The ecology and evolution of the *Methylophilaceae* family across aquatic salinity gradients

Arthi Ramachandran

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

August 2023

© Arthi Ramachandran, **2023**

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared

By: *Arthi Ramachandran*

Entitled: *The ecology and evolution of the Methylophilaceae family across aquatic salinity gradients*

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. Yves Gélinas

______________________________ External Examiner
Dr. Etienne Yergeau

______________________________ Examiner
Dr. William Zerges

______________________________ External
Dr. Malcolm Whiteway

______________________________ External
Dr. Eric Pedersen

______________________________ Thesis Supervisor
Dr. David A. Walsh

Approved by ____
Dr. Robert Weladji

03/08/2023 ____
Dr. Pascale Sicotte
Faculty of Arts and Science

Abstract

The ecology and evolution of the *Methylophilaceae* family across aquatic salinity gradients

Arthi Ramachandran, PhD
Concordia University, 2023

Anthropogenic pressures on aquatic environments are leading to rapid environmental change. One of the main environmental stressors in marine environments is change in salinity. It is a major factor that controls the distribution and diversity of organisms. As microbes are essential to maintaining ecosystem services and biogeochemical cycling, it is important to understand the diversity, distribution, ecology, and metabolism of bacteria living in these rapidly changing environments. In this thesis, bacterial adaptation and evolution is investigated across salinity gradients spatially and temporally. A targeted approach, using the *Methylophilaceae* as a model family to monitor change, allowed us to explore and increase our understanding of niche differentiation, genome streamlining, and ecological adaptations.

We first studied the diversity and distribution of the *Methylophilaceae* family across a spatial salinity gradient where we detected the clades from freshwater to brackish-marine waters in the Saint Lawrence Estuary and Bedford Basin. These distinct populations correlated with specific salinity concentrations, which suggests different ecological roles in response to environmental gradients. This led us to also investigate the adaptation and evolution of *Methylophilaceae* in response to changing environmental conditions over time and space in the increasingly freshening Arctic Ocean. We found evidence for endemic Arctic Ocean taxa as well as a novel clade, BS01, which was adapted to the unique lower salinity surface waters of the Arctic Ocean. We also discovered that over time, the *Methylophilaceae* family showed an increase in specific taxa associated to the years experiencing the highest amount of freshening.

The work in this thesis has increased our understanding of the *Methylophilaceae* family and their niche differentiation to distinct environmental conditions. Overall, this work has larger implications for understanding how climate change is affecting northern aquatic ecosystems and how microbes are evolving and adapting to this change.

Acknowledgements

Firstly, I would like to thank to my supervisor Dr. David A. Walsh for his continued support, guidance, and patience throughout my time in his lab. Your mentorship helped shape me into the scientist and science communicator that I am today. I would also like to thank my committee members Dr. William Zerges and Dr. Malcolm Whiteway for their valuable input and support during our meetings.

I would like to thank my lab mates and friends at Concordia for sharing this experience with me, for being there whenever I needed help, and for taking many coffee breaks with me. I would also like to thank my colleagues at NRCan, OCSA, and Arctic Basecamp for having kind words of encouragement while I was writing my thesis.

I would like to thank all my family and friends for being there for me during my whole graduate school experience and sharing in my successes. A special thanks to Wayne Brake for taking the time out of his busy schedule to help me with the thesis writing process.

Lastly but certainly not least, I would like to thank and express my sincerest gratitude to my parents, sister, and husband for their continuous support. My parents, Jayanthy and Raman, always supported me and were the first ones to encourage me to explore and ask questions. My sister, Lalitha, always believed in me and never failed to pick up the phone when I needed her input. My husband, Benjamin, went on this journey with me while finishing up his own PhD so I always had an amazing and understanding partner by my side.

I cannot express in words how amazing my support system was during my graduate studies and I would like to thank you all from the bottom of my heart for shaping me into the scientist and science advisor that I am today.

Dedication

I dedicate this thesis to my son, Adhvik. Ever since I found out you were coming into our lives, I've tried my best to make the world a slightly better place for you to grow up in. For the past two and a half years, I've watched you grow up and had the opportunity to see the world through your eyes. It helped me regain that feeling of wonderment and beauty amongst all the chaos and destruction from climate change. I want to encourage your curiosity and penchant for asking questions. I'd like to foster your interest in science and to think creatively. My hope is that one day you will read this thesis in a better world where I'll be able to take you to the Arctic and experience my journey again with you by my side.

I love you.

