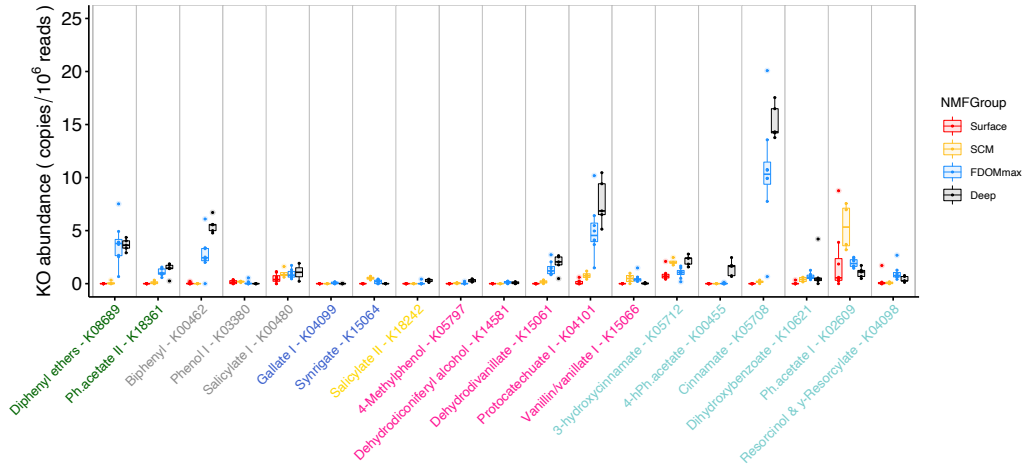

Figure 2.10: *Normalized abundance per water column feature of KO numbers markers of aromatic compounds degradation pathways annotated from metatranscriptomes.*

A number of aromatic compounds (e.g. salicylate, 3-hydroxycinnamate or benzoate) can originate from lignin as well as other sources such as marine phytoplankton. Within the pathways involved in the degradation of aromatic compounds from possible marine or terrestrial origin, benzoate CoA-ligase (K04110), salicylate monooxygenase (K00480) and 3-hydroxycinnamate hydroxylase (K05712) were the most abundant in metagenomes (20, 24 and 17 copies/10^6 reads respectively) but not in metatranscriptomes (7.5, 5 and 2 copies/106 reads). The most

common funneling pathway for benzoate was through benzoyl-CoA as evidenced by the lower abundance of genes (5 copies/10^6 reads) and transcripts (2 copies/10^6 reads) encoding benzoate 1,2-dioxygenase (K05549) compared to benzoate CoA-ligase. Accordingly, the ring-fission benzoyl-CoA 2,3-epoxidase (K15512) was significantly more abundant (22 copies/10^6 reads) than the ring-fission marker genes catechol 1,2-dioxygenase (K03381) and catechol 2,3-dioxygenase (K00446) (3 and 7 copies/10^6 reads in the deep and SCM respectively). However, both benzoyl-CoA 2,3-epoxidase and catechol 2,3-dioxygenase were among the most abundant genes in the metatranscriptomes (20 copies/10^6 reads), but with maximum abundance in the SCM and the deep, respectively (Figure 2.8b). Of the ring-fission pathway marker genes, gentisate 1,2-dioxygenase (K00450) was one of the most abundant in metagenomes (15 copies/10^6 reads in the FDOMmax) and metatranscriptomes (20 copies/10^6 reads in the deep).

2.3.5 Taxonomic identity of aromatic compound degradation genes and their distribution across the microbiomes

We estimated the fraction of bacterial genomes harboring each marker gene by comparing the total number of gene variants for select aromatic compound degradation pathway markers to the number of the single copy universally-distributed *recA* genes (Figure 2.8c, Figure 2.11). The estimated fraction of bacterial genomes with aromatic compound degradation genes increased with depth, reaching a maximum in the FDOMmax (8-75%, Figure 2.8c) and then decreased in the deep water (5-25%). The genes present in the highest fraction of bacterial genomes were involved in the degradation of benzoate through benzoyl-CoA (50% for K04110 and 45% for K15521 in the FDOMmax), gentisate (65% for K00450 in the FDOMmax), vanillate (75% for K03862 in the FDOMmax), salicylate and 3-hydroxycinnamate (45% for K00480 and 40% for K05712 in the SCM) (Figure 2.11). These numbers may be overestimated as they assume only a single gene copy per genome, whereas multiple paralogs may be present in a single genome.

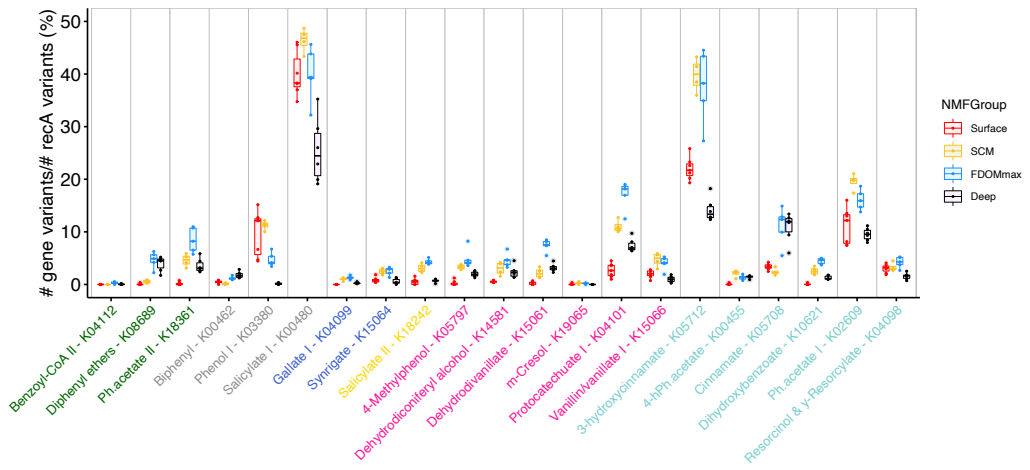

Figure 2.11: *Estimated fraction of the microbiome harboring genes annotated with KO marker of aromatic compounds.*

Taxonomic analysis of aromatic compound degradation marker genes revealed that the number of gene clusters generally increased continuously with depth (Figure 2.8d). Surface gene

clusters were predominantly affiliated with *Rhodobacterales* (more than 50% of the gene clusters for K03381, K00446, K00481 and K03862) and unclassified *Alphaproteobacteria* (up to 50% for K15065 gene clusters), with a significant contribution from *Gammaproteobacteria* for K05549 (40%) and K15065 (25%) (Figure 2.8d). In the SCM and FDOMmax, unclassified *Alphaproteobacteria* dominated the taxonomic affiliations of aromatic compound degradation gene clusters (10-55%) and *Rhodospirillales* contributed significantly to all gene clusters (10-30%), except for genes involved in the degradation of methylgallate (K15065), which was primarily encoded by Chloroflexi (30% in the FDOMmax) (Figure 2.8d). We generally observed more gene clusters in the deep than in the FDOMmax (Figure 2.8d), while these genes were present in a smaller fraction of the deep communities than the FDOMmax communities (Figure 2.8c), suggesting a broader phylogenetic diversity of aromatic compound degradation genes in the deep than in the FDOMmax. This is supported by the large contribution of other taxa (taxa contributing individually to <5%) in the deep microbiomes. The contribution of taxa such as *Rhodospirillales* and *Rhodobacterales* may be underestimated due to the large fraction of *Alphaproteobacteria* genes that could only be assigned at the class level.

2.3.6 Aromatic compound degradation pathways captured in metagenome assembled genomes

We reconstructed metagenome-assembled genomes (MAGs) from our metagenomic data. We performed metagenomic binning of each of the 22 metagenome assemblies individually to reconstruct a total of 1,772 MAGs. After filtering for genomes with greater than 30% completeness and less than 10% contamination, 823 genomes remained (Figure 2.12). Thirty-one of the 32 marker genes involved in aromatic compound degradation pathways were identified (only dihydroxyphenylacetate 2,3-dioxygenase – K00455 was not detected) across 59% (482 of 823) of the MAGs (Figure 2.13 and 2.14). The highest percentage of MAGs harboring aromatic compound degradation genes was in the FDOMmax (64%) and SCM (67%), and the lowest percentage in the surface (47%) and deep waters (54%). In general, the taxonomic diversity of MAGs increased with depth. Marker genes were identified in a broad taxonomic diversity of MAGs, including *Alphaproteobacteria*, *Gammaproteobacteria*, and *Dehalococcoidiia*, among others (Figure 2.13). *Alphaproteobacteria* were common in the SCM and FDOMmax, while *Gammaproteobacteria* were common in the surface.

To investigate the ecology of bacterial taxa most implicated in the degradation of aromatic compounds, we further examined MAGs with complete or near-complete aromatic compound degradation pathways. We selected 46 MAGs most enriched in near-complete aromatic compound degradation pathways (see methods, Figure 2.14). Of the 46 MAGs, 24 were recovered from metagenomes originating from the FDOMmax and 16 from the SCM layers. 38 of the MAGs were assigned to *Alphaproteobacteria* (Figure 2.15), 3 MAGs belonged to the *Dehalococcoida*, 4 MAGs to the *Gammaproteobacteria* and one MAG to the class *Binatia*.

Given the large representation of *Alphaproteobacteria* in the MAGs most implicated in aromatic compound degradation, we investigated the evolutionary origins and phylogenetic rela-

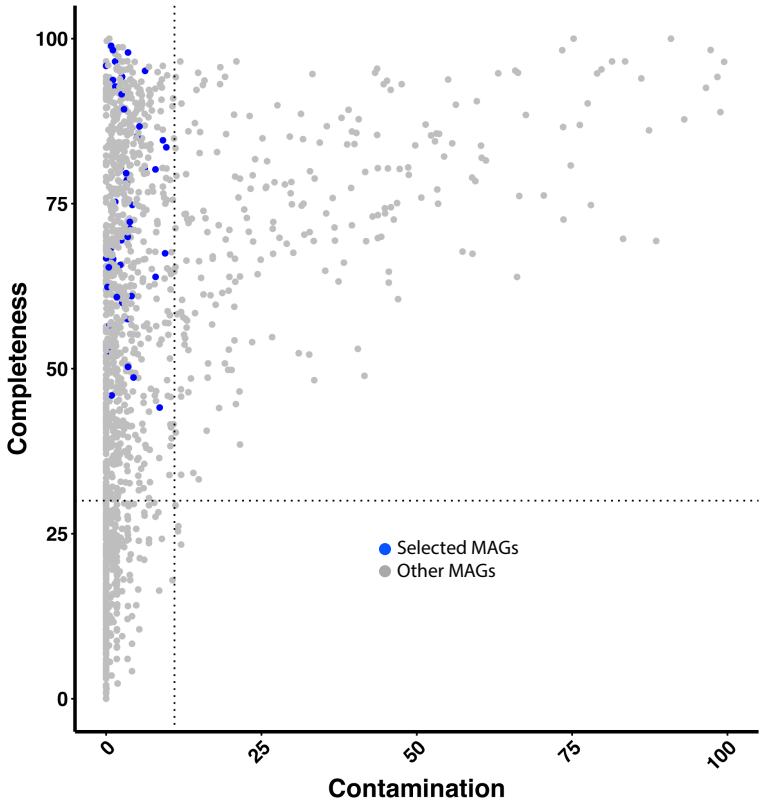

Figure 2.12: *Completeness and contamination of the set of 1772 MAGs reconstructed from 22 individual metagenomes. The vertical and horizontal dotter lines represent 10% contamination and 30% completeness respectively. Selected MAGs are the 46 MAGs that have been selected as most implicated in lignin-derived aromatic compound degradation based on the number and completeness of their aromatic compound degradation pathway.*

tionships between our 38 *Alphaproteobacteria* MAGs and reference genomes available from the Genome Taxonomy Database (GTDB)[48]. Based on an average nucleotide identity threshold of 95%, our 38 *Alphaproteobacteria* MAGs belonged to 16 genomospecies, that were phylogenetically associated with 12 clades (Figure 2.17a, Figure 2.16). The clades were within the *Rhodobacterales*, *Thallassobaculales*, *Rhodospirillales*, *Defluviicoccales* and five GTDB orders of uncultured *Alphaproteobacteria* (UBA8366, UBA6615, GCA2731375, UBA2966, UBA828). Each clade was comprised of Canada Basin MAGs as well as a basal branch consisting of genomes of marine origin. These results demonstrate that the *Alphaproteobacteria* MAGs were phylogenetically distinct but shared recent common ancestry with genomes from other oceanic zones.

To investigate the distribution of the 12 clades represented by the *Alphaproteobacteria* MAGs across the Arctic water column features, we performed fragment recruitment of both the metagenomic and metatranscriptomic reads against the MAGs representing the 16 genomospecies (Figure 2.17b). Overall, the distribution in metagenomes and metatranscriptomes was similar and all the MAGs were most abundant and active either in the FDOMmax or the SCM (Figure 2.17b). We identified four general patterns of distribution across water column features consisting of 1) restriction to the FDOMmax (Defluvii-CB9_331, UBA2966), 2) common to the SCM

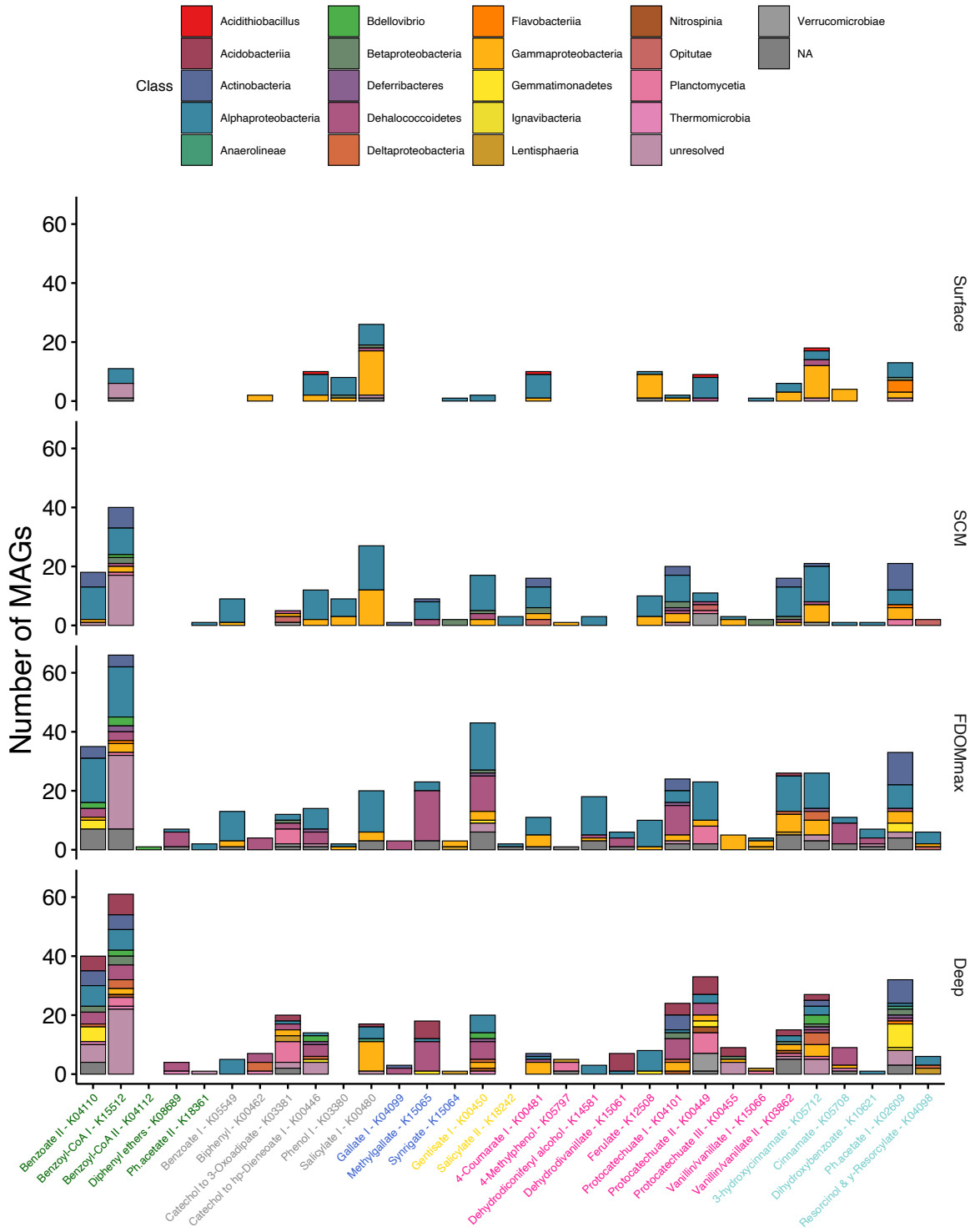

Figure 2.13: *Taxonomic identity of the MAGs harboring genes annotated with KO marker of aromatic compounds degradation pathways. The taxonomy is displayed at the class level.*

and FDOMmax (UBA8366, UBA6615 clade, *Thalassobaculales* clade, *Defluviiccocales* genomes CB21_SCM.1 and CB4_SCM.1), 3) common in the FDOMmax and deeper waters (UBA828 clade genomes, GCA 2732375 genomes and *Rhodospirillales* genomes) and 4) restricted to the SCM (Rhodosp-CB4_SCM and Rhodobact-CB21_SCM). These results show that MAGs with near-complete aromatic compound degradation pathways are strongly associated with and active in HS-rich regions of the water column.

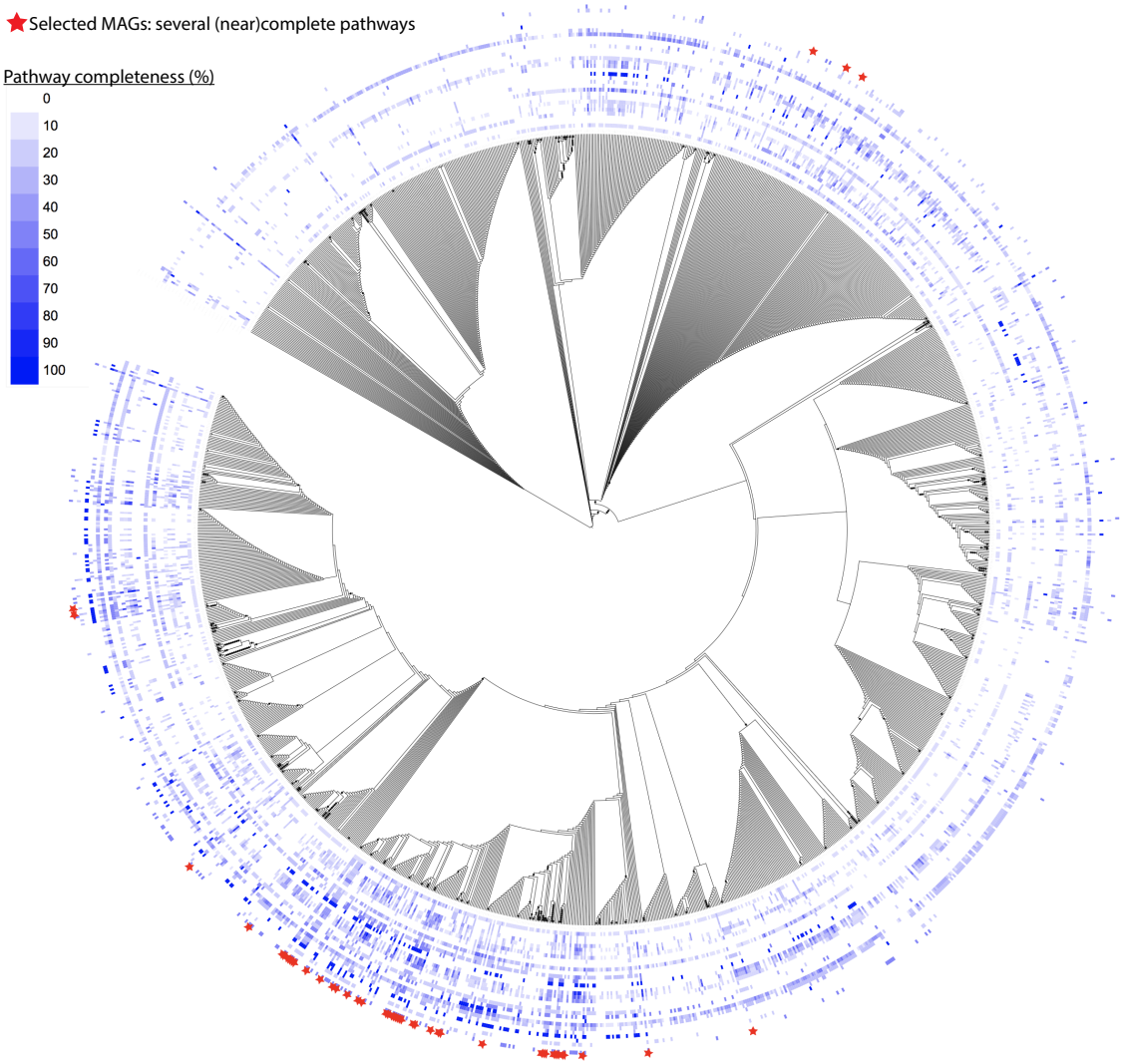

Figure 2.14: *Phylogenetic tree reconstructed from the concatenation of 120 conserved genes for the bacterial genomes of the MAGs dataset. The heatmap represents the completeness of the lignin-derived aromatic compounds degradation funneling pathways. Red stars correspond to the MAGs selected based on the amount and completeness of pathways they harbor.*

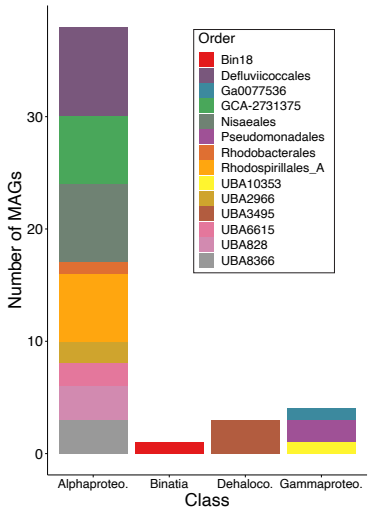

Figure 2.15: *Taxonomic identity of the MAGs most implicated in the degradation of lignin-derived aromatic compounds. Taxonomy is displayed at the order level and based on the tree placement of the MAGs, using the GTDB.*

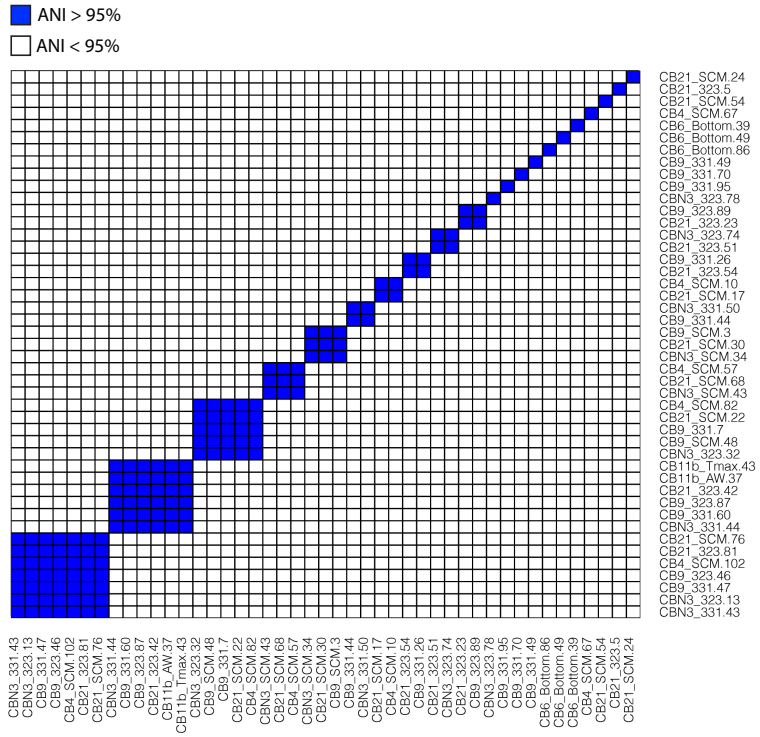

Figure 2.16: *Heatmap of the average nucleotide identity for the 46 MAGs most implicated in the degradation of lignin-derived aromatic compounds.*

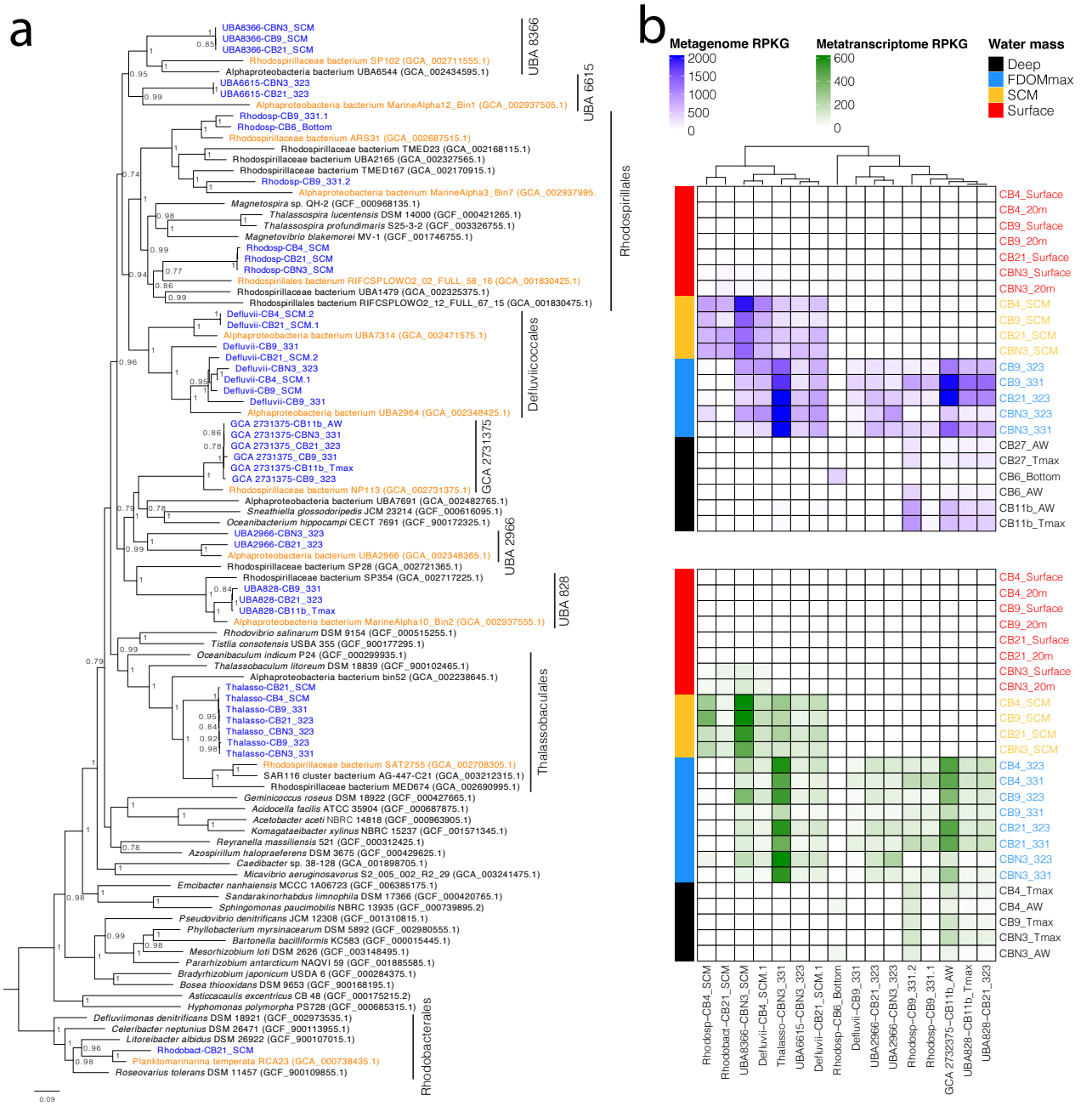

Figure 2.17: *a) Phylogenetic tree of the Alphaproteobacteria MAGs identified as most implicated in the degradation of aromatic compounds in the Arctic Ocean. Blue: MAGs reconstructed in this study; orange: genomes from other studies identified as the closest relatives to the MAGs reconstructed in this study; black: other genomes. b) geographic distribution of the Alphaproteobacteria MAGs identified as most implicated in the degradation of aromatic compounds within the Canada basin. Top panel: mapping of the metagenomes on the MAGs. Bottom panel: mapping of the metatranscriptomes on the MAGs.*

2.3.7 Global ocean distribution of Alphaproteobacteria MAGs from Canada Basin

We sought to determine if the MAGs most implicated in aromatic compound degradation were more broadly distributed beyond the Arctic Ocean. We therefore investigated the distribution of the *Alphaproteobacteria MAGs* by fragment recruitment against a set of 151 metagenomes broadly representative of the global ocean microbiome (Figure 2.18). Of the 16 representative MAGs, two were commonly detected outside of the Arctic Ocean. Rhodosp-

CB9_331.2 was identified in the mesopelagic metagenomes from all oceanic regions, but not the Mediterranean Sea or Red Sea. The *Rhodobacterales* MAG (Rhodobact-CB21_SCM) was also identified in surface water metagenomes, most notably from the Southern Ocean. Although several other MAGs were detected at low frequency in the Southern Ocean (e.g. Rhodosp-CB6_bottom and Defluvii-CB9_331) the majority (12 MAGs, 75% of the MAGs) were not detected outside of the Arctic Ocean. Likewise, the vast majority of the most closely related marine reference genomes were not detected in Canada Basin metagenomes. The exceptions were *Planktomarinamarina* (*Rhodobacterales*) and the reference genome within UBA828.

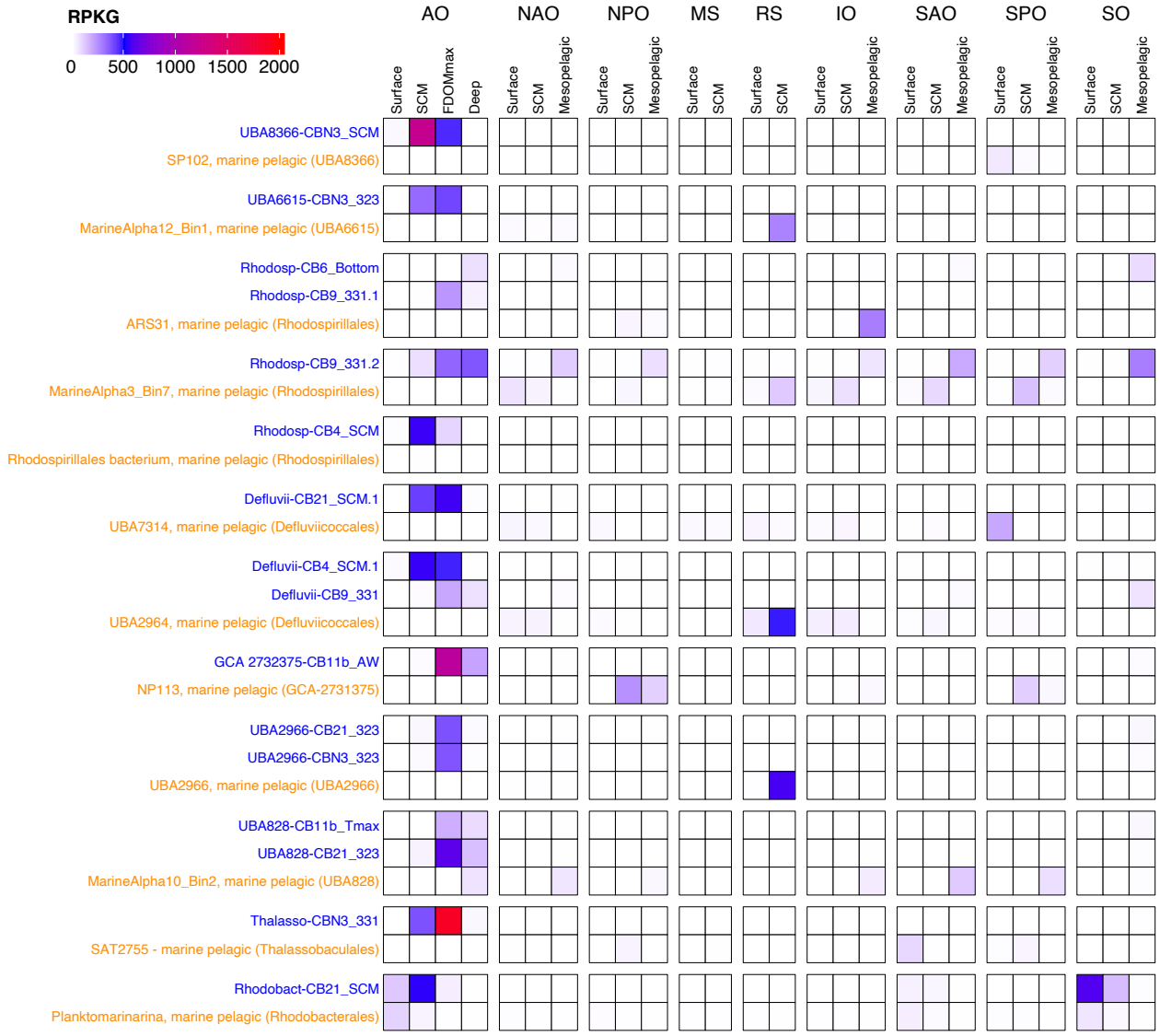

Figure 2.18: *Fragment recruitment of global oceans metagenomes on the Alphaproteobacteria MAGs identified as most implicated in the degradation of aromatic compounds. Blue: MAGs identified in this study and representative of the 16 genomospecies; orange: genomes from other studies identified as the closest relatives to the MAGs reconstructed in this study. AO: Arctic Ocean; IO: Indian Ocean; MS: Mediterranean Sea; NAO: North Atlantic Ocean; NPO: North Pacific Ocean; RS: Red Sea; SAO: Southern Atlantic Ocean; SO: Southern Ocean; SPO: Southern Pacific Ocean.*

2.3.8 Enhanced aromatic compound degradation capacity in Alphaproteobacteria MAGs restricted to Canada Basin

We hypothesized that an enhanced capability for aromatic compound degradation may be implicated in the evolutionary adaptation of the Arctic Ocean populations. We therefore compared the abundance and diversity of aromatic compound degradation genes between the Arctic MAGs and the set of their closest relatives.

Generally, the Arctic Ocean genomes possessed a similar diversity (4-18 gene families per genome), but more copies of marker genes (4-66 copies per genome) for the degradation of aromatic compounds compared to their closest relatives from other oceans (0-16 genes per genome) (Figure 2.19). Between 1.5% and 4% of the genes with an EC annotation was annotated with an EC from lignin-derived aromatic compound degradation pathways for both the Arctic MAGs and genomes from other oceans (Figure 2.19). However, out of the 16 Arctic Ocean MAGs, 10 possessed a higher diversity of marker genes and 14 possessed a higher number of marker gene copies than their sister taxa from other oceans (Figure 2.19). This is despite the difference in genome completeness estimated for the reference MAGs (64-98%) compared to the Arctic Ocean MAGs (range of 42-95% completeness). Some marker genes were found exclusively in our Arctic *Alphaproteobacteria* MAGs, including those for the degradation of catechol to 3-oxoadipate, phenol I, dehydrodivanillate, protocatechuate I, cinnamate and resorcinol (Figure 2.19). Other marker genes were found exclusively in the reference MAGs, such as genes for the degradation of benzoyl-CoA II, gallate I and III, m-cresol and gamma-resorcylate.

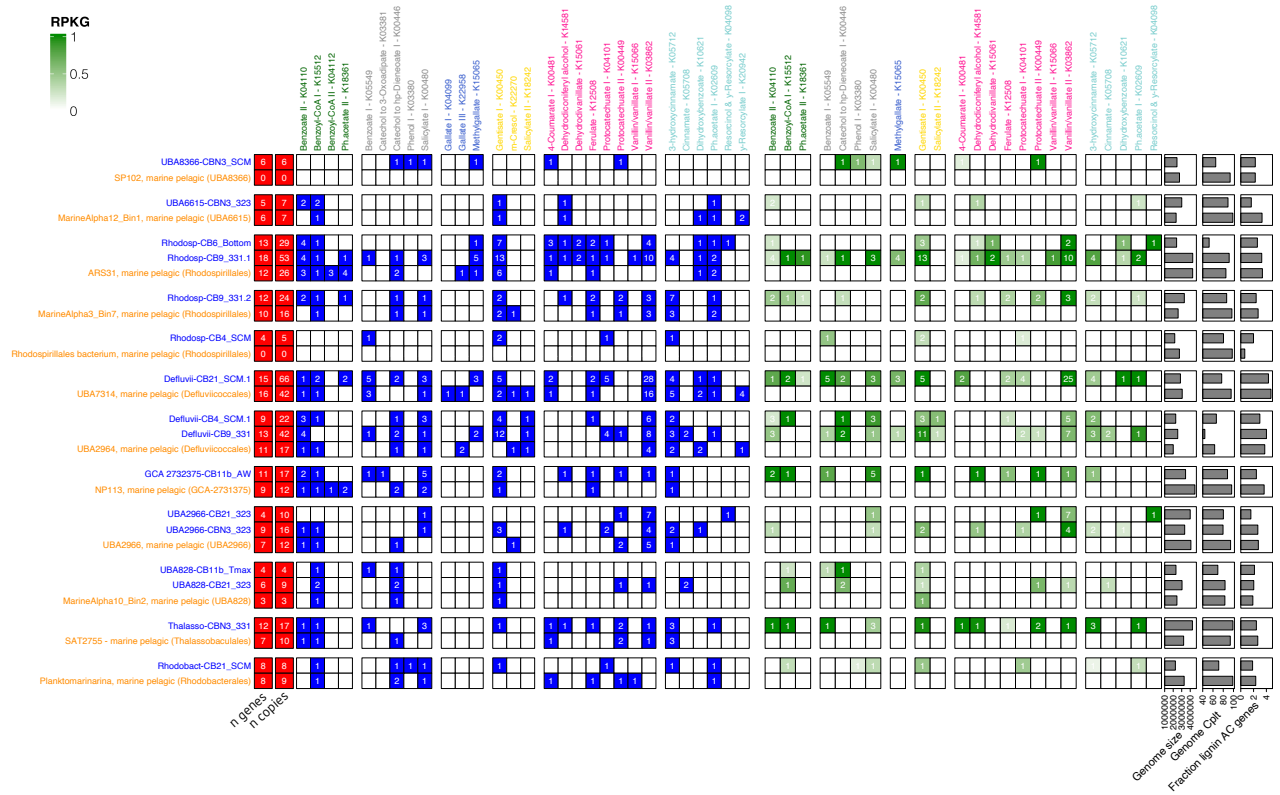

Figure 2.19: *Comparative genomics of the the Arctic Ocean Alphaproteobacteria MAGs identified as most implicated in the degradation of aromatic compounds (blue) and their closest relative genomes (orange) from other studies. Left panel: graphical table of the presence (blue cells) and absence (white cells) of genes annotated with Kos marker of aromatic compound degradation pathways. Number in the cells represent the number of genes annotated with a KO. Right panel: graphical table of the mean expression (green cells) and absence of expression (white cells) of genes annotated with KOs marker of aromatic compound degradation pathways. Number in the cells represent the number of genes expressed.*

Every single Arctic MAG contained at least one and up to seven marker genes that were not found in its closest related genome from other oceans (Figure 2.19). These included gentisate 1,2-dioxygenase (K00450) in Rhodosp-CB4_SCM, Defluvii-CB9_331, UBA2966-CBN3_323, Thalasso-CBN3_331 and Rhodobact-CB21_SCM, as well as vanillate O-demethylase (K03862) in Rhodosp-CB6_Bottom and Rhodosp-CB9_331.1 or benzoate CoA-ligase (K04110) in UBA6615-CBN3_323 and Rhodosp-CB9_331.2. These genes were the most abundant in the metagenomes and metatranscriptomes and most common throughout the Arctic Ocean microbiomes (Figure 4a-c). The three MAGs with the highest number of aromatic compound degradation gene copies (Defluvii-CB21_SCM.1, Rhodosp-CB9_331.1 and Defluvii-CB9_331) all possessed more copies of the vanillate O-demethylase (K03862) gene than their sister genomes. In addition, all of their sister genomes lacked the protocatechuate ring-opening genes (K04101 or K00449). A vast majority of the marker genes were also expressed in our Arctic Ocean MAGs (Figure 7, right panel), and we found only gentisate 1,2-dioxygenase (K00450) slightly expressed in the MarineAlpha10_Bin2 of the reference genomes.