Contribution of Authors

In Chapter 2, Bill Li, Kevin Pauley and Jeff Anning provided logistical support in collecting samples from Bedford Basin. Paul del Giorgio provided the bacterial abundance data and Roxane Maranger provided the bacterial production data for the SLE. Ludmila Chistoserdova provide the M. mobilis culture. Arthi Ramachandran performed all experiments and associated data analysis for this manuscript. Ludmila Chistoserdova and Hans-Peter Grossart commented on an earlier version of the manuscript. Arthi Ramachandran and David Walsh contributed to writing the manuscript.

In Chapter 3, the samples were collected by David A. Walsh. David Colatriano extracted the environmental genomic DNA. Susan McLatchie performed the Quantitative PCR (qPCR) targeting the BS01 or OM43 xoxF4 genes. All other sequence, data, and bioinformatic analyses was performed by Arthi Ramachandran. Arthi Ramachandran and David A. Walsh contributed to writing the manuscript.

In Chapter 4, William K.W. Li and his team designed the sampling strategy and collected the samples. Arthi Ramachandran and David A. Walsh selected the samples and designed the quality control process. Arthi Ramachandran carried out the filter-PCRs and gDNA PCRs and prepared the samples for subsequent 16S rRNA sequencing. Arthi Ramachandran performed the amplicon data processing, hierarchical clustering, taxonomic assignment, and all *Methylophilaceae* analyses. Susanne Kraemer performed the Chao 1 richness, alpha diversity variance, general dissimilarity models, and the TITAN analyses. Vera Onana performed the Random Forest analysis. David A. Walsh, Susanne Kraemer, and Arthi Ramachandran contributed to writing the manuscript. All authors commented on the manuscript.

Table of Contents

List of Figures ... *xi*

List of Supplementary Figures ... *xv*

List of Supplementary Tables ... *xvi*

List of Appendices Figures and Tables ... *xvii*

List of abbreviations ... *xix*

Chapter 1. Introduction ... *1*

1.1. Methylotrophy ... **2**

1.1.1. Microbial methylotrophy ... 2
1.1.2. Sources of methanol ... 3

1.2. OM43/LD28 methylotrophs ... **4**

1.2.1. Marine methylotrophy ... 4
1.2.2. Freshwater methylotrophy ... 7
1.2.3. Diversity of Beta-proteobacterial methylotrophs across ecosystems ... 7
1.2.4. Metabolic versatility in OM43/LD28 methylotrophs ... 8

1.5. Northern aquatic ecosystems ... **10**

1.5.1. The Saint-Lawrence Estuary ... 10
1.5.2. The Bedford Basin ... 10
1.5.3. The Arctic Ocean ... 11

1.6. Rationale ... **12**

2. Investigation of XoxF methanol dehydrogenases reveals new methylotrophic bacteria in pelagic marine and freshwater ecosystems ... *14*

2.1. Abstract ... **14**

2.2. Introduction ... **15**

2.3. Materials and Methods ... **17**

2.3.1. Sampling and environmental DNA and RNA extraction ... 17
2.3.2. *xoxF4* PCR primer design and amplification ... 18
2.3.3. Cloning and sequencing of *xoxF4* fragment ... 19
2.3.4. Quantification of *xoxF4* gene and mRNA abundances ... 20
2.3.5. Metagenomic analysis of PQQ-dependent methanol dehydrogenases ... 21

2.4. Results ... **21**

2.4.1. *xoxF4* PCR primer design and specificity ... 21
2.4.2. *xoxF4* diversity in aquatic environments ... 22
2.4.3. Abundance and expression of *xoxF4* genes ... 24
2.4.4. Comparative metagenomics of *mxaF* and *xoxF* diversity and biogeography ... 25

2.5. Discussion ... **28**

2.5.1. OM43 methylotroph abundance and activity in coastal habitats ... 28
2.5.2. Expanding the diversity of putative marine methylotrophs ... 30
2.5.3. Rare-earth elements and their relationship to methylotrophs ... 31
2.5.4. Freshwater and estuarine pelagic methanol-oxidizing bacteria ... 31

2.6. Acknowledgements ... **33**

2.7. Funding ... **33**

2.8. Supplementary Figures and Tables ... **34**

Bridging text ... ***39***

Chapter 3. A novel freshwater to marine evolutionary transition revealed within Methylophilaceae bacteria from the Arctic Ocean ... ***40***

3.1. Abstract ... **40**

3.2. Importance ... **41**

3.3. Introduction ... **41**

3.4. Results ... **44**

3.4.1. Environmental context ... 44
3.4.2. *Methylophilaceae* in the Canada Basin ... 45
3.4.3. Genomic characteristics of BS01 ... 49
3.4.4. Gene content variation and marine adaptation in BS01 ... 51
3.4.5. BS01 energy and nutrient metabolism ... 52
3.4.6. BS01 biogeography across aquatic ecosystems ... 55