2.4 Discussion

Aromatic compound degradation capacity of the Arctic Ocean microbiomes reflects the ability to degrade humic-rich DOM from terrestrial and sediment origin

The degradation of aromatic compounds emerged as a central metabolism of the Canada Basin microbiomes, which is consistent with a capacity to degrade the humic-rich DOM in the Arctic Ocean. We found that the aromatic compound degradation genes contributed more to the total metabolic system of the Arctic Ocean microbiomes than in other oceans, paralleling the higher concentrations of HS in the Arctic Ocean[18]. This enhanced aromatic compound degradation capacity was associated with the humic-rich tOM layer of the Canada Basin (FDOMmax). Individually, most of the aromatic compound degradation genes were more abundant in the FDOMmax (Figure 3a), where humic-rich DOM concentrations are maximal, as evidenced by the distribution of the FDOM C1 fraction (Figure 2.1c). The lower abundance of aromatic compound degradation genes in the surface (Figure 2.6a) may be explained by a preference of the surface microbiomes to process the non-aromatic tOM fraction as sunlit waters are generally poor in aromatic compounds. This is supported by the higher contribution of the non-photoreactive (non-aromatic) FDOM C4 fraction and very low contribution of the aromatic C1 fraction in the surface (Figure 2.1c).

We found that the distribution and diversity of aromatic compound degradation genes in the Canada Basin matched the diversity of aromatic compounds expected from the Arctic Ocean watershed organic matter. Lignin is a polymer of three aromatic monolignols, which form the H, G and S unit (defined by 0, 1 and 2 methoxy group on the aromatic ring, respectively) when cross-linked. Coniferous trees dominate the boreal forest of the Arctic Ocean watershed, and are characterized by a high G/S ratio in their lignin polymer[49]. Vanillin (G-unit) is therefore expected in the tOM of the riverine input to the Arctic Ocean. This has been shown in the Mackenzie River, the major river draining to the Canada Basin: the OM of suspended sediments in the Mackenzie was dominated by 1-methoxy aromatic compounds (G), among which vanillin and vanillate contributed the most[50]. The Mackenzie River also contained a significant contribution of benzoate derivatives (H) and smaller amounts of syringate derivatives (S). Our results showed that the genes involved in the degradation of vanillate, benzoate and methylgallate (syringate derivatives) were among the most abundant and expressed genes (Figure 2.6a-b) in the Canada Basin microbiomes, paralleling the aromatic compound composition of the McKenzie River DOM input.

Sediments also contribute to the humic-rich OM reaching the Canada Basin, by exchanging OM with brine sinking along the shelf[16]. The OM on the Mackenzie shelf, bordering the Canada Basin, contained higher levels of vanillyl, syringyl and cinnamyl phenols than any other North American Arctic shelf[51]. The high abundance and expression of genes involved in the degradation of vanillate, syringate and 3-hydroxycinnamate (Figure 2.6a-b) suggests that the Canada Basin microbiomes access the variety of aromatic compounds from shelf sediments sources. This demonstrates that the Canada Basin microbiomes can access humic-rich DOM from terrestrial and sediment origin as growth substrates.

Rhodospirillales are implicated as aromatic compound degraders in the Arctic Ocean

We showed that a few clades of *Alphaproteobacteria* were strongly implicated in aromatic

compounds degradation (Figure 2.8, 2.12-2.13). Previous work in the Arctic Ocean showed that *Gammaproteobacteria* are associated with humic-rich DOM degradation. For example, in an experiment adding humic-rich OM derived from a thermokarst to coastal Arctic water (Chukchi sea) microbiomes, *Gammaproteobacteria* taxa *Colwelliaceae* (order *Alteromonadales*) rapidly dominated the microbial communities[52]. Similarily, *Alteromonadales* (genera *Glaciecola*, *SAR92* clade) were associated with humic-rich riverine-derived OM consumption in an Arctic fjord[53]. These studies may seem incongruent with our results. However, these studies focused on the microbiomes of the surface waters only. In our study, we observed a significant *Gammaproteobacteria* signal in the taxonomy of aromatic compound degradation genes within the surface samples (Figure 2.6, 2.11), which supports earlier studies. In addition, the tOM used in previous experiments contained many other compounds than aromatic compounds, including more labile protein-like compounds. The *Gammaproteobacteria* in the surface could then be adapted to a fast consumption of pulses of labile OM, while the *Rhodospirillales* taxa of our study were more adapted to a slow degradation of more refractory and steadier amount of aromatic compound-rich humic substances in the FDOMmax. This would be in line with the high variability in seasonal conditions and DOM concentrations in the surface waters of the Arctic Ocean compared to more stable DOM concentrations and conditions in the FDOMmax throughout the year[54,55].

All but one of the MAGs most implicated in aromatic compound degradation belonged to closely related *Alphaproteobacteria* clades, based on the GTDB. Based on NCBI taxonomy, all but one of these MAGs belonged to the *Rhodosprillales* order. Here we used the NCBI taxonomy to be able to relate our findings to previous reports in the literature. We concluded that the capacity to degrade aromatic compounds is phylogenetically concentrated in *Rhodosprillales* within the Arctic Ocean. In the global ocean, a few studies previously identified *Rhodospirillales* taxa in aromatic compound- and HS-degrading consortia. A *Rhodospirillales* strain (*Thalassospira profundimaris*) was one of six taxa isolated from the East China Sea surface microbiomes enriched with vanillic acid[56]. This strain was able to grow on benzoic acid, 4-hydroxybenzoic acid, and to a lesser extent on syringate and ferulate. However, this *Rhodospirillales* strain is a member of a different clade (also including *Magnetospira* and *Magnetovibrio*) than the *Rhodospirillales* genomes we report in our study (Figure 2.8a). *Rhodospirillales* were also identified in flow-through experiments in which marine microbiomes were exposed to riverine HS as a sole carbon source[8]. *Rhodospirillales* represented 6% of the taxa identified and were only reported in the low salinity experiment (14 PSU). Taxa were found in the *Thalassospira* (4 taxa) and *Thalassobaculales* (1 taxa) clades. We also reported *Thalassobaculales* genomes within the taxa implicated in aromatic compound degradation. However, the genomes we reported were located in the water column at salinities $>$ 30 PSU. The studies reporting *Rhodospirillales* focused only on the surface water microbiomes, while we investigated the whole water column. The focus on surface waters in other studies is usually based on the assumption that aromatic compounds and HS have a terrestrial origin and are transported to the ocean with freshwater input, therefore concentrated in the surface layers. The focus on surface waters microbiomes may therefore explain why *Rhodospirillales* have not yet been reported as most implicated in aromatic compounds and HS degradation within the ocean. Based on the distribution of HS

in the Arctic Ocean, we investigated the whole water column and specifically the FDOMmax, which allowed us to identify *Rhodospirillales* as strongly implicated in aromatic compound degradation. Further work within other oceans, as well as experimental work using HS as sole carbon sources in the microbiomes of the FDOMmax will be necessary to fully elucidate the role of *Rhodospirillales* in the degradation of HS in the Arctic Ocean, and the global ocean.

Evolutionary adaptation of Rhodospirillales in the Arctic Ocean

The phylogenetic divergence of our *Rhodospirillales* MAGs from relatives in other oceanic regions and their restricted distribution to the Arctic Ocean suggests the Arctic populations are evolutionarily adapted to life in the Arctic Ocean. The high number of aromatic compound degradation gene copies in the MAGs compared to their closest relatives in other oceans suggest that the capacity to use aromatic compounds as a growth substrate played a role in their evolutionary adaptation. The disproportionately high amount of HS in the FDOMmax may then act as a selective pressure on these MAGs. Previous studies have demonstrated evolutionary adaptation of microbial MAGs restricted to the Canada Basin: a new *Methylophilaceae* clade[57] as well as *SAR11*[58] and *Chloroflexi* ecotypes[44]. The *Methylophilaceae* clade evolved via a freshwater to marine transition, highlighting the importance of the terrestrial-marine interface in shaping the Canada Basin microbiomes. However, HS does not appear as the main selective pressure for *Methylophilaceae* as their distribution is restricted to the surface. The Arctic *SAR11* and *Chloroflexi* clades were restricted to the SCM and FDOMmax, where humic-rich DOM is enriched within the Canada Basin. The *Chloroflexi* Arctic ecotype was replete with aromatic compound degradation genes, some of these acquired by lateral gene transfer from terrestrial taxa. This *Chloroflexi* ecotype was found within the water masses rich in HS (FDOMmax), similarly to the *Rhodospirillales* MAGs of our study. The preference for humic-rich water masses coupled to an enhanced capacity to degrade aromatic compounds in *Chloroflexi* and *Rhodospirillales* suggest that the ability to use aromatic compounds as growth substrate provides an evolutionary advantage in the humic-rich environment of the Canada Basin FDOMmax.

2.5 Conclusion

The dissolved organic matter of the Arctic Ocean is characterized by a disproportionately high contribution of HS compared to other oceans. With the increasing terrestrial input of humic-rich OM to the Arctic Ocean as a result of escalating permafrost thawing and river runoff, it is predicted that the contribution of HS to the Arctic Ocean OM will increase[25]. The fate of this carbon is important to consider with respect to changing biogeochemical cycles of the Arctic Ocean. In this study, we showed that the metabolic pathways involved in the degradation of HS were widespread, abundant and expressed in the microbiomes of the Canada Basin. The capacity to degrade humic-rich OM in the Arctic Ocean microbiomes was enhanced compared to the microbiomes of the global ocean in the upper water column. The diversity and distribution of the aromatic compound degradation machinery revealed that the Arctic Ocean microbiomes were equipped to use OM from terrestrial sources as growth substrates. We identified that the

aromatic compound degradation capacity was concentrated phylogenetically in *Rhodospirillales*. The phylogeny, comparative genomics and biogeographic distribution of these *Rhodospirillales* suggest an evolutionary adaptation driven by the disproportionately high amount of HS in the Arctic Ocean. Overall, this study demonstrates that the Arctic Ocean microbiomes are capable of processing OM of terrestrial origin. Our study predicts that OM of terrestrial origin can be remineralized in the Arctic Ocean and that *Rhodospirillales* will gain importance as tOM inputs continue to increase in the Arctic Ocean.

2.6 Methods

Sampling, DNA and RNA extraction

Samples were collected in September 2017 during the Joint Ocean Ice Study cruise to the Canada Basin. We analyzed 22 metagenomes and 25 metatranscriptomes generated from samples collected across the water column of the Canada Basin. Eight specific water masses were sampled: the surface mixed layer (surface: 5 m and 20 m depth) characterized by fresher water due to riverine input and ice melt, the subsurface chlorophyll maximum (SCM), in the halocline (FDOMmax at salinity of 32.3 and 33.1 PSU, referred as 32.3 and 33.1), and deeper water from Atlantic origin at the temperature maximum (referred as Tmax), 1000 m depth (Atlantic water, further referred as AW) and 10 or 100 m above the bottom (further referred as bottom).

We filtered 14L of seawater for DNA samples and 7L of seawater for RNA samples sequentially through a 3 μm pore size polycarbonate track etch membrane filter (AMD manufacturing, ON, Canada) and a 0.22 μm pore size Sterivex filter (Millipore, MA, USA). Filters were stored in RNALater (ThermoFischer, MA, USA), and kept frozen at -80^{0}C until processing in the lab. DNA was extracted following the method described in Colatriano *et al.*[59]. Briefly, the preservation solution was expelled and replaced by a SDS solution (0.1 M Tris-HCl pH 7.5, 5% glycerol, 10 mM EDTA, 1% sodium dodecyl sulfate) and incubated at room temperature for 10 min and then at 95^{0}C for 15 min. The cell lysate was then centrifuged at 3,270 x g. Proteins were removed by precipitation with MCP solution (Lucigen, WI, USA) and the supernatant was collected after centrifugation at 17,000 x g for 10 min at 4^{0}C. DNA was precipitated with 0.95 volume of isopropanol and rinsed twice with 750 μL ethanol before being air dried. The DNA was resuspended in 25 μL of low TE buffer, pH 8 (10 mM Tris-HCl, 0.1mM EDTA) and stored at -80^{0}C.

The RNA extraction procedure was adapted from the mirVana RNA extraction kit (ThermoFisher, MA, USA). RNAlater was expelled from the Sterivex and replaced by 1.5 mL of Lysis buffer and Sterivex was vortexed. 150 μL of miRNA homogenate were added, the Sterivex vortexed and incubated on ice for 10 min. The cell lysate was expelled from the Sterivex, 0.9x the volume of acid-phenol-chloroform was added and the solution was vortexed for 30-60 sec. The mix was centrifuged at 10,000 x g for 5 min, and the top aqueous phase gently removed and transferred to a fresh tube. 1.25 volume of 100% ethanol was added to the aqueous phase and vortexed to mix. The mix was filtered through mirVana Filter Cartridges by centrifugating at

10,000 x g for 10 s, and the flow through discarded. The RNA was rinsed with 700 μL of Wash Solution 1 and then with 500 μL Wash solution 2/3 by centrifugating at 10,000 x g for 10 sec. RNA was then eluted with 50 μL of Elution solution (0.1 M EDTA) warmed at 95^0C. 700 μL of RTL buffer and 500 μL 100% ethanol were added to the RNA suspension and the suspension was centrifuged for 15 sec at 10,000 x g on a RNeasy MinElute column. RNA was washed first with RPE buffer by centrifuging 500 μL for 15 sec at 10,000 x g and then 80% ethanol for 2 min at 10,000 x g. The empty column was then centrifuged at 12,000 x g for 5 min to discard the excess liquid. The RNA was finally eluted by centrifugation of 28 μL and then 10 μL of RNase free water for 1 min at 12,000 x g and stored at -80^0C.

Dissolved organic matter samples collection, analysis of fluorescence measurements

DOM samples were collected in Niskin bottles mounted on a conductivity-temperature-depth rosette profiler and immediately filtered using pre-combusted 0.3 μm glass fiber filters (GF75, Advantec) into pre-combusted amber glass vials. Fluorescence spectra were measured using a Fluoromax 4 Jobin Yvon fluorometer[25]. Parallel factor analysis was applied to decompose the fluorescence signal into their main components following the procedures outlined in Murphy *et al.*[60]. The PARAFAC model validated 7 components including 5 humic-like (C1-C2, C4-C5, and C7) and 2 protein-like components (C3 and C6) in 4,483 samples collected from surface to 10 m above bottom sediment in the Canada Basin between 2007 and 2017[25].

Metagenomic sequencing, assembly and annotation

Sequencing, assembly and annotation were performed by the Joint Genome Institute (CA, USA). Each individual metagenome and metatranscriptome were sequenced on the Illumina NovaSeq platform, generating paired-end reads of 2x150 bp for all libraries. Single assemblies were created for each individual sample using SPAdes[61] with kmer sizes of 33,55,77,99,127 bp. Gene prediction and annotation was performed using the DOE Joint Genome Institute Integrated Microbial Genomes Annotation Pipeline v.4.16.5[62].

Building of EC and KO abundance matrices

The number of copies of genes annotated to each Enzyme Commission (EC) number or KEGG Orthology (KO) number in a metagenome or metatranscriptome was calculated by summing the depth of coverage of all genes or transcripts annotated with this EC or KO. To obtain the final EC and KO abundances (number of gene or transcript copies/106 reads for this EC or KO) the total number of copies were then normalized by the library size (number of reads) with the TMM method[63], using the calcNormFactors function of the edgeR package in R[64]. Genes were assigned to a total of 3,102 EC numbers and 12,018 KO identifiers for the metagenomes and 2,830 EC numbers and 10,556 KO identifiers for the metatranscriptomes. Before multivariate analysis, EC and KO abundances were transformed with a Hellinger transformation (decostand function from the R vegan package[65]).

Multivariate analysis

Non-metric multidimensional scaling (NMDS) was performed on the EC number abun-

dance matrices with the metaMDS function from the R vegan package, using two dimensions and the Bray-Curtis dissimilarity metric. Non-negative matrix factorization (NMF) was performed with the nmf function from the NMF package in R[66]. NMF decomposes the abundance matrix into two matrices: a coefficient matrix that describes the overall structure of the abundance matrix with a limited number of descriptors (called sub-metagenomes and sub-metatranscriptomes in this study, their number being the rank), and a basis matrix the provides the weights of each original descriptors (EC number) on the new descriptors (sub-metagenomes, sub-metatranscriptomes). The advantage of NMF is to directly link the overall structure of the abundance matrix to the individual elements (EC number) driving this structure. We first performed the NMF analysis with rank values ranging from 3 to 7, 100 runs, and various algorithms (”nsnmf ”, “Brunet”, ”KL”). We obtained the optimal results for the nsNMF algorithm, random seed of the factorized matrices, and a rank value of 4. We performed the final analysis with 200 runs, rank of 4, random seed and nsNMF algorithm.

Calculation of gene and pathway indices

The indices were calculated for EC number annotated from metagenomes and metatranscriptomes by combining two methods described in Jiang *et al.*[67] and Kim textitet al.[68]. We first used the EC number abundance matrices (annotated from metagenomes and metatranscriptomes) and the coefficient matrices (SMG/SMT x samples) to calculate both the spearman correlation coefficient and the multidimensional projection between all pairs of EC number annotated from metagenomes (ECMG) and SMG as well as all pairs of EC number annotated from metatranscriptomes (ECMT) and SMT. The spearman correlation coefficient between an ECMG/ECMT (i) and a SMG/SMT (k) $\rho_{i,k}$ was calculated using the abundance profile of a ECMG/ECMT and a SMG/SMT along all the samples. The multidimensional projection between an ECMG/ECMT and a SMG/SMT was calculated as the cosine of the angle between the vectors represented by an ECMG/ECMT abundance in the samples space and the vector represented by SMG/SMT in the samples space. The abundance profiles of ECMG/ECMT and SMG/SMT were first normalized, and the multidimensional projection was calculated as:

$$Cos\Theta_{i,k} = \sum_{j=1}^{n} a_{i,j} \times s_{k,j} \tag{2.1}$$

Where $\text{Cos}\Theta_{i,k}$ is the multidimensional projection between the ECMG/ECMT i and the SMG/SMT k, n is the number of samples, $a_{i,j}$ is the normalized abundance of the ECMG/ECMT i in the sample j, and $s_{k,j}$ is the normalized abundance of the SMG/SMT k in the sample j. We then used the basis matrix to calculate the score of each ECMG/ECMT:

$$Score(i) = 1 + \frac{1}{log_2(q)} \sum_{k=1}^{q} p(i,k) \times log_2(p(i,k)) \tag{2.2}$$

Where i is the ECMG/ECMT, q is the number of SMG/SMT (4 in our study), k is the SMG/SMT, p(i,k) is the probability of finding the ECMG/ECMT i in the SMG/SMT k We calculated the final EC index (annotated from metagenome and metratranscriptome) on each

SMG/SMT by multiplying the spearman correlation coefficient, cos theta and the EC score:

$$I_{i,k} = \rho_{i,k} \times Cos\Theta_{i,k} \times Score(i) \tag{2.3}$$

This allowed us to calculate an index for each pair of ECMG/ECMT and SMG/SMT.

Aromatic compound degradation gene and pathway selection

To select the metabolic pathways and enzymes involved in the degradation of lignin-derived aromatic compounds, we first surveyed the literature for the various lignin breakdown compounds reported to be degraded by bacteria[36,69,70]. We then retrieved the MetaCyc (https://metacyc.org) pathways that were involved in the degradation of these compounds, as well as the EC numbers involved in these pathways. For analysis of marker genes for each pathways, we first selected one key reaction in each pathway. We then retrieved genes annotated with KO numbers corresponding to the EC numbers associated with the key reactions. We first chose reactions involved in aromatic ring-opening, aromatic ring-oxidation, and aromatic ring-reduction steps. If the EC numbers of these reactions were not specific to a pathway, we chose reactions involved in the addition of CoA to the aromatic ring or reactions involved in oxidoreduction steps of the side chains of the aromatic ring. If 2 aromatic compound degradation pathways possessed the same EC number associated with the selected key reaction, and if this EC number was not found on any other Metacyc pathways, we used the EC as marker for only one of the two pathways. We could not retrieve any marker EC number for several pathways.

Calculation of the fraction of genes involved in the degradation of aromatic compounds within the pool of metabolic genes

To calculate the percentage of the gene pool associated with the degradation of aromatic compounds in each sample, we summed the total number of genes copies annotated with EC number within our selection of aromatic compounds degradation pathways. We then divided this number by the total number of gene copies annotated with EC numbers.

Statistical analyses

To compare the means of EC indices distribution as well as the percentage of aromatic compound degradation genes in metagenomes, we first performed a permutational ANOVA (PERMANOVA) using the perm.anova function of the RVAideMemoire package in R. When the p-value of the PERMANOVA test was less than 0.05, we performed pairwise comparison between groups with Student t-test, using the PERMANOVA residuals variance as the variance for the Student t-tests. Two groups were considered different if their p-value was less than 0.05.

Calculation of aromatic compound degradation pathways completeness

Pathway completeness in a sample was calculated by dividing the number of EC numbers belonging to this specific pathway and present in a sample by the total number of EC number of this specific pathway. The marker genes abundance was obtained using the normalized abundance of a specific KO as calculated to obtain the KO abundance matrices (see above).

Estimation of the fraction of taxa harboring aromatic compound degradation genes

The estimated percentage of genomes harboring marker genes in a sample was calculated by dividing the total number of gene variants annotated with the KO of interest by the total number of gene variants annotated with the KO corresponding to the recA gene (K03553).

Taxonomic assignment of aromatic compound degradation genes

Genes annotated with KO identified as marker for selected lignin-derived aromatic compounds degradation pathways were grouped by water column feature and dereplicated with CD-hit (v.4.6)[71] at 95% identity. The dereplicated set of genes were searched against the NCBI nr protein database (downloaded 21/08/27) using DIAMOND (v.0.9.30.131)[72]. To assign a taxonomic identity to these genes, the DIAMOND (v.0.9.30.131) output was imported in MEGAN[72] using the January 2021 mapping file ("megan-map-Jan2021.db"). The lowest common ancestor parameters were set at minimum e-value of $1x10^{-20}$ and at top percent of 1%. The file containing taxonomic identity of the genes was then exported from MEGAN and processed with a custom-made R script.

Metagenome binning

Metagenomic binning was performed on each individual assembly with Metabat2 (v.2.12.1)[73] using scaffold longer than 2500 bp. Contamination and completeness of the metagenome-assembled genomes (MAGs) were estimated with CheckM (v.1.0.7)[74]. MAGs greater than 30% completeness and less than 10% contamination were selected for further analysis. Phylogenetic placement of MAGs was performed based on the concatenation of 120 conserved genes for bacteria and 122 conserved gene for archaea using the Genome Database Taxonomy Database toolkit (GTDB-Tk – v.1.3.0)[48,75].

Metabolic reconstruction and MAG selection based on the capacity to degrade aromatic compounds

To select MAGs enriched in aromatic compound degradation capacity, we selected all the genes annotated with EC numbers belonging to pathways involved in the degradation of aromatic compounds within the MAGs. As ring-fission pathways can be involved in the degradation of non-lignin aromatic compounds, we used only the funneling pathways to select for MAGs enriched in the degradation of lignin aromatic compounds. For each MAG, we calculated the completeness of all funneling aromatic compounds degradation pathways. We considered a pathway complete if a MAG contained genes annotated with all the EC number of this pathway. The pathway completeness percentage was obtained by dividing the number of EC numbers involved in a pathway within a MAG by the number of reactions of this pathway. This number was normalized by the MAG completeness. For each MAG, we then calculated the median of all pathway completeness. Based on the distribution of the medians of all MAGs, we selected 4% as the median threshold above which a MAG was selected as having a high capacity to degrade aromatic compounds. We obtained a total of 46 MAGs. We calculated the average nucleotide identity (ANI) between these 46 MAGs using fastANI (v.1.3)[76] and grouped the MAGs with an ANI>95% as the same genomospecies, obtaining a total of 22 genomospecies.

Phylogenetic analyses of MAGs

To reconstruct the phylogeny of the 38 Alphaproteobacteria MAGs (16 genomospecies), we first manually investigated their phylogenetic placement within the GTDB. For each of our selected MAGs, we picked the most closely related genomes from the GTDB, as well as genomes representative of distinct families. We then reconstructed a phylogeny with our selected MAGs and the selected genomes from GTDB using concatenation of 120 conserved genes and FastTree[48]

Metagenome and metatranscriptome fragment recruitment

In order to evaluate the abundance and overall expression of our selected MAGs across the samples, we mapped the reads of the metagenomes and metatranscriptomes to our selected MAGs using bbmap (v.35) and a minimum sequence identity of 98%. We then calculated final RPKG values (reads per MAG kilo base pairs per metagenome giga base pairs), by dividing the total number of reads mapped to each MAG, by the size of the MAG (kbp) and the size of the metagenome/metatranscriptome (Gbp).

In order to evaluate the distribution of selected Alphaprotebacteria MAGs and their most closely related reference genomes across oceans, we performed fragment recruitment as in Kraemer *et al.*[58]. The 16 MAGs representative of the genomospecies enriched with the capacity to degrade aromatic lignin moieties, as well as 12 closely related reference genomes were searched against the metagenomic dataset using blastall (v.2.2.25) (e-value=0.00001). The recruited reads were extracted from the metagenomes and searched against a database consisting of the concatenation of all 28 genomes (16 Arctic Alphaproteobacteria MAGs and 12 reference genomes) using blastall (v.2.2.25). We selected the best hit, filtered for a minimum of 100 bp alignment and 98% sequence identity. We then calculated the RPKG values by normalizing the number of reads recruited by kilobase of genome and gigabase of metagenome.

Annotation of publicly available genomes

Gene sequences were retrieved from publicly available genomes at NCBI (Table S4) and translated to proteins. Ribosomal RNA genes were predicted in Infernal v. 1.1.2[77] against Rfam v. 14.2[78]. Gene functions were annotated in KofamScan using default settings and a bitscore-to-threshold ratio of 0.7[79].

2.7 Acknowledgments

The data were collected aboard the CCGS Louis S. St-Laurent in collaboration with researchers from Fisheries and Oceans Canada at the Institute of Ocean Sciences and Woods Hole Oceanographic Institution's Beaufort Gyre Exploration Program and are available at http://www.whoi.edu/beaufortgyre. We would like to thank both the Captain and crew of the CCGS Louis S. St-Laurent and the scientific teams aboard.

Bibliography

[1] S. Dou, J. Shan, X. Song, R. Cao, M. Wu, C. Li, et al. *Are humic substances soil microbial residues or unique synthesized compounds? A perspective on their distinctiveness*. **Extremophiles**, 30:159–165, 2020.

[2] M. Kida, N. Fujitake, V. Suchewaboripont, S. Poungparn, M. Tomotsune, et al. *Are humic substances soil microbial residues or unique synthesized compounds? A perspective on their distinctiveness*. **Aquat. Sci.**, 80(3):1–11, 2018.

[3] H.J. Park, N. Chae, W.J. Sul, B.Y. Lee, Y.K Lee, et al. *Temporal Changes in Soil Bacterial Diversity and Humic Substances Degradation in Subarctic Tundra Soil*. **Microb. Ecol.**, 80(3):668–75, 2015.

[4] V. Kisand, S. Gebhart, J. Rullköter, and M. Simon. *Significant bacterial transformation of riverine humic matter detected by pyrolysis GC-MS in serial chemostat experiments*. **Mar. Chem.**, 149:23–31, 2013.

[5] E.C. Esham, W. Ye, and M.A. Moran. *Identification and characterization of humic substances-degrading bacterial isolates from an estuarine environment*. **FEMS Microbiol. Ecol.**, 34(2):103–11, 2000.

[6] S. Opsahl and R. Benner. *Distribution and cycling of terrigenous dissolved organic matter in the ocean*. **Nature**, 386:480–2, 1997.

[7] V. Kisand, D. Rocker, and M. Simon. *Significant decomposition of riverine humic-rich DOC by marine but not estuarine bacteria assessed in sequential chemostat experiments*. **Aquat Microb Ecol.**, 53(2): 151–60, 2008.

[8] D. Rocker, V. Kisand, B. Scholz-Böttcher, T. Kneib, A. Lemke, et al. *Differential decomposition of humic acids by marine and estuarine bacterial communities at varying salinities*. **Biogeochemistry**, 111(1-3):331–46, 2012.

[9] Y. Shen, R. Benner, L.L. Robins, and J.G Wyn. *Sources, distributions, and dynamics of dissolved organic matter in the Canada and Makarov Basins*. **Front. Mar. Sci.**, 3:198, 2016.

[10] R.M. Holmes, J.W. McClelland, B.J. Peterson, S.E. Tank, E. Bulygina, et al. *Seasonal and Annual Fluxes of Nutrients and Organic Matter from Large Rivers to the Arctic Ocean and Surrounding Seas*. **Estuaries and Coasts**, 35(2):369–82, 2012.

[11] P.A. Raymond, J.W. McClelland, R.M. Holmes, A.V. Zhulidov, K. Mull, et al. *Flux and age of dissolved organic carbon exported to the Arctic Ocean: A carbon isotopic study of the five largest arctic rivers*. **Global Biogeochem. Cycles**, 21(4):1–9, 2007.

[12] R.W. Macdonald, Z.A. Kuzyk, and S.C. Johannessen. *It is not just about the ice: a geochemical perspective on the changing Arctic Ocean*. **J. Environ. Stud. Sci.**, 5(3):288–301, 2015.

[13] C.M. Duarte, T.M. Lenton, P. Wadhams, and P. Wassmann. *Abrupt climate change in the Arctic*. **Nat. Clim. Chang.**, 2(2):60–2, 2012.

[14] K.E. Frey and J.W. McClelland. *Impacts of permafrost degradation on arctic river biogeochemistry*. **Hydrol. Process.**, 23:169–82, 2009.

[15] P.J Mann, R.G.M. Spencer, P.J. Hernes, J. Six, G.R. Aiken, et al. *Pan-Arctic Trends in Terrestrial Dissolved Organic Matter from Optical Measurements*. **Front. Earth Sci.**, 4:1–18, 2016.

[16] L.G. Anderson and R.W. Macdonald. *Observing the Arctic Ocean carbon cycle in a changing environment*. **Polar Res.**, 34(1):26891, 2015.

[17] C.T. Connolly, M.B. Cardenas, G.A. Burkart, R.G.M. Spencer, and J.W. McClelland. *Groundwater*

as a major source of dissolved organic matter to Arctic coastal waters. **Nat. Commun.**, 11(1):1–8, 2020.

[18] R. Gonçalves-Araujo, M.A. Granskog, A. Bracher, K. Azetsu-Scott, P.A. Dodd, et al. *Using fluorescent dissolved organic matter to trace and distinguish the origin of Arctic surface waters*. **Sci. Rep.**, 6: 1–12, 2016.

[19] V Polyakov, K. Orlova, and E. Abakumov. *Soils of the lena river delta, yakutia, russia: Diversity, characteristics and humic acids molecular composition*. **Polarforschung**, 88(2):135–50, 2018.

[20] A. Muscolo, M. Sidari, and S. Nardi. *Humic substance: Relationship between structure and activity. deeper information suggests univocal findings*. **J. Geochemical Explor.**, 129:57–63, 2012.

[21] A.A. Chupakova, A.V. Chupakov, N.V. Neverova, and O.S. Shirokova, L.S. Pokrovsky. *Photodegradation of river dissolved organic matter and trace metals in the largest European Arctic estuary*. **Sci. Total Environ.**, 622-623:1343–52, 2018.

[22] J. Jung, J.E. Son, Y.K Lee, K.H Cho, Y. Lee, et al. *Tracing riverine dissolved organic carbon and its transport to the halocline layer in the Chukchi Sea (western Arctic Ocean) using humic-like fluorescence fingerprinting*. **Sci. Total Environ.**, 772:145542, 2021.

[23] M. Chen, J.H. Kim, S.I. Nam, F. Niessen, W.L. Hong, et al. *Tracing riverine dissolved organic carbon and its transport to the halocline layer in the Chukchi Sea (western Arctic Ocean) using humic-like fluorescence fingerprinting*. **Sci. Rep.**, 6:39213, 2016.

[24] L.L. Belicka and H.R. Harvey. *The Sequestration of Terrestrial Organic Carbon in Arctic Ocean Sediments: A Comparison of Methods and Implications for Regional Carbon Budgets*. **Geochimica et Cosmochimica Acta**, 73(20):6231–48, 2009.

[25] C. DeFrancesco and C. Guéguen. *Tracing riverine dissolved organic carbon and its transport to the halocline layer in the Chukchi Sea (western Arctic Ocean) using humic-like fluorescence fingerprinting*. **J. Geophys. Res. Ocean**, 126(2):e2020JC016578, 2021.

[26] Z. Gao and C. Gueguen. *Distribution of thiol, humic substances and colored dissolved organic matter during the 2015 Canadian Arctic GEOTRACES cruises*. **Mar. Chem.**, 203:1–9, 2018.

[27] R. Sutton and G. Sposito. *Molecular structure in soil humic substances: The new view*. **Environ. Sci. Technol.**, 39(23):9009–15, 2005.

[28] J. Gerke. *Concepts and misconceptions of humic substances as the stable part of soil organic matter: A review*. **Agronomy**, 8(76), 2018.

[29] V.I. Esteves, M. Otero, and A.C. Duarte. *Comparative characterization of humic substances from the open ocean, estuarine water and fresh water*. **Org. Geochem.**, 40(9):942–50, 2009.

[30] T. Dittmar and G. Kattner. *The biogeochemistry of the river and shelf ecosystem of the Arctic Ocean: A review*. **Mar. Chem.**, 83(3-4):103–20, 2003.

[31] J.I. Hedges, G. Eglinton, P.G. Hatcher, et al. *The molecularly-uncharacterized component of nonliving organic matter in natural environments*. **Organic Geochemistry**, 31:945–58, 2000.

[32] M.W.I. Schmidt, M.S. Torn, S. Abiven, et al. *Persistence of soil organic matter as an ecosystem property*. **Nature**, 478(7367):49–56, 2011.

[33] K. Ekschmitt, E. Kandeler, C. Poll, et al. *Soil-carbon preservation through habitat constraints and biological limitations on decomposer activity*. **J Plant Nutrition Soil Sci**, 171(1):27–35, 2008.

[34] D. Kim, H.J. Park, S. Nam, S.C. Kim, and H. Lee. *Humic substances degradation by a microbial consortium enriched from subarctic tundra soil*. **Korean J. Microbiol.**, 55(4):367–76, 2019.

[35] N. Hertkorn, H. Claus, P. Schmitt-Kopplin, E.M. Perdue, and Z. Filip. *Utilization and transformation*

of aquatic humic substances by autochthonous microorganisms. **Environ. Sci. Technol.**, 36(20): 4334–45, 2002.

[36] D.P. Brink, K. Ravi, G. Liden, and M.F. Gorwa-Grauslund. *Mapping the Diversity of Microbial Lignin Catabolism: Experiences from the ELignin Database*. **Applied Microbiology and Biotechnology**, 103(10):6231–48, 2019.

[37] G. Janusz, A. Pawlik, J. Sulej, et al. *Lignin degradation: Microorganisms, enzymes involved, genomes analysis and evolution*. **FEMS Microbiology Reviews**, 41(6):941–62, 2017.

[38] G.J. Herndl, H. Agogué, F. Baltar, et al. *Regulation of aquatic microbial processes: The 'microbial loop' of the sunlit surface waters and the dark ocean dissected*. **Aquatic Microbial Ecology**, 53(1): 59–68, 2008.

[39] A.M. Comeau, W.F. Vincent, and C. Lovejoy. *Novel chytrid lineages dominate fungal sequences in diverse marine and freshwater habitats*. **Sci Rep**, 6:1–6, 2016.

[40] A.S. Gladfelter, T.Y. James, and A.S. Amend. *Marine fungi*. **Curr Biol**, 29:R191–R195, 2019.

[41] A.B. Bochandsky, M.A. Clouse, and G.J. Herndl. *Novel chytrid lineages dominate fungal sequences in diverse marine and freshwater habitats*. **Sci Rep**, 6:1–6, 2016.

[42] D. Ilicic and H.P. Grossart. *Basal Parasitic Fungi in Marine Food Webs—A Mystery Yet to Unravel*. **J Fungi**, 8(114), 2022.

[43] D. Kim, H.J. Park, W.J. Sul, and H. Park. *Transcriptome analysis of Pseudomonas sp. from subarctic tundra soil: pathway description and gene discovery for humic acids degradation*. **Folia Microbiol.**, 63(3):315–23, 2018.

[44] D. Colatriano, P.Q Tran, C. Gueguen, W.J. Williams, C. Lovejoy, et al. *Genomic evidence for the degradation of terrestrial organic matter by pelagic Arctic Ocean Chloroflexi bacteria*. **Commun. Biol.**, 1(1)):90, 2018.

[45] K.R. Murphy, C.A. Stedmon, T.D. Waitze, and G.M. Ruiz. *Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy*. **Mar. Chem.**, 108(1-2):40–58, 2008.

[46] C. Guéguen and P. Kowalczuk. *Colored Dissolved Organic Matter in Frontal Zones. In: Chemical Oceanography of Coastal Zones*. Springer-Verlag Berlin Heidelberg, 2013.

[47] H.S. Seung and D.D. Lee. *Learning the parts of objects by non-negative matrix factorization*. **Nature**, 401(6755):788–91, 1999.

[48] P.A. Chaumeil, A.J. Mussig, P. Hugenholtz, and D.H. Parks. *GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database*. **Bioinformatics**, 36(6):1925–7, 2020.

[49] R.M.W. Amon, A.J. Rinehart, S. Duan, P. Louchouarn, A. Prokushkin, et al. *Dissolved Organic Matter Sources in Large Arctic Rivers*. **Geochimica and Cosmochimica**, 94:217–237, 2012.

[50] M.A. Goñi, M.B. Yunker, R.W. MacDonald, and T.I. Eglinton. *Distribution and sources of organic biomarkers in arctic sediments from the Mackenzie River and Beaufort Shelf*. **Mar. Chem.**, 71(1-2): 23–51, 2000.

[51] M.A. Goñi, A.E. O'Connor, Z.Z. Kuzyk, M.B. Yunker, C. Gobeil, et al. *Distribution and sources of organic matter in surface marine sediments across the North American Arctic margin*. **J. Geophys. Res. Ocean.**, 118(9):4017–35, 2013.

[52] R.E. Sipler, C.T.E. Kellogg, T.L. Connelly, Q.N. Roberts, P.L. Yager, et al. *Microbial community response to terrestrially derived dissolved organic matter in the coastal Arctic*. **Front. Microbiol.**, 8: 1–19, 2017.

[53] M. Lund Paulsen, O. Müller, A. Larsen, et al. *Biological transformation of Arctic dissolved organic matter in a NE Greenland fjord*. **Limnol. Oceanogr.**, 64(3):1014–33, 2019.

[54] J. Davis and R. Benner. *Seasonal trends in the abundance, composition and bioavailability of particulate and dissolved organic matter in the Chukchi/Beaufort Seas and western Canada Basin*. **Deep Res. Part II Top Stud. Oceanogr.**, 52(24-26):3396–410, 2005.

[55] J.É Tremblay, K. Simpson, J. Martin, et al. *Vertical stability and the annual dynamics of nutrients and chlorophyll fluorescence in the coastal, southeast Beaufort Sea*. **J. Geophys. Res. Ocean**, 113(7): 1–14, 2008.

[56] P. Lu, W. Wang, G. Zhang, et al. *Isolation and characterization marine bacteria capable of degrading lignin-derived compounds*. **PLoS One**, 15:1–19, 2020.

[57] A. Ramachandran, S. McLatchie, and D.A. Walsh. *A Novel Freshwater to Marine Evolutionary Transition Revealed within Methylophilaceae Bacteria from the Arctic Ocean*. **MBio**, 12(3):e0130621, 2021.

[58] S. Kraemer, A. Ramachandran, D. Colatriano, et al. *Diversity and biogeography of SAR11 bacteria from the Arctic Ocean*. **ISME J**, 14(1):79–90, 2020.

[59] D. Colatriano and D.A. Walsh. *An aquatic microbial metaproteomics workflow: From cells to tryptic peptides suitable for tandem mass spectrometry-based analysis*. **J. Vis. Exp.**, 103:e52827, 2015.

[60] K. Murphy, C.A. Stedmon, D. Graeber, et al. *Fluorescence spectroscopy and multi-way techniques. PARAFAC*. **Anal. Methods**, 5(23):6557–66, 2013.

[61] A. Bankevich, S. Nurk, D. Antipov, et al. *SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing*. **J. Comput. Biol.**, 19(5):455–77, 2012.

[62] M. Huntemann, N.N. Ivanova, K. Mavromatis, et al. *The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v . 4)*. **Stand. Genomic Sci**, 1:1–5, 2016.

[63] M.D. Robinson and A.A. Oshlack. *A scaling normalization method for differential expression analysis of RNA-seq data*. **Genome Biol.**, 11(3), 2010.

[64] M.D. Robinson, D.J. McCarthy, and G.K. Smyth. *edgeR: A Bioconductor package for differential expression analysis of digital gene expression data*. **Bioinformatics**, 26(1):139–40, 2009.

[65] J. Oksanen, K. Roeland, P. Legendre, et al. *The vegan package*. , pages 1–130, 2007.

[66] R. Gaujoux and C. Seoighe. *A flexible R package for nonnegative matrix factorization*. **BMC Bioinformatics**, 11, 2010.

[67] X. Jiang, M.G.I. Langille, R.Y. Neches, et al. *Functional Biogeography of Ocean Microbes Revealed through Non-Negative Matrix Factorization*. **PLoS One**, 7(9):1–9, 2012.

[68] H. Kim and H. Park. *Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis*. **Bioinformatics**, 23(12):1495–502, 2007.

[69] T.D.H. Bugg, M. Ahmad, E.M. Hardiman, and other. *Pathways for degradation of lignin in bacteria and fungi*. **Nat. Prod. Rep.**, 28(12):1883–96, 2011.

[70] N. Kamimura, K. Takahashi, K. Mori, et al. *Bacterial catabolism of lignin-derived aromatics: New findings in a recent decade: Update on bacterial lignin catabolism*. **Environ. Microbiol. Rep.**, 9(6): 679–705, 2017.

[71] W. Li and A. Godzik. *Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences*. **Bioinformatics**, 22(13):1658–9, 2006.

[72] C. Bagci, S. Patz, and D.H. Huson. *DIAMOND+MEGAN: Fast and Easy Taxonomic and Functional*

Analysis of Short and Long Microbiome Sequences. **Current Protocols**, 1(3):1–29, 2021.

[73] D.D. Kang, F. Li, E. Kirton, et al. *MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies*. **PeerJ.**, 7:e7359, 2019.

[74] M. Imelfort, C.T. Skennerton, D.H. Parks, et al. *CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes*. **Genome Res.**, 25:1043–55, 2015.

[75] D.H. Parks, M. Chuvochina, D.W. Waite, et al. *A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life*. **Nat. Biotechnol.**, 36(10):996, 2018.

[76] C. Jain, L.M. Rodriguez, A.M. Phillippy, et al. *High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries*. **Nat. Commun.**, 9(1):1–8, 2018.

[77] E.P. Nawrocki and S.R. Eddy. *Infernal 1.1: 100-fold faster RNA homology searches*. **Bioinformatics**, 29(22):2933–5, 2013.

[78] I. Kavalri, E.P. Nawrocki, N. Ontiveros-Palacios, et al. *Rfam 14: Expanded coverage of metagenomic, viral and microRNA families*. **Nucleic Acids Res.**, 49:192–200, 2021.

[79] T. Aramaki, R. Blanc-Mathieu, H. Endo, et al. *KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold*. **Bioinformatics**, 36(7):2251–2, 2020.

phosphorus

Chapter 3

Is the Arctic Ocean fat? Meta-omics reveal a new possible important role for neutral lipid metabolism in the Arctic Ocean microbiomes

3.1 Abstract

Background: The ability to store neutral lipids is common across the tree of life, enabling the accumulation of carbon reserves for periods of resource scarcity. Microbial lipids, of which neutral lipids may constitute a significant fraction, are of major importance to sustain Arctic Ocean higher trophic levels. The seasonality of the Arctic Ocean represents an extreme case of alternating periods of abundance and scarcity of energy and nutrients. During the spring and summer, long days of light coupled with increased riverine input bring abundant resources for the growth of phytoplankton primary producers and microbial heterotrophs. In contrast, during winter, primary productivity is at its lowest due to thick ice cover blocking the already limited amount of light from short days, subsequently depriving the microbial heterotrophs from their main resource. The microbes of the Arctic Ocean may therefore use the storage of neutral lipids as a strategy to survive winter conditions. However, the capacity of the Arctic Ocean microbiomes to synthesize and use neutral lipids, and how the neutral lipid metabolism may affect the carbon cycle in a changing Arctic Ocean is relatively unknown. In this study, we used a combination of metagenomics and metatranscriptomics to determine the prevalence, biogeography and phylogenetic diversity of neutral lipids metabolic pathways in the microbiomes of the Canada Basin in the Arctic Ocean.

Results: A metagenomic and metatranscriptomic community-scale analysis revealed that the biosynthesis of the neutral lipids triacylglycerols (TAGs) and polyhydroxyalkanoates (PHAs) were both common across the global ocean prokaryotes and picoeukaryotes, but the TAG biosynthetic pathway was more abundant in the Canada Basin. Taxonomic assignment of marker genes attributed most of the TAG biosynthesis capacity to picoeukaryotic phytoplankton while PHA biosynthesis capacity was restricted to prokaryotes. Genome-resolved analyses showed that PHA and TAG biosynthesis capacity was phylogenetically separated among bacterial taxa, unraveling an unexpected phylogenetic diversity of TAG biosynthetic marine bacterial taxa. Taxonomic assignment of marker genes, coupled to genome reconstruction showed that the

genomic capacity to degrade TAGs was more prevalent than the TAG biosynthesis capacity in bacterial populations. We further showed that the TAG and PHA biosynthesis capacity co-occurred with the potential capacity to process different carbon sources: carbohydrates and aromatic compounds respectively.

Conclusion: Overall, this study highlights a potential important role for the metabolism of neutral lipids in the microbiomes and carbon cycling of the Arctic Ocean. The greater picoeukaryotic-based genomic capacity to synthesize NLs in the Arctic Ocean combined with the prevalence of NL degradation pathways in the prokaryotic community supports the hypothesis that eukaryotic NL are used as an important growth resource by the heterotrophic prokaryotic community. The discovery of an unexpected phylogenetic diversity of NL producers in the bacterial communities lead us to postulate that NLs from bacterial origin may play an important role to support the Arctic Ocean food web during the polar night.

3.2 Introduction

Storage is the accumulation of chemical resources by an organism for future use and is an important process in the life strategy of organisms subjected to variable conditions and nutrient supplies[1]. Among the classes of compounds used to store organic carbon (C), lipids have the highest energy density (up to 9.3 kcal/g) compared to proteins and carbohydrates (4.1 kcal/g)[2]. Lipid storage is therefore a common trait across the domains of life[3–5].

Neutral lipids (NLs) are the common storage form of lipids, and can be differentiated from the charged lipids that comprise cellular membranes[6]. There are three NLs produced by eukaryotes: triacylglycerols (TAGs), steryl esters (SEs), and wax esters (WEs). All are synthesized through the esterification of a fatty acid (FA) with a second moiety. Diacylglycerol is used to form TAGs, a sterol to form SEs, or a second FA to form WEs[4]. In the ocean, TAG biosynthesis is widespread and well-studied in eukaryotic phytoplankton[7] while SE and WE biosynthesis are restricted to fewer eukaryotic phytoplankton species[8]. TAGs, SEs and WEs can all accumulate in lipid droplets within the cell and act as storage for FAs. These FAs can later be used as component for the cell membrane or to generate acetyl-CoA through beta-oxidation[4].

Although TAG biosynthesis, and to a lesser extent WE and SE[5,9,10] biosynthesis, has been observed in bacteria, polyhdroxyalkanaoates (PHAs) are a more widespread bacterial energy storage compound[5]. Short chain PHA such as polyhydroxybutanoate (PHB) are produced from the condensation of two acetyl-CoA molecules followed by their polymerization[11]. Longer chain PHA are polymerized from modified FAs diverted from beta-oxidation. In the ocean, PHA-producing bacteria have been described and isolated from oil-degrading communities[12]. Although studies have expanded the diversity of known marine bacteria able to store PHA[13], the phylogenetic extent and distribution of the PHA-synthesizing bacterial taxa in the ocean

is still unknown.

Conceptually, carbon storage can be distinguished into two different modes termed surplus storage and reserve storage[3]. Surplus storage is supply-dependent where the cell stores resources that are in excess of the current metabolic requirement for cell growth[3]. Surplus storage is therefore not in competition for resources with other metabolic processes. As the synthesis of FAs incorporated in NLs requires energy input, the abundance of an energy source (in the form of light or labile C) is required for NL surplus storage[14]. In addition, the C pool must be in excess of what is needed for other metabolic processes. This condition is optimal for microorganisms when other nutrients (such as nitrogen - N - and phosphate - P) are limited as funneling of C to protein synthesis to sustain growth is favoured over NL storage when both C and N are abundant[14]. The accumulation of TAGs and PHAs in microorganisms is therefore favored under labile C excess and nutrient (N, P) limitation[14–16]. However, microbial cells can considerably reduce the energy demand for NL storage by using exogenous FAs that are incorporated in NLs[15] instead of FA synthesis. In addition, TAGs can accumulate in microbial cells when both C and N sources are sufficient for cell growth[17]. In contrast to surplus storage, reserve storage occurs under conditions when carbon resources are limited, diverting resources from other metabolic processes[18]. Reserve storage may prove advantageous, when using the resource in the future provides more value than using it immediately. For example using reserve storage when resources are declining may provide resistance against imminent starvation[19]. Evidence of NL reserve storage includes *Pseudomonas putida* accumulating PHA up to 26% of dry cell mass[20] and *Rhodococcus opacus* accumulating TAG up to 21% of its dry cell mass[21] both in C-limited conditions.

Given the important role of carbon storage in dynamic energy environments, the cycling of NLs may be important in sustaining Arctic Ocean microbial communities, which are faced with intense seasonal changes in energy and nutrient availability. In winter, the Arctic Ocean receives little light due to short days and a thick ice-layer. From spring to fall, a large part of the ocean is ice-free and exposed to long light days. The nutrients released by ice melt and river discharge during spring, combined with longer light days, trigger phytoplankton growth at the surface mixed layer[22]. Phytoplankton deplete the water of nutrients during the late spring and summer[23]. For phytoplankton, spring is a period of inorganic carbon (CO_2), energy (light) and nutrient (N, P) abundance. In summer and fall, C and energy are still abundant, but nutrients are limited, while in winter, both energy and nutrients are limited. Accumulation of energy reserves may therefore be used to survive the Arctic winter, as it has been shown for some Arctic phytoplankton taxa[24].

In addition, the Arctic Ocean phytoplankton communities are dominated by eukaryotic algae, with a general absence of *Cyanobacteria*[25]. Under nutrient limitation, eukaryotic phytoplankton store energy in the form of neutral lipids and carbohydrates[7]. The low nutrients and dominance of eukaryotic phytoplankton may favour a higher fraction of the primary productivity allocated to lipids rather than other macromolecules in some regions of the Arctic Ocean compared to other oceans[26,27]. Considering that phytoplankton lipids contribute signif-

icantly to the transfer of energy to the Arctic Ocean food web[28] and that NLs can account for up to 80% of the lipid pool of the Arctic Ocean phytoplankton[17,29], phytoplankton NLs are of major importance for the Arctic Ocean food web. However, the extent, biogeography and phylogenetic distribution of microorganisms involved in NL biosynthesis is still poorly understood in the Arctic Ocean.

Similar to phytoplankton, the strong Arctic seasonality may favor the accumulation of NLs in heterotrophic prokaryotes. The spring phytoplankton blooms, combined with riverine discharge, constitute a pulse of organic matter for Arctic Ocean heterotrophs. These prokaryotic heterotrophs must then live on diminishing primary produced OM as summer and fall progress[30] and finally on the recycling of the leftover stock of primary-produced OM and decaying living matter during the polar night[31]. The metabolic activity of ocean microbial communities during the Arctic polar night shows that prokaryotes use energy[32]. It has been shown that storage of energy reserve helps prokaryotes to survive winter in other systems such as soils[33]. However, the knowledge on NL storage in ocean prokaryotes is mostly limited to oil-degrading consortia[34], raising important questions, such as *(i) what is the global biogeography of NL metabolism pathways in the global ocean microbiomes and how does the Arctic Ocean microbiomes compare to the rest of world oceans? (ii) what is the abundance, distribution and phylogenetic diversity of microbial taxa capable of synthesizing and degrading NLs in the Arctic Ocean? (iii) What sources of carbon may feed the NL synthesis in the Arctic Ocean microbiomes?*

In this work, we generate coupled metagenomic/metatranscriptomic profiles across the highly stratified waters of the Canada Basin, Arctic Ocean including the sunlit surface water, subsurface layer (constituted by the chlorphyll-a maximum and maximum of fluorescent dissolved organic matter) and deep ocean. We show that the metabolic pathways involved in NL biosynthesis and degradation are more common in the Arctic Ocean microbiomes compared to the rest of the world ocean microbiomes. The biosynthesis of TAGs is dominated by phototrophic picoeukaryotes, while PHA biosynthesis is restricted to prokaryotes. Our results demonstrate that a component of the prokaryotic community is able to use exogenous TAGs as growth substrates. In addition, the reconstruction of genomes from metagenomes unraveled an unexpected diversity of bacterial taxa found mostly in deep waters able to synthesize TAGs. The taxonomy of TAG synthesis taxa contrasted with the taxonomy of bacterial taxa able to synthesize PHAs, which were predominantly found in the subsurface layer. The higher number of genes encoding CAzymes in genomes from the deep water suggests that these genomes use recycled OM to feed NL biosynthesis. In contrast, the higher number of aromatic compound degradation genes in genomes found in the subsurface layers lead us to propose that these genomes used OM of terrestrial origin to feed their NL synthesis. Our findings shed a new light on the importance of NL cycle in the Arctic Ocean microbiomes.

3.3 Results

3.3.1 Physicochemical structure of the Canada Basin water column

The study focused on the microbiomes of the Canada Basin sampled along a latitudinal transect from 73^{0}N to 81^{0}N. We collected samples during the summer and fall (September) through the water column of the salinity-stratified Canada Basin, targeting four distinct water column features (Figure 1). The surface (5 m and 20 m depth) and subsurface chlorophyll maximum (SCM; 55-95 m) features within the photic zone. The surface samples were characterized by lower salinity (25-29 PSU), and high phytoplankton cell abundance while the SCM had high level of fluorescent dissolved organic matter (FDOM) and the highest bacterial cell abundance (Figure 3.1, 3.2). Below the photic zone, we retrieved samples from Pacific-origin water, at salinities of 32.3 and 33.1 PSU. These waters were characterized by the highest levels of fluorescent dissolved organic matter (FDOM), representing the FDOMmax feature (Figure 3.1, 3.2). The water below 400 m depth is of Atlantic-origin. In this water, referred as deep water in this study, we collected samples at the temperature maximum (Tmax), at 1000 m depth and at the bottom.

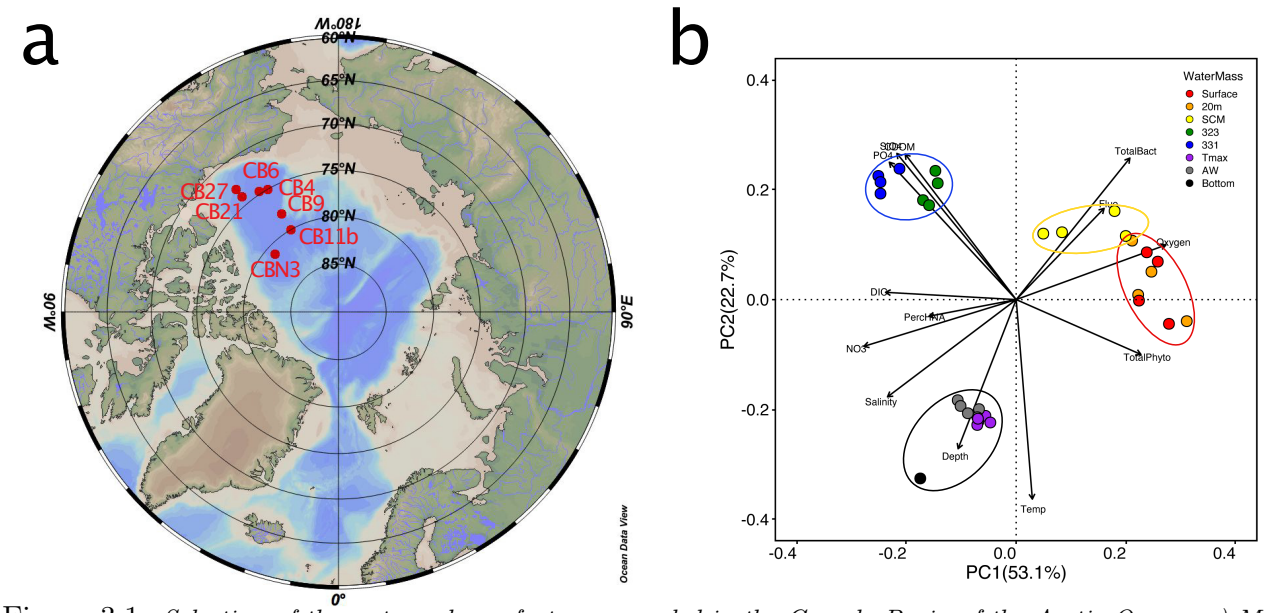

Figure 3.1: *Selection of the water column features sampled in the Canada Basin of the Arctic Ocean. a) Map of the station sampled in the Canada Basin. b) Principal component analysis of the samples based on their physicochemical parameters that show the differences among the four water column features: the surface (red circle), the SCM (yellow circle), the FDOMmax (blue circle) and the deep samples (black circles).*

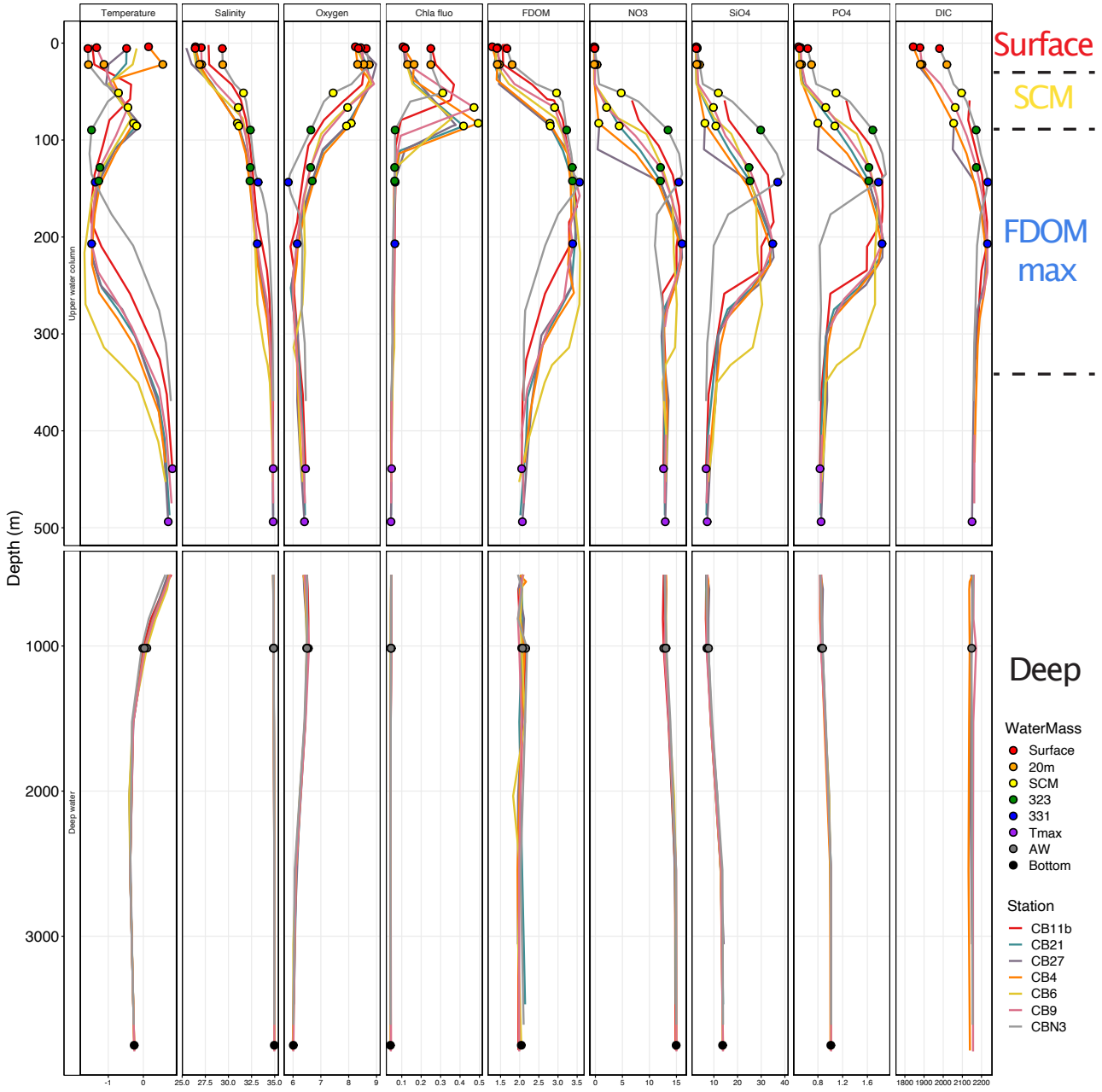

Figure 3.2: *Depth profiles of environmental variables for the 7 stations sampled in the Canada Basin for this study. Points on the profiles represent the location of the metagenomes sampled.*

3.3.2 Biosynthesis of neutral lipids

We investigated if the Canada Basin metagenomes and metatranscriptomes encoded NL biosynthetic pathways (Figure 3.3a). Pathways for the biosynthesis of TAGs and WEs were complete in the surface, SCM and FDOMmax metagenomes (Figure 3.4). In addition, complete pathways were identified in metatranscriptomes from the surface and the SCM, but not the FDOMmax (Figure 3.5). In contrast, for the deep metagenomes, we did not find the full

set of enzymes for TAG and WE biosynthesis, but we found evidence for the complete WE biosynthesis pathway in deep metatranscriptomes. The pathways for both polyhydroxyalkanoates (PHAs) polyhydroxybutanoate (PHB) and medium chain length polyhydroxyalkanoates (mclPHAs) biosynthesis were complete in metagenomes and metatranscriptomes. We did not find any evidence for the presence of the genes involved in the SE biosynthesis pathway in metagenomes or metatranscriptomes.

To explore how NL biosynthesis was distributed within the Canada Basin, we measured the abundance of genes coding for enzymes that catalyze key steps of NL biosynthesis (Figure 3.3a) and compared their abundance with other ocean metagenomes (Figure 3.6). We also measured the expression of these genes. The key step of TAG biosynthesis is the esterification of diacylglycerol (DAG) with a FA. This step can be achieved in two ways: by the acyl-CoA independent transfer of a FA to DAG catalyzed by the PDAT (phospholipid:diacylglycerol acyltransferase, EC:2.3.1.158) or by the esterification of an acyl-CoA with DAG catalyzed by DGAT (DAG O-acyltransferase, EC:2.3.1.20). The PDAT was encoded by one gene family (K00679) characterized only in eukaryotes[35]. We found 4 gene families (K00635, K11155, K11160, K18851) in our data, characterized in prokaryotes[36] or eukaryotes[37], predicted to encode enzymes with DGAT activity (Figure 3.3a). The key step of WE biosynthesis, the esterification of an acyl-CoA and a fatty alcohol (EC:2.3.1.75), is catalyzed by the enzymes encoded by three of these gene families (K00635, K11155, K11160). The surface was characterized by the highest abundance of PDAT in the metagenomes (0.05 copies/cell) while the expression was the highest in the SCM (up to 3 transcripts/gene copy) (Figure 3.3b). In contrast PDAT genes were barely detected in other oceans (Figure 3.3c). To a smaller extent, the DGAT gene abundance peaked in the SCM metagenomes (0.01 copies/cell) (Figure 3.3b) but was only expressed in the surface (up to 2 transcripts/gene copy). This contrasted with other ocean metagenomes, where DGAT genes were more abundant and mostly found in deep waters (0.08 copies/cell) (Figure 2c).

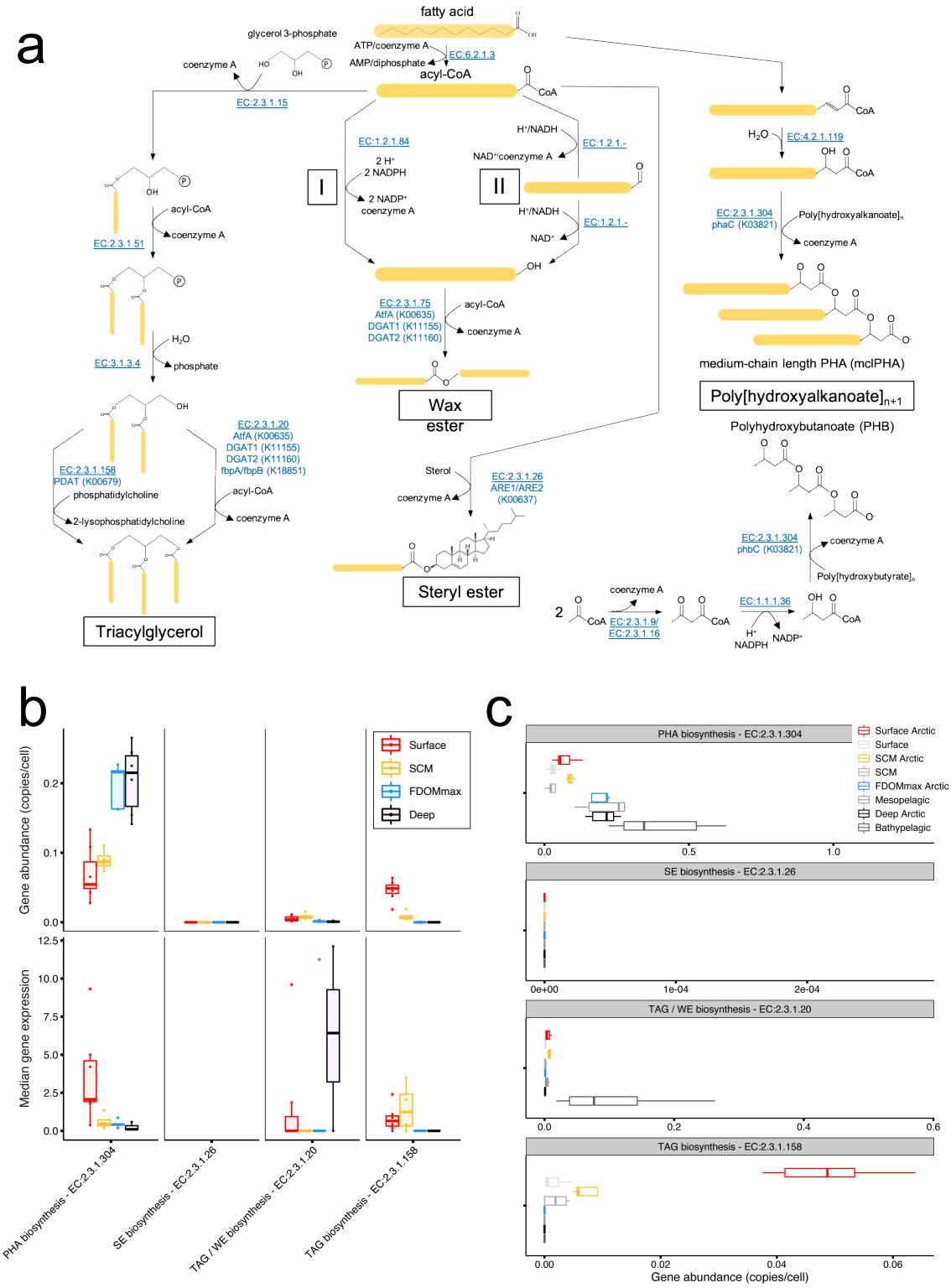

Figure 3.3: *Biosynthesis of neutral lipids in the Canada Basin and the global ocean microbiomes. a) Schematics of biosynthetic pathways used to produce the neutral lipids triacylglycerols, wax esters, steryl esters and polyhydroxyalkanoates. b) Abundance of genes coding for protein families that catalyze key steps of the neutral lipid biosynthetic pathways within metagenomes and metatranscriptomes. c) Fraction of the gene pool allocated to genes coding for protein families that catalyze key steps of the neutral lipid biosynthetic pathways comparing the water column of the Canada Basin (red, yellow, blue and black) and water column of the global ocean (shades of grey).*

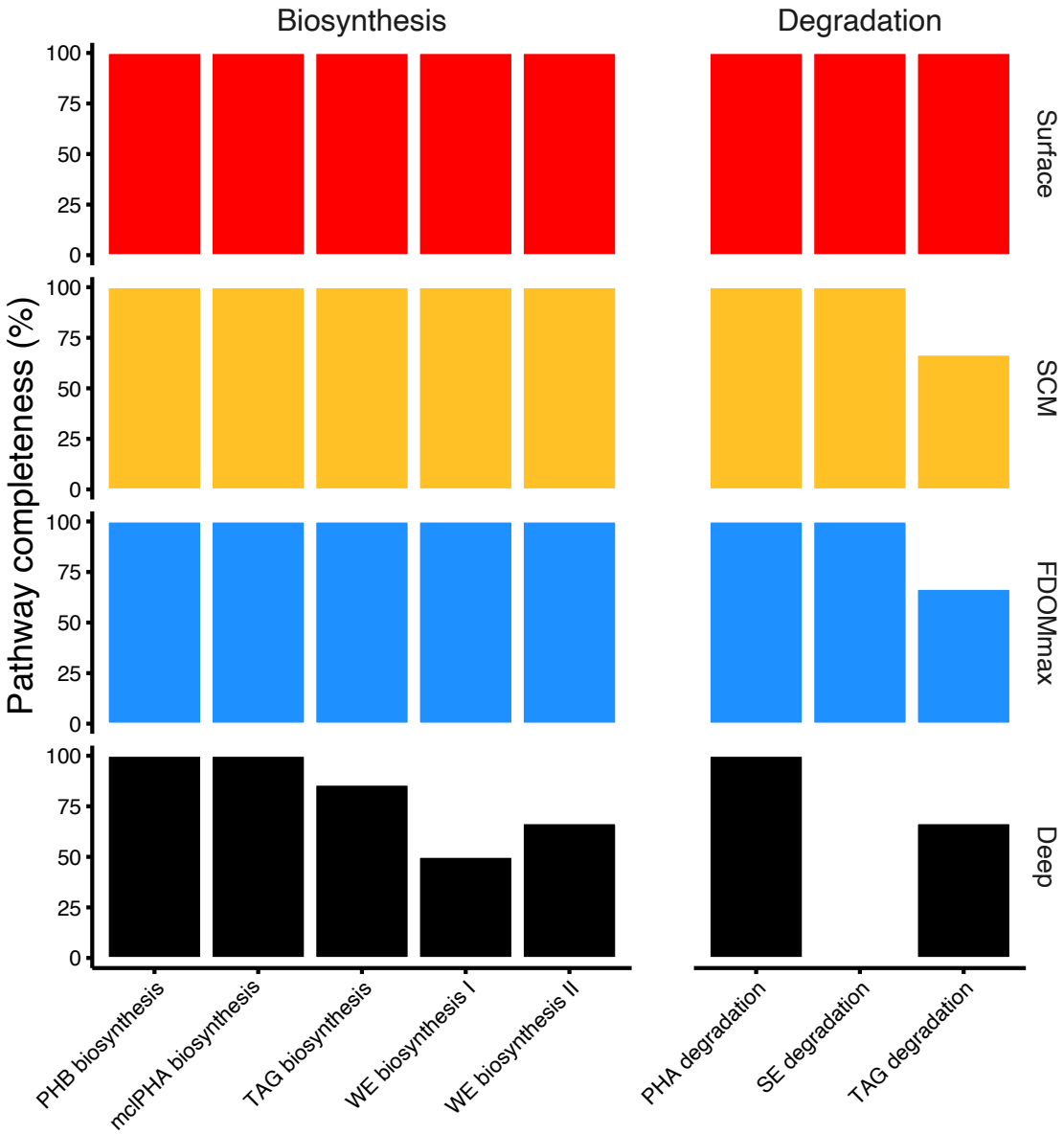

Figure 3.4: *Completeness of the neutral lipid biosynthesis and degradation pathways in each water feature in the metagenomes.*

The key step for PHA biosynthesis is the polymerization of an hydroxyalkanoate through esterification (EC:2.3.1.304). The PHA synthase that catalyzes the reaction is encoded by the *phbC* gene involved in PHB biosynthesis or by *phaC* gene involved in mclPHA biosynthesis. Both *phaC* and *phbC* genes belong to the same protein family (K03821). Within the Arctic Ocean, the PHA synthase genes were the most abundant of all NL biosynthesis marker genes, peaking in the aphotic zone metagenomes (0.21 copies/cell in the FDOMmax and the deep) similar to other oceans (Figure 3.3b,c). Interestingly, the abundance of the PHA synthase genes was higher in surface and SCM of the Arctic than other oceans and its expression was maximum in the surface (2.4 transcripts/gene copy). Altogether, within the global ocean microbiomes, these results show a greater allocation of the gene pool to TAG and WE biosynthesis particular to the microbiomes of the Arctic Ocean photic zone. They also show a genomic potential for PHA biosynthesis in the aphotic zone, similarly to the rest of the world oceans.

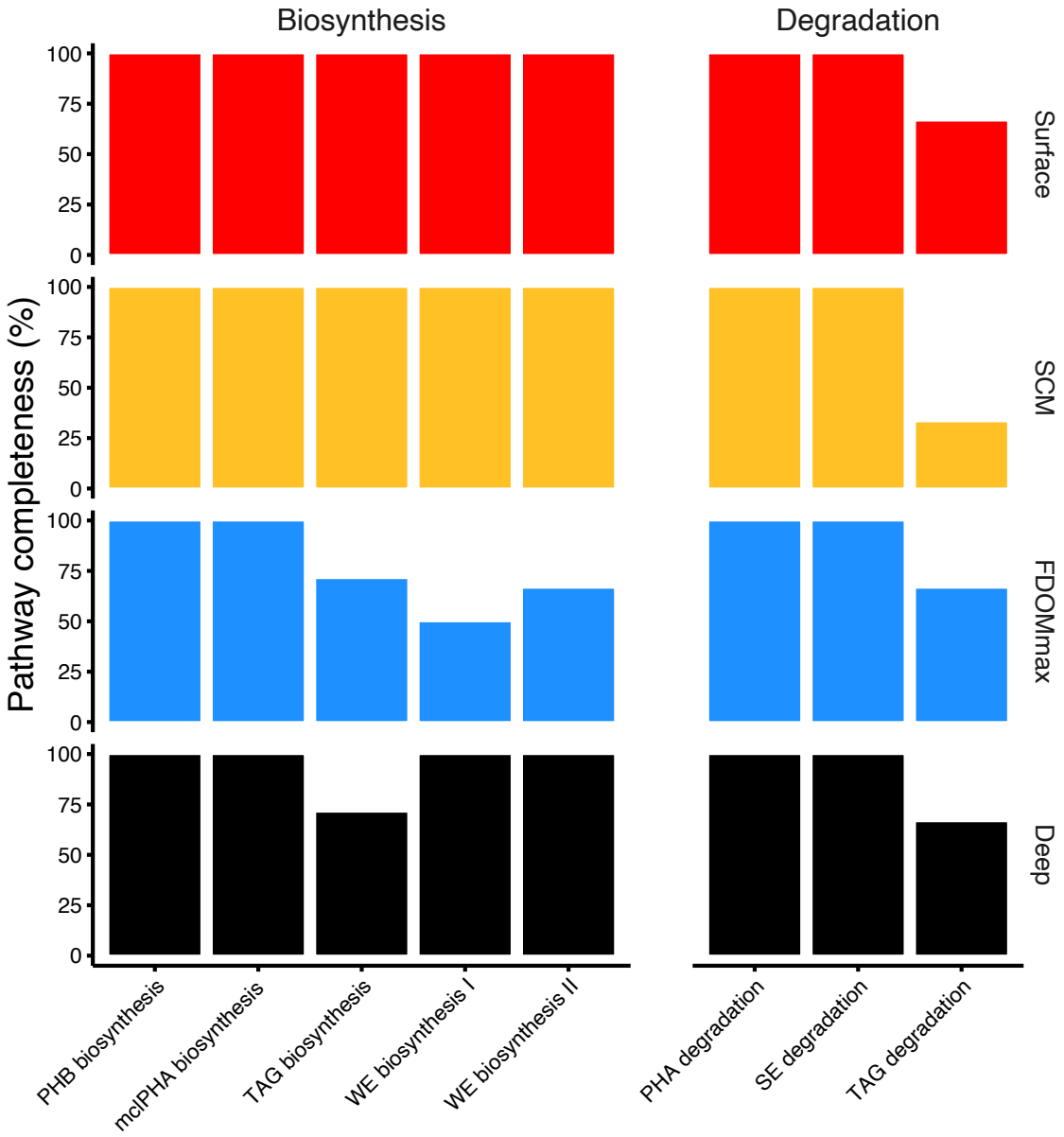

Figure 3.5: *Completeness of the neutral lipid biosynthesis and degradation pathways in each water feature in the metatranscriptomes.*

3.3.3 Degradation of neutral lipids

We sought to determine if the Arctic Ocean microbiomes possessed the suite of genes encoding NL degradation pathways (Figure 3.7a). The TAG degradation pathway was complete only in the surface samples (Figure 3.4) while we found genes for the PHA degradation pathway in all water column features for metagenomes and metatranscriptomes (Figure 3.4, 3.5). The PHA degradation pathway consisted of the depolymerization step to free a hydroxyalkanoate from the PHA. We found the *phaZ* genes (K05973) encoding the enzyme catalyzing the depolymerization of PHB (EC:3.1.1.75) but not *phaZ* genes (K22249, K22250) encoding the enzyme catalyzing the depolymerization for mclPHAs (EC:3.1.1.76) for both metagenomes and metatranscriptomes. Both WE and SE degradation pathways consisted of the hydrolysis of the ester bond to free a FA. We found genes coding for the SE esterase, catalyzing SE degradation

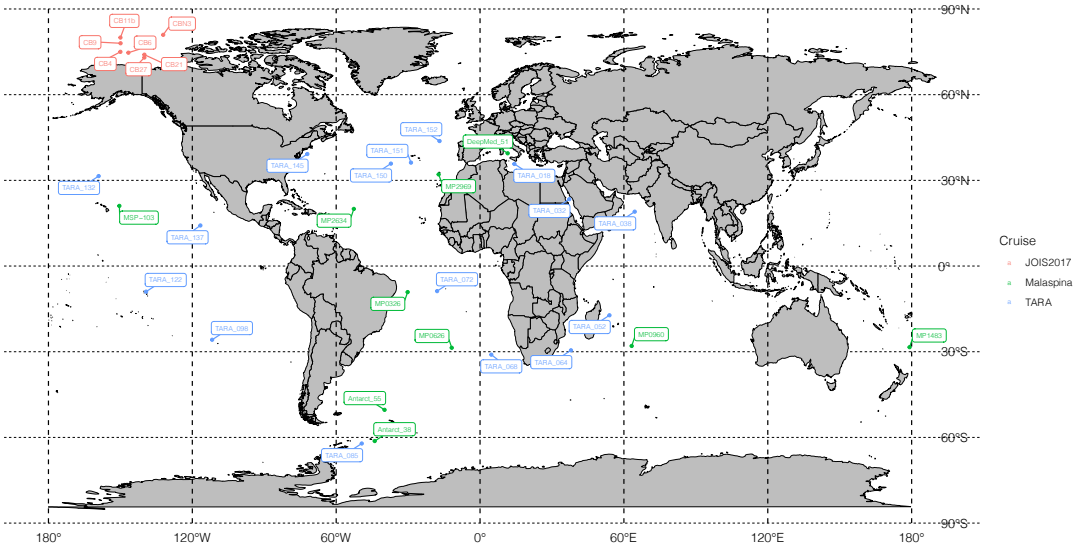

Figure 3.6: *Map of the Arctic and Global ocean metagenomes used in this study.*

(EC:3.1.1.13) in both metagenomes and metatranscriptomes (Figure 3.4, 3.5). However, we did not find genes coding for the WE hydrolase.