3.5. Discussion ... **57**

3.5.1. *Methylophilaceae* diversity in the salinity-stratified Arctic Ocean ... 57
3.5.2. Freshwater-marine transitions within the *Methylophilaceae* ... 58
3.5.3. Niche differentiation of BS01 and OM43 ... 60
3.5.4. Bacterial evolution in a changing Arctic Ocean ... 62

3.6. Materials and Methods ... **63**

3.6.1. Sampling and metagenomic data generation ... 63
3.6.2. 16S rRNA gene and ITS analysis ... 63
3.6.3. MAG generation and analysis ... 64
3.6.4. Concatenated protein phylogeny ... 65
3.6.5. Comparative genomics ... 65
3.6.6. Metatranscriptomic analysis ... 65
3.6.7. Fragment recruitment ... 67
3.6.8. *xoxF4* primer design and qPCR ... 67

3.7. Acknowledgements ... **67**

3.8. Supplementary Figures and Tables ... **68**

Bridging text ... ***73***

Chapter 4. A 2004-2012 time series of bacterial and archaeal community dynamics in a changing Arctic Ocean ... ***74***

4.1. Abstract ... **74**

4.2. Main Text ... **74**

4.3. Materials and Methods ... **80**

4.3.1. Environmental data and sample collection ... 80
4.3.2. Filter PCR of 16S rRNA gene amplicons ... 81
4.3.3. Amplicon data processing ... 82
4.3.4. Statistical analyses ... 82

4.3. Acknowledgments ... **84**

4.4. Competing Interests ... **84**

4.5. Supplementary Figures and Tables **84**

Chapter 5. Conclusions, discussions, and future directions* *91

5.1. *Methylophilaceae*: Sentinels of Change? **91**

5.2. Importance of Ecotypes in Understanding Change **92**

5.3. Microbial Evolution Associated to a Changing Arctic Ocean and the Potential Global Implications **93**

5.4. Future Directions: Linking Microbial Research in Whole Ecosystem Studies **95**

References* *97

Appendix A: Quality Control and Assessment of the filter-PCR method, and Additional Analyses of the Community Composition* *107

A.1. Quality Control and Assessment of the Methodology **107**

A.1.1. Assessment of the filter PCR approach 107

A.1.2. Community structure of filter vs gDNA replicates at Stations CB9 and CB15 in 2017 108

A.1.3. Community structure of filter replicates at Stations CB9 and CB15 in 2004 and 2012 109

A.1.4. PCR variation using genomic DNA from 2017 samples 110

A.2. Additional Analyses of the Community Composition **111**

A.2.1. Overall Community Composition 111

A.2.2. Niche Breadth Analysis of the Whole Community Composition 114

A.2.3. Diversity and distribution of the Methylophilaceae family 118

A.3. Discussion **121**

Appendix B: Characterization of novel methylotrophs isolated from the Western Arctic Ocean* *124

B.1. Introduction **124**

B.2. Enrichment and Cultivation **127**

B.3. Sequencing and Exploring the Metagenome Assembled Genomes (MAGs) **129**

B.3.1. NAD-dependent alcohol dehydrogenase: Potential for methanol oxidation? 131

B.4. Conclusion and Future Studies **133**

List of Figures

Figure 1.1 Overview of methanol production and pathways into the ocean. Black lines show potential sources of methanol and red lines show the loss of potential methanol. 4

Figure 2.1 Map of the coastal Atlantic Ocean and St Lawrence Estuary showing station locations (circles) used in this study. The inset map at the lower left shows the study region (dashed box) in the broader context of North America. 18

Figure 2.2 Phylogenetic analysis of partial (~500 bp) xoxF4 gene sequences PCR-amplified from Bedford Basin and Stations B and 22 in the St Lawrence Estuary using primer pair 974F/1441R. The tree was inferred using maximum likelihood and a GTR + Gamma distribution (four categories) model of evolution. Clone identifiers for sequences produced in this study are presented in parentheses. 23

Figure 2.3 Abundance of xoxF4 genes and mRNA transcripts from seasonal samples collected in Bedford Basin and along the salinity gradient of the St Lawrence Estuary. Error bars indicate standard deviation. 24