To explore if the genomic capacity to degrade NLs was differentially abundant in Arctic Ocean metagenomes compared to metagenomes from other regions of the global ocean, we determined the global biogeography of key genes for NL degradation pathways. The key step in TAG degradation is the hydrolysis of a FA from TAGs (EC:3.1.1.3) (Figure 3.7a). A few gene families (K01046, K14675, K14674) encode the TAG acylhydrolase enzyme catalyzing this reaction. The abundance of TAG acylhydrolase was slightly more elevated in the SCM and FDOMmax (0.2 copies/cell) within the Arctic Ocean (Figure 3.7b) and in the bathypelagic waters (0.3 copies/cell) within other regions of the global ocean (Figure 3.7c). The *phaZ* genes were most abundant in the surface of the Arctic Ocean (0.05 copies/cell) and the deep waters of other regions from the global ocean (0.08 copies/cell). One gene coding for the TAG hydrolase (K14674) also had the capacity to hydrolyze SEs (EC:3.1.1.13). In addition, we found two other gene families (K12298, K01052) coding for SE hydrolase. SE hydrolase genes were significantly more abundant in the microbiomes of the Arctic Ocean surface (up to 0.01 copies/cell) than in the microbiomes of any other water column features of other world oceans (<0.0025 copies/cell).

These results show that marker genes involved in NL degradation are more abundant in the surface water than those for NL synthesis. In addition, we demonstrate that the capacity to use TAGs and SEs as C sources is represented in a higher fraction of the metabolic genes pool in the Arctic Ocean microbiomes than in other oceans.

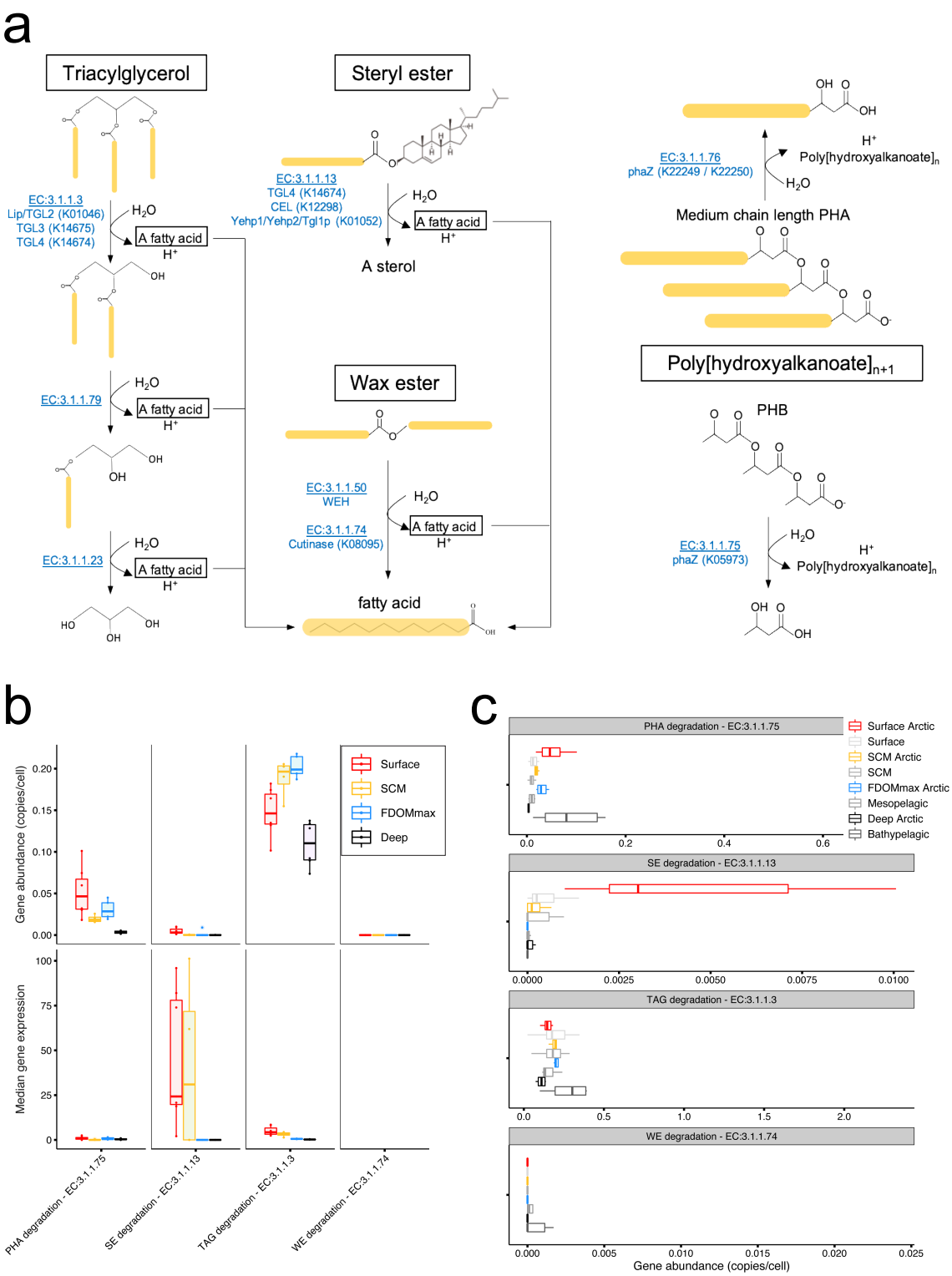

Figure 3.7: *Degradation of neutral lipids in the Canada Basin and the global ocean microbiomes. a) Schematics of pathways involved in the degradation of neutral lipids triacylglycerols, wax esters, steryl esters and polyhydroxyalkanoates. b) Abundance of genes coding for protein families that catalyze key steps of the neutral lipids degradation pathways within metagenomes and metatranscriptomes of the Canada Basin. c) Fraction of the gene pool allocated to genes coding for protein families that catalyze key steps of the neutral lipids degradation pathways comparing the water column of the Canada Basin (red, yellow, blue and black) and water column of the global ocean (shades of grey).*

3.3.4 Fatty acid import in Arctic Ocean microbiomes

Since our results showed that the Arctic Ocean microbiomes were enriched in NL degradation genes in comparison to NL biosynthesis genes, we went on to investigate if NLs released by producers could be imported and utilized by another fraction of the microbiome. A necessary step for the use of exogenous NLs is FA import. The transmembrane import system most specific to FAs and most studied is the *fadL/fadD/fadR* system (Figure 3.8a). The role of the FA import system is to direct exogenous FAs to the β-oxidation pathway. Import of long chain FAs is coded by the *fadL* gene (K06076). Import of FAs by the *fadL* transporter is coupled to its activation with CoA by the long chain FA acyl-CoA synthase (*fadD*, K01897)[38,39]. The *fadR* gene is then activated and represses FA synthesis while activating β-oxidation. Small and medium chain FAs can passively diffuse through the membrane and are activated with CoA by the FA acyl-CoA synthetase (*fadK*, K12507). In the Arctic Ocean, the *fadL*, *fadD* and *fadK* genes were generally more abundant in SCM and FDOMmax while the *fadR* gene was more abundant in the FDOMmax and deep (Figure 3.8b). We found similar abundances for the *fadD* gene in the Arctic FDOMmax (2.5 copies/cell) and the bathypelagic waters of other regions of the global ocean (2 copies/cell). The FA transporter gene *fadL* was significantly more abundant in the bathypelagic waters outside of the Arctic Ocean (0.5 copies/cell) than in any other waters of the global region. Although the abundance of the *fadL* gene was low in the Arctic (0.05 copies/cell in the surface and SCM), this gene was highly expressed in the Arctic Ocean photic zone (20 transcripts/gene copy in the surface and 35 transcripts/gene copy in the SCM).

As an alternative to β-oxidation, exogenous FAs can be incorporated into membrane phospholipids (PLs) (Figure 3.8a). The membrane-bound acyl-phospholipid O-acyltransferase gene (*aas*, K05939) participates to the incorporation of exogenous FAs to PLs in Escherichia coli[40]. Exogenous FAs can also be incorporated into PLs through phosphorylation by FA kinase (*fakA*/K07030, *fakB*/K25232). Within the global ocean, the abundance of the *aas* genes was enriched in the Arctic SCM (0.45 copies/cell) while *fakA* abundance was enriched in the Arctic FDOMmax (0.28 copies/gene) (Figure 4b,c).

Other membrane proteins, less specific to FAs, have been shown to facilitate the exogenous FA import such as the *mce1* complex (K02066), an ABC FA transporter (K24820) and a porin (*ompF*, K09476)[38] (Figure 3.8a). We only detected the presence of the *mce1* complex that was enriched in the bathypelagic zone outside of the Arctic Ocean (0.4 copies/gene) (Figure 3.8c). Overall, these results show that genes involved in the import of exogenous FAs are more abundant in the Arctic Ocean than the other world oceans.

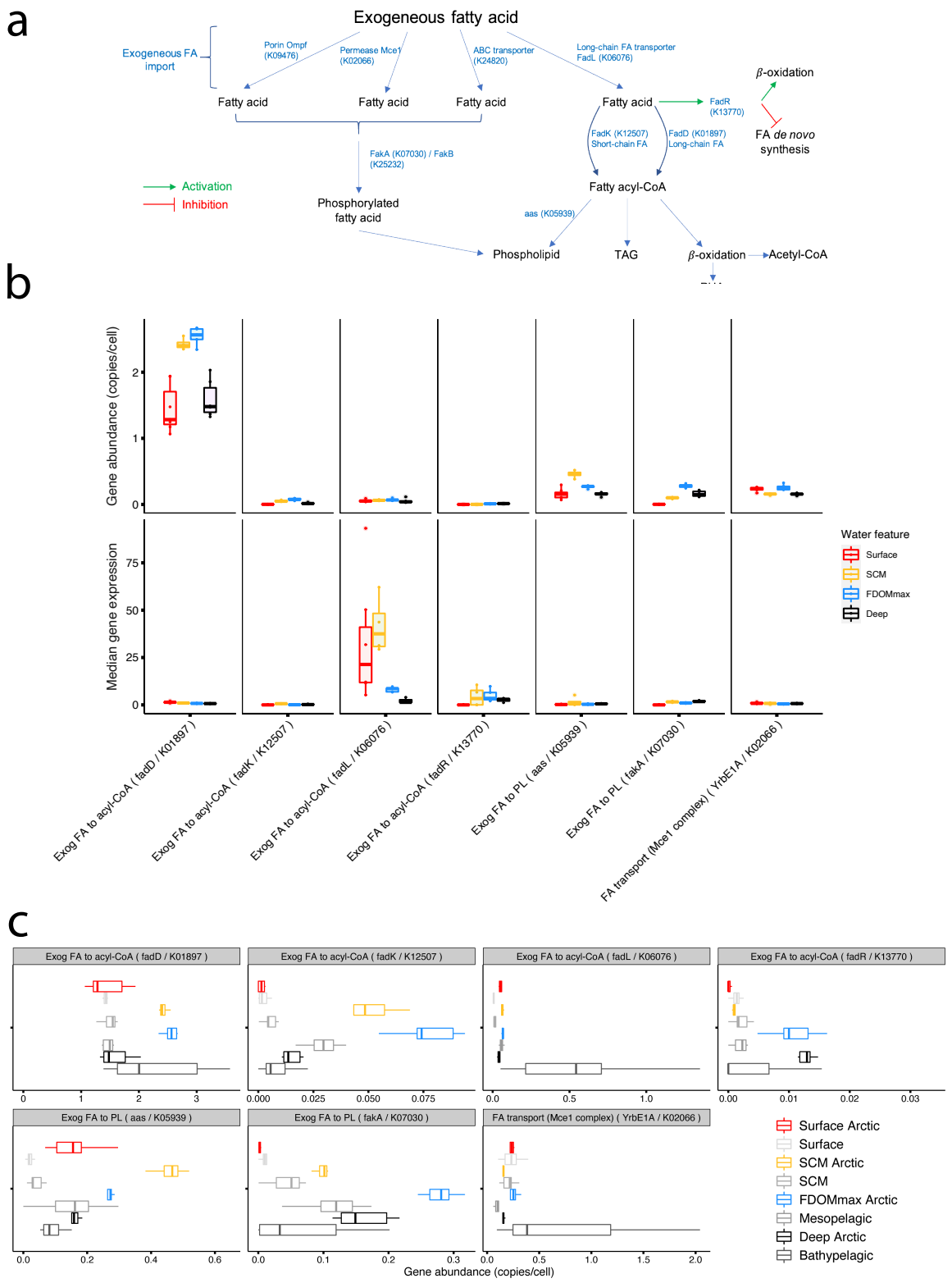

Figure 3.8: *Import of exogenous fatty acids (FAs) in the Canada Basin and the global ocean microbiomes. a) Schematics of various transporters that have been reported to selectively import FAs or facilitate the import of FA non-selectively. b) Abundance of genes coding for protein families involved in exogenous FA import within metagenomes and metatranscriptomes of the Canada Basin.*

3.3.5 Taxonomic distribution of marker genes for NL synthesis and degradation pathways

The capacity of the ocean microbiomes to synthesize and degrade NLs has mostly been studied using isolation and cultures of specific bacterial strains[12,41,42]. Consequently, we only possess scarce information regarding the extent and phylogenetic breadth of NL biosynthesis and degradation in the ocean microbiomes. This lack of knowledge prompted us to determine the taxonomic affiliation of NL biosynthesis and degradation genes. Genes were clustered at 95% identity to reduce redundancy and searched against the NCBI nr database. The PDAT genes (EC:2.3.1.158) as well as the SE hydrolase genes (EC:3.1.1.13), were exclusively assigned to eukaryotes, while the *phaC* and *phbC* genes (EC:2.3.1.304) were assigned to bacteria and archaea (Figure 3.9). In the photic zone (surface and SCM), phototrophic eukaryotes in the order *Mamiellophyceae* dominated the taxonomic affiliation of the PDAT genes (EC:2.3.1.158). Different eukaryotic taxa (with no order assigned, as well as *Eumetazoa*, *Prymnesiophyceae* and *Ciliophora*) were affiliated with DGAT genes (EC:2.3.1.20). DGAT genes were also assigned to bacteria in the surface (mostly *Gammaproteobacteria*) and SCM (orders *Acidimicrobia* and *Actinobacteria* as well as *Deltaproteobacteria*) (Figure 3.9). In the photic zone, we found more TAG acylhydrolase gene clusters (EC:3.1.1.3, 128 in the surface and 199 in the SCM) and *fadL* gene clusters (K06076, 89 in the surface and 73 in the SCM) than DGAT gene clusters (41 in the surface and 52 in the SCM) assigned to bacterial taxa. Bacterial TAG acylhydrolase gene clusters in the photic zone were assigned to *Rhizobiales*, and various *Gammaproteobacteria* orders. In addition, no *fadL* genes were assigned to eukaryotic taxa. This suggests that a subset of the bacterial community uses NLs as a growth substrate without the ability to synthesize them.

In the deep, synthesis and degradation of NLs was dominated by PHAs and almost entirely assigned to bacteria (Figure 3.9). The number of *phaC* gene clusters (K03821) and *phaZ* gene clusters (EC:3.1.1.75) increased with depth (55-366 and 42-49 respectively) and were characterized by similar taxonomic distribution, falling overwhelmingly in *Alphaproteobacteria* orders (Figure 3.9). *Rhodobacterales* and unclassified *Alphaproteobacteria* were most represented in the surface while unclassified *Alphaproteobacteria*, *Rhodospirillales* and unclassified *Proteobacteria* were most represented in the rest of the water column (SCM, FDOMmax and deep waters). The number of bacterial DGAT gene clusters involved were lower in the FDOMmax (16) and deep waters (38) than in the photic zone (41 in the surface and 52 in the SCM). The TAG acylhydrolase gene clusters were also more numerous than the DGAT gene clusters in the FDOMmax (301) and deep waters (425).

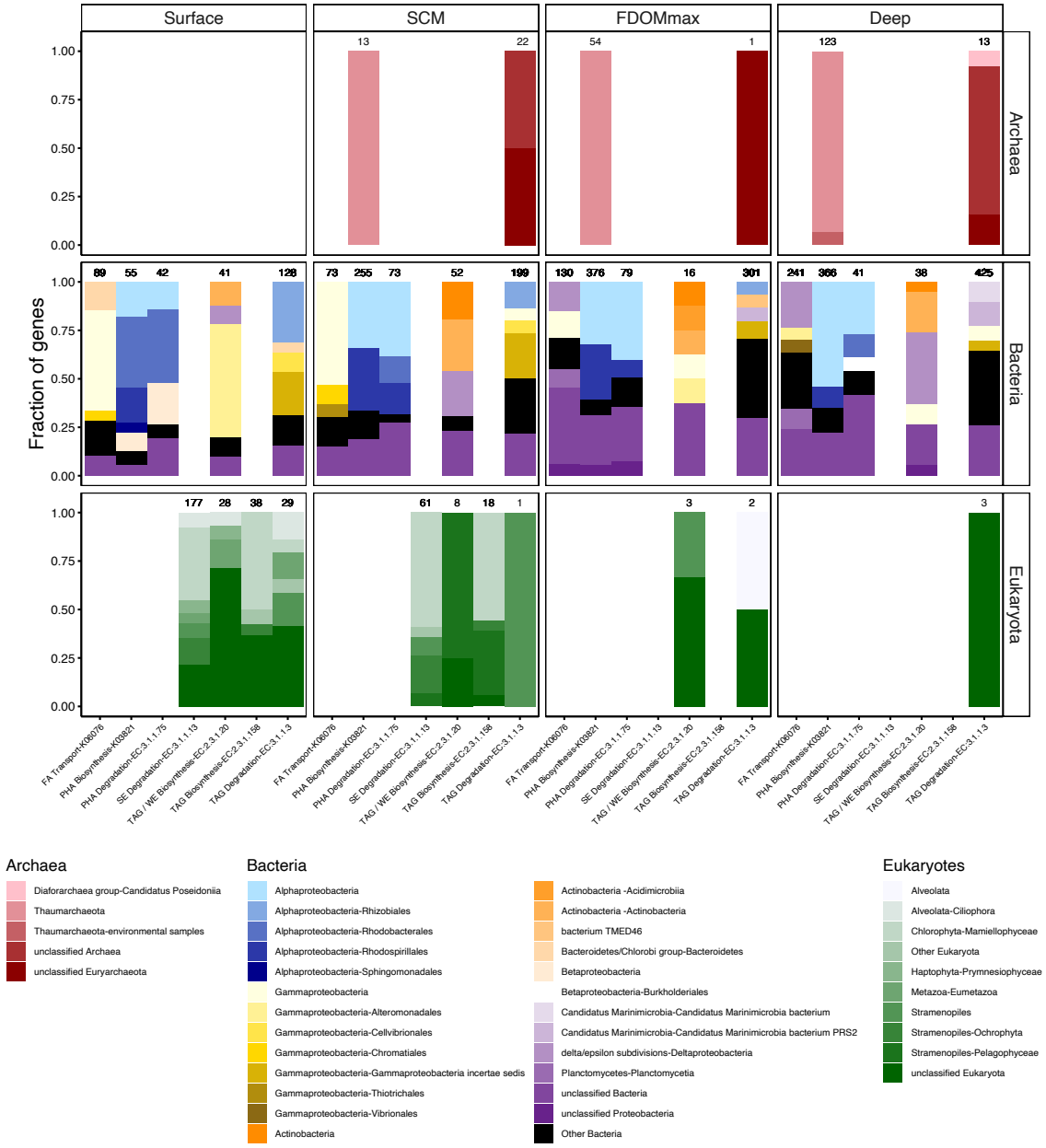

Figure 3.9: *Taxonomic affiliation of genes coding for protein families that catalyze key step of neutral lipid biosynthetic and degradation pathways. The taxonomy is separated by domain and water column features of the Canada Basin. The numbers at the top of each bar represent the total number of gene clusters (clustered at 95% identity) obtained for a protein family (KO number) in a water column feature.*

3.3.6 Identification of storage compounds metabolic genes in metagenome-assembled genomes

We observed a discrepancy between the abundance and taxonomic identity of NL biosynthesis and degradation marker genes. The taxonomic affiliation of DGAT genes with phototrophic eukaryotes in the ocean was expected. However, the taxonomic affiliation of DGAT genes to bacteria as well as the high number of genes clusters for TAG acylhydrolase affiliated with bacteria was more surprising. We therefore asked if a subset of the bacterial population could use NLs as a growth substrate as opposed to another subset using biosynthesis and degradation

of NLs as a survival strategy. To explore the co-occurrence of NL biosynthesis and degradation genes in the bacterial genomes of the Canada Basin, we reconstructed metagenome-assembled genomes (MAGs) from our metagenomic data. After filtering for genomes greater than 50% completeness and less than 10% contamination and strain heterogeneity, 664 genomes remained.

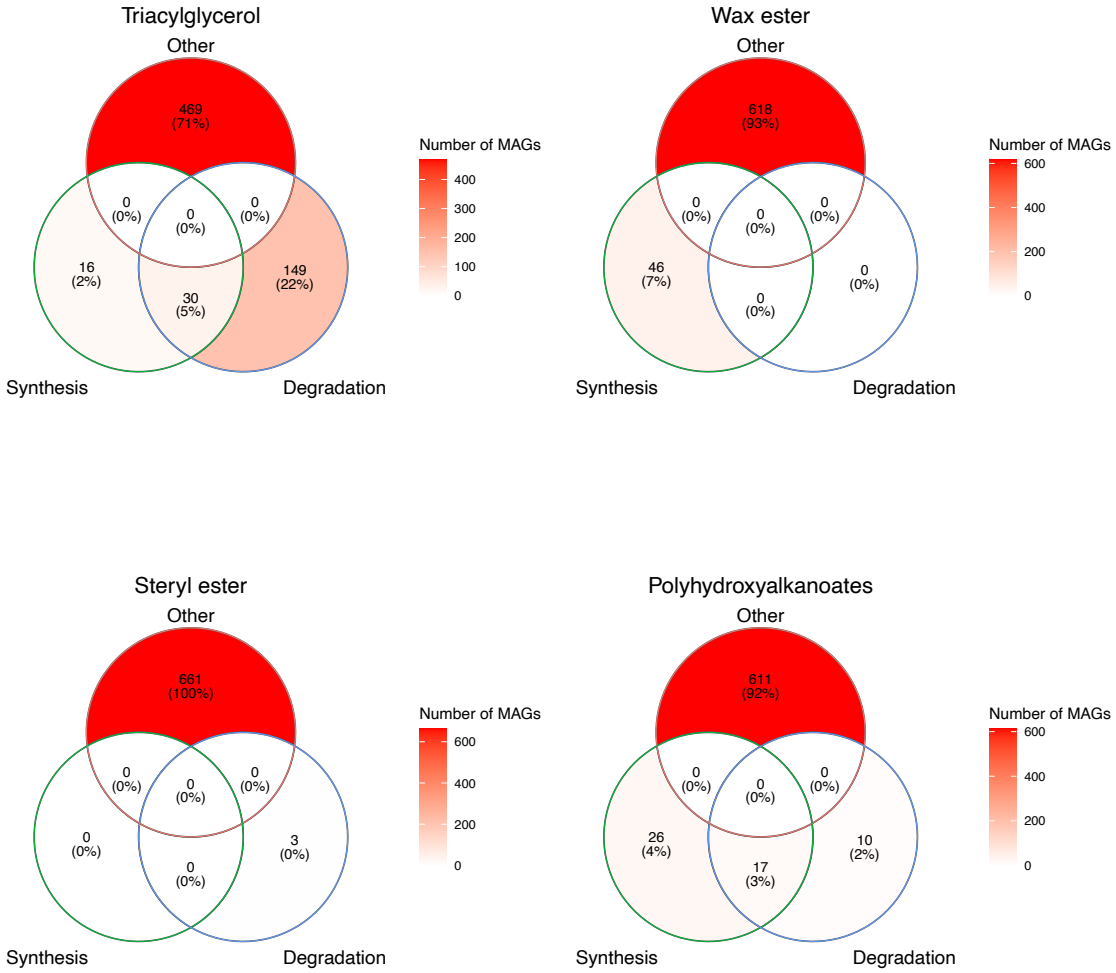

Figure 3.10: *Venn diagrams of the number of metagenomes assembled genomes (MAGs) that contain gene coding for protein families catalyzing key steps of NL biosynthesis (Synthesis), NL degradation (Degradation) or that don't contain either genes (Other).*

Seven percent of MAGs (46 MAGs) showed evidence for TAG biosynthesis (DGAT gene, EC:2.3.1.20, Figure 3.10) while 22% of MAGs (149 MAGs) showed evidence for TAG degradation (TAG acylhydrolase gene, EC:3.1.1.3). MAGs with TAG biosynthesis capacity were classified mostly into *Acidimicrobiia* (7 *Acidimicrobiales* MAGs), *Planctomycetes* (9 *Pirerullales* MAGs), and various orders of uncultivated bacteria and archaea (UBA), UBA9160 being the most represented (8 MAGs) (Figure 3.12, 3.11). The taxonomic identity of MAGs with TAG degradation capacity was different, belonging to *Gammaproteobacteria* (among which 10 *Pseudomonadales* MAG), *Verrucomicrobiae* (12 MAGs), *Planctomycetes* (27 MAGs) and various UBA orders (Figure 3.12, 3.11).

The taxonomy of MAGs with PHA biosynthesis capacity (*phaC* and *phbC* genes, EC:2.3.1.304) was significantly different than MAGs with TAG biosynthesis capacity (Figure 3.12, 3.11).

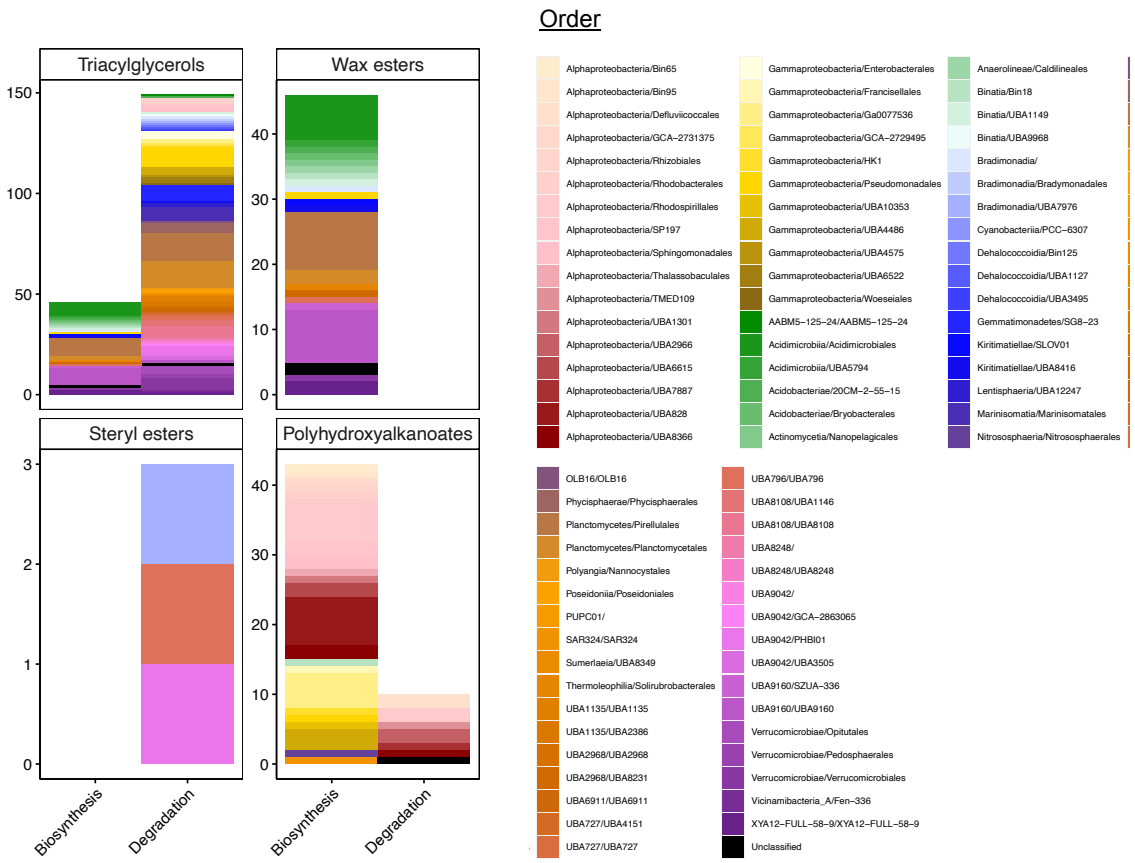

Figure 3.11: *Taxonomic identity of metagenomes assembled genomes that harbour genes coding for protein families catalyzing key steps of NL biosynthesis (Biosynthesis) or NL degradation (Degradation).*

While DGAT genes were found mostly in *Myxococcota*, *Planctomycetota* and *Actinobacteria*, the *phaC* and *phbC* genes were found in *Alpha-* and *Gammaproteobacteria*. PHA biosynthesis MAGs were dominated by *Alphaproteobacteria* mostly within the orders *Rhodospirillales* (7 MAGs) and UBA828 (7 MAGs) as well as by *Gammaproteobacteria*, mostly within the orders Ga0077536 (5 MAGs) and UBA4486 (3 MAGs). A smaller subset of the MAGs (10 MAGs, 2%), all within *Alphaproteobacteria*, had the capacity to degrade PHAs (*phaZ* gene, Figure 3.12, 3.11). We found 3 MAGs with the SE degradation capacity gene (SE hydrolase gene, EC:3.1.1.13). These results show that different subsets of the Canada Basin microbiomes can store TAGs/SEs and PHAs and that a large fraction of the community may use exogenous TAGs as a growth substrate.

3.3.7 Co-incidence of exogenous FA transport and NL metabolism genes in metagenome-assembled genomes

Our results demonstrated an enhanced capacity to synthesize and degrade NLs in the Arctic Ocean prokaryotes and picoeukaryotes compared to other oceans. These NLs can be used by their producers as a survival strategy but can also be used by others as a growth resource. The ability to import exogenous FAs is necessary for bacteria to generate acetyl-CoA from an exogenous source of NLs. Alternatively, bacteria can use exogenous FAs to integrate into NLs

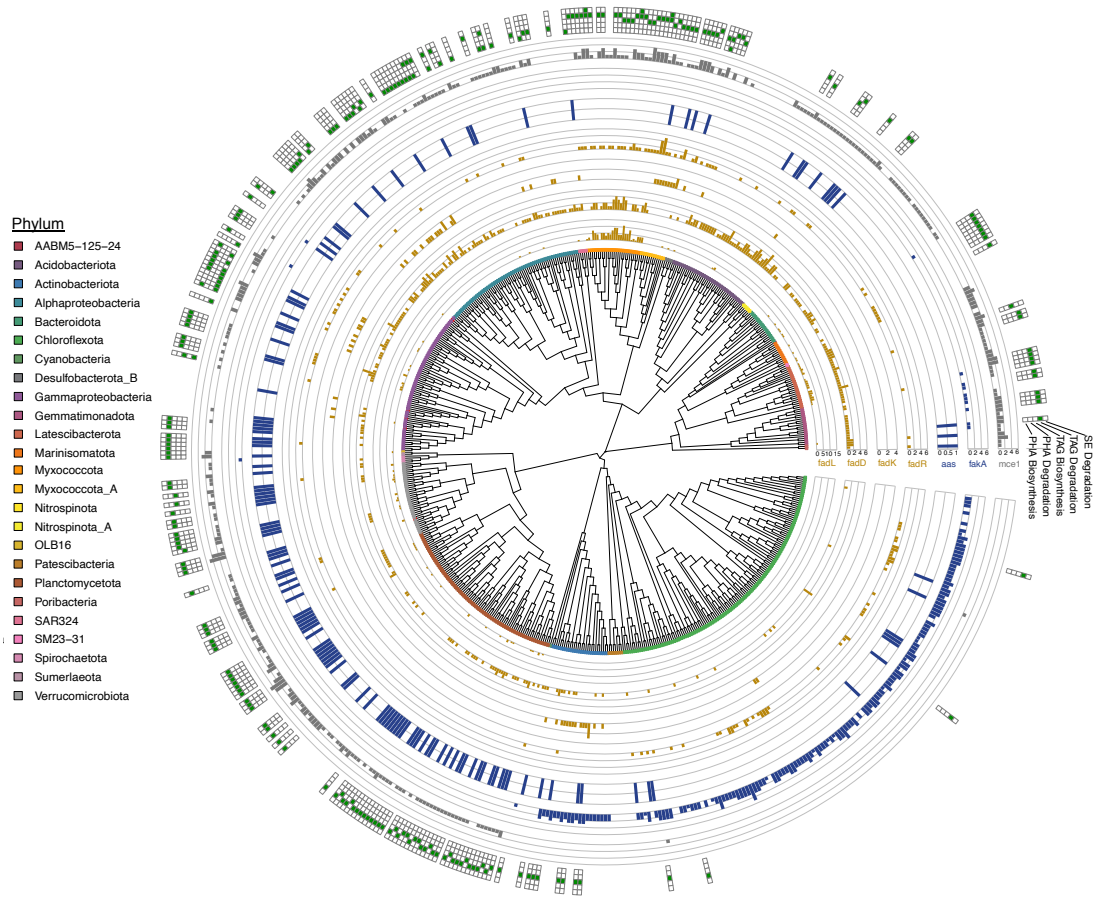

Figure 3.12: *Phylogenetic tree of bacterial metagenome assembled genomes (MAGs) displaying neutral lipid metabolism and exogenous FA import genes. The presence of gene coding for protein families catalyzing key steps of NL biosynthesis and degradation in a MAG is represented by a green cell. The number of genes involved in the import of FA and funneled to β-oxidation (dark yellow) or to membrane lipids (blue) in each MAG is represented by a vertical bar..*

or PLs to circumvent the energy expense of *de novo* FA synthesis. To further examine how exogenous FA could contribute to be used as a growth substrate or incorporated in NLs and PLs, we explored how FA transport genes were distributed through MAGs with NL biosynthesis and degradation capacity.

MAGs with different NL metabolic capacity all had very close mean genome completeness (Figure 3.13). The *fadL* long-chain FA transporter co-occurred with both TAG degradation and TAG biosynthesis gene in MAGs but not with PHA metabolism genes (Figure 3.12, 3.14). However, the long-chain FA-acylating gene *fadD* only co-occurred with TAG degradation, while the short-chain FA acylating gene *fadK* co-occurred with both TAG and PHA biosynthesis (Figure 3.12, 3.14). The *mce1* gene that imports exogenous FAs co-occurred with TAG metabolism and PHA biosynthesis genes. The *aas* gene, incorporating acylated FAs to PLs only co-occurred with TAG degradation and PHA biosynthesis. Oppositely, the *fakA* gene, that directly phosphorylates exogenous FA to incorporate into PLs was incongruent with both TAG degradation and PHA biosynthesis. All the genes involved in the import and acetylation of exogenous FAs were incongruent with MAGs that were not involved in NL metabolism. Only the *fakA* gene co-occurred with these MAGs. Altogether, these results show that MAGs with NL metabolism

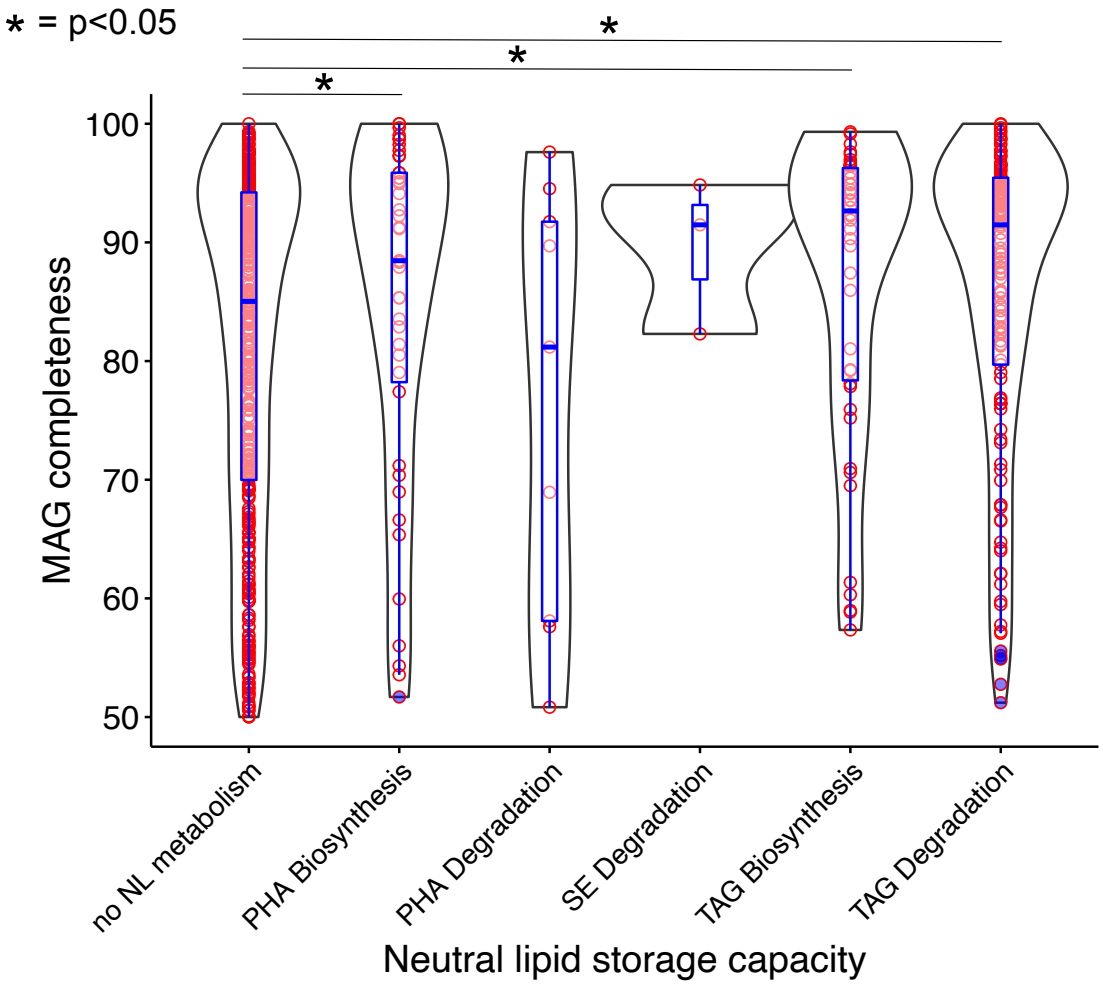

Figure 3.13: *Completeness of metagenomes assembled genomes that harbor genes coding for protein families catalyzing key steps of NL biosynthesis (PHA biosynthesis, TAG biosynthesis) or NL degradation (SE degradation, PHA degradation, TAG degradation) and MAGs that don't harbor any NL biosynthesis or degradation genes (no NL metabolism).*

generally possess the genes to funnel exogenous FAs either to NL or β-oxidation, while MAGs that are not involved in NL metabolism possess genes to funnel exogenous FAs to PLs.

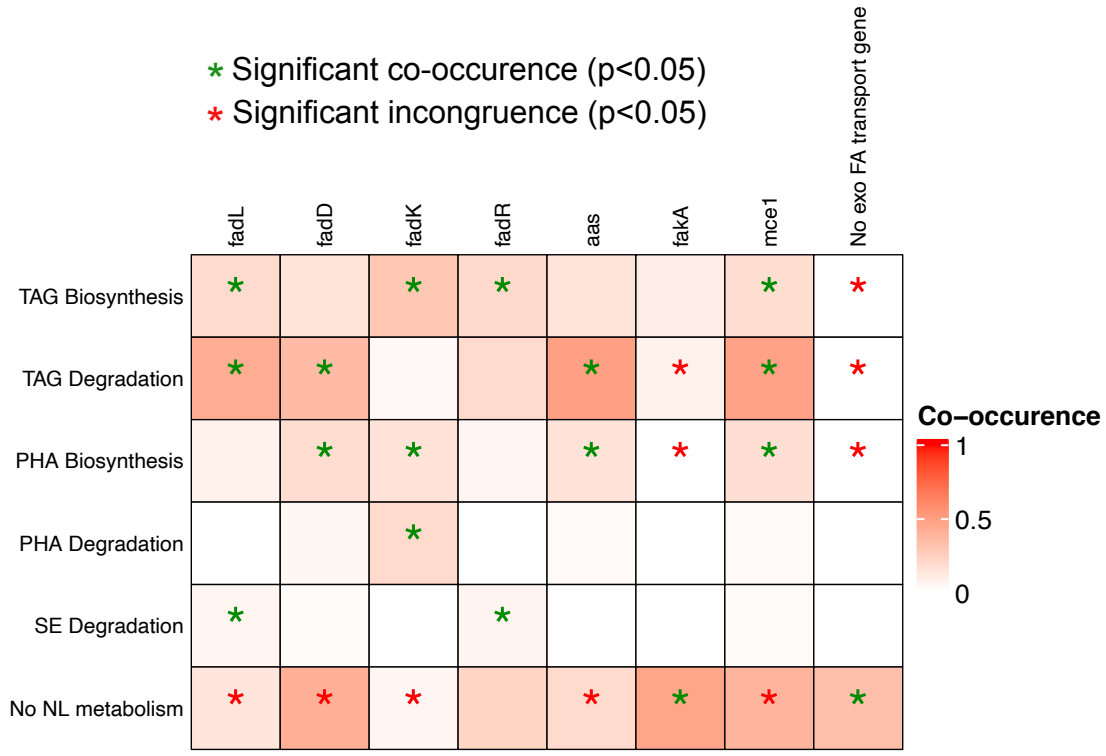

Figure 3.14: *Co-occurrence of genes involved in NL metabolism and exogenous FA import in our MAG dataset. The co-occurrence is calculated based on the Sorensen index. Green stars indicate statistically significant co-occurrence of genes in MAGs. Red stars indicate a significant absence of one gene if the second gene is present (incongruence of genes).*

3.3.8 Organic carbon sources for TAG and PHA biosynthesis in Canada Basin microbiomes

Our results showed that PHAs were the most common NLs employed by bacteria in the Canada Basin (K03821, Figure 3.3, 3.9), but also revealed their potential to store TAGs and WEs (EC:2.3.1.20, Figure 3.3,3.9). Our results also highlighted that MAGs with the capacity to store PHAs and TAGs could import exogenous FAs and possibly use them to incorporate into NLs (Figure 3.8-3.12, 3.14). We investigated the diversity of organic compounds that bacteria with the capacity to store NLs use for heterotrophic growth and NL storage. Studies investigating the storage of NLs in model species of prokaryotes have shown that they can use sugars [43] or aromatic compounds (AC) [44] as both a C and energy source to feed NL accumulation.

Our results show that on average, MAGs harbouring the DGAT gene possessed more genes coding for CAZymes and less aromatic compound degradation genes than MAGs harbouring the *phaC* and *phbC* genes (Figure 3.15, 3.16). While the MAGs able to synthesize PHAs are found mostly in the SCM and FDOMmax, the MAGs capable of synthesizing TAGs are most abundant in the deep waters, at the exception of *Actinobacteria* that were preferentially abundant in the SCM (Figure 3.15). MAGs found in the FDOMmax and SCM contained a higher number of aromatic compound degradation genes, while MAGs found in the deep had more CAZyme-coding genes. MAGs with more CAZyme-coding genes also possessed higher numbers

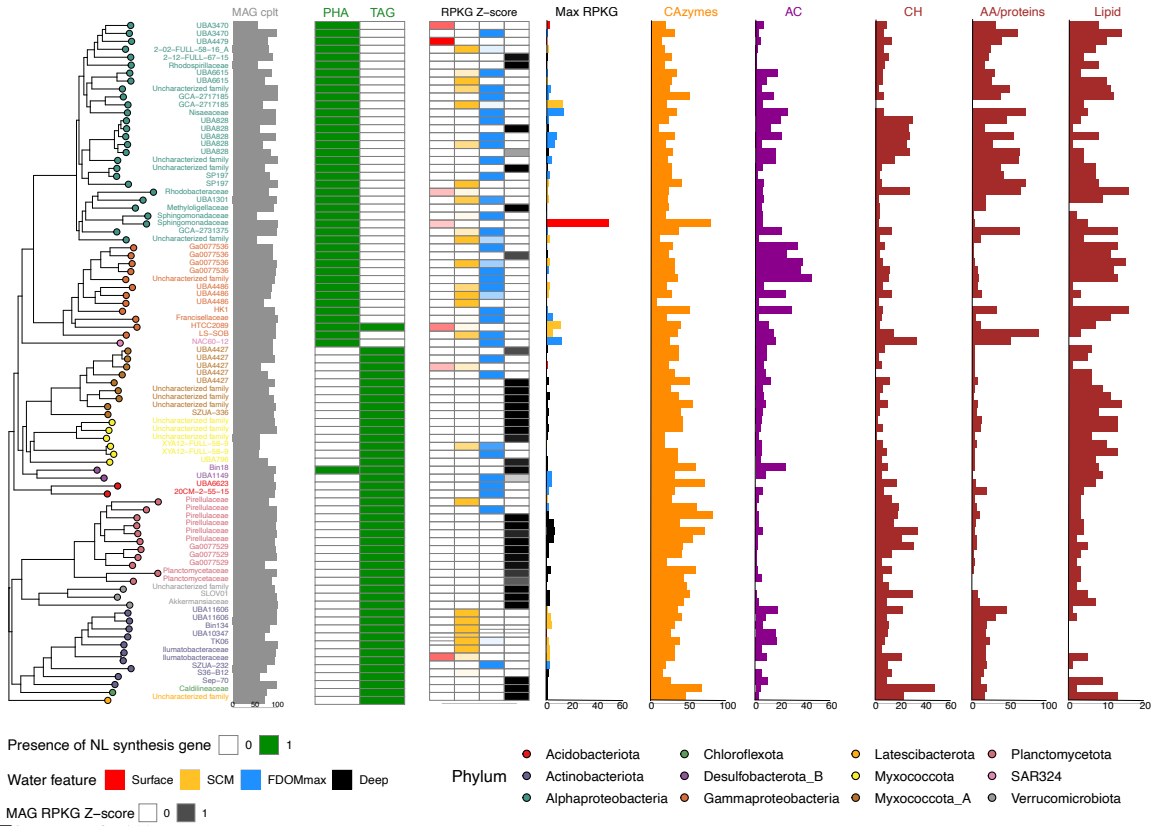

Figure 3.15: *Phylogenetic tree of bacterial metagenome assembled genomes (MAGs) displaying genomic capacity for NL synthesis and use of various carbon sources. Grey bars represent the MAG completeness (%). Green cells represent the presence of genes coding for the protein families catalyzing key steps of triacylglycerol or polyhydroxyalkanoate cycling. The red/yellow/blue/black heatmap represents the z-score of a MAG abundance over all the samples, averaged for each water feature. The red/yellow/blue/black bar plot represents the maximum abundance of each MAG, color-coded by the water feature of the samples in which the MAG was the most abundant (red=Surface, yellow= SCM, blue= FDOMmax, black=deep). The orange, purple and dark red bar plot represent, for each MAG, the number of genes coding for CAZymes (orange), AC degradation enzymes (purple) and carbohydrates (CH), amino acids/proteins (AA/proteins) and lipid transporters (dark red).*

of carbohydrate transporters. The number of genes coding for proteins and lipid transporters did not match any pattern of NL synthesis in the MAGs. *Gammaproteobacteria* and *Planctomycetota* were the taxonomic group most enriched in aromatic compound degradation genes and CAZymes coding genes, respectively. Altogether, these results show that the capacity to store different NLs is segregated among different phyla living in different depth of the water column and co-occur with gene involved in the use of different sources of C and energy.

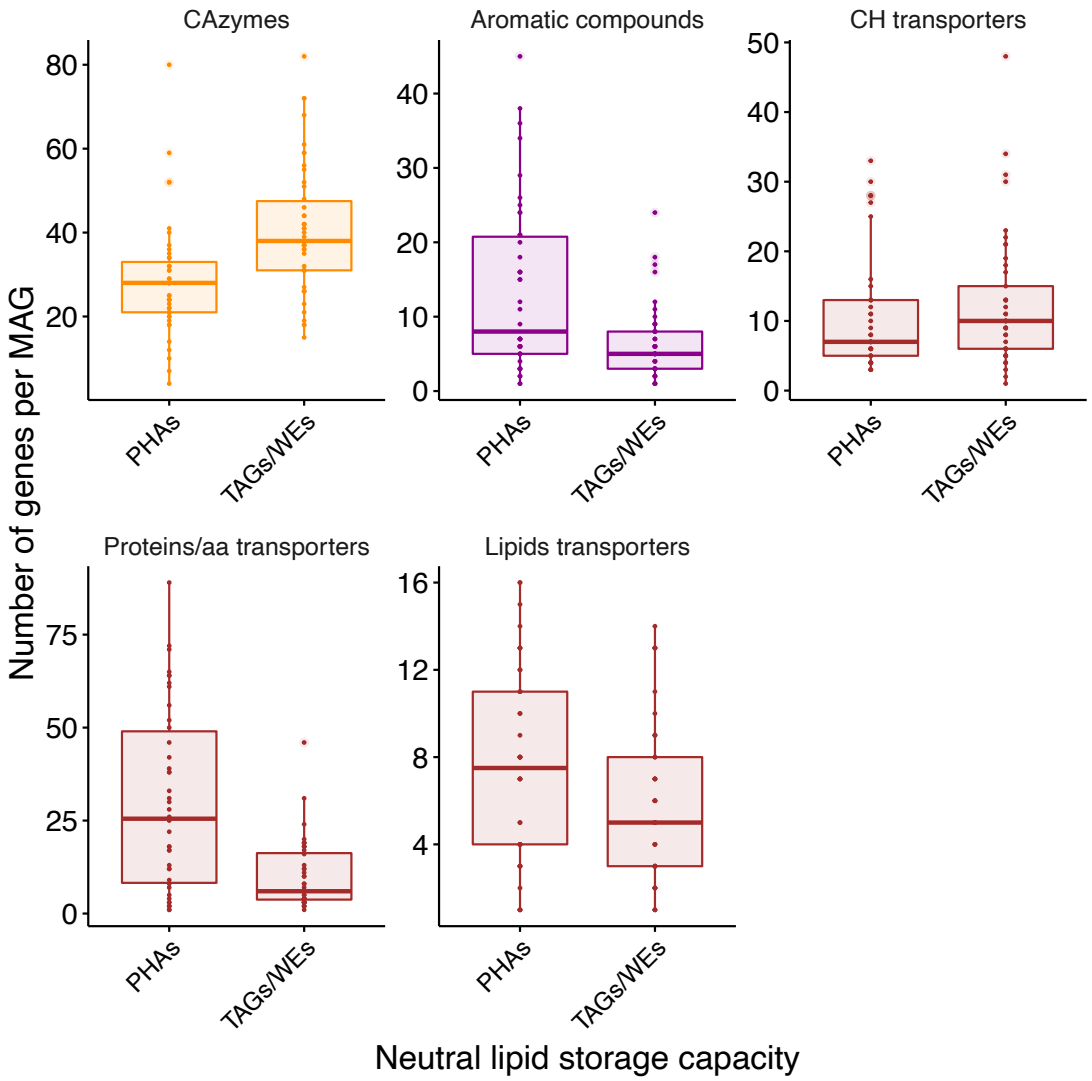

Figure 3.16: *Comparison of the average number of genes between metagenome-assembled genomes involved in PHA and TAG metabolism for genes coding for CAZymes (orange), genes involved in AC degradation (purple) and genes coding for carbohydrates, amino acids/proteins and lipids transporters (dark red).*

3.4 Discussion

The importance of phototrophic eukaryotes for neutral lipid storage in the Arctic Ocean

In this study, we highlighted the importance of NL metabolism in the microbiomes of the Canada Basin. The dominance of picoeukaryotic primary producers in the Arctic Ocean photic zone may explain the higher abundance of NL biosynthesis genes that we observed in the Arctic Ocean microbiomes compared to microbiomes from other oceans. The storage of TAGs is common in eukaryotes, but restricted to a few taxa in prokaryotes[5]. In the ocean, phototrophic eukaryotes are the most efficient photosynthetic TAG storers, as *Cyanobacteria* favour glycogen and PHAs[45] and TAG storage in ocean heterotrophic prokaryotes is limited by the primary produced C supply[46]. The hypothesis of picoeukaryotes dominating the storage of NLs is sup-

ported by our results showing that the most abundant and expressed NL biosynthesis genes in the photic zone were the PDAT genes (EC:2.3.1.158). These genes were exclusively from a eukaryotic origin, overwhelmingly from phototrophic eukaryotes *Mamiellophyceae* (*Micromonas* and *Bathycoccus*) in the surface and *Ochrophyta* (*Pelagophytes*) in the SCM. The storage of NLs in the eukaryotic phototrophs of the Arctic Ocean basins (such as the Canada Basin) photic zone may be further enhanced due the strong stratification. The photic zone is N depleted[22] after the phytoplankton blooms triggered by ice retreat in the spring. The stratification of the Canada Basin impedes a resupply of N from deeper waters in summer[47]. The low N prevents phytoplankton growth, but coupled to long periods of intense light in the Arctic summer, constitute optimal conditions for the accumulation of C reserve in the form of NLs[48].

The NLs stored during the Arctic summer may be key to eukaryotic phytoplankton survival during winter. The use of accumulated NLs coupled with reduced metabolism is indeed a strategy for some Arctic phytoplankton to overwinter during the dark polar night[24]. *Bathycoccus* populations have been shown to dominate the winter prasinophyte taxa[49]. Our data showing the prevalence of *Bathyccocus* in the TAG-biosynthetic taxa suggest that they use TAG storage to maintain some metabolic activity through winter. Some eukaryotic algae can sense decreasing light levels and trigger lipid storage biosynthesis, as demonstrated in Antarctic sea ice diatoms[50]. In the Arctic Ocean, this could be used as a strategy for phytoplankton to stock up reserves for winter and may explain the dominance of eukaryotes in the primary producer populations of the Arctic Ocean.

Environmental changes occurring in the Arctic Ocean due to a warming climate may enhance the storage of NLs by eukaryotic phytoplankton. Our results showed that *Micromonas* was one of the dominant TAG biosynthesizing taxa in the surface waters of the Arctic Ocean. *Micromonas* is a dominant, pan-Arctic phototrophic taxon[51]. The biosynthesis of TAGs in *Micromonas* taxa could therefore be highly important in sustaining the food web throughout the Arctic Ocean. Interestingly, it has been shown that the populations of *Micromonas* increased as the Arctic ocean water warmed over a 5 year period[51], and that *Micromonas polaris* has the potential to evolutionary adapt to warming waters[52]. The storage of NLs may therefore become more and more prevalent as the Arctic Ocean warms due to climate change. In addition, climate change observations and current scenarios indicate a longer period of phototrophic activity in the Arctic Ocean due to higher light intensity in the photic zone with sea ice loss[23]. Observations have also shown that the stratification of the Arctic basins has strengthened, decreasing the resupply of nutrients from deep to surface waters[47]. The increasing light availability coupled to the decreasing nutrients supply are susceptible to considerably promote the storage of NLs in the Arctic Ocean phytoplankton in the future, increasing the transfer of energy from phytoplankton lipids to the rest of the food chain.

TAGs serve as a carbon source for a "lipotrophic" subset of Arctic Ocean microbiomes

We suspect that the higher fraction of lipids in the primary productivity C pool of the Arctic Ocean compared to other oceans may be used as an important C source for the heterotrophic

microbial populations. Lipids produced by phytoplankton can be released in the water column through various mechanisms such as exudation[53], viral lysis[54,55], autolysis[7] or algicidal bacteria[56,57]. The number of prokaryotic gene clusters annotated with the TAG acylhydrolase genes (EC:3.1.1.3) was higher that the number of prokaryotic gene clusters annotated with the DGAT (EC:2.3.1.20). This supports the existence of a "lipotrophic" fraction of the microbiome not able to store NLs but able to use exogenous TAGs as C and energy source in the Arctic Ocean. Results showing the existence of a fraction (22%) of our MAG dataset that harbor TAG hydrolase genes (EC:3.1.1.3) but not the DGAT enzyme (EC:2.3.1.20) further reinforces this hypothesis. In addition, a higher fraction of these "lipotrophic" MAGs contained genes involved in the import and use of exogenous FAs compared to MAGs with no NL metabolism genes. As the TAG acylhydrolase is secreted extracellularly[58], the co-occurrence of the TAG hydrolase and genes involved in exogenous FA import in the "lipotrophic" MAGs points to their capacity to use exogenous TAGs as growth substrate. The ability of ocean microbiomes to metabolize exogenous TAGs has so far only been documented in the Antarctic Ocean[59], suggesting that the "lipotrophic" microbiomes may be prevalent in cold oceanic zones.

Unlike TAGs, PHAs seem to be exclusively consumed by its prokaryotic producers. Within a water mass, the number of *phaC* and *phbC* gene clusters (K03821) was consistently higher that the number of *phaZ* gene clusters (EC:3.1.1.75). Further supporting the consumption of PHAs only by its producers is the observation of only a small number of MAGs (10) that only contained the *phaZ* gene, although this may result from gene absence in incomplete MAGs.

Expanded taxonomic diversity of prokaryotic TAG and PHA storer in the ocean

We expected that the capacity to store PHAs would be common in Arctic bacterial taxa as it is commonly reported that the ability to store PHAs is widespread across prokaryotes[5]. Our results show a large diversity of taxa with the ability to polymerize PHAs as evidenced by the large number of genes clusters involved in PHA biosynthesis (annotated with K03821). The phylogenetic restriction of PHA biosynthesis within *Alpha-* and *Gammaproteobacteria* that we observe in our data is consistent with what was has been reported in other world oceans[60]. However, taxa from marine environment reported to store PHAs were restricted to those that could be cultivated[60]. To our knowledge, no studies attempted to systematically assess the capacity of ocean microbiomes to synthesize PHAs. Our work is therefore the first report to systematically survey the capacity of ocean microbiomes to synthesize PHAs. Only one study reported the global taxonomic distribution of PHA degradation taxa by surveying the taxonomic affiliation of the *phaZ* gene in various environments[61]. This study showed that *phaZ* gene in marine environment was overwhelmingly assigned to *Gamma-* and *Alphaproteobacteria*. Our results showed that most taxa with *phaZ* gene also possessed the *phaC* or *phbC* gene. In addition, we showed that the abundance of both *phaC/phbC* and *phaZ* genes represented similar fraction of the total gene pool in the Arctic and other oceans. This led us to think that PHA storage is widely distributed in the global ocean and phylogenetically concentrated in *Gamma-* and *Alphaproteobacteria*. Our study can therefore serve as a basis to further explore PHA storage capacities in the global ocean microbiome.

Our study considerably expands the known diversity of bacterial taxa potentially storing TAGs in the ocean. Most of the bacterial taxa with this ability were previously identified in C-rich environment such as municipal and agro-industrial waste waters[7,62], or in environments with long periods of C scarcity such as desert soils[2,43]. These taxa are largely within the *Actinobacteria* (e.g. *Rhodococcus opacus, Rhdococococcus jostii, Mycobacterium bovis, Streptomyces avermitilis*) and *Gammaproteobacteria* groups (e.g. *Acinetobacter baylyi, Alcanivorax borkumensis*)[6,63,64] but genes coding for the DGAT were also found in members of *Alpha-*, *Beta-* and *Deltaproteobacteria* as well as in *Bacteroidetes*[65]. In the ocean, TAG biosynthesis has frequently been reported for bacterial species of the *Gammaproteobacteria* genera *Alcanivorax*[66] and *Marinobacter*[67] in oil spills[34,41,68]. The diversity of taxa able to store TAGs that we report in this study was therefore unexpected. In addition to the expected *Actinobacteria* (10 MAGs), we also report genomes in the group *Myxoccocota* (15 MAGs), *Planctomycetota* (11 MAGs), *Verrucomicrobia* (3 MAGs), *Acidobacteria* (2 MAGs), *Desulfobacterota* (2 MAGs), *Chloroflexota* (1 MAG) and *Latescibacteria* (1 MAG). A *Myxoccocota* species has been shown to produce TAGs[69] and TAG biosynthesis was identified in a phagocytic *Planctomycete*[70]. However, to our knowledge, our study considerably expands the diversity of bacterial taxa known to synthesize TAGs/WEs in the ocean. Further experimental work such as microscopy will be needed to confirm the accumulation of TAGs in these groups within the Arctic Ocean.

The ecology of NL storage for the Arctic microbiomes

The storage of NLs may be used by a subset of the Arctic Ocean bacterial populations as a life strategy to survive the highly variable seasonal conditions of the Arctic Ocean. The storage of NLs allows cells to decouple their metabolic activity from the supply of carbon sources[3]. This could prove an adaptive strategy in the highly fluctuating Arctic Ocean. The spring and summer are characterized by an intense phytoplanktonic primary productivity, while in the dark winter, photosynthesis is not active[22]. Bacterial taxa with the ability to store NLs may use primary-produced carbon to constitute reserves in spring, summer and early winter and consume these reserves to maintain a metabolic activity during the unproductive Arctic winter. The storage of compounds in summer and its mobilization in winter is a strategy that has been observed in soil bacteria to survive through winter[33]. In addition, NLs may help bacterial cell to resist other stressors. As highly reduced compounds, PHAs help maintain the redox state of the cell under oxidative stress[11,71]. PHAs could therefore participate to protect the Arctic Ocean bacteria from cold-induced oxidative stress, similarly to what has been shown in the Antarctic bacterium *Pseudomonas sp. 14-3.*[72,73].

In the Arctic Ocean, bacteria most probably accumulate NLs slowly under a *reserve* storage scenario. In the Arctic Ocean, the optimal conditions for NL storage with an excess of C and energy and N limitation is achieved during phytoplankton blooms. During these blooms, eukaryotic phytoplankton leak excess C in the water column, and outcompete bacteria for inorganic N sources[74]. These blooms are limited in the central Arctic[22], such as the Canada Basin, and not likely to provide enough C to enable the *surplus* storage of NLs in bacteria. Our results showing very low abundance of MAGs able to store TAGs or PHAs in the surface samples, where phytoplankton blooms occur, support this hypothesis. Rather, bacteria in the

Arctic Ocean may store NLs with limited resources, hence under a *reserve* storage scenario. Under that scenario, the accumulation of NLs competes with other metabolic processes and would be slow. The storage of NLs would therefore be favored by strategies limiting the energy demand of NLs biosynthesis. Using exogenous FAs instead of synthesizing FAs to incorporate FAs into NLs greatly reduces the energy cost of NL storage. Our results show that a higher percentage of MAGs able to store TAGs and PHAs contained genes involved in exogenous FA import and use than MAGs with no NL metabolism genes. Exogenous FAs can be incorporated directly to TAGs. FAs can also be incorporated into PHAs either by being diverted from β-oxidation (to form mclPHA) or by going through β-oxidation and producing acetyl-CoA that can be used to synthesize PHB (short-chain PHA). The absence of the medium-chain PHA hydrolase gene in our data suggests that the second option is preferred.

In addition to using exogenous FAs, bacteria may use different substrate to feed the pool of acetyl-CoA that can be used for either TAG of PHB synthesis. As C and energy sources are limited in the Arctic Ocean, the pool of acetyl-CoA "set aside" from other metabolic processes for NL synthesis must be limited, hence NL accumulation is slow. However, even a limited but lasting C and energy source may be sufficient for slow NL accumulation. This limited but lasting C source may come from terrestrial organic matter in the Arctic Ocean. Spring and summer freshwater input carry a disproportionately high amount of terrestrial organic matter to the Arctic Ocean compared to other oceans[75]. This terrestrial carbon accumulates and peaks in the FDOMmax by sinking with brine in winter[76], and is also abundant in the SCM. Studies from our group showed that the potential ability to use aromatic compounds reflects the capacity to use this terrestrial organic matter as a growth substrate[77]. MAGs with the ability to store TAGs and PHAs that were abundant in the SCM and FDOMmax harboured more aromatic compound degradation genes than those more abundant in deep waters. This suggests that these MAGs may use terrestrially-derived aromatic compounds to feed the pool of acetyl-CoA necessary for PHB and de novo FA synthesis for TAGs. As an example, the degradation of vanillin, an aromatic compound found in the Canada Basin FDOMmax, generates 1 NADPH, 1 NADH, 4 H+ and 2 pyruvates. The production of one molecule of acetyl-CoA from one pyruvate generates a NADPH. The formation and polymerization of one hydroxybutanoate monomer into PHB requires two acetyl-CoA molecules and one NADPH and one H+. The use of one vanillin molecules to add one monomer to PHB therefore generates a positive budget of 1 NADH, 2 NADPH and 3 H+. In contrast, MAGs with the capacity to synthesize TAGs and abundant in deep samples possessed more CAZymes genes. This may reflect the capacity to recycle carbohydrates from dying bacteria or phytoplankton-derived marine snow[78]. The carbon and nutrients concentration of marine snow and dying bacteria are up to 4 orders of magnitude more elevated than the background water concentrations[79]. The encounter of a marine snow particle or a dying bacterium may therefore represent a surplus storage scenario for the taxa able to store TAGs. They may use the TAGs accumulated during these encounters to survive until the next encounter. Interestingly, studies showed that storage of TAGs is favoured over PHAs in cyclic feast/fast conditions when there is an overlap in time between the supply of C and N during the feast stage[15]. As marine snow and dying bacteria also contain nitrogen, this may explain why we only find TAG and not PHA biosynthesis genes in MAGs that are

abundant in deep samples.

A possible role for prokaryotic NL storage to sustain the winter Arctic Ocean ecosystem

We hypothesize that NL storage in the upper water column of the Arctic Ocean shifts from phototrophic eukaryotic based production in spring through fall towards heterotrophic prokaryotic based production during the winter polar night (Figure 3.17). There is evidence that the Arctic Ocean is not an unproductive desert during the polar night but still retains significant levels of biological activity[31,32]. It implies some light independent levels of metabolic activity at the base of the food web, hence in the microbiomes[80]. The storage of PHAs and TAGs by the bacterial taxa during the spring, summer, fall and throughout winter may provide the carbon pool to maintain this level of metabolic activity through the polar night. The C stored in PHAs and TAGs accumulating bacteria could then go up and support the food chain through bacterivory. Bacterivory is a common strategy for phototrophic eukaryotes to overwinter in polar seas[81]. Reports have shown that, in the winter, phototrophic *Micromonas pusilla*, one of the most abundant phototrophs in the Arctic Ocean, can survive by grazing on bacteria[82,83] and that winter prokaryotic populations are controlled by grazing[84]. This suggests a new and important role for the NL storage metabolism in the Arctic prokaryotic communities in sustaining the ecosystem during the polar night. To confirm this hypothesis would require winter sampling to identify storage of NLs in winter bacterial populations, and isotopes experiments to track NLs from prokaryotes to macro-organisms.

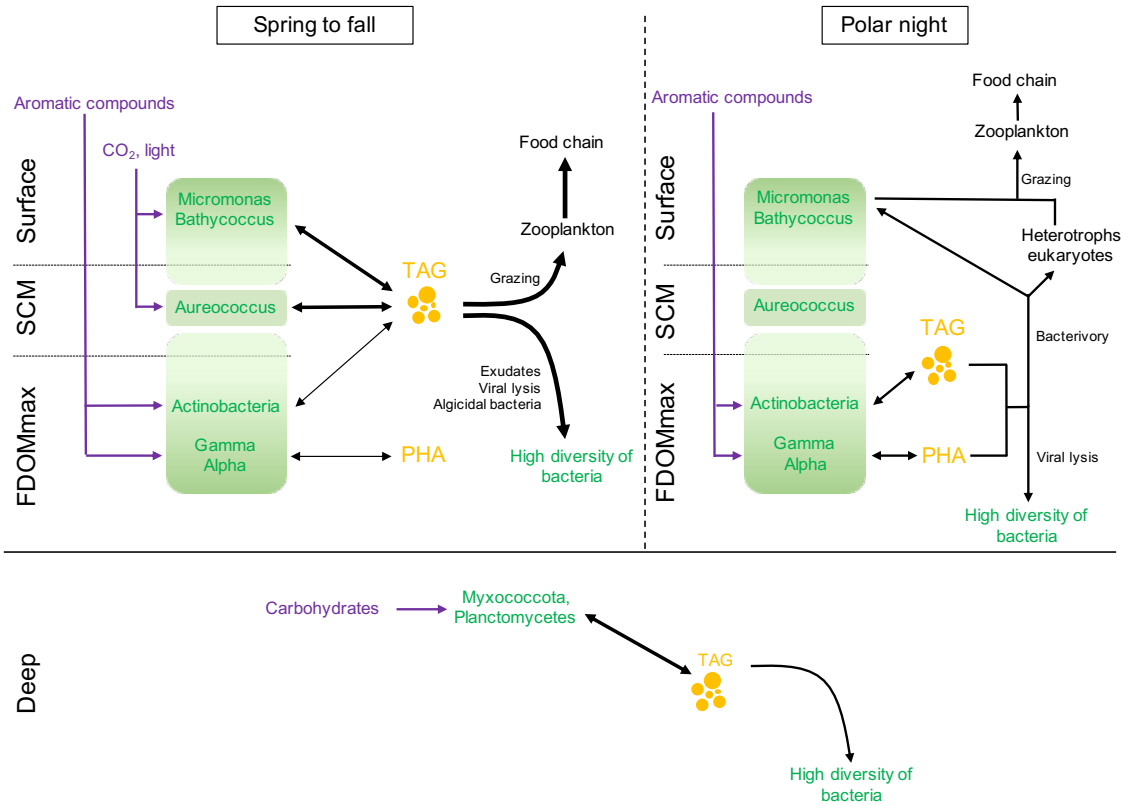

Figure 3.17: *Proposed model for the metabolism of neutral lipids in the microbiomes of the Arctic Ocean.*

3.5 Conclusion

The Arctic Ocean is characterized by an intense seasonality that drives strong changes in conditions, nutrients, and energy availability for the microbiomes. In this context, the accumulation of NLs as energy and carbon reserve may play an important role as a strategy to survive through yearly period of resource scarcity during winter. Given that lipids from microbial origin are an important growth resource for the Arctic Ocean food chain, and that NLs may constitute a significant fraction of microbial lipids, it is crucial to understand the metabolism of neutral lipids in the Arctic Ocean microbiome. In this study, we show that the pathways to synthesize and degrade various NLs were complete and expressed in the Arctic Ocean microbiome. PHA biosynthetic pathways were homogeneously distributed across the Arctic and global ocean microbiomes. The enhanced capacity to synthesize the TAGs in the microbiomes of the Arctic Ocean photic zone compared to the microbiomes of the global ocean was attributed to the prevalence of eukaryotic phytoplankton in the Arctic Ocean. The higher abundance and taxonomic distribution of TAG degradation compared with TAG biosynthesis genes, coupled to an enhanced capacity to degrade TAGs and import exogenous FAs in the bacterial populations of the Arctic Ocean compared to the global ocean suggested that a subset of the bacterial populations uses eukaryotic primary-produced NLs as a growth substrate. The identified diversity of bacterial genomes encoding the capacity to synthesize PHAs as well as the unexpected diversity of bacterial genomes able to synthesize TAGs in the photic zone prompted us to propose that NLs from bacterial origin may sustain the food web during the unproductive winter polar night. Overall, this study unravels an important role for microbial NLs and unveils their importance as a survival strategy in the Arctic Ocean.

3.6 Methods

Sampling, DNA and RNA extraction

Samples were collected in September 2017 during the Joint Ocean Ice Study cruise to the Canada Basin. We analyzed 22 metagenomes and 25 metatranscriptomes generated from samples collected across the water column of the Canada Basin. Throughout the water column, 8 depths corresponding to specific water masses were sampled: the surface mixed layer (surface: 5 m and 20 m depth) characterized by fresher water due to riverine input and ice melt, the subsurface chlorophyll maximum (SCM), 2 samples in the fluorescent dissolved organic matter maximum, in the halocline (FDOMmax: at salinity of 32.3 and 33.1 PSU, referred as 32.3 and 33.1), and deeper water from Atlantic origin at the temperature maximum (referred as Tmax), 1000 m depth (Atlantic water, further referred as AW) and 10 or 100 m above the bottom (further referred as bottom).

We filtered 14L of seawater for DNA samples and 7L of seawater for RNA samples sequentially through a 3 μm pore size polycarbonate track etch membrane filter (AMD manufacturing, ON, Canada) and a 0.22 μm pore size Sterivex filter (Millipore, MA, USA). Filters were stored

in RNALater (ThermoFischer, MA, USA), and kept frozen at -80^{0}C until processing in the lab. DNA was extracted following the method described in Colatriano *et al.*[85]. Briefly, the preservation solution was expelled and replaced by a SDS solution (0.1 M Tris-HCl pH 7.5, 5% glycerol, 10 mM EDTA, 1% Sodium Dodecyl Sulfate) and incubated at room temperature for 10 min and then at 95^{0}C for 15 min. The cell lysate was then ultracentrifuged at 3,270 x g. Proteins were precipitated with the protein precipitation solution MCP (Lucigen, WI, USA), and supernatant was collected after centrifugation at 17,000 x g for 10 min at 40^{0}C. DNA was then precipitated with 0.95 volume of isopropanol and rinsed twice with 750 μL ethanol before being air dried. The DNA was resuspended in 25 μL of low TE buffer, pH8 (10 mM Tris-HCl, 0.1mM EDTA) and stored at -80^{0}C.

The RNA extraction procedure was adapted from the mirVana RNA extraction kit (ThermoFisher, MA, USA). RNAlater was expelled from the Sterivex and replaced by 1.5 mL of Lysis buffer and Sterivex was vortexed. 150 μL of miRNA homogenate were added, the Sterivex vortexed and incubated on ice for 10 min. The cell lysate was expelled from the Sterivex, 0.9x the volume of Acid-phenol-Chloroform was added and the solution and vortexed for 30-60 sec. The mix was then centrifuged at 10,000 x g for 5 min, and the top aqueous phase gently removed and transferred to a fresh tube. 1.25 volume of ethanol 100% were added to the aqueous phase and vortexed to mix. The mix was then filtered through mirVana Filter Cartridges by centrifugating at 10,000 x g for 10 s, and the flow through discarded. The RNA was rinsed with 700 μL of Wash Solution 1 and then with 500 μL Wash solution 2/3 by centrifugating at 10,000 x g for 10 sec. RNA was then eluted with 50 μL of Elution solution (0.1 M EDTA) warmed at 95^{0}C. 700 μL of RTL buffer and 500 μL 100% ethanol were added to the RNA suspension and the suspension was centrifuged for 15 sec at 10,000 x g on a RNeasy MinElute column. RNA was washed first with RPE buffer by centrifuging 500 μL for 15 sec at 10,000 x g and then 80% ethanol for 2 min at 10,000 x g. The empty column was then centrifuged at 12,000 x g for 5 min to discard the excess liquid. The RNA was finally eluted by centrifugating first 28 μL and then 10 μL of RNase free water for 1 min at 12,000 x g and stored at -80^{0}C.

Metagenomic sequencing, assembly and annotation

Sequencing, assembly, and annotation were performed by the Joint Genome Institute (CA, USA). Metagenomes and metatranscriptomes were sequenced on the Illumina NovaSeq platform, generating paired-end reads of 2x150 bp for all libraries. Single assemblies were created by JGI for each individual sample using SPAdes[86] with kmer sizes of 33, 55, 77, 99, 127 bp. Gene prediction and annotation was performed using the DOE Joint Genome Institute Integrated Microbial Genomes Annotation Pipeline v.4.16.5[87].

Computation of gene abundance, expression, and transcripts abundance profiles

Metagenomics and metatranscriptomics data files containing genes IDs, gene annotations, gene depth of coverage and other gene information were retrieved from the DOE Joint Genome Institute Integrated Microbial Genomes (JGI/IMG, https://img.jgi.doe.gov) repository. Metagenomes and metatranscriptomes from the Canada Basin originate from samples collected by our lab. Metagenomes from other world oceans originate from samples collected by others and an-

notated by JGI/IMG. The abundance of a KEGG ortholog number (KO, gene family) in a metagenome or a metatransciptome was calculated by summing the depth of coverage of all genes annotated with this KO. KO abundance matrices therefore represent the metagenomic or metatranscriptomic profiles across the samples. As samples vary in terms of depth of sequencing (or library size), we normalized the metagenomic and metatranscriptomic profiles to the relative cell number to get a per-cell number of copies using the approach detailed in Milanese *et al.*[88]. Specifically, the KO abundances were divided by the median abundance of 10 universal single-copy phylogenetic marker genes (K06942, K01889, K01887, K01875, K01883, K01869, K01873, K01409, K03106, and K03110) for both the metagenomes and metatranscriptomes. These normalized KO abundances can therefore be interpreted as the per-cell number of gene copies for a given protein family (KO). Previous studies[88,89] showed that the metagenomic and metatranscriptomic abundances of the universal single-copy phylogenetic marker genes correlate. In addition, these genes are expressed over many different conditions and are therefore suitable to normalize the metatranscriptomic profiles. Similarly, the normalized KO abundance in metatranscriptomes can be interpreted as a per-cell number of transcripts for this KO. The gene expression, expressed as the relative number of transcripts per gene copy, was calculated for each KO by calculating the ratio between normalized KO abundance from metatranscriptome and metagenome.

Selection of NL storage and degradation pathways and marker genes

Metabolic pathways for the biosynthesis of TAGs, WEs, SEs and PHAs as well as metabolic pathways for the degradation of TAGs, SEs and PHAs were retrieved from MetaCyc (https://metacyc.org). To evaluate the abundance of various pathways, we used the abundance (see above for abundance calculation) of genes coding for various protein families (KO) that catalyze key reactions (EC) in the NL biosynthesis and degradation pathways. The key reactions for TAGs (EC:2.3.1.20 – diacylglycerol O-acyltransferase/wax synthase and EC:2.3.1.158 – phospholipid:diacylglycerol acyltransferase), WEs (EC:2.3.1.75 – wax synthase) and SEs (EC:2.3.1.26 – sterol O-acyltransferase) biosynthesis were selected as the esterification reaction between a FA and diacylglycerol, a fatty alcohol and sterol respectively. The marker enzyme for the degradation of TAGs (EC:3.1.1.3 – TAG acylhydrolase) and SEs (EC:3.1.1.13 – SE esterase) was responsible for the cleavage of the ester bond to free a FA from TAGs and SEs respectively. The key reaction chosen for the synthesis of PHAs (EC:2.3.1.304 – PHA synthase) was the esterification involved in the elongation of polyhydroxyalkanoate while the de-esterification (EC:3.1.1.75 – PHA hydrolase) was chosen as the key reaction for the PHA degradation pathway. As EC:2.3.1.304 was created in 2021, after the annotation of our metagenomes in 2018, we could not find any genes annotated with EC:2.3.1.304 in our dataset. Instead, we used the KEGG ortholog annotation K03821. There was no metabolic pathway for the degradation of WE in Metacyc. We therefore selected enzymes previously reported to cleave the ester bond of wax ester as marker enzymes: wax ester hydrolase (EC:3.1.1.50) and cutinase (EC:3.1.1.74).