Figure 2.4 Phylogenetic analysis of PQQ-dependent methanol dehydrogenase MxaF/XoxF proteins. Protein sequences generated from our study are in red, while those from lake and ocean metagenomes are in blue. Some methylotrophs have multiple homologs of XoxF that can branch into different clades. The following parameters were used: maximum-likelihood method, 100 bootstrap iterations, JTT substitution model, gamma distribution model for the rate variation with four discrete gamma categories. 27

Figure 2.5 Survey of PQQ-dependent methanol dehydrogenase genes across aquatic metagenomes.. 27

Figure 3.1 Phylogenetic analysis of 16S rRNA genes from Methylophilaceae from Canada Basin metagenomes and a diversity of aquatic ecosystems. The tree was inferred using maximum likelihood (500 bootstraps) and GTR + gamma distribution (four categories) with invariant site model of evolution and the nearest-neighbor interchange heuristic search method. The tree was rooted using Methylibium as an outgroup to the Methylophilaceae. Sequences from the current study are highlighted in red. Only bootstrap values of >60 are included in the tree. 46

Figure 3.2 Diversity and biogeography of Methylophilaceae based on ITS variants recovered from Canada Basin metagenomes. (A) Phylogenetic analysis of the Methylophilaceae group across various aquatic regions and depths using the ITS region. The tree was inferred using maximum likelihood (500 bootstraps) and a GTR + gamma distribution (four categories) with invariants sites model of evolution and the nearest-neighbor interchange heuristic search method. Sequences from the current study are highlighted in red. Only bootstrap values of >60 are included in the tree. (B) Abundance of six ITS variants based on summed coverage in metagenome assemblies. (C) Principal coordinate analysis ordination of Bray-Curtis dissimilarities of Arctic samples based on summed coverage of six ITS variants. 48

Figure 3.3 Phylogenomic comparison of BS01 with representative Methylophilaceae genomes. (A) Maximum likelihood phylogenetic analysis of a concatenated alignment of 48 orthologs shared between all Methylophilaceae genomes. Values at the nodes are bootstrap values (100 pseudoreplicates). (B) Scatterplot comparing G+C content and genome size. (C) Whole-proteome pI values versus relative frequency in select Methylophilaceae genomes from freshwater and marine habitats. 50

Figure 3.4 Reconstruction of methylotrophic metabolism in BS01 and comparison to other Methylophilaceae. (A) Distribution of central one-carbon metabolism genes. (B) Gene expression pattern for central carbon metabolism pathways in Canada Basin surface waters revealed through fragment recruitment of metatranscriptomics against Met-BS01-1 and HTCC2181 genomes. (C) Quantification of BS01 and OM43 abundances in the Canada Basin using qPCR analysis of xoxF4 gene abundance. Error bars indicate standard deviation. DCM, deep chlorophyll maximum; PWW, Pacific winter water. 53

Figure 3.5 Reconstruction of nitrogen metabolism in BS01 and comparison to other Methylophilaceae. (A) Distribution of central nitrogen metabolism genes. (B) Gene expression pattern for central nitrogen metabolism pathways in Canada Basin surface waters revealed through fragment recruitment of metatranscriptomics against Met-BS01-1 and HTCC2181 genomes. ... 55

Figure 3.6 Biogeography of BS01 across the global ocean and estuaries. (A) Phylogenetic analysis of the Methylophilaceae family across multiple aquatic regions and depths using the XoxF4 methanol dehydrogenase protein recovered from metagenomes. The tree was inferred using maximum likelihood (500 bootstraps) and JTT + gamma distributed with invariants (four categories) sites model of evolution, with nearest-neighbor interchange heuristic search method. Colored sequences are those from the Arctic Ocean (red), the Antarctic (green), or estuaries (blue). Only bootstrap values of >60 are included in the tree. (B) Distribution of BS01 and OM43-A revealed through fragment recruitment of aquatic metagenomes against Met-BS01-1 and HTCC2181 genomes reported as reads per kilobase of the MAG per gigabase of metagenome (RPKG). The diagonal lines are to signify no metagenome data for that water column feature.. 56

Figure 4.1 The change in environmental conditions in the Canada Basin over the 2004-2012 time-series: (a) Map of the Arctic Ocean showing station locations (circles) included in this study; (b) Variability in summer sea ice extent in the Arctic Ocean, data is from the National Snow and Ice Data Center; (c) the latitudinal gradient in sea surface salinity over the time series; (d) stratification index, error bars indicate among station standard deviation; (e) ratio of picophytoplankton to nanophytoplankton; (f) bacterioplankton. In panels e-f, values are the average of all measurements >15 m in depth and salinity < 31 PSU, and error bars indicate among depth and among-station standard error. 76