Selection of genes involved in exogenous fatty acid transport

To retrieve genes involved in the import of exogenous FAs in prokaryotes, we surveyed the literature of exogenous FA use by prokaryotes[38–40,58,90,91]. Based on the literature survey, we

retrieve genes belonging to the well-known fadL/fadD/fadR (K06076/K01897/K13770) system of long-exogenous FA import as well as the short-chain FA transporter fadK (K12507). We also retrieved genes for import system that were non-specific for exogenous FAs, but have been reported to facilitate the import of exogenous FAs: the porin ompF (K09476), permease mce1 (K02066) and an ABC transported (K24820). We also retrieved genes that we implicated in the funneling of exogenous FA to phospholipids: the fakA/fakB (K07030/K25232) and the aas gene (K05939).

Taxonomic assignment of neutral lipid metabolism marker genes

To assign a taxonomy to genes annotated to marker reactions of neutral lipid metabolic pathways, we first grouped genes per water column and used CD-hit [92] to dereplicate the genes at 95% identity. We searched the dereplicated set of genes against the NCBI nr database (download on August 27th 2021) using DIAMOND blastp [93]. The DIAMOND output was imported to MEGAN 95 using the January 2021 mapping file ((“megan-map-Jan2021.db”). In MEGAN, we set the lower common ancestors’ parameters at a minimum e-value of 1x10-20 and a top precent of 1%. We exported the file containing the taxonomic affiliation of the genes and processed it with a custom-made R code.

Binning

Single sample metagenomic assemblies and metagenomic co-assemblies of samples from the Canada Basin and Amundsen Gulf (Table 1) were performed to recover a greater taxonomic diversity of metagenome assembled genomes (MAGs). Co-assemblies of 24 samples from the Canada Basin, 11 samples from Canada Basin surface, 7 samples from the Amundsen Gulf, and individual assemblies of all 31 samples from the Canada Basin and Amundsen Gulf were performed. All metagenomic assemblies were generated using MEGAHIT (v.1.2.7) [94] with k-mer sizes of 27,37,47,57,67,77,87. The input metagenome reads for each assembly were mapped to the assembly with BWA (v.0.7.17) [95] using the mem option. The mapping results were processed using ”jgi _summarize _bam _contig _” depths from MetaBAT2 v.2.12.1 [96] and metagenomic binning was performed for each of the assemblies using MetaBAT2 (v.2.12.1) with default settings, resulting in 4824 genome bins. Genome quality was evaluated using CheckM v.1.0.114 [97] with the ”lineage _wf” workflow. The 924 bins with >50% completeness and <10% contamination and strain heterogeneity, were considered at least medium quality MAGs [98]. The MAGs with >50% completeness and <10% contamination and strain heterogeneity were dereplicated with dRep [99] using 95% ANI cut-off to remove species level redundancy, resulting in 664 representative MAGs. Dereplicated MAGs with >50% completeness and <10% contamination and strain heterogeneity were taxonomically classified with the GTDB-tk (v.1.3.0) [100] using the ”classify _wf” workflow.

Annotation of MAGs

We used Prokka (v.1.12) [101], implementing Prodigal (v.2.6.3) [102], to predict and annotate MAGs. INFERNAL (v.1.1.2) [103] was used to predict ribosomal RNA genes against Rfam (v.14.2) [104], KEGG, KofamScan/KofamKOALA [105] and Metacyc, using default settings and a bitscore-to-threshold ratio of 0.7.

Metabolic reconstruction in MAGs

The presence of marker genes involved in the metabolism of neutral lipids were retrieved based on their annotation with EC and KO numbers associated with markers reactions (see above). The number of genes coding for enzymes that assemble, modify and breakdown oligo- and polysaccharides in MAGs were retrieved using genes annotated with EC numbers corresponding to carbohydrate-active enzymes (CAZymes) as described in Lonbard *et al.*[106]. The number of genes coding for enzymes involved in AC degradation were retrieved in MAGs using genes annotated with dioxygenases (EC:1.13.11.-) and monooxygenases (EC:1.14.13.-) from aromatic compound degradation pathways within Metacyc (https://metacyc.org/).

Statistical tests

To compare the difference between the mean of MAG completeness for different groups we used a Welsch t-test between pairs of groups. The p-values were calculated using permutations (9999) and considered significant under 0.05. To calculate the co-occurrence between NL metabolism marker genes and FA transport genes in MAGs, we computed the Sorensen index using the presence and absence of these genes in MAGs. The p-value were calculated using 9999 permutations. P-values were considered significant < 0.05.

3.7 Acknowledgments

The data were collected aboard the CCGS Louis S. St-Laurent in collaboration with researchers from Fisheries and Oceans Canada at the Institute of Ocean Sciences and Woods Hole Oceanographic Institution's Beaufort Gyre Exploration Program and are available at http://www.whoi.edu/beaufortgyre. We would like to thank both the Captain and crew of the CCGS Louis S. St-Laurent and the scientific teams aboard.

Bibliography

[1] O. Varpe. *Life history adaptations to seasonality*. **Integr. Comp. Biol.**, 57:943–960, 2017.

[2] P. Hauschild, A. Röttig, M.H. Madkour, et al. *Lipid accumulation in prokaryotic microorganisms from arid habitats*. **Appl. Microbiol. Biotechnol.**, 101:2203–16, 2017.

[3] K. Mason-Jones, S.L. Robinson, G.F. Veen, et al. *Microbial storage and its implications for soil ecology*. **ISME J.**, pages 1–13, 2021.

[4] K. Athenstaedt and G. Daum. *The life cycle of neutral lipids: Synthesis, storage and degradation*. **Cell. Mol. Life Sci.**, 63:1355–69, 2006.

[5] L.S. Serafim, A.M.R.B. Xavier, and P.C. Lemos. *Storage of Hydrophobic Polymers in Bacteria*. **Biog. Fat. Acids, Lipids Membr.**, pages 483–507, 2019.

[6] A.F. Alvarez and D. Georgellis. *Bacterial Lipid Domains and Their Role in Cell Processes*. **Biog. Fat. Acids, Lipids Membr.**, pages 575–92, 2019.

[7] K.W Becker, J. R. Collins, B.P. Durham, et al. *Daily changes in phytoplankton lipidomes reveal mechanisms of energy storage in the open ocean*. **Nat. Commun.**, 9(1):1–9, 2018.

[8] I.A. Guschina and J.L. Harwood. *Lipids in Aquatic Ecosystems*. Springer Science-Business Media, 2009.

[9] H.M. Alvarez. *Triacylglycerol and wax ester-accumulating machinery in prokaryotes*. **Biochimie**, 120:28–39, 2016.

[10] J. Holert, K. Brown, A. Hashimi, et al. *Steryl ester formation and accumulation in steroid-degrading bacteria*. **Biochimie**, 86(2):e02353–19, 2020.

[11] V. Kumar, S. Kumar, and D. Singh. *Microbial polyhydroxyalkanoates from extreme niches: Bioprospection status, opportunities and challenges*. **Int. J. Biol. Macromol**, 147:1255–67, 2020.

[12] H.M. Alvarez, O.H. Pucci, and A. Steinbüchel. *Lipid storage compounds in marine bacteria*. **Appl. Microbiol. Biotechnol.**, 47(2):1132–9, 1997.

[13] R.Y.U. Takahashi, N.A.S. Castilho, M.A.C. da Silva, et al. *Prospecting for marine bacteria for polyhydroxyalkanoate production on low-cost substrates*. **Bioengineering**, 4(60):1132–9, 2017.

[14] L.A. Garay, Boundy-Mills K.L., and J.B. German. *Accumulation of high-value lipids in single-cell microorganisms: A mechanistic approach and future perspectives*. **J. Agric. Food Chem.**, 62: 2709–27, 2014.

[15] L. Argiz, D. Correa-Galeote, Á. Val del Rio, et al. *Valorization of lipid-rich wastewaters: A theoretical analysis to tackle the competition between polyhydroxyalkanoate and triacylglyceride-storing populations*. **Sci. Total Environ**, page 150761, 2022.

[16] S. Qadeer, A. Khalid, S. Mahmood, et al. *Utilizing oleaginous bacteria and fungi for cleaner energy production*. **J. Clean. Prod.**, 168:917–28, 2017.

[17] D. Morales-Sánchez, P.S.C. Schulze, V. Kiron, and R.H. Wijfels. *Temperature-Dependent Lipid Accumulation in the Polar Marine Microalga Chlamydomonas malina RCC2488*. **Front. Plant Sci.**, 11:619064, 2020.

[18] F. Stuart Chapin III, E-D. Schulze, and H.A. Mooney. *The ecology and economics of storage in plants*. **Annu. Rev. Ecol. Syst.**, 21:423–47, 1990.

[19] A. Matin, C. Veldhuis, V. Stegeman, and M. Veenhuis. *Selective advantage of a Spirillum sp. in a carbon-limited environment. Accumulation of poly-beta-hydroxybutyric acid and its role in starvation*.

J. Gen. Microbiol., 112:349–55, 1979.

[20] I. Poblete-Castro, I.F. Escapa, C. Jäger, et al. *The metabolic response of P. putida KT2442 producing high levels of polyhydroxyalkanoate under single- and multiple-nutrient-limited growth: Highlights from a multi-level omics approach*. **Microb. Cell Fact.**, 11:34, 2012.

[21] H.M. Alvarez, F. Mayer, D. Fabritius, and A. Steinbüchel. *Formation of intracytoplasmic lipid inclusions by Rhodococcus opacus strain PD630*. **Arch. Microbiol.**, 165:377–86, 1996.

[22] M Ardyna, C.J. Mundy, N. Mayot, et al. *Under-Ice Phytoplankton Blooms: Shedding Light on the "Invisible" Part of Arctic Primary Production*. **Front. Mar. Sci.**, 7:608032, 2020.

[23] M Ardyna and K.R. Arrigo. *Phytoplankton dynamics in a changing Arctic Ocean*. **Nat. Clim. Chang.**, 10:892–903, 2020.

[24] I. Schaud, H. Wagner, M. Graeve, and U. Karsten. *Effects of prolonged darkness and temperature on the lipid metabolism in the benthic diatom Navicula perminuta from the Arctic Adventfjorden, Svalbard*. **Polar Biol.**, 40:1425–39, 2017.

[25] R. Terrado, K. Scarcella, M. Thaler, et al. *Small phytoplankton in Arctic seas: vulnerability to climate change*. **Biodiversity**, 14(1):2–18, 2013.

[26] M.S. Yun, H.T. Joo, J.W. Park, et al. *Lipid-rich and protein-poor carbon allocation patterns of phytoplankton in the northern Chukchi Sea, 2011*. **Cont. Shelf Res.**, 158:26–32, 2018.

[27] P.S. Bhavya, B.K. Kim, N. Jo, et al. *A Review on the Macromolecular Compositions of Phytoplankton and the Implications for Aquatic Biogeochemistry*. **Ocean Sci. J.**, 154(1):1–14, 2019.

[28] D. Kohlbach, M. Graeve, B.A. Lange, et al. *The importance of ice algae-produced carbon in the central Arctic Ocean ecosystem: Food web relationships revealed by lipid and stable isotope analyses*. **Limnol. Oceanogr.**, 61:2027–44, 2016.

[29] S. Falk-Petersen, J.R. Sargent, J. Henderson, et al. *Lipids and fatty acids in ice algae and phytoplankton from the Marginal Ice Zone in the Barents Sea*. **Polar Biol.**, 20:41–7, 1998.

[30] K.R. Arrigo and G.L. van Dijken. *Secular trends in Arctic Ocean net primary production*. **J. Geophys. Res. Ocean**, 116:C09011, 2011.

[31] J. Berge, M. Daase, P.E. Renaud, et al. *Unexpected Levels of Biological Activity during the Polar Night Offer New Perspectives on a Warming Arctic*. **Curr. Biol.**, 25:2555–61, 2015.

[32] J. Berge, P.E. Renaud, G. Darnis, et al. *In the dark: A review of ecosystem processes during the Arctic polar night*. **Prog. Oceanogr.**, 139:258–71, 2015.

[33] L. Zifcakova, T. Vetrovsky, V. Lombard, et al. *Feed in summer, rest in winter: microbial carbon utilization in forest topsoil*. **Microbiome**, 5:122, 2017.

[34] M.M. Yakimov, K.N. Timmis, and P.N. Golyshin. *Obligate oil-degrading marine bacteria*. **Curr. Opin. Biotechnol.**, 18:257–66, 2007.

[35] A. Dahlqvist, U. Stahl, M. Lenman, et al. *Phospholipid:diacylglycerol acyltransferase: An enzyme that catalyzes the acyl-CoA-independent formation of triacylglycerol in yeast and plants*. **Proc. Natl. Acad. Sci. U. S. A.**, 97(12):6487–92, 2000.

[36] M. Wälterman, A. Hinz, H. Robenek, et al. *Mechanism of lipid-body formation in prokaryotes: how bacteria fatten up*. **Mol. Microbiol.**, 55(3):750–63, 2005.

[37] A.C. Turchetto-Zolet, F.S. Maraschin, G.L. de Morais, et al. *Evolutionary view of acyl-CoA diacylglycerol acyltransferase (DGAT), a key enzyme in neutral lipid biosynthesis*. **BMC Evol. Biol.**, 11: 263, 2011.

[38] J.M. Salvador López and I.N.A. Van Bogaert. *Microbial fatty acid transport proteins and their biotechnological potential*. **Biotechnol. Bioeng.**, 118:2184–2201, 2021.

[39] J. Yao and C.O. Rock. *Exogenous fatty acid metabolism in bacteria*. **Biochimie**, 2017.

[40] J. Yao and C.O. Rock. *How Bacterial Pathogens Eat Host Lipids: Implications for the Development of Fatty Acid Synthesis Therapeutics*. **J. Biol. Chem.**, 290(10):5940–6, 2015.

[41] E. Manilla-Pérez, A.B. Lange, S. Hetzler, and A. Steinbüchel. *Occurrence, production, and export of lipophilic compounds by hydrocarbonoclastic marine bacteria and their potential use to produce bulk chemicals from hydrocarbons*. **Appl. Microbiol. Biotechnol.**, 86:1693–1706, 2010.

[42] C.C.C.R. de Carvalho and M.J. Caramujo. *Lipids of Prokaryotic Origin at the Base of Marine Food Webs*. **Mar. Drugs**, 10:2698–2714, 2012.

[43] A. Röttig, P. Hauschild, M.H. Madkour, et al. *Analysis and optimization of triacylglycerol synthesis in novel oleaginous Rhodococcus and Streptomyces strains isolated from desert soil*. **J. Biotechnol.**, 225:48–56, 2016.

[44] M. Salmela, T. Lehtinen, E. Efimova, et al. *Alkane and wax ester production from lignin-related aromatic compounds*. **Biotechnol. Bioeng.**, 116:1934–45, 2019.

[45] M.C. Posewitz. *Engineering pathways to biofuels in photoautotrophic microorganisms*. **Biofuels**, 5(1):67–78, 2014.

[46] S. Wagner, F. Schubotz, K. Kaiser, et al. *Soothsaying DOM: A Current Perspective on the Future of Oceanic Dissolved Organic Carbon*. **Front. Mar. Sci**, 7:341, 2020.

[47] Y. Zhuang, H. Jin, W-J. Cai, et al. *Freshening leads to a three-decade trend of declining nutrients in the western Arctic Ocean*. **Environ. Res. Lett.**, 16:054047, 2021.

[48] T.B. Bittar, Y. Lin, L.R. Sassano, and B.J. Wheeler. *Carbon allocation under light and nitrogen resource gradients in two model marine phytoplankton*. **J. Phycol.**, 49:523–35, 2013.

[49] N. Joli, A. Monier, R. Logares, and C. Lovejoy. *Seasonal patterns in Arctic prasinophytes and inferred ecology of Bathycoccus unveiled in an Arctic winter metagenome*. **ISME J.**, 11:1372–85, 2017.

[50] T. Mock and B.M.A. Kroon. *Photosynthetic energy conversion under extreme conditions - II: The significance of lipids under light limited growth in Antarctic sea ice diatoms*. **Phytochemistry**, 61: 53–60, 2002.

[51] N.J. Freyria, N. Joli, and C. Lovejoy. *A decadal perspective on north water microbial eukaryotes as Arctic Ocean sentinels*. **Sci. Rep.**, 11:8413, 2021.

[52] I. Benner, A.J. Irwin, and Z.V. Finkel. *Capacity of the common Arctic picoeukaryote Micromonas to adapt to a warming ocean*. **Limnol. Oceanogr. Lett.**, 5:221–7, 2020.

[53] D.C.O. Thornton. *Dissolved organic matter (DOM) release by phytoplankton in the contemporary and future ocean*. **Eur. J. Phycol.**, 49(1):20–46, 2014.

[54] S.M. Short. *The ecology of viruses that infect eukaryotic algae*. **Environ. Microbiol.**, 14(9):2253–71, 2012.

[55] C. Lonborg, M. Middelboe, and C.P.D. Brussaard. *Viral lysis of Micromonas pusilla: impacts on dissolved organic matter production and composition*. **Biogeochemistry**, 116:231–40, 2013.

[56] N. Meyer, A. Bigalke, A. KaulfuB, and G. Pohnert. *Strategies and ecological roles of algicidal bacteria*. **FEMS Microbiol. Rev.**, 41:880–99, 2017.

[57] B. Barati, F.F. Zafar, and P.F. Rupani. *Bacterial pretreatment of microalgae and the potential of novel nature hydrolytic sources*. **Environ. Technol. Innov.**, 21:101362, 2021.

[58] W. Yao, K. Liu, H. Liu, et al. *A Valuable Product of Microbial Cell Factories: Microbial Lipase*. **Front. Microbiol.**, 12:743377, 2021.

[59] M. Goutx, C. Guige, and L. Striby. *Triacylglycerol biodegradation experiment in marine environmental conditions: definition of a new lipolysis index*. **Org. Geochem.**, 34:1465–73, 2003.

[60] M. Suzuki, Y. Tachibana, and K-I. Kasuya. *Biodegradability of poly(3-hydroxyalkanoate) and poly(ϵ-caprolactone) via biological carbon cycles in marine environments*. **Polym. J.**, 53:47–66, 2021.

[61] V.R. Vijakainen and L.A. Hug. *The phylogenetic and global distribution of bacterial polyhydroxyalkanoate bioplastic-degrading genes*. **Environ. Microbiol.**, 22(3):1717–31, 2021.

[62] M. Dourou, A. Aggeli, S. Papanikolaou, and G. Aggelis. *Critical steps in carbon metabolism affecting lipid accumulation and their regulation in oleaginous microorganisms*. **Appl. Microbiol. Biotechnol.**, 102:2509–23, 2018.

[63] D. Jendrossek. *Bacterial Organelles and Organelle-like Inclusions*. Springer, 2020.

[64] M. Wältermann, T. Stöveken, and A. Steinbüchel. *Key enzymes for biosynthesis of neutral lipid storage compounds in prokaryotes: Properties, function and occurrence of wax ester synthases/acyl-CoA:diacylglycerol acyltransferases*. **Biochimie**, 89:230–42, 2007.

[65] R. Kalscheuer. *Handbook of Hydrocarbon and Lipid Microbiology, Chapter 40: Genetics of Wax Ester and Triacylglycerol Biosynthesis in Bacteria*. Springer, 2010.

[66] B. Bredemeier, R. Hulsch, J.O. Metzger, and L. Berthe-Corti. *Submersed Culture Production of Extracellular Wax Esters by the Marine Bacterium Fundibacter jadensis*. **Mar. Biotechnol.**, 5: 579–83, 2003.

[67] J-F. Rontani, P.C. Bonin, and J.K. Volkman. *Production of Wax Esters during Aerobic Growth of Marine Bacteria on Isoprenoid Compounds*. **Appl. Environ. Microbiol.**, 65(1):221–30, 1999.

[68] V. Cafaro, V. Izzo, E. Notomista, and A. Di Donato. *Marine Enzymes for Biocatalysis: Sources, Biocatalytic Characteristics and Bioprocesses of Marine Enzymes, Chapter 14: Marine hydrocarbonoclastic bacteria*. Woodhead publishing, 2013.

[69] T. Ahrendt, H. Wolff, and H.B. Bode. *Neutral and Phospholipids of the Myxococcus xanthus Lipodome during Fruiting Body Formation and Germination*. **Appl. Environ. Microbiol.**, 81(19): 6538–47, 2015.

[70] T. Shiratori, S. Suzuki, Y. Kakizawa, and K-i. Ishida. *Phagocytosis-like cell engulfment by a planctomycete bacterium*. **Nat. Commun.**, 10:5529, 2019.

[71] P.K. Obulisamy and S. Mehariya. *Polyhydroxyalkanoates from extremophiles: A review*. **Bioresour. Technol.**, 325:124653, 2021.

[72] N.D. Ayub, P.M. Tribelli, and N.I. López. *Polyhydroxyalkanoates are essential for maintenance of redox state in the Antarctic bacterium Pseudomonas sp. 14-3 during low temperature adaptation*. **Extremophiles**, 13:59–66, 2009.

[73] M.J. Ayub, N.D. Pettinari, B.S. Méndez, and N.I. López. *The polyhydroxyalkanoate genes of a stress resistant Antarctic Pseudomonas are situated within a genomic island*. **Plasmid**, 58:240–8, 2007.

[74] M.M. Mills, Z.W. Brown, S.R. Laney, et al. *Nitrogen Limitation of the Summer Phytoplankton and Heterotrophic Prokaryote Communities in the Chukchi Sea*. **Front. Mar. Sci.**, 5:362, 2018.

[75] L.G. Anderson and R.M.W. Amon. *Biogeochemistry of Marine Dissolved Organic Matter: Second Edition, Chapter 14: DOM in the Arctic Ocean*. Elsevier, 2014.

[76] J. Jung, J.E. Son, Y.K. Lee, et al. *Tracing riverine dissolved organic carbon and its transport to the halocline layer in the Chukchi Sea (western Arctic Ocean) using humic-like fluorescence fingerprinting*.

Sci. Total Environ., 772:145542, 2021.

[77] D. Colatriano, P.Q. Tran, C. Guéguen, et al. *Genomic evidence for the degradation of terrestrial organic matter by pelagic Arctic Ocean Chloroflexi bacteria*. **Commun. Biol.**, 1:90, 2018.

[78] S. Acinas, P. Sánchez, G. Salazar, et al. *Deep ocean metagenomes provide insight into the metabolic architecture of bathypelagic microbial communities*. **Commun. Biol.**, 4:604, 2021.

[79] R. Stocker and J.R. Seymour. *Ecology and Physics of Bacterial Chemotaxis in the Ocean*. **Microbiol. Mol. Biol. Rev.**, 76(4):792–812, 2012.

[80] M-E. Garneau, S. Roy, C. Lovejoy, et al. *Seasonal dynamics of bacterial biomass and production in a coastal arctic ecosystem: Franklin Bay, western Canadian Arctic*. **J. Geophys. Res. Ocean.**, 113: C07S91, 2008.

[81] D.K. Stoecker and P.J. Lavrentyev. *Mixotrophic plankton in the polar seas: A pan-Arctic review*. **Front. Mar. Sci.**, 5:292, 2018.

[82] R.W. Sanders and R.J. Gast. *Bacterivory by phototrophic picoplankton and nanoplankton in Arctic waters*. **FEMS Microbiol. Ecol.**, 82:242–53, 2012.

[83] F. Unrein, R. Massana, L. Alonso-Saéz, and J.M. Gasol. *Significant year-round effect of small mixotrophic flagellates on bacterioplankton in an oligotrophic coastal system*. **Limnol. Oceanogr.**, 52(1):456–69, 2007.

[84] D. Vaque, O. Guadayol, F. Peters, et al. *Seasonal changes in planktonic bacterivory rates under the ice-covered coastal Arctic Ocean*. **Limnol. Oceanogr.**, 53(6):2427–38, 2008.

[85] D. Colatriano and D.A. Walsh. *An aquatic microbial metaproteomics workflow: From cells to tryptic peptides suitable for tandem mass spectrometry-based analysis*. **J. Vis. Exp.**, 103:e52827, 2015.

[86] A. Bankevich, S. Nurk, D. Antipov, et al. *SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing*. **J. Comput. Biol.**, 19(5):455–77, 2012.

[87] M. Huntemann, N.N. Ivanova, K. Mavromatis, et al. *The standard operating procedure of the DOE-JGI Metagenome Annotation Pipeline (MAP v . 4)*. **Stand. Genomic Sci**, 1:1–5, 2016.

[88] A. Milanese, D.R. Mende, L. Paoli, et al. *Microbial abundance, activity and population genomic profiling with mOTUs2*. **Nat. Commun.**, 10:1014, 2019.

[89] S. Sunagawa, D.R. Mende, G. Zeller, et al. *Metagenomic species profiling using universal phylogenetic marker genes*. **Nat. Methods**, 10(12):1196–9, 2013.

[90] K.N. Timmis, M. Boll, O. Geiger, et al. *Aerobic Utilization of Hydrocarbons , Oils , and Lipids*. Springer, 2018.

[91] M. Rodríguez-Moya and R. Gonzalez. *Proteomic analysis of the response of Escherichia coli to short-chain fatty acids*. **J. Proteomics**, 122:86–99, 2015.

[92] W. Li and A. Godzik. *Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences*. **Bioinformatics**, 22(13):1658–9, 2006.

[93] C. Bagci, S. Patz, and D.H. Huson. *DIAMOND+MEGAN: Fast and Easy Taxonomic and Functional Analysis of Short and Long Microbiome Sequences*. **Current Protocols**, 1(3):1–29, 2021.

[94] D. Li, C-M. Liu, R. Luo, et al. *MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph*. **Bioinformatics**, 31(10):1674–6, 2015.

[95] H. Li and R. Durbin. *Fast and accurate short read alignment with Burrows–Wheeler transform*. **Bioinformatics**, 25(14):1754–60, 2009.

[96] D.D. Kang, F. Li, E. Kirton, et al. *MetaBAT 2: An adaptive binning algorithm for robust and*

efficient genome reconstruction from metagenome assemblies. **PeerJ.**, 7:e7359, 2019.

[97] D. Parks, M. Imelfort, C.T. Skennerton, et al. *CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes*. **Genome Res.**, 25:1043–55, 2015.

[98] R.M. Bowers, N.C. Kyrpides, R. Stepanauskas, et al. *Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea*. **Nat. Biotechnol.**, 35(8):725–31, 2017.

[99] M.R. Olm, C.T. Brown, B. Brooks, and J.F. Banfield. *dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication*. **ISME J.**, 11:2864–8, 2017.

[100] P.A. Chaumeil, A.J. Mussig, P. Hugenholtz, and D.H. Parks. *GTDB-Tk: A toolkit to classify genomes with the genome taxonomy database*. **Bioinformatics**, 36(6):1925–7, 2020.

[101] T. Seemann. *Prokka: Rapid prokaryotic genome annotation*. **Bioinformatics**, 30(14):2068–9, 2014.

[102] D. Hyatt, G-L. Chen, P.F. LoCascio, et al. *Integrated nr Database in Protein Annotation System and Its Localization*. **Nat. Commun.**, 6(1):1–8, 2010.

[103] E.P. Nawrocki and S.R. Eddy. *Infernal 1.1: 100-fold faster RNA homology searches*. **Bioinformatics**, 29(22):2933–5, 2013.

[104] I. Kavalri, E.P. Nawrocki, N. Ontiveros-Palacios, et al. *Rfam 14: Expanded coverage of metagenomic, viral and microRNA families*. **Nucleic Acids Res.**, 49:192–200, 2021.

[105] T. Aramaki, R. Blanc-Mathieu, H. Endo, et al. *KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold*. **Bioinformatics**, 36(7):2251–2, 2020.

[106] V. Lombard, H.G. Ramulu, E. Drula, et al. *The carbohydrate-active enzymes database (CAZy) in 2013*. **Nucleic Acids Res..**, 42:D490–D495, 2014.

phosphorus

Chapter 4

Metagenomics and molecular composition of the dissolved organic matter reveal Arctic particularities among the metabolism of the global ocean microbiome

4.1 Abstract

Background: The Arctic is subjected to a strong seasonal cycle with acute variations in temperatures, ice and light regime. In addition, with its important watershed, the organic matter of the Arctic Ocean has a strong terrestrial signature. The Arctic Oceans therefore possesses a characteristic physicochemical environment. The diversity of microbial metabolic processes are key actors in shaping the geochemical cycles by transforming a wide variety of organic and inorganic compounds. However, despite their important roles, the biogeography and phylogenetic distribution of metabolic processes across microbial taxa as well as how they are linked to the physicochemical environment remain largely unexplored. This information is especially relevant given the fast changes in the biogeochemical cycles of the Arctic Ocean.

Results: In this study, we undertook to systematically unravel the diversity of metabolic processes favoured by the Arctic Ocean microbiomes and link them to the composition of the dissolved organic matter. Using global ocean comparative metagenomics, we discovered that the microbiomes of the polar oceans' photic zone were metabolically distinct than in temperate oceans. Microbiomes from the polar ocean photic zone demonstrated a strong eukaryotic signal. In addition, we uncovered that these microbiomes favoured the metabolism of lipids. The most striking metabolic feature of the polar ocean photic zone's microbiomes, however, was the prevalence of microbial genes and pathways involved in the metabolism of glycans that might reflect their role in cold adaptation. Importantly, the biogeography of glycan genes and pathways corresponded to an enrichment of transformations involving sugar moieties in the Arctic Ocean photic zone. In addition, we found that the metabolism of the Arctic subsurface water microbiomes was distinct to the rest of the global ocean. The main distinguishing feature was the prevalence of aromatic compound degradation genes that was concurrent with a strong aromaticity signal in the organic matter of the Arctic Ocean subsurface waters.

Conclusion: Overall, for the first time, this study systematically uncovers the diversity of metabolic processes favoured by the microbiomes of the Arctic Ocean. In addition, we demonstrated that these distinct metabolic processes are linked to the composition of the Arctic Ocean dissolved organic matter. This study is a cornerstone in highlighting metabolic processes of importance for the Arctic Ocean microbiomes and will serve as a guide for future work investigating in detail the evolutionary origin and the role of these microbial processes in shaping the biogeochemical cycles of the Arctic Ocean.

4.2 Introduction

The biogeochemistry of the Arctic Ocean is distinct from other oceans. With a watershed area that is twice the size of its area[1], the Arctic Ocean is heavily influenced by land. The Arctic Ocean receives 11% of global freshwater input, while representing only 1% of the global ocean volume[2]. Coupled to ice melting, this creates a relatively fresh surface mixed layer. Fresh water input also imparts the Arctic Ocean organic matter with the highest terrestrial signature within the global ocean[2]. The Arctic Ocean is also characterized by extreme seasonal dynamics, creating strong variations in physicochemical conditions[3]. During the winter, the very low solar radiation of the polar night is blocked by a thick ice layer[4]. In the spring, the increasing light levels, coupled with ice melt and increasing riverine input provide both the energy and nutrients to trigger phytoplankton blooms[5]. The phytoplankton activity decreases during summer and smaller blooms generally appear during the fall[5]. The impact of land and the unique seasonal dynamics of the Arctic Ocean result in a distinct physicochemical microbial habitat.

The metabolic capacity of microbial communities reflects their ability to transform and process the inorganic and organic matter they find in their environment[6,7]. As a consequence, the physicochemical landscape and microbial metabolism influence each other[8]. In previous chapters, we focused on two microbial metabolisms within the Arctic Ocean, specifically the degradation of aromatic compounds (chapter 2) and the production and degradation of neutral lipids (chapter 3). We demonstrated the prevalence of theses metabolic processes in Arctic Ocean microbiomes compared to the rest of the global ocean, as well as unraveled the phylogenetic diversity of organisms involved in these processes. In addition, we discussed how the prevalence of these metabolic processes was linked to the unique biogeochemical setup of the Arctic Ocean. However, despite these examples of distinct metabolic features in Arctic Ocean microbiomes, and the importance of microbial metabolism in shaping the ocean biogeochemical cycles, we still lack a global view of the ensemble of metabolic processes favoured by microbial communities of the Arctic Ocean compared to the rest of the global ocean and how they are linked to their physicochemical environment.

In this final research chapter, we aimed to fill this knowledge gap by characterizing the biogeographical partitioning of a much broader range of metabolic processes involved in the

metabolism of inorganic and organic matter in the global ocean microbiomes. These metabolic processes include all the genes and pathways involved in the use and production of inorganic (*e.g.* CO_2, ammonia) and organic (*e.g.* proteins, glycans, humic substances) matter as well as genes and pathways involved in the physical access to inorganic and organic matter (*e.g.* transporters, motility, biofilm formation). In addition, we sought to link the genes and pathways involved in the metabolism of inorganic and organic matter to the composition of DOM that the microbes are producing, consuming and transforming in the Arctic Ocean.

In chapter 2 and 3, we used NMF (non-negative matrix factorization) on metagenomes to identify metabolic genes and pathways specifically associated with each water column feature within the Arctic Ocean and studied their phylogenetic distribution across bacteria. In this chapter we now extend the statistical method to encompass a wider range of genes and pathways involved in the metabolism of DOM and dissolved inorganic matter (DIM) in the prokaryotic and picoeukaryotic communities of the global ocean. We found metabolic features common to northern and southern polar surface waters that were distinct from the surface waters at lower latitudes. Polar surface waters were characterized by a signal for eukaryotic pathways. In addition, glycan and lipid metabolism pathways were characteristic of polar surface waters. In general, the deep water of the Arctic Ocean was metabolically similar to the mesopelagic zone of other oceans, but we found a unique metabolic signature in the subsurface layers of the Arctic Ocean. The unique metabolic signature of the Arctic Ocean subsurface was centred around the maximum of fluorescent dissolved organic matter and composed of pathways involved in the degradation of aromatic compounds, confirming and strengthening the previous findings reported in chapter 2.

We further aimed to link metabolic capacities encoded in metagenomes identified as characteristic to the Arctic Ocean's microbiomes to the molecular composition of DOM in the Arctic Ocean. Fourier-transform ion cyclotron resonance mass spectrometry (FT ICR MS) can retrieve the water DOM composition of various samples at the molecular level by inferring the molecular formula of DOM compounds based on the mass of peaks in the mass spectra[9,10]. Analysis of mass difference between peaks of the mass spectra can also reveal reactivity between DOM compounds[11]. Through the combination of metagenomics and FT ICR MS, we identified a close relationship between the genes and pathways favoured by the microbiomes and the molecular composition of DOM in the Arctic Ocean. The molecular composition of the DOM in the Arctic Ocean was vertically stratified, paralleling the vertical structure of microbial metabolism. Specifically, in the subsurface waters we highlighted a strong aromaticity of the DOM that corresponded to the enrichment of aromatic compound degradation genes while in the surface, we observed an enhanced reactivity involving sugar moieties between DOM compounds, corresponding to the prevalence of glycan metabolism genes and pathways.

Altogether, this study reveals that within the global ocean, the unique biogeochemical environment of the Arctic Ocean shapes distinct metabolic features in the microbial communities that are linked to the molecular composition of DOM. Ultimately, this study will serve as a basis to explore individual metabolic processes of importance in the Arctic Ocean, phylogenetic

diversity of the microbes involved in these processes, and how these metabolic processes shape the Arctic Ocean biogeochemical cycles.

4.3 Results

4.3.1 Biogeography of the global ocean microbiome's metabolism

We sought to determine how microbial metabolism was distributed across the global ocean. As a representation of the global ocean microbiomes, we collected publicly available metagenomes from the Pacific, Atlantic, Indian and Southern Oceans as well as the Red Sea and Mediterranean Sea. Metagenomes represented surface, subsurface chlorophyll maximum, mesopelagic and bathypelagic waters. The global dataset was analyzed along with the metagenomes we sampled in the Canada Basin (Figure 4.1a). To generate metabolic profiles genes were assigned to protein families using the KEGG orthology (KO) database. The KO abundances within metagenome was calculated as the sum of coverage of gene sets assigned to each KO. The abundance was normalized by the median abundance of 10 conserved single-copy marker genes to get a final per-cell KO abundance profile. To capture the genomic capacity of the global ocean microbiome to metabolize organic and inorganic, we focused our analysis on the KO involved in metabolic pathways. As a result, we obtained a matrix of KO profiles over the metagenomes sampled across the global ocean.

We analyzed the KO profiles from the compiled global ocean metagenomic dataset using NMF analysis. The NMF revealed that six SMGs optimally described the biogeography of the global ocean microbial metabolism (Figure 4.2a-b). Two SMGs contributed importantly to the samples of the photic zone (surface and SCM) across the global ocean. The first SMG contributed most to the surface and subsurface chlorophyll maximum (SCM) (Figure 4.1b) of temperate oceans and decreased with increasing latitude (Figure 4.3), and hence was called temperate photic zone SMG. In contrast, the second SMG contributed most to the polar ocean surface and SCM samples (Figure 4.1b) and decreased from the poles to the equator (Figure 4.3), and was coined polar photic zone SMG. The polar photic zone SMG decreased sharply with temperature and salinity while the temperate photic zone SMG increased with both salinity and temperature (Figure 4.3). Interestingly, the polar photic zone SMG increased sharply with oxygen concentration, while the temperate photic zone SMG was maximum around an oxygen optimum of 200 μmol.kg^{-1}(Figure 4.3).

We identified four SMGs whose contribution was most important in samples of the dark ocean. The SMGs whose contribution was the highest in the mesopelagic samples and the deep samples were named global mesopelagic SMG and global deep SMG respectively. Interestingly the global deep SMG barely contributed to the deep waters of the Arctic Ocean (Figure 4.1b). Instead, the global mesopelagic SMG was the main contributor in these waters (Figure 4.1b). Our analysis also revealed an Antarctic deep SMG, whose contribution was mostly restricted to the deep waters of the Antarctic Ocean. The last SMG was named the Arctic subsurface

SMG as its contribution was mostly restricted to Arctic waters found below the surface. Its contribution peaked in the fluorescent dissolved organic matter maximum (FDOMmax) but was also important in the Arctic SCM and deep waters (Figure 4.1b).

Altogether, these results show that in the global ocean, the metabolism of the surface water microbiomes is partitioned between the poles and temperate oceans, while the mesopelagic microbiomes metabolism is homogeneous across the globe. Importantly, these results also demonstrate that the microbiomes of the Arctic Ocean are characterized by particular metabolic capacities all throughout the water column.