Figure 4.2 The change in bacterial and archaeal communities in the Canada Basin over the 2004-2012 time series: (a) Partial dependency plots of changes in Chao1 richness predicted across the time series by the random forest approach; (b) The magnitude and rate of change in community composition (taxonomic turnover) across the time series as predicted by generalized dissimilarity modelling; (c) the number and taxonomic composition of bacterial and archaeal ASVs that

significantly increased or decreased across the time series. SML, surface mixed layer; UAW, upper Arctic water; PW, Pacific water... 77

List of Tables

Table 2.1 Environmental data from sample locations in BB and SLE 23

Table 3.1 Location and environmental characteristics of samples collected for metagenomic analyses 45

List of Supplementary Figures

Supplementary Figure 2.1 Primer design technique 34

Supplementary Figure 3.1 Map of the study location in the Canada Basin, Arctic Ocean........ 68

Supplementary Figure 3.2 Comparative genomics of shared gene content among Methylophilaceae strains visualized using Anvi'o. 69

Supplementary Figure 3.3 Summary of the distribution of metabolism modules across Methylophilaceae genomes.. 70

Supplementary Figure 4.1 A hierarchical clustering analysis of samples based on the ASV distributions using the hclust function in stats package for R with a Bray-Curtis dissimilarity metric. The relative abundance of taxa that were 3.5% or more of the community composition were plotted next to the clustered samples. 85

Supplementary Figure 4.2 (A) Phylogenetic analysis of 16S rRNA genes from Methylophilaceae from Canada Basin time series, metagenomes and a diversity of aquatic ecosystems. (B) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic samples based on the relative abundance of Methylophilaceae ASVs. 86

Supplementary Figure 4.3 The relative abundance of Methylophilaceae ASVs across the three biomes (SML, UAW, PW), nine years (2004-2012), and four sites (CB9, CB15, CB21, CB29).87

List of Supplementary Tables

Supplementary Table 2.1 Metagenomic sample sites 34

Supplementary Table 3.1 Characteristics of Methylophilaceae genomes. 70

Supplementary Table 3.2 List of aquatic metagenomes used in the fragment recruitment analysis. 71

Supplementary Table 3.3 Gene expression pattern for central carbon metabolism genes in Canada Basin surface waters revealed through fragment recruitment of metatranscriptomics against Met-BS01-1 and HTCC2181 genomes. 71

Supplementary Table 3.4 Gene expression pattern for central nitrogen metabolism genes in Canada Basin surface waters revealed through fragment recruitment of metatranscriptomics against Met-BS01-1 and HTCC2181 genomes. 72

Supplementary Table 4.1 Sample information and environmental data. Values in red were missing from the original dataset and were interpolated as average values from the same depths or water masses. 88

Supplementary Table 4.2 Sum of squares based on anova of Chao1 richness as a function of sampling year, latitude, and depth for each water mass. Bold values indicate significate covariates. 88

Supplementary Table 4.3 RF results indicating relative importance (as % of explained variation) of the explanatory variables on bacterioplankton diversity. 89

Supplementary Table 4.4 GDM results including the model deviance, deviance explained, and predictor importance. Bold values indicate significant predictors. 89

Supplementary Table 4.5 Summary of threshold indicator analysis with year as the environmental gradient variable. 90

List of Appendices Figures and Tables

Figure A.1 A hierarchical clustering analysis of filter-PCR vs gDNA PCR for 3 sampling depths, based on the ASV distributions using the hclust function in stats package for R with a Bray-Curtis dissimilarity metric. The relative abundance of taxa was plotted next to the clustered samples. 109

Figure A.2 (A) A beta diversity analysis showing the variation between the samples and within the samples (B) A hierarchical clustering analysis of surface replicate filter-PCRs for 2 sampling years, based on the ASV distributions using the hclust function in stats package for R with a Bray-Curtis dissimilarity metric. The relative abundance of taxa greater than or equal to 0.01% was plotted next to the clustered samples. 110

Figure A.3 A hierarchical clustering analysis of gDNA PCR replicates based on the ASV distributions using the hclust function in stats package for R with a Bray-Curtis dissimilarity metric. The relative abundance of taxa was plotted next to the clustered samples. 111

Figure A.4 (A) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic samples based on the relative abundance of ASVs (B) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic SML samples based on the relative abundance of ASVs (C) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic UAW samples based on the relative abundance of ASVs (D) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic PW samples based on the relative abundance of ASVs........ 112

Figure A.5 A hierarchical clustering analysis of samples based on the ASV distributions using the hclust function in stats package for R with a Bray-Curtis dissimilarity metric. The relative abundance of taxa that were 3.5% or more of the community composition were plotted next to the clustered samples. 114

Figure A.6 Specialist-generalist classification of ASVs was based on Levin's Index (B). The function spec.gen from R package EcolUtils was used to calculate B for random 100 permutations of the ASV table and categorized ASVs into generalists or specialists. (A) Niche-breadth analysis of the SML (B) Niche-breadth analysis of the UAW (C) Niche-breadth analysis of the PW.... 116

Figure A.7 Specialist-generalist classification of ASVs was based on Levin's Index (B). The function spec.gen from R package EcolUtils was used to calculate B for random 1000 permutations of the ASV table and categorized ASVs into generalists or specialists........ 118

Figure A.8 (A) Phylogenetic analysis of 16S rRNA genes from Methylophilaceae from Canada Basin time series, metagenomes and a diversity of aquatic ecosystems. (B) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic samples based on the relative abundance of Methylophilaceae ASVs........ 120

Figure A.9 The relative abundance of Methylophilaeae ASVs across the three biomes (SML, UAW, PW), nine years (2004-2012), and four sites (CB9, CB15, CB21, CB29)........ 121

Figure B.1 General workflow for isolating and sequencing methylotroph strains. 128

Figure B.2 General workflow: (A) Sampling sites (B) Incubation of seawater from the surface waters and the subsurface chlorophyll maximum (SCM), supplemented with methanol or methanol and urea. (C) 1mL of supplemented seawater with the addition of a rare earth metal plated on 48-well plates and incubated at 4°C. (D) Growth of cultures is measured using flow cytometry... 128

Figure B.3 Phylogenetic analysis of 16S rRNA genes from the Sulfitobacter MAG from Canada Basin and a diversity of aquatic ecosystems. The tree was inferred using maximum likelihood (500 bootstraps) and GTR 1 gamma distribution (four categories) with invariant site model of evolution and the nearest-neighbor interchange heuristic search method. Sequences from the current study are highlighted in red. Only bootstrap values of .60 are included in the tree. 129

Figure B.4 Phylogenetic analysis of 16S rRNA genes from the Thalassospira MAG from Canada Basin and a diversity of aquatic ecosystems. The tree was inferred using maximum likelihood (500 bootstraps) and GTR 1 gamma distribution (four categories) with invariant site model of evolution and the nearest-neighbor interchange heuristic search method. Sequences from the current study are highlighted in red. Only bootstrap values of .60 are included in the tree. 130

Figure B.5 Phylogenetic analysis of the NAD-dependent methanol dehydrogenase protein recovered from the MAGs. The tree was inferred using maximum likelihood (500 bootstraps) and JTT 1 gamma distributed with invariants (four categories) sites model of evolution, with nearest-neighbor interchange heuristic search method. Colored sequences are those from this study. Only bootstrap values of .60 are included in the tree. 132

Table B.1 Assembled and binned metagenome assembled genomes (MAGs) from Arctic Ocean cultures 131

List of abbreviations

AAI	Average amino acid identity
ADH	Alcohol dehydrogenase
ASV	Amplicon sequence variant
AOA	Ammonia-oxidizing Archaea
BB	Bedford Basin
BLAST	Basic local alignment search tool
C1	One-carbon
CB	Canada Basin
CCGS	Canadian Coast Guard Ship
cDNA	Complementary deoxyribonucleic acid
CTD	Conductivity, Temperature and Density
DAPI	4',6-diamidino-2-phenylindole
dbRDA	Distance-based redundancy analysis
DCM	Deep chlorophyll maximum
DGGE	Denaturing gradient gel electrophoresis
DMSP	Dimethylsulfoniopropionate
DNA	Deoxyribonucleic acid
DNA-SIP	Deoxyribonucleic acid Stable isotope probing
dNTPs	Deoxynucleoside triphosphates
DOM	Dissolved organic matter
DOE	Department of Energy
FISH	Fluorescence in situ hybridization
gDNA	Genomic deoxyribonucleic acid
GS/GOGAT	Glutamine synthetase/glutamine oxoglutarate aminotransferase
GTR	General time reversible
HMW	High molecular weight
H_4F	Tetrahydrofolate
H_4MPT	Tetrahydromethanopterin
IMG	Integrated Microbial Genomes
IMG/M	Integrated Microbial Genomes / Metagenomes