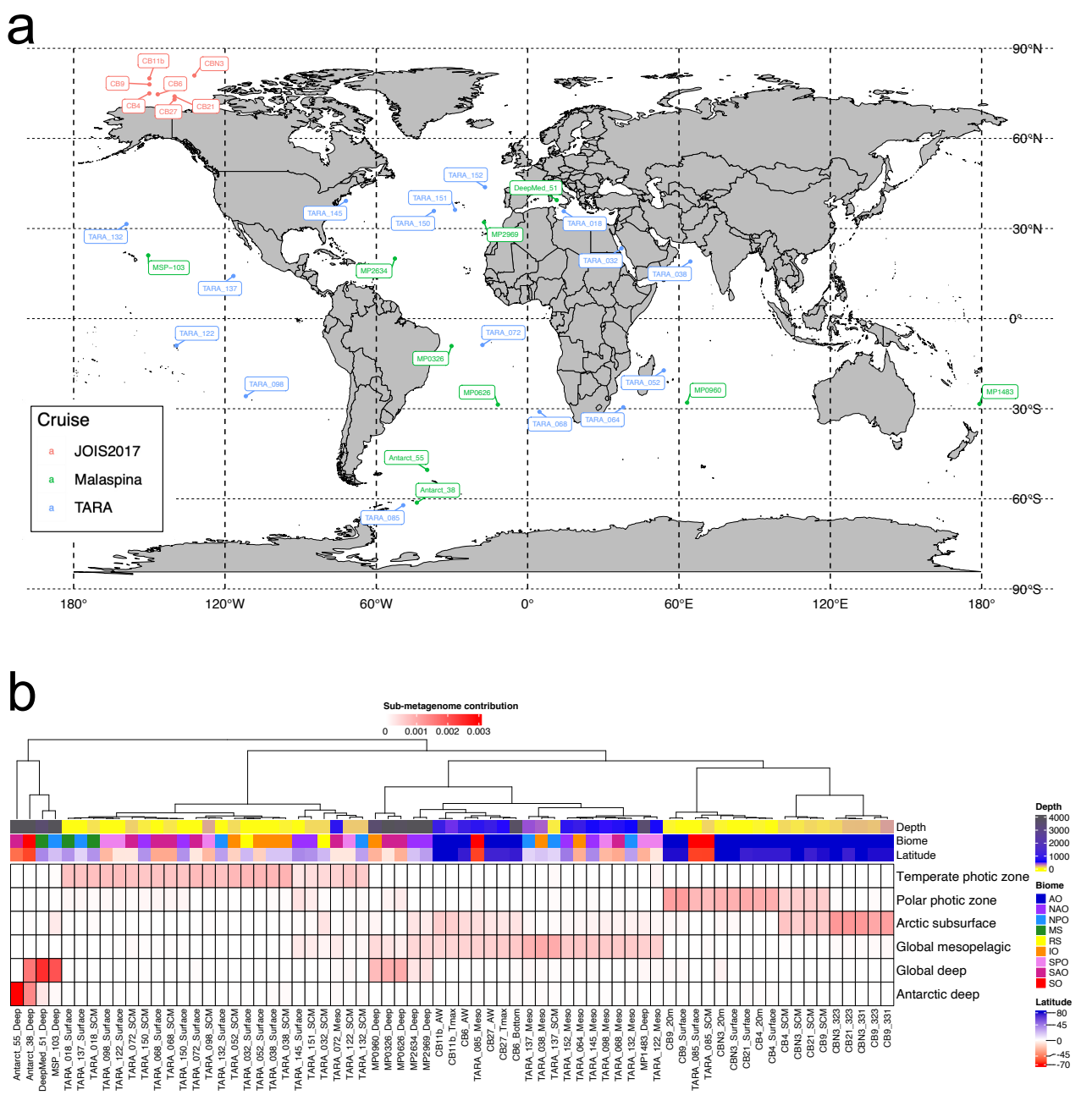

Figure 4.1: *Biogeography of the global ocean microbiomes's metabolism. a) Map of the stations sampled for metagenomes in this study (red) and retrieved from IMG/JGI (green and blue). b) Contribution of the sub-metagenomes to the global ocean metagenomes.*

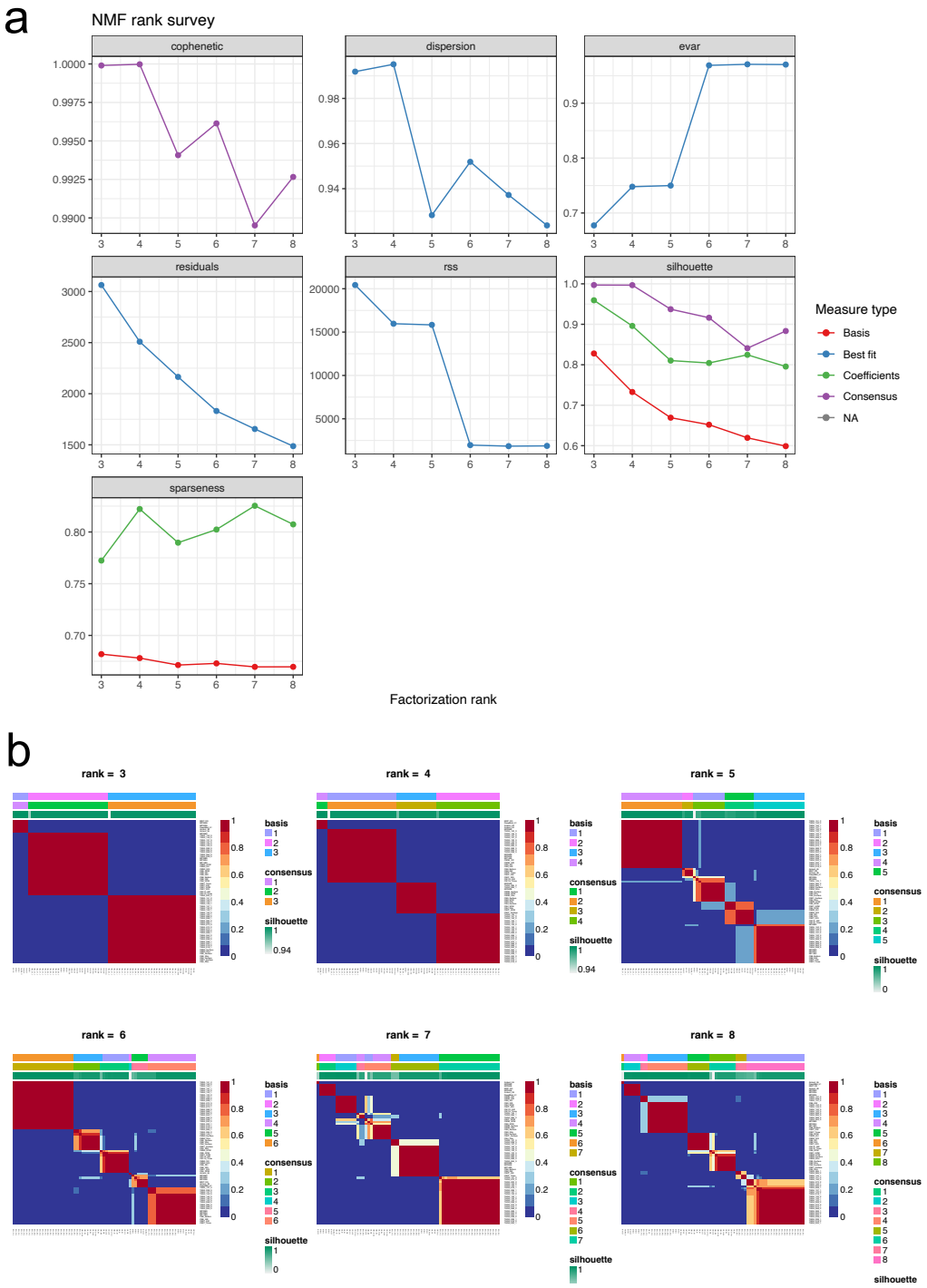

Figure 4.2: *a) Evolution of various metrics used to quantify the optimal rank to be used for the non-negative matrix factorization analysis. Cophenetic correlation represents the correlation between the sample distances from the consensus matrix and the cophenetic distance between these samples when they are clustered. The dispersion is defined as 1-rss/$\Sigma_{i,j}(V_{i,j})^2$ (Vi,j are the entries of the KO abundance matrix). Evar estimates the fraction of variance of the KO abundance matrix explained by the NMF results. Residuals is the sum of residuals between the original KO abundance matrix and the matrix estimated using the NMF. The rss is the residual sum of squares between the original KO abundance matrix and its estimate using the NMF algorithm. Sparseness is equal to 1 if all the elements of a vector are null but for 1. Oppositely, the sparseness is equal to 0 is all the element of a vector are equals. The sparseness of the basis and coefficient matrices are calculated as the mean sparseness of its element vectors. b) Consensus matrices obtained from non-negative matrix factorization of the global ocean KO abundance matrix using various rank values and 100 runs.*

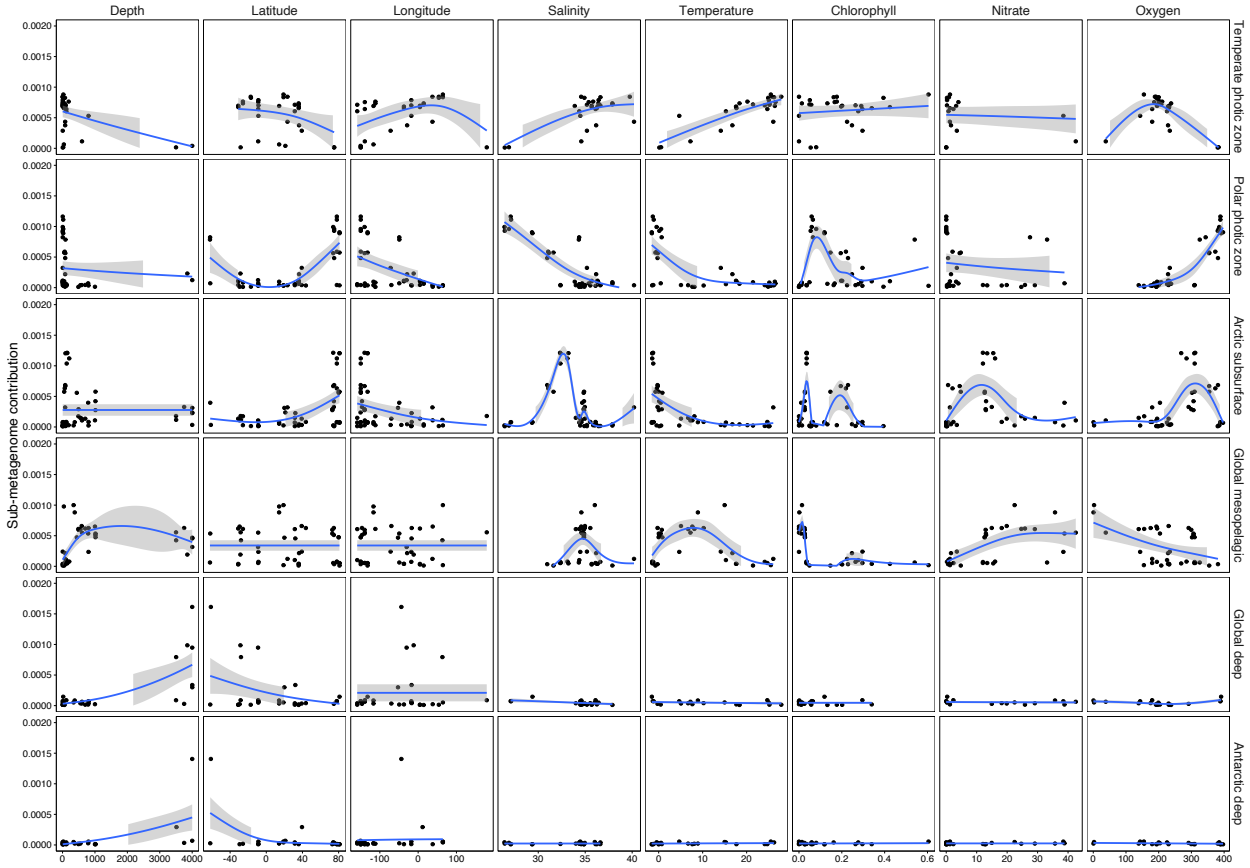

Figure 4.3: *Evolution of the contribution of the six sub-metagenomes along physicochemical gradients*

4.3.2 Metabolic characteristics of the global ocean sub-metagenomes

Based on the partitioning of the global ocean microbial metabolism into biogeography-related SMGs, we sought to determine which KOs and associated pathways differentiated the six SMGs. We calculated a KO index that quantified the specificity of a KO for each of the six SMGs. This index ranges from -1 (a KO is not found in a SMG and equally represented in all the other SMGs), to 0 (a KO is equally represented across all SMGs), to 1 (a KO is represented in only one SMG). In addition, using the KO index allowed us to calculate a median index per KO pathway and KO modules for each of the six SMGs.

As expected from sunlit photic zone of the temperate global ocean, most of KOs, KEGG modules and KEGG pathways with highest indices in the temperate photic zone SMG were involved in photosynthesis (Figure 4.4, 4.5-4.5). Within the KOs with the 50 highest indices (Figure 4.4), we found many protein families constituting photosystems I and II (4 and 6 respectively), as well as photosynthetic pigments (phycoerythrin) and protein families involved in the transformation of these pigments and their intermediates (e.g. protochlorophyllide). Thirteen protein families (NAD(P)H-quinone oxidoreductase, cytochromes) involved in the oxidative phosphorylation of photosynthetic microorganisms (cyanobacteria, chloroplasts of photosynthetic eukaryotes) were also represented in the KOs with the top 50 indices within the temperate photic zone SMG (Figure 4.4).

In the polar photic zone SMG, we found a strong eukaryotic signal. In the KEGG modules with the highest 50 median indices, there were 10 modules specific to eukaryotes or eukaryotic organelles (mitochondria) (Figure 4.5). Triacylglycerol biosynthesis and acylglycerol degradation were in the 50 highest ranking KEGG modules (Figure 4.5). We also found their key enzymes (phospholipid:diacylglycerol acyltransferase - K00679 and degradation TAG lipase - K14674) within the 50 highest ranking protein families (Figure 4.4), confirming the importance of NL metabolism in the Arctic Ocean microbiomes (chapter 3). But we also found additional lipid metabolism modules and pathways in the top 50 ranking, specifically those involved in fatty acid biosynthesis (modules: fatty acid metabolism in mitochondria, fatty acid elongation in endoplasmic reticulum; pathways: fatty acid elongation, biosynthesis of unsaturated fatty acid, linoleic and alpha-linoleic acid metabolism) and degradation (beta-oxidation).

The most striking feature of the polar photic zone SMG was the prevalence of protein families, modules and pathways involved in the metabolism of glycans (polysaccharides). We found 15 protein families, 9 KEGG modules and 12 KEGG pathways involved in glycan metabolism within the top 50 ranking in the polar photic zone SMG. The biosynthesis of N- and O-glycans were highly represented, but the 2 highest scoring protein families (K12977, K03949) were involved in the biosynthesis of the lipopolysaccharide system (Kdo-lipid A) (Figure 4.4). The biosynthesis of glycosylphosphatidylinositol (GPI)-anchor, a cell membrane structure that uses glycan to anchor protein to a phosphatidylinositol unit also emerged as a particularity of the polar photic zone SMG. Interestingly, other than glycan metabolism, we found protein families involved in the synthesis of photosynthetic pigments (K15747, K09839) and the modules involved in the synthesis of photosynthetic pigments biosynthesis regulator (jasmonic acid biosynthesis) within the top 50 ranking.

The global ocean comparative analysis recovered and strengthened our previous observations from chapter 2 on the enrichment of aromatic compound degradation genes in the Arctic Ocean within the global ocean. Indeed, the Arctic subsurface SMG was overwhelmingly dominated by aromatic compound degradation metabolism. Out of the 50 highest ranking protein families in the Arctic subsurface SMG, 23 were involved in aromatic compound degradation (Figure 4.4). This aromatic compound degradation signal was also detected by the prevalence of aromatic compound degradation KEGG modules (Figure 4.5) (11 out of 50 highest ranking modules) and KEGG pathways (Figure 4.6) (16 out of 50 highest ranking pathways - under the Xenobiotics biodegradation and metabolism pathway family) in the Arctic subsurface SMG.

In the mesopelagic ocean, the sources of energy and organic carbon are limited. The metabolic features of the global mesopelagic SMG reflected these limitations. KEGG modules for the use of inorganic substrates such as ammonia as energy sources (nitrification, complete nitrification, ammonia oxidation) were among the highest-ranking modules in the global mesopelagic SMG (Figure 4.5). Consequently, we found the methane/ammonia monooxygenases subunit A and B in the 50 highest ranking protein families of the global mesopelagic SMG (Figure 4.5). Similarly, we found a strong signal for the non-phototrophic carbon fixa-

tion metabolism in the global mesopelagic SMG. We found three KEGG modules of carbon fixation in the 50 highest ranking modules (Figure 4.6) (incomplete reductive citrate cycle, hydroxypropionate-hydroxybutyrate cycle and reductive acetyl-CoA pathway, reductive citrate cycle). Accordingly, 7 protein families involved in carbon fixation pathways (K18954, K18953, K18602, K18603, K18604, K18605, K15019) were among the 50 highest ranking protein families in the global mesopelagic SMG.

Both global deep and Antarctic deep SMGs were characterized by high median indices of KEGG pathways for cellular motility and biofilm formation. However, these two SMGs also showed marked differences. High indices of proteins from the type IV secretion system, involved in exchange of DNA through conjugation but also protein delivery to other cells was striking in the global deep SMG. Our results also show the presence of multiple KOs (K02661, K06597) involved in pilus assembly, necessary to establish a physical contact between cells during the conjugation process, among the top 50 highest indices of the global deep SMG (Figure 4.4). In the Antarctic deep SMG, we observed few proteins involved in quorum sensing (K18304, K10926, K10927, K10925, K10924) and bacterial motility (K02399, K02413).

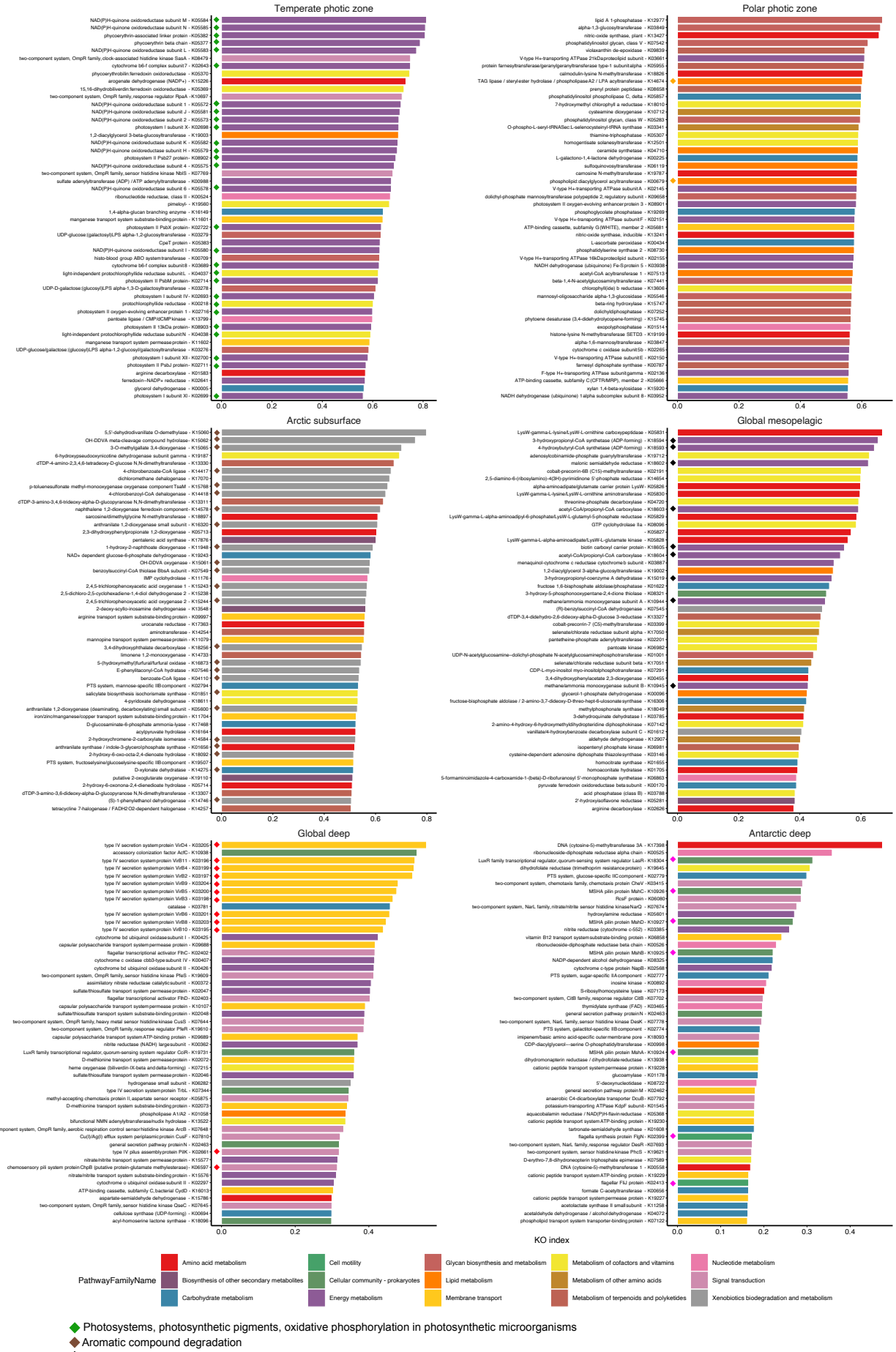

Figure 4.4: *Fifty highest ranking indices for protein families (KO) for each of the six sub-metagenomes*

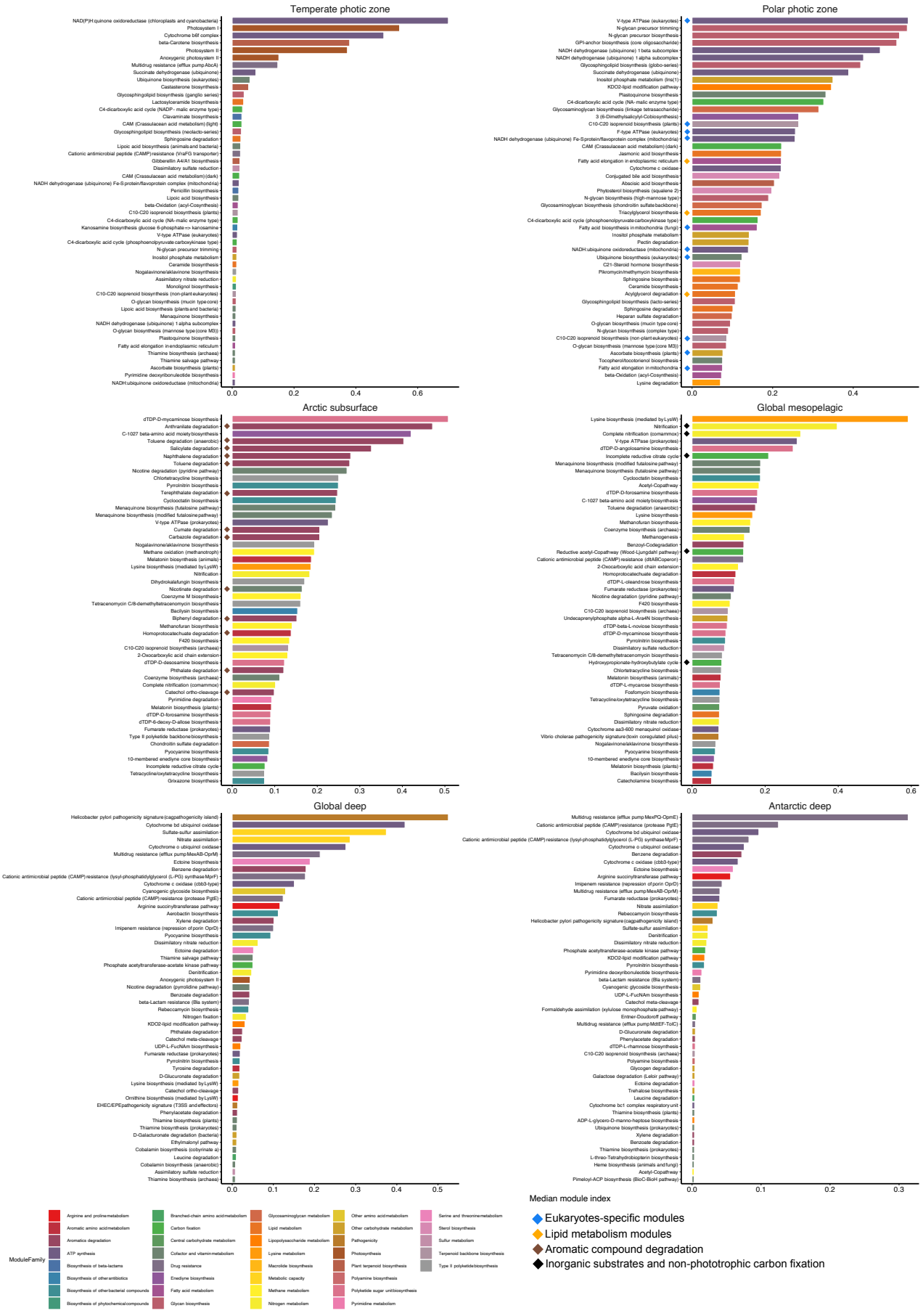

Figure 4.5: *Fifty highest ranking median indices for KEGG modules for each of the six sub-metagenomes*

Figure 4.6: *Fifty highest ranking median indices for KEGG pathways for each of the six sub-metagenomes*

4.3.3 Molecular composition of the Arctic Ocean water column dissolved organic matter

Based on the contribution of the six SMGs, the Arctic samples fell into three different clusters. The samples constituting these three clusters were found in different water column features: the surface, the SCM/FDOMmax and the deep waters. We asked if the partition of samples based on metabolic processes of microbiomes was reflected in differences in the composition of organic matter within the Arctic Ocean. More specifically, we sought to determine if we found relationships between the distinct metabolic features of the Arctic Ocean microbiomes and the molecular composition of DOM. We therefore characterized the molecular composition of the dissolved organic matter (DOM) in the water column of the Arctic Ocean using Fourier transform ion cyclotron resonance mass spectrometry (FT ICR MS).

Non-metric multidimensional scaling (NMDS) analysis of DOM molecular characterization data showed that the grouping of samples based on DOM (Figure 4.7a) paralleled the grouping of samples based on SMGs contribution into the three water features: surface, SCM and deep waters. Overlaid chemical classifications on the plot indicated that a higher contribution of oxygenated compounds, peptides, and amino sugars in surface samples is driving the separation with the SCM and deep samples. On the other hand, a higher contribution of condensed hydrocarbons and lower contribution of lignin (CRAM)-like compounds in the SCM samples were driving the separation with deep samples.

Through the multiple and diverse metabolic pathways used to transform DOM, microbial communities considerably complexify the molecular composition of DOM[12]. Complexity, or diversity, of DOM is therefore an indicator of which compounds are accessed, used and transformed by microbial communities. Functional diversity of the DOM calculated based on elemental composition and aromaticity/double bond equivalency indices indicated that the stratification of the water column resulted in significant changes in the reactivity and overall geochemical signature of the DOM (Figure 4.7b). The elemental composition diversity index of the SCM DOM were significantly ($p < 0.05$) higher than both surface and deep DOM, suggesting more diverse DOM composition in the SCM than in the surface and the deep. The unsaturation/aromaticity diversity index was maximal in the SCM DOM where we found the stronger contribution of the Arctic subsurface SMG and hence aromatic compound degradation genes and pathways. Elemental composition diversity was significantly ($p < 0.05$) higher in surface compared to deep. However, the unsaturation and aromaticity diversity index was not significantly different between surface and deep.

As we observed a coherence between the composition of DOM and the metabolic features of the Arctic Ocean microbiomes, we sought to examine if the reactivity of DOM would be coherent with the distinct metabolic genes and pathways found in different water features. To examine the reactivity of DOM, we performed a network analysis of putative biochemical transformations. A putative transformation corresponded to the loss or gain of a chemical group between two mass peaks. We systematically compared peaks for mass differences that

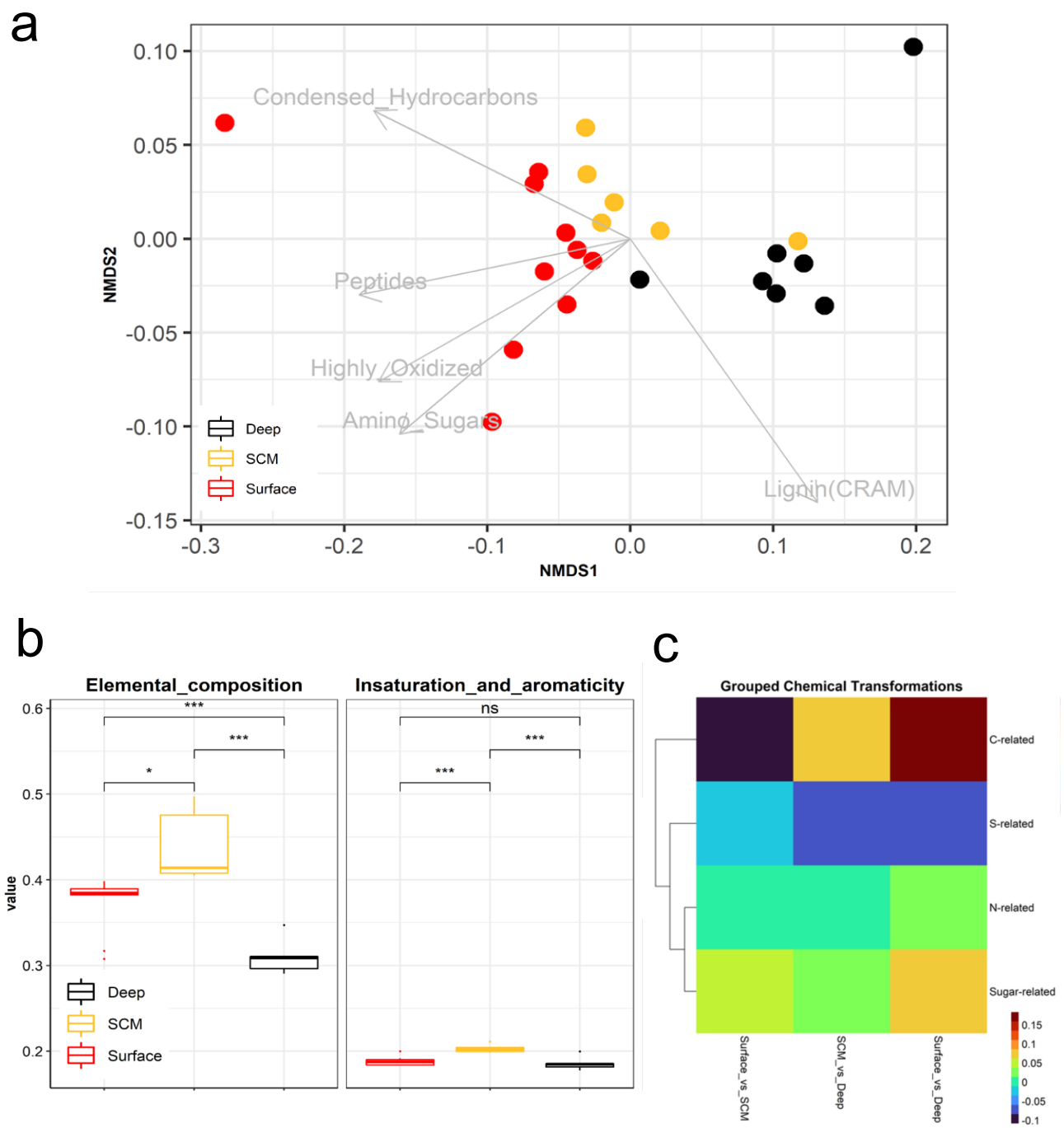

Figure 4.7: *Molecular characterization of the dissolved organic matter in the water column of the Arctic Ocean. a) Non-metric multi-dimensional Scaling (NMDS) analysis of all peaks with assigned molecular formula generated via FT ICR MS. The chemical classes were overlaid on the NMDS plot to investigate the ocean stratification in terms of dissolved organic matter (DOM) chemical classes. Depths are coloured as surface (red), SCM (yellow), and deep (black). b) Functional diversity analyses on all peaks with assigned molecular formula generated via FT ICR MS. The elemental composition diversity index is calculated based on the mass and elemental (C,H,N,O,S,P) composition. Insaturation and aromaticity index is calculated based on the aromaticity (AI) and double bond equivalency (DBE-O) indices. Statistical significance was determined by the Kruskal-Wallis test (non-parametric) with Dunn post-hoc. The "ns" symbol indicates $p > 0.05$. All $p < 0.05$ and 0.001 are summarized with one and three asterisks, respectively. c) Log 2 fold change (Log_2FC) of the relative abundance of biogeochemical transformation based on all detected peaks from FT ICR MS. The Log_2FC was calculated to do a pairwise comparison among depths. The transformations were grouped into 4 main categories : Carbon (C), sulfur (S), nitrogen (N), and sugar related transformation*

corresponded to the loss or gain of 125 chemical groups. The top 25 transformations that were

significantly different across water column features were grouped into 4 categories depending on the molecular composition of the chemical group exchanged in the reaction: carbon (C), sugar (CH2O), nitrogen (N), and sulfur (S) related. We found important differences between the photic (surface and SCM) and aphotic (deep samples) zone (Figure 4.7c). Sugar-related transformations exhibited the strongest differences, with surface and SCM samples depicting more enriched sugar-related transformation compared to deep ($\log_2$FC > 0.02). The trends in sugar-related transformation paralleled the contribution of the polar photic zone SMG and hence the glycan metabolism genes and pathways in the surface (high contribution), the SCM (lower contribution) and the deep samples (no contribution). N-related transformations were also more enriched in surface samples ($\log_2$FC > 0.007), specifically when comparing to deep ($\log_2$C > 0.01). Sulfur-related transformations, on the other hand, were depleted in surface and SCM ($\log_2$FC < -0.01) compared to deep, with surface waters exhibiting more depleted S-related transformations compared to SCM ($\log_2$FC < -0.01). However, we also observed reactivity differences within the photic zone: the C-related transformations were enriched in both surface and SCM compared to deep ($\log_2$FC > 0.01), but they were depleted in surface compared to SCM ($\log_2$FC < 0).

4.4 Discussion

Strong eukaryotic cold adaptation signature in the metabolic capacity of the Arctic Ocean photic zone

The phytoplankton communities of both the Arctic Ocean[13] and the Southern Ocean[14] are dominated by eukaryotic taxa, with very low abundance or absence of cyanobacteria. Our results capture this eukaryotic signal in the microbiomes metabolism as evidenced by multiple KEGG modules from eukaryotes having high indices in the polar ocean photic zone SMG. Both cold temperatures throughout the year and the elevated levels of solar radiation during the spring and summer favor the accumulation or reactive oxygen species (ROS) at toxic levels in phytoplankton cells[15]. For phytoplankton to persist and thrive in the surface of polar oceans would therefore require strategies to mitigate the effect of ROS. Compared to the temperate ocean's surface, in the polar ocean's surface, we found many protein families and KEGG pathways involved in photoprotection and cold adaptation. Phytoplankton can modulate their photosynthetic pigments levels to adapt to the amount of solar radiation. Jasmonic acid has been described as an important hormone that regulates pigments and proteins levels[16]. The high index of the jasmonic acid biosynthesis module in the polar ocean photic zone SMG may indicate that phytoplanktonic communities in polar oceans use it to decrease their pigment levels and mitigate the production of ROS in response to the large amount of solar radiation during the long spring and summer days. The beta-ring hydroxylase (K15747) and violaxanthin de-epoxidase (K09839) both had high indices in the polar photic zone SMG and were involved in the biosynthesis of pigments (lutein and zeaxanthin) used to quench high energy chlorophyll intermediates produced during high solar radiation levels, preventing the formation

of ROS. The presence of two nitric oxide (NO) synthases amongst the 50 highest ranking protein families in the polar photic zone SMG, reinforces the ROS protection, as NO acts as an antioxidant in phytoplankton[17]. Similarly, tocopherol a derivative of vitamin E, acts as an antioxidant against cold stress in Antarctic phytoplankton[18] and its synthesis was observed in the 50 highest KEGG modules. Altogether, these results suggest that the polar ocean phytoplankton communities have developed an arsenal of metabolic processes for optimal growth under stress in cold environment with intense photoperiods.

The enrichment of glycan metabolism in the microbial communities of the polar oceans photic zone reveals multi-level cold adaptation mechanisms

The cold tolerance of polar ocean photic zone microbial communities may enhance their carbohydrate metabolism. Many vital cell functions are affected by the rigidification of cell envelopes (cell membrane and cell wall) under low temperatures. The fluidity of the cell membrane in cold environments is maintained by an increase in the fraction of unsaturated, polyunsaturated and branched fatty acids (FAs) in the membrane[19]. The use of strategies to fluidize the membrane in polar ocean microbes is evidenced in our results by the relatively high indices of KEGG pathways for the biosynthesis of unsaturated FAs or synthesis of linoleic and alpha-linoleic acid (an unsaturated FAs) in the polar oceans photic zone. Lipopolysaccharide (LPS) is the main component of the outer membrane of gram-negative bacteria[20]. LPS is anchored to the membrane through the lipid A. The core glycan is linked to the lipid A and can bear a longer glycan called the O-chain. The two top protein families of the polar ocean photic zone, were involved int the synthesis and modification of LPS: lipid A 1-phosphatase (K12977) and alpha-1,3-glucosyltransferase (K03849). Similarly, we found the Kdo2-lipid modification pathway in the top 10 highest ranking KEGG pathways. We therefore suspect that modifications of LPS, either through the saturation of the fatty acid chains in lipid-A or the modification of the core and O-chain glycans participate to the cold-tolerance response of the microbial communities from the polar ocean photic zone, as has been shown in various studies[20].

In addition to the modification of FA structure, cells can use other compounds to modulate membrane fluidity. Bacteria have been shown to use carotenoids, a class of isoprenoids, to improve membrane fluidity in cold environments[21,22]. The high indices of C10-C20 isporenoid biosynthesis KEGG modules, as well as protein farnesyl transferase (K05955, involved in isoprenoid biosynthesis) in the polar ocean photic zone SMG may indicate that microbial taxa use carotenoids as an aid to regulate membrane fluidity. In addition, carotenoids are involved in scavenging ROS[23].

Microbial taxa can modulate other component of their envelopes to circumvent the detrimental effects of very low temperatures. Gram-positive bacteria increase the thickness of their peptidoglycan walls to resist freeze-thaw cycles, osmotic imbalance, or the disruptive effects of ice crystals[15]. The biosynthesis of glycoproteins requires the attachment of a glycan to proteins through nitrogen or oxygen atom, hence being called a N- or O-glycan. The numerous high indices KEGG modules involved in N- and O-glycan biosynthesis may indicate an

enhanced capacity to synthesize proteoglycans in the microbial taxa of the polar oceans photic zone. They could also indicate an enhanced capacity to produce exopolysaccharides (EPSs). EPSs are polymers of sugars and are usually substituted with proteins[24]. Microbes can indeed release EPSs in their vicinity (covalently or loosely bound the cell) to dampens the deleterious effects of low temperatures[25]. EPSs, among other functions, are osmoprotectants, can act as diffusion barrier or as a barrier for ice[15]. The modifications of LPS, as well as the enhanced synthesis of cell wall glycoproteins and EPS may significantly increase the amount of glycans in the microbiomes of the polar ocean's photic zone compared to other oceanic zones.

The enriched metabolism of glycan in the Arctic Ocean my leave a strong signature in the DOM. We found that the sugar-related transformation in DOM were enriched in the surface compared to the SCM and the deep, and enriched in the SCM compared to the deep. This trend followed the contribution of the polar photic zone SMG and therefore of glycan metabolism: high in the surface, lower in the SCM and not contributing to the deep waters. Lots of glycan metabolism genes and pathways of the polar photic zone SMG were involved in the biosynthesis of polysaccharides in the cell wall, membrane or extracellularly. These polysaccharides may therefore be more accessible and used as a growth substrate by other heterotrophic species. The enrichment of sugar-related transformation in the surface supports this hypothesis. In this scenario, polysaccharides (or glycans) would represent a more important carbon and energy source for the microbial loop in the polar oceans than in temperate oceans. More investigation would be however be needed, such as the biogeography of carbohydrate active enzymes or the quantification of polysaccharides. Nevertheless, our results highlight a potential and new important role of glycans for microbiomes to thrive in the Arctic Ocean.