ITS	Internal transcribed spacer
JOIS	Joint Ocean Ice Study
JTT	Jones-Taylor-Thornton
LB	Lysogeny broth
LGT	Lateral gene transfer
MAG	Metagenome Assembled Genome
MDH	Methanol dehydrogenase
MeOH	Methanol
MG	Marine group
mRNA	Messenger ribonucleic acid
miRNA	Micro ribonucleic acid
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide + hydrogen
NADP	Nicotinamide adenine dinucleotide phosphate
NCBI	National Center for Biotechnology Information
NDH	NADH dehydrogenase
NNI	Nearest-neighbor interchange
NQR	NADH:quinone oxidoreductase
NSERC	Natural Science and Engineering Research Council
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
pI	Isoelectric point
PQQ	Pyrroloquinoline quinone
PSU	Practical salinity units
PW	Pacific waters
PWW	Pacific winter waters
qPCR	Quantitative polymerase chain reaction
REE	Rare earth element
RF	Random Forest
RFLP	Restriction fragment length polymorphism
RPKG	Reads per kilobase of the MAG per gigabase of metagenome

rRNA	Ribosomal ribonucleic acid
RT-qPCR	Reverse transcriptase quantitative polymerase chain reactionc
RuMP	Ribulose monophosphate
SCM	Subsurface chlorophyll maximum
SIP	Stable isotope probing
SLE	Saint Lawrence Estuary
SML	Surface mixed layer
TMAO	Trimethylamine N-oxide
UAW	Upper Arctic waters

Chapter 1. Introduction

The ability of aquatic ecosystems to persist through disturbances, while maintaining essential ecosystem functions, is being increasingly tested as the world faces high rates of environmental change associated to climate change and other anthropogenic stressors (Philippot, Griffiths and Langenheder 2021). A key part of shaping ecosystem persistence is microbial resilience, adaptation, and evolution. Microbes provide critical ecosystem services in aquatic environments. They are essential for nutrient cycling, carbon sequestration, decomposition, primary production, water purification, and climate regulation (Ducklow 2008; Gupta, Gupta and Singh 2017). Therefore, understanding the complex links between microbial change and the services they provide will help ecosystem functioning.

Coastal ecosystems are amongst the most productive and have been heavily subjected to unprecedented environmental changes. Over the past decade, the Saint-Lawrence Estuary in eastern Quebec has been facing milder, almost ice-free winters, early springs, and increased warming of summer surface waters (Galbraith *et al.* 2022). The Bedford Basin, a temperate North-West Atlantic inlet located in Nova Scotia, has also been experiencing higher rates of warming and changes to the phytoplankton community composition (Robicheau *et al.* 2022). The Arctic Ocean has been undergoing rapid changes, warming four times faster than the global rate, leading to increased freshening (Timmermans, Toole and Krishfield 2018; Wang *et al.* 2018; Fu *et al.* 2020). One of the main environmental stressors in marine environments is change in salinity. It is a major factor that controls the distribution and diversity of organisms. The Saint-Lawrence Estuary allows us a unique environment to study salinity across a spatial gradient. The Arctic Ocean surface waters are undergoing long-term freshening due to increase sea ice melt and river

runoff. This provides us with the opportunity to study adaptation and evolution linked to salinity over a temporal gradient.

To better study the consequences of environmental changes, it is important to understand the diversity, distribution, ecology, and metabolism of bacteria living in these rapidly changing environmental conditions. In this thesis, methylotrophic Beta-proteobacteria, along with other important bacterial communities, were investigated in three different aquatic ecosystems. They are of particular interest because of their known ecotypic diversity and distribution across various habitats. The discovery of unique ecotypes across a salinity gradient in the Saint-Lawrence Estuary, the Bedford Basin, and the Arctic Ocean allows us to decipher patterns of adaptation and evolution.

1.1. Methylotrophy

1.1.1. Microbial methylotrophy

Methylotrophs are a taxonomically diverse group of microorganisms that use reduced one-carbon (C1) compounds as their sole carbon and energy source (Chistoserdova and Lidstrom 2013). Methylotrophs degrade C1 compounds, like methane, methanol, methylamine, and formaldehyde, using different pathways for their oxidation, demethylation, and assimilation into biomass (Neufeld *et al.* 2008a; Chistoserdova, Kalyuzhnaya and Lidstrom 2009; Chistoserdova 2011a; Halsey, Carter and Giovannoni 2012; Chistoserdova and Lidstrom 2013). Methanotrophs have received most of the attention because of its implication in methane, a potent greenhouse gas, metabolism, even though they only account for a fraction of methylotroph diversity. Marine methylotrophs play an important role in the metabolism and assimilation of C1 compounds like methanol and methylated compounds containing amino, halide, and/or sulfur moieties in the oceans (Neufeld *et al.* 2008a). Most of the previous work done on methylotrophs has focused on

cultured representatives in the Alpha- and Gamma-proteobacterial groups. However, this early work overlooked a group of marine Beta-proteobacteria within the *Methylophilaceae* family which play an important role in the metabolism and assimilation of C1 compounds in the world's oceans (Giovannoni *et al.* 2008; Huggett, Hayakawa and Rappé 2012; Jimenez-Infante *et al.* 2015).