Signature of the capacity to process terrestrial organic matter in the microbial communities of the Arctic Ocean

The metabolism of the microbial communities from the Arctic Ocean subsurface waters reveals a strong signal for the degradation of organic matter from terrestrial origin. The tOM discharge to the Arctic Ocean is maximum during the spring freshet and stays in the surface with the fresh water that carries it. During the winter and ice formation, this tOM sinks and accumulates in the halocline, forming a maximum of tOM[26]. The tOM maximum is particular to the Arctic Ocean and has been identified due to its spectrophotometric properties, forming a maximum of fluorescent dissolved organic matter (FDOMmax). Indeed, tOM absorbs light and fluoresces at different wavelengths than autochthonous marine organic matter (OM)[27]. This particular spectrophotometric characteristic stems from aromatic compounds, derived from soil humic substances that compose a large fraction of tOM[28]. The large number of proteins families and KEGG modules in the top 50 highest indices within the Arctic subsurface SMG but no other SMGs suggests that the capacity to degrade tOM is enhanced in the microbial communities of the Arctic Ocean subsurface. This is further supported by the contribution of the Arctic subsurface SMG being maximum in the FDOMmax, where we find the highest concentration of tOM. These findings strongly support and strengthen the results of chapter 2. In Chapter 2, we demonstrated that, within the Arctic Ocean, the microbiomes of the FDOMmax are enriched

in aromatic compound degradation genes. We also showed that, within the global ocean, it was in the Arctic Ocean FDOMmax that aromatic compound degradation genes made up the highest fraction of the metabolic gene pool. In addition, chapter 2 and a previous study from our lab have associated the unusually high number of aromatic compound degradation genes in Arctic taxa to an adaption to the high levels of tOM in the Arctic Ocean[29]. The results of the current work reported in this chapter go a step further by showing that the enrichment of aromatic compound degradation genes in the microbiomes of the FDOMmax and the SCM within the global ocean is important enough to differentiate the metabolism of the Arctic subsurface waters' microbiomes from microbiomes of the rest of the global ocean. Moreover, within the Arctic Ocean, it was in the FDOMmax and SCM that the DOM unsaturation/aromaticity diversity index was maximal. This high diversity of compounds with an elevated degree of unsaturation may reflect both the abundance of aromatic compounds in the SCM/FDOMmax, but also the diversification of these compounds through microbial transformations. This finding is another element reinforcing the importance of terrestrially-derived aromatic compounds for the microbiomes of the Arctic Ocean subsurface. Ultimately, we would need to determine the contribution of terrestrially-derived OM to the energy flow of the Arctic Ocean microbiomes, but also if its fate is remineralization to CO_2 or transformation to refractory DOM and long-term storage in the Arctic Ocean.

Unique molecular signature in the water column of the Arctic Ocean dissolved organic matter

We observed a strong stratification of the Arctic Ocean DOM between the surface, subsurface and deep samples. This stratification was also observed in other oceans[30,31]. In these studies, the surface was characterized by higher H/C ratio, indicative of freshly produced, phytoplankton-derived OM. The H/C ratio decreased with depth, and the deep samples contained higher fraction of CRAM compounds. Our results show similarities with these studies: the surface had a higher fraction of peptides, amino-sugars and highly-oxidized compounds, all characterized by high H/C ratios, while the deep contained the low H/C ratio CRAM compounds. However, in studies from other oceans, the aromaticity increased with depth, reaching its maximum in the deep samples[31,32]. The Arctic Ocean showed a different trend, with the aromaticity being maximum in the SCM. This was strikingly different than other oceanic regions in which the aromaticity of the SCM DOM is among the lowest of the water column. This high aromaticity index of the Arctic SCM most probably stems from the strong terrestrial signature in Arctic DOM, as evidenced by the important FDOM signal in the SCM (chapter 2). Our DOM analysis is preliminary and we will need more in-depth characterization of the DOM bulk properties to compare to what has been found in other oceans.

4.5 Conclusion

The Arctic Ocean is characterized by a unique biogeochemical landscape. Microbial metabolic processes shape and are shaped by the ocean biogeochemistry. Despite the importance of microbial metabolic processes, we still lack a global view of all the metabolic characteristics of the Arctic Ocean within the global ocean. In this study, we systematically surveyed the metabolism of microbial communities in the global ocean. We discovered that microbiomes from the surface waters of polar oceans exhibit similarities with each other but differ markedly from microbiomes found in temperate and equatorial waters. Specifically, we found a strong eukaryotic signature and processes involved in the metabolism of glycans and lipids. Our study demonstrated that the microbiomes of the Arctic subsurface waters was characterized by a uniquely strong signal of aromatic compound degradation processes. By characterizing the molecular composition of the water DOM, we showed that the DOM stratification followed the stratification of microbial metabolic processes in the Arctic Ocean. In addition, we found a strong aromaticity of DOM in the Arctic subsurface waters, corresponding to the prevalence of aromatic compounds degradation genes and an enrichment of chemical reactions involving sugar moieties in the Arctic surface waters, that corresponded to the strong contribution of glycan metabolic genes and pathways. Altogether, we could, for the first time, unravel the particularities of the Arctic Ocean microbiomes metabolism within the global ocean and relate it to the composition of the DOM. In next steps, we will seek to find degradation pathways for the glycan compounds in the Arctic surface and aromatic compounds in the Arctic subsurface to confirm the putative role of microbial communities in glycan metabolism and aromatic compound degradation respectively. Finally, we will seek to link metabolic genes and pathways to molecular formulae through network analysis to highlight the main actors in glycan metabolism and aromatic compound degradation in the Arctic Ocean. Nevertheless, our study serves as a base to further investigate the phylogenetic distribution of metabolic processes of importance in the Arctic Ocean, and their impact on the Arctic Ocean biogeochemical cycles.

4.6 Methods

Sampling, DNA and RNA extraction

Samples were collected in September 2017 during the Joint Ocean Ice Study cruise to the Canada Basin. We analyzed 22 metagenomes and 25 metatranscriptomes generated from samples collected across the water column of the Canada Basin. Eight specific water masses were sampled: the surface mixed layer (surface: 5 m and 20 m depth) characterized by fresher water due to riverine input and ice melt, the subsurface chlorophyll maximum (SCM), in the halocline (FDOMmax at salinity of 32.3 and 33.1 PSU, referred as 32.3 and 33.1), and deeper water from Atlantic origin at the temperature maximum (referred as Tmax), 1000 m depth (Atlantic water, further referred as AW) and 10 or 100 m above the bottom (further referred as bottom).

We filtered 14L of seawater for DNA samples and 7L of seawater for RNA samples sequentially through a 3 μm pore size polycarbonate track etch membrane filter (AMD manufacturing, ON, Canada) and a 0.22 μm pore size Sterivex filter (Millipore, MA, USA). Filters were stored in RNALater (ThermoFischer, MA, USA), and kept frozen at -80^{0}C until processing in the lab. DNA was extracted following the method described in Colatriano *et al.*[33]. Briefly, the preservation solution was expelled and replaced by a SDS solution (0.1 M Tris-HCl pH 7.5, 5% glycerol, 10 mM EDTA, 1% sodium dodecyl sulfate) and incubated at room temperature for 10 min and then at 95^{0}C for 15 min. The cell lysate was then centrifuged at 3,270 x g. Proteins were removed by precipitation with MCP solution (Lucigen, WI, USA) and the supernatant was collected after centrifugation at 17,000 x g for 10 min at 4^{0}C. DNA was precipitated with 0.95 volume of isopropanol and rinsed twice with 750 μL ethanol before being air dried. The DNA was resuspended in 25 μL of low TE buffer, pH 8 (10 mM Tris-HCl, 0.1mM EDTA) and stored at -80^{0}C.

The RNA extraction procedure was adapted from the mirVana RNA extraction kit (ThermoFisher, MA, USA). RNAlater was expelled from the Sterivex and replaced by 1.5 mL of Lysis buffer and Sterivex was vortexed. 150 μL of miRNA homogenate were added, the Sterivex vortexed and incubated on ice for 10 min. The cell lysate was expelled from the Sterivex, 0.9x the volume of acid-phenol-chloroform was added and the solution was vortexed for 30-60 sec. The mix was centrifuged at 10,000 x g for 5 min, and the top aqueous phase gently removed and transferred to a fresh tube. 1.25 volume of 100% ethanol was added to the aqueous phase and vortexed to mix. The mix was filtered through mirVana Filter Cartridges by centrifugating at 10,000 x g for 10 s, and the flow through discarded. The RNA was rinsed with 700 μL of Wash Solution 1 and then with 500 μL Wash solution 2/3 by centrifugating at 10,000 x g for 10 sec. RNA was then eluted with 50 μL of Elution solution (0.1 M EDTA) warmed at 95^{0}C. 700 μL of RTL buffer and 500 μL 100% ethanol were added to the RNA suspension and the suspension was centrifuged for 15 sec at 10,000 x g on a RNeasy MinElute column. RNA was washed first with RPE buffer by centrifuging 500 μL for 15 sec at 10,000 x g and then 80% ethanol for 2 min at 10,000 x g. The empty column was then centrifuged at 12,000 x g for 5 min to discard the excess liquid. The RNA was finally eluted by centrifugation of 28 μL and then 10 μL of RNase free water for 1 min at 12,000 x g and stored at -80^{0}C.

Generation of KO abundance tables

Bioinformatics metagenomics files containing genes IDs, gene annotations, gene depth of coverage and other gene information were retrieved from the DOE Joint Genome Institute Integrated Microbial Genomes (https://img.jgi.doe.gov) repository. The abundance of a KEGG ortholog number (KO, gene family) in a metagenome was calculated by summing the depth of coverage of all genes annotated with this KO. KO abundance matrices therefore represent the metagenomic profiles in the samples. As samples vary in terms of depth of sequencing (or library size), we normalized the metagenomic profiles to the relative cell number to get a per-cell number of copies[34]. Concretely, the KO abundances were divided by the median abundance of 10 universal single-copy phylogenetic marker genes (K06942, K01889, K01887, K01875,

K01883, K01869, K01873, K01409, K03106, and K03110). These normalized KO abundances can therefore be interpreted as the per-cell number of gene copies for a given protein family (KO).

Non-negative matrix factorization of the KO abundance matrix

Non-negative matrix factorization (NMF) was performed with the nmf function form the NMF package in R[35]. NMF analysis decomposes the KO abundance matrix in a product of two matrices: the basis matrix and the coefficient matrix. The coefficient matrix contains a reduced number of new descriptors whose profiles describes the overall structure of the metagenomic profiles across the samples. The basis matrix contains the weights of the original descriptors (KO) on the new set of descriptors. Different subsets of KO have different weights on each of the new descriptors. In our study, the new descriptors therefore represent a subset of specific KO. We therefore named the new descriptors as sub-metagenomes (SMGs).

The number of SMGs (rank) is arbitrary and was chosen based on a way to optimize a number of metric and with biological consideration of the structure of the global ocean. NMF was first ran with different ranks (Slide 2-3), with a number of 100 runs for each rank. The consensus matrices represent (Slide 2), for each rank, over the 100 runs, the fraction of times 2 samples are clustered together based on the coefficient matrix. Various metrics were also calculated (Slide 3) based on the NMF results.Cophenetic represents the finish. Rss is the square root of the summed squared differences between the original KO abundance matrix and the product of the basis and coefficient matrix. The dispersion is defined as $1 - rss/\sum_{i,j} j(V_{i,j})^2$ ($V_{i,j}$ are the entries of the KO abundance matrix) and estimates the fraction of variance of the KO abundance matrix explained by the product of the basis and coefficient matrices. Residuals is the sum of differences between the original KO abundance matrix the matrix approximated by the product of the basis and coefficient matrices. Sparseness represents the homogeneity of a vector. Sparseness is equal to 1 if all the elements of a vector are null but for 1. Oppositely, the sparseness is equal to 0 is all the element of a vector are equals. The sparseness of the basis and coefficient matrices are calculated as the mean sparseness of its element vectors. A rank of 6 was chosen for following analysis as its metrics were very similar to other ranks. In addition, clustering based on both the coefficient and basis matrices produced the same clusters of samples. More importantly, the samples clustering corresponded to meaningful physical and biological oceanic regions: (i) photic zone of the global ocean minus the poles (ii) photic zone of the poles (iii) subsurface water masses of the Arctic Ocean (iv) mesopelagic waters of the global ocean (v) bathypelagic waters of the global ocean (vi) bathypelagic waters of the Antarctic ocean.

Calculation of protein families, KEGG module and KEGG pathway indices

The indices were calculated for KO number annotated from metagenomes by combining two methods described in Jiang *et al.*[36] and Kim textitet al.[37]. We first used the KO number abundance matrices and the coefficient matrices (SMG x samples) to calculate both the spear-

man correlation coefficient and the multidimensional projection between all pairs of KO number annotated from metagenomes and SMG. The spearman correlation coefficient between an KO (i) and a SMG (k) $\rho_{i,k}$ was calculated using the abundance profile of a KO and a SMG along all the samples. The multidimensional projection between an KO and a SMG was calculated as the cosine of the angle between the vectors represented by an KO abundance in the samples space and the vector represented by SMG in the samples space. The abundance profiles of KO and SMGT were first normalized, and the multidimensional projection was calculated as:

$$Cos\Theta_{i,k} = \sum_{j=1}^{n} a_{i,j} \times s_{k,j} \tag{4.1}$$

Where $\text{Cos}\Theta_{i,k}$ is the multidimensional projection between the KO i and the SMG k, n is the number of samples, $a_{i,j}$ is the normalized abundance of the KO i in the sample j, and $s_{k,j}$ is the normalized abundance of the SMG k in the sample j. We then used the basis matrix to calculate the score of each KO:

$$Score(i) = 1 + \frac{1}{log_2(q)} \sum_{k=1}^{q} p(i,k) \times log_2(p(i,k)) \tag{4.2}$$

Where i is the KO, q is the number of SMG (6 in our study), k is the SMG, p(i,k) is the probability of finding the KO i in the SMG k. We calculated the final KO index on each SMG by multiplying the spearman correlation coefficient, cos theta and the KO score:

$$I_{i,k} = \rho_{i,k} \times Cos\Theta_{i,k} \times Score(i) \tag{4.3}$$

This allowed us to calculate an index for each pair of KO and SMG

FT ICR MS data acquisition and data analysis

A 12 T Bruker SolariX FTICR mass spectrometer located at EMSL, a DOE-BER national user facility in Richland, WA, was used to collect high-mass resolution spectra of the organic molecules in the samples collected from the basin. A standard Bruker ESI source was used to generate negatively charged molecular ions. The ion accumulation time (IAT) was varied between 0.01 and 0.05 seconds to account for variations in carbon concentrations in each sample. One hundred forty-four individual scans were averaged for each sample and internally calibrated using organic matter homologous series separated by 14 Da. The mass measurement accuracy was typically within 1 ppm for singly charged ions across a broad m/z range (200-1200 Da). Instrument settings were optimized using Suwannee River Fulvic Acid (IHSS) used as a standard. The instrument was flushed between samples using double deionized water and HPLC grade methanol. Blanks were analyzed at the beginning and the end of the day to monitor background contaminants and day to day . BrukerDaltonik version 4.2 data analysis software converted raw spectra (obtained from each sample) to a list of m/z values applying FT-MS peak picker module with a signal-to-noise ratio (S/N) threshold of 7 and absolute intensity threshold of 106. Formularity[38] software was used to assign chemical formulae based on the criteria described in Tfaily *et al.*[10]

The assigned molecular formulae were visualized on van Krevelen diagrams[39] based on their ratios of hydrogen to carbon (H/C) and oxygen to carbon (O/C). The van Krevelen diagrams provide a means to compare the average properties of OM and enable the identification of the major chemical classes (lipids, peptides, lignin (CRAM), carbohydrates, unsaturated hydrocarbons, condensed hydrocarbons, and highly oxygenated compounds) present in the samples. The H/C and O/C ranges for chemical classification are follows: Amino sugar and carbohydrates (O/C: 0.5-0.7; H/C:0.8-2.5), Condensed hydrocarbons (O/C: 0.0-0.4; H/C:0.2-0.8), Lignin(CRAM) (O/C: 0.29-0.65; H/C:0.7-1.5), Lipids (O/C: 0.0-0.3; H/C:1.5-2.5), Peptides (O/C: 0.3-0.6; H/C:1.5-2.3), Amino sugar and Highly oxygenated compounds (O/C: 0.65-1.0; H/C:0.8-1.5), Unsaturated hydrocarbons (O/C: 0.0-0.3; H/C:1.0-1.6).

The molecular properties of the filtered peaks that received a molecular formula assignment by Formularity were determined by calculating molecular indexes based on each peak's elemental composition. These include: modified Aromaticity Index (AImod) that reflects the "density" of carbon-to-carbon double bonds within a molecule; and finally double bond equivalence (DBE) that represents the amount of unsaturation in a molecule and can indicate the presence of aromatic structures[40].

Investigation of biochemical transformations through dissolved organic matter network analysis

Network analysis was employed to determine DOM biochemical transformations or decomposition pathways across different samples. Pairwise mass differences in an all-vs-all approach between all m/z peaks within a sample were calculated. These mass differences were compared to a list of 86 commonly observed biochemical transformations adopted from Breitling *et al.*[11] included reactions that could occur by biotic and/or abiotic means depending on the environmental conditions and loss or gain of phosphate or an amine group.

The R package RCy3[41] was used to build and analyze these transformation networks using Cytoscape v.3.8.1[42] via R. In the networks, each m/z peak corresponds to nodes, and the mass difference between two m/z peaks is represented by the edges of the networks[43]. Network heterogeneity was calculated using Cytoscape's core application NetworkAnalyzer version 4.4.6[44]. The network heterogeneity statistic is based on the clustering coefficient parameter. A high clustering coefficient value indicates that the nodes on the network have a high number of neighbouring nodes, *i.e.*, diversity of biochemical transformations and low reaction specificity, in this context. Inversely, a low clustering coefficient value indicates that the nodes on the network have a low number of shared connections with neighbouring nodes, hence a low diversity of transformations and a high biochemical reaction specificity[45,46].

Statistical analysis

The log-transformed peak intensities were transformed into Bray-Curtis distances using the

"vegdist" function for the vegan package and ordination of the data, based on the normalized intensities, was calculated using non-metric multidimensional scaling (NMDS). The percentages of chemical classes were overlaid on the NMD plot to investigate the depth effect.

4.7 Acknowledgments

The data were collected aboard the CCGS Louis S. St-Laurent in collaboration with researchers from Fisheries and Oceans Canada at the Institute of Ocean Sciences and Woods Hole Oceanographic Institution's Beaufort Gyre Exploration Program and are available at http://www.whoi.edu/beaufortgyre. We would like to thank both the Captain and crew of the CCGS Louis S. St-Laurent and the scientific teams aboard.

Bibliography

[1] C. Michel, J. Hamilton, E. Hansen, et al. *Arctic Ocean outflow shelves in the changing Arctic: A review and perspectives*. **Prog. Oceanogr.**, 139:66–88, 2015.

[2] F-J.W. Parmentier, T.R. Christensen, S. Rysgaard, et al. *A synthesis of the arctic terrestrial and marine carbon cycles under pressure from a dwindling cryosphere*. **Ambio**, 46(S1):S53–S69, 2017.

[3] K.A. Brown, J.M. Holding, and E.C. Carmack. *Understanding Regional and Seasonal Variability Is Key to Gaining a Pan-Arctic Perspective on Arctic Ocean Freshening*. **Front. Mar. Sci.**, 7:606, 2020.

[4] J. Berge, P.E. Renaud, G. Darnis, et al. *In the dark: A review of ecosystem processes during the Arctic polar night*. **Prog. Oceanogr.**, 139:258–71, 2015.

[5] M Ardyna and K.R. Arrigo. *Phytoplankton dynamics in a changing Arctic Ocean*. **Nat. Clim. Chang.**, 10:892–903, 2020.

[6] Z. Xie, Z. Yu, Y. Li, et al. *Soil microbial metabolism on carbon and nitrogen transformation links the crop-residue contribution to soil organic carbon*. **Biofilms Microbiomes**, 8:14, 2022.

[7] J. McCarren, J.W. Becker, D.J. Repeta, et al. *Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea*. **Proc. Natl. Acad. Sci.**, 107(38):16420–7, 2010.

[8] E.K. Hall, E.S. Bernhardt, R.L. Bier, et al. *Understanding how microbiomes influence the systems they inhabit*. **Nat. Microbiol**, 3:977–82, 2018.

[9] E.B. Graham, A.R. Crump, D.W. Kennedy, et al. *Multi 'omics comparison reveals metabolome biochemistry, not microbiome composition or gene expression, corresponds to elevated biogeochemical function in the hyporheic zone*. **Sci. Total Environ.**, 642:742–53, 2018.

[10] M.M. Tfaily, R.M. Wilson, W.T. Cooper, et al. *Vertical Stratification of Peat Pore Water Dissolved Organic Matter Composition in a Peat Bog in Northern Minnesota*. **J. Geophys. Res. Biogeosciences**, 123:479–94, 2018.

[11] R. Breitling, S. Titchie, D. Goodenowe, et al. *Ab initio prediction of metabolic networks using Fourier transform mass spectrometry data*. **Metabolomics**, 2(3):155–64, 2006.

[12] B.E. Noriega-Ortega, G. Wienhausen, A. Mentges, et al. *Does the chemodiversity of bacterial exometabolomes sustain the chemodiversity of marine dissolved organic matter?* **Front. Microbiol.**, 10:215, 2019.

[13] R. Terrado, K. Scarcella, M. Thaler, et al. *Small phytoplankton in Arctic seas: vulnerability to climate change*. **Biodiversity**, 14(1):2–18, 2013.

[14] F. Flaviani, D.C. Schroeder, K. Lebret, et al. *Distinct oceanic microbiomes from viruses to protists located near the Antarctic Circumpolar current*. **Front. Microbiol.**, 9:1474, 2018.

[15] C. Lauritano, C. Rizzo, A. Lo Giudice, and M. Saggiomo. *Physiological and molecular responses to main environmental stressors of microalgae and bacteria in polar marine environments*. **Microorganisms**, 8:1957, 2020.

[16] B. Barati, S-Y. Gan, J. Deardall, and S-M. Phang. *Green algal molecular responses to temperature stress*. **Acta Physiol. Plant.**, 41:26, 2019.

[17] P. Li, C-Y. Liu, H. Liu, et al. *Protective function of nitric oxide on marine phytoplankton under abiotic stresses*. **Nitric Oxide - Biol. Chem.**, 33:88–96, 2013.

[18] W.C. Dunlap, A. Fujisawa, Y. Yamamoto, et al. *Notothenioid fish, krill and phytoplankton from Antarctica contain a vitamin E constituent (α-tocomonoenol) functionally associated with cold-water*

adaptation. **Comp. Biochem. Physiol. - B Biochem. Mol. Biol.**, 133:299–305, 2002.

[19] M.F. Siliakus, J. van der Oost, and S.W.M. Kengen. *Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure*. **Extremophiles**, 21:651–70, 2017.

[20] M.M. Corsaro, A. Casillo, E. Parrilli, and M.L. Tutino. *Psychrophiles: From Biodiversity to Biotechnology: Second Edition, Chapter 13: Molecular Structure of Lipopolysaccharides of Cold-Adapted Bacteria*. Springer International Publishing, 2017.

[21] W. Seel, D. Baust, D. Sons, and others. *Carotenoids are used as regulators for membrane fluidity by Staphylococcus xylosus*. **Sci. Rep.**, 10:330, 2020.

[22] M.V. Jagganadham, M.K. Chattopadhyay, Subbalakshmi, et al. *Carotenoids of an Antarctic psychrotolerant bacterium, Sphingobacterium antarcticus, and a mesophilic bacterium, Sphingobacterium multivorum*. **Arch. Microbiol.**, 173:418–24, 2000.

[23] M. Zuluaga, Gueguen V., G. Pavon-Djavid, and D. Letourneur. *Carotenoids from microalgae to block oxidative stress*. **BioImpacts**, 7(1):1–3, 2017.

[24] J. Angelin and M. Kavitha. *Exopolysaccharides from probiotic bacteria and their health potential*. **Int. J. Biol. Macromol.**, 162:853–65, 2020.

[25] A. Casillo, R. Lanzetta, M. Parrilli, and M.M. Corsaro. *Exopolysaccharides from marine and marine extremophilic bacteria: Structures, properties, ecological roles and applications*. **Mar. Drugs**, 16:69, 2018.

[26] J. Jung, J.E. Son, Y.K Lee, K.H Cho, Y. Lee, et al. *Tracing riverine dissolved organic carbon and its transport to the halocline layer in the Chukchi Sea (western Arctic Ocean) using humic-like fluorescence fingerprinting*. **Sci. Total Environ.**, 772:145542, 2021.

[27] K.R. Murphy, C.A. Stedmon, T.D. Waitze, and G.M. Ruiz. *Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy*. **Mar. Chem.**, 108(1-2):40–58, 2008.

[28] J. Gerke. *Concepts and misconceptions of humic substances as the stable part of soil organic matter: A review*. **Agronomy**, 8(76), 2018.

[29] D. Colatriano, P.Q Tran, C. Gueguen, W.J. Williams, C. Lovejoy, et al. *Genomic evidence for the degradation of terrestrial organic matter by pelagic Arctic Ocean Chloroflexi bacteria*. **Commun. Biol.**, 1(1)):90, 2018.

[30] A.M. Martinez-Pérez, H. Osterholz, M. Nieto-Cid, et al. *Molecular composition of dissolved organic matter in the Mediterranean Sea*. **Limnol. Oceanogr.**, 62:2699–2712, 2017.

[31] P. Li, J. Tao, J. Lin, et al. *Stratification of dissolved organic matter in the upper 2000m water column at the Mariana Trench*. **Sci. Total Environ.**, 668:1222–31, 2019.

[32] P.M. Medeiros, M. Seidel, L.C. Powers, et al. *Dissolved organic matter composition and photochemical transformations in the northern North Pacific Ocean*. **Geophys. Res. Lett.**, 42:863–70, 2015.

[33] D. Colatriano and D.A. Walsh. *An aquatic microbial metaproteomics workflow: From cells to tryptic peptides suitable for tandem mass spectrometry-based analysis*. **J. Vis. Exp.**, 103:e52827, 2015.

[34] A. Milanese, D.R. Mende, L. Paoli, et al. *Microbial abundance, activity and population genomic profiling with mOTUs2*. **Nat. Commun.**, 10:1014, 2019.

[35] R. Gaujoux and C. Seoighe. *A flexible R package for nonnegative matrix factorization*. **BMC Bioinformatics**, 11, 2010.

[36] X. Jiang, M.G.I. Langille, R.Y. Neches, et al. *Functional Biogeography of Ocean Microbes Revealed through Non-Negative Matrix Factorization*. **PLoS One**, 7(9):1–9, 2012.

[37] H. Kim and H. Park. *Sparse non-negative matrix factorizations via alternating non-negativity-constrained least squares for microarray data analysis*. **Bioinformatics**, 23(12):1495–502, 2007.

[38] N. Tolíc, Y. Liu, A. Liyu, et al. *Formularity: Software for Automated Formula Assignment of Natural and Other Organic Matter from Ultrahigh-Resolution Mass Spectra*. **Anal. Chem.**, 89:12659–65, 2017.

[39] D. Van Krevelen. *Graphical-statistical method for the study of structure and reaction processes of coal*. **Fuel**, 29:269–84, 1950.

[40] B.P. Koch and T. Dittmar. *From mass to structure: An aromaticity index for high-resolution mass data of natural organic matter*. **Geochim. Cosmochim. Acta**, 20:926–32, 2006.

[41] J.A. Gustavsen, S. Pai, R. Isserlin, et al. *Rcy3: Network biology using cytoscape from within R*. **F1000Research**, 8:1774, 2019.

[42] P. Shannon, A. Markiel, O. Ozier, et al. *Cytoscape: A software environment for integrated models of biomolecular interaction networks*. **Genome Research**, 13(11):2498–504, 2003.

[43] K. Longnecker and E.B. Kijawinski. *Using network analysis to discern compositional patterns in ultrahigh-resolution mass spectrometry data of dissolved organic matter*. **Rapid Commun. Mass Spectrom.**, 30(22):2388–94, 2016.

[44] Y. Assenov, F. Ramírez, S-E. Schelhorn, et al. *Computing topological parameters of biological networks*. **Bioinformatics**, 24(2):282–4, 2008.

[45] A-S. Barabási and Z.N. Oltvai. *Network biology: Understanding the cell's functional organization*. **Nat. Rev. Genet.**, 5:101–13, 2004.

[46] J. Hu and Y. Zhang. *Discovering the interdisciplinary nature of Big Data research through social network analysis and visualization*. **Scientometrics**, 112:91–109, 2017.

Chapter 5

Conclusion and perspectives

The myriad of metabolic processes that microbiomes can perform enable them to thrive in an incredible variety of environments[1], including the most extreme ones[2]. Their metabolic capacity to use and transform all kind of inorganic and organic matter is at the core of their role in controlling the major biogeochemical cycles[3]. It is therefore crucial to understand the ecology of microbial processes and how they are affected by perturbations to seize how they shape their ecosystem and physicochemical environment. This is especially relevant in northern ecosystem, which are subjected to the fastest and most important climate change-induced perturbations[4]. In this thesis we unraveled the biogeography and phylogenetic diversity of metabolic genes and pathways of importance for Arctic Ocean microbiomes and linked them to their unique ecological environment.

In this thesis, we first sought to highlight the metabolic preferences of the microbiomes within each water column features of the Arctic Ocean. Non-negative matrix factorization revealed a strong signature of aromatic compound degradation genes in the microbiomes of the maximum of fluorescent dissolved organic matter (FDOMmax), characteristic of the accumulation of tDOM in the Arctic Ocean[5]. By comparing with microbiomes of the global ocean, we discovered that the abundance of aromatic compound degradation genes was higher in microbiomes of the Arctic water column features compared to similar water column features from other oceans, and across the global ocean, was the highest in the Arctic FDOMmax microbiomes. The diversity of aromatic compound degradation pathways in the Arctic Ocean microbiomes reflected the capacity to degrade the variety of aromatic compounds expected from organic matter of the Arctic Ocean watershed. Using metagenome-based genome reconstruction, we could identify that the capacity to degrade aromatic compounds was phylogenetically concentrated in *Rhodospirillales*. The global distribution of these *Rhodospirillales* showed that most were restricted to the Arctic Ocean, while comparative genomics demonstrated that those *Rhodospirillales* were enriched in aromatic compound degradation genes compared to their closest related taxa from other oceanic zones. This led us to postulate that the acquisition of aromatic compound degradation genes may play a role for the adaptation of these *Rhodospirillales* in the Arctic Ocean. This study therefore highlights the capacity of the microbiomes to process the disproportionately vast amount of tDOM input to the Arctic Ocean.

In future works, it would be valuable to determine how the capacity to degrade aromatic compounds manifests in the cycling of terrestrial organic matter in the Arctic Ocean. Specifically, we would need to measure the quantities and qualities of tDOM cycled by the Arctic Ocean microbiomes, the metabolic routes and the fate of the tDOM to complete remineralization or to recalcitrant fractions. In a first step, incubation experiments could be undertaken, by amending microbial communities from various water column features of the Arctic Ocean with tDOM sampled from the major rivers flowing to the Arctic. We envision these experiments as the combination of meta-omic techniques to characterize the shift in microbiomes taxa, gene and pathways composition and activities with measurement of bulk (total organic and inorganic C and N content, etc) and molecular (mass spectrometry) DOM properties. This would enable us to understand what taxa actively degrade tDOM, and through which metabolic routes. In addition, model compounds identified in tDOM could be radiolabeled to determine with exactitude its fate to remineralization, or to refractory DOM. This is an especially important question as the increasing permafrost thawing in the Arctic watershed[6] combined with intensifying riverine input[7] to the ocean carries ever-increasing amount of tDOM to the Arctic Ocean[8]. This tDOM could exacerbate climate change perturbations by being remineralized to CO_2 or oppositely buffers their effects by being stored long-term as refractory DOM.

As the storage of fat is widespread for macroorganisms of polar regions[9], we suspected that it would be an important metabolic feature for the Arctic Ocean microbiomes. We demonstrated that the storage of neutral lipids, particularly triacylglycerols was highly enriched in the photic zone of the Arctic Ocean compared to other water column features of the Arctic and the rest of the global ocean. We could link the enriched triacyglycerol biosynthesis capacity to the dominance of eukaryotic phytoplankton in the Arctic Ocean as opposed to the dominance of cyanobacteria in temperate oceans. We proposed that the storage of these neutral lipids favour the survival of eukaryotic phytoplankton during the Arctic polar night. Based on the presence of triacylglycerol degradation genes and missing triacylglycerol biosynthesis genes in a large part of the genomes of the prokaryotic community, we suspected that the eukaryotic phytoplankton-based triacylglycerols production could serve as an important growth resource for the heterotrophic prokaryote populations. Through metagenome-based genome reconstruction, we also uncovered an unexpected diversity of bacterial taxa able to synthesize triacylglycerol and most abundant in the deep Arctic waters, as well as taxa able to synthesize poyhydroxyalkanoates in the subsurface waters (SCM and FDOMmax). The deep water and subsurface taxa were enriched in carbohydrate active enzymes and aromatic compound degradation genes respectively, suggesting the use of different carbon sources to feed the neutral lipid biosynthesis in different water column features. The aromatic compound degradation capacity in the subsurface taxa suggests the use of tDOM to feed neutral lipid biosynthesis. As tDOM concentration in the subsurface is constant over time[10], and could feed the prokaryotes neutral lipid biosynthesis, we proposed that the neutral lipid production used to sustain higher levels of the food chain could switch from eukaryotic phytoplankton production in spring, summer and fall to prokaryotic-based production in the polar night.

Future work would be needed to confirm the role of prokaryotic-based NL production to

sustain higher levels of the food web during the winter polar night. In a first step, it would be needed to confirm the storage of neutral lipids within lipid droplet inclusions using microscopy techniques. As the accumulation and use of neutral lipids is a highly dynamic process, and varies with the availability of resources, it would be interesting to perform sampling campaign throughout the year. As the Arctic Ocean is hard to access during winter, a solution would be to use autonomous samplers[11]. We could then follow, throughout the year, the evolution of the neutral lipid metabolism through metatranscriptomics and the systematic characterization and quantification of lipids (lipidomics[12]). More generally, the seasonal dynamics of the Arctic Ocean microbiomes and their metabolism is poorly understood as seasonal sampling in the Arctic Ocean is very scarce, and most studies, such as ours are limited to periods of easy access to the Arctic Ocean such as summer and fall. There is therefore an important need for studies investigating the dynamics of microbiomes and their metabolism throughout the year in the Arctic Ocean. Ultimately, we think that a comprehensive study including meta-omics techniques, as well as the measurement of physico-chemical parameters, bulk and molecular properties of DOM should be undertaken to tackle important questions: *what is the succession of microbiomes throughout the year, and the ecological factors involved? What is the connectivity between members of the Arctic Ocean microbiomes? What is the partition of organic matter fractions and metabolic processes between members of the Arctic Ocean microbiomes, and the fate of this organic matter fractions?*

In this thesis, for the first time, we systematically uncovered the ensemble of metabolic processes that were favoured by the Arctic Ocean microbiomes within the global ocean. We showed that the metabolism of the polar ocean photic zone microbiomes differed markedly from the metabolic capacity of microbiomes from temperate oceans. We demonstrated a strong signature of various metabolic genes and pathways involved in cold adaptation within the polar ocean photic zone. Within these metabolic genes and pathways, the metabolism of glycans was highly represented. The prevalence of glycan metabolic genes and pathways was further supported by a high occurrence of putative reactions involving sugar moieties within the DOM of the Arctic Ocean photic zone. As the glycan metabolic genes were mainly involved in the biosynthesis of polysaccharides in the membrane or cell wall, they may be more easily accessible for consumption by other taxa. Combined with the signature of reactions involving sugar moieties in the DOM, we suspect that this enrichment in glycan metabolism pathways may favour a high production of glycans that may serve as a growth resource for the heterotrophic microbiome. Our analysis of microbial metabolism in the global ocean also allowed us to show that, metabolically, the microbiomes of the subsurface water of the Arctic Ocean (FDOMmax and SCM) represent a unique feature within the global ocean. This was mainly attributed to the prevalence of aromatic compound degradation genes. The prevalence of aromatic compound degradation genes corresponded to a strong aromaticity in the DOM of the Arctic Ocean subsurface waters. These results strengthen the link between the capacity to process the vast amount of tDOM input to the Arctic Ocean to the degradation of aromatic compounds in the microbiomes.

The molecular composition of DOM, as opposed to the microbiome metagenomics or meta-

transcriptomics, has not been elucidated in standardized study at the scale of the global ocean. We deeply think that such a study would be necessary to clearly establish a link between the biogeography of metabolic genes and the composition of DOM, as well as contextualize our findings on the prevalence of certain metabolic genes and pathways associated with corresponding features of the DOM within the Arctic Ocean. This thesis represents important progress in placing the metabolism of the Arctic Ocean on the global ocean map, and understanding what features distinguish its microbiomes from the microbiomes of other oceans. The million-dollar question now is to predict what will happen to the Arctic microbiomes and their metabolic genes and pathways under a rapidly changing climate. In that regard, we think that future research should be oriented in understanding how the Arctic Ocean microbiomes and their metabolism are affected under climate-change perturbations. Time-series are a great tool to do that. We can link changes in the microbiomes to evolution of conditions in the Arctic Ocean and specifically to striking events such as the 2012 sea ice minimum. However, if any exists, it would be too long to start multidecade time series from scratch. That is where, large scale, controlled experiments become interesting. Mesocosm experiments are particularly appealing as they allow one to perturb an ecosystem in its natural environment in a controlled way. Mesocosms already exist in the Arctic Ocean[13], and it would be invaluable to perform experiments mimicking predicted climate change-induced perturbations (loss of sea ice, warming of the epipelagic waters, increased tDOM influx), while investigating microbiomes and their metabolism with meta-omics techniques, but also the rates of important biogeochemical processes. In this context, the findings of our thesis will guide the design of these kind of experiments towards metabolic processes of importance and the taxa that operate them.

Bibliography

[1] B. Bajic and A. Sanchez. *The ecology and evolution of microbial metabolic strategies*. **Curr. Opin. Biotechnol.**, 62:123–8, 2020.

[2] G.N. Gupta, S. Srivastas, S.K. Khare, and V. Prakash. *Extremophiles: An Overview of Microorganism from Extreme Environment*. **Int. J. Agric. Environ. Biotechnol.**, 7(2):371–80, 2014.

[3] P.G. Falkowski, T. Fenchel, and E.F. Delong. *The Microbial Engines That Drive Earth 's Biogeochemical Cycles*. **Science**, 320:1034–9, 2008.

[4] J.E. Box, W.T. Colgan, T.R. Christensen, et al. *Key indicators of Arctic climate change: 1971-2017*. **Environ. Res. Lett.**, 14:045010, 2019.

[5] J. Jung, J.E. Son, Y.K Lee, K.H Cho, Y. Lee, et al. *Tracing riverine dissolved organic carbon and its transport to the halocline layer in the Chukchi Sea (western Arctic Ocean) using humic-like fluorescence fingerprinting*. **Sci. Total Environ.**, 772:145542, 2021.

[6] K.E. Frey and J.W. McClelland. *Impacts of permafrost degradation on arctic river biogeochemistry*. **Hydrol. Process.**, 23:169–82, 2009.

[7] P.J Mann, R.G.M. Spencer, P.J. Hernes, J. Six, G.R. Aiken, et al. *Pan-Arctic Trends in Terrestrial Dissolved Organic Matter from Optical Measurements*. **Front. Earth Sci.**, 4:1–18, 2016.

[8] L.G. Anderson and R.W. Macdonald. *Observing the Arctic Ocean carbon cycle in a changing environment*. **Polar Res.**, 34(1):26891, 2015.

[9] A.S. Blix. *Adaptations to polar life in mammals and birds*. **J. Exp. Biol.**, 219:1093–105, 2016.

[10] C.A. Stedmon, R.M.W. Amon, D. Bauch, et al. *Insights Into Water Mass Origins in the Central Arctic Ocean From In-Situ Dissolved Organic Matter Fluorescence*. **JGR Oceans**, 126(7), 2021.

[11] M. Wietz, C. Bienhold, K. Metfies, et al. *The polar night shift: seasonal dynamics and drivers of Arctic Ocean microbiomes revealed by autonomous sampling*. **ISME Commun.**, 1, 2021.

[12] T. Züllig, M. Trötzmüller, and H.C. Köfeler. *Lipidomics from sample preparation to data analysis: a primer*. **Anal. Bioanal. Chem.**, 412:2191–209, 2020.

[13] T.M. Tsagaraki, B. Pree, O. Leiknes, et al. *Bacterial community composition responds to changes in copepod abundance and alters ecosystem function in an Arctic mesocosm study*. **ISME J.**, 12: 2694–705, 2018.