1.1.2. Sources of methanol

Methanol is an abundant oxygenated volatile organic compound in the ocean where it can serve as a carbon and energy source for methylotrophic microorganisms. In the oceans, methanol concentrations range from 50-400 nM with turnover times as low as one day (Dixon, Beale and Nightingale 2011; Dixon and Nightingale 2012). Potential sources of methanol in the ocean include production in anoxic bottom waters and sediments (Millet *et al.* 2008). Methanol is also produced through the degradation of methylated sugars and hydrolysis of methyl halides (Millet *et al.* 2008). Methanol is linked to primary production with C1 units either being directly released by phytoplankton or as a result of bacterial breakdown of algal carbohydrates (Heikes 2002; Mincer and Aicher 2016). Methanol is a main degradation product of pectin and lignin and is therefore associated with plant growth and decay (Heikes 2002; Millet *et al.* 2008). These C1 compounds are also transported to the surface waters from terrestrial sources such as biomass burning, atmospheric production, and anthropogenic sources (Heikes 2002; Millet *et al.* 2008; Dixon, Beale and Nightingale 2011) (**Figure 1.1**). In fact, a study conducted in the Northwest Passage showed that the high Arctic waters are a sink for atmospheric methanol, which could potentially increase as sea-ice extent decreases (Sjostedt *et al.* 2012).

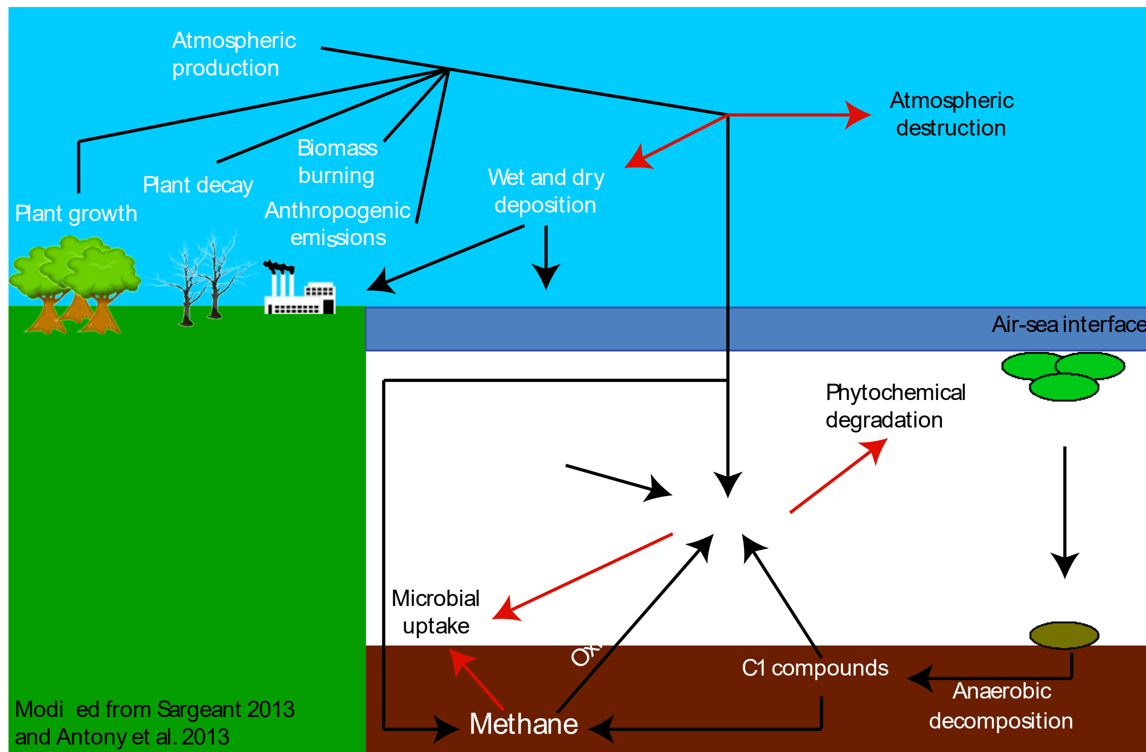


Figure 1.1 *Overview of methanol production and pathways into the ocean. Black lines show potential sources of methanol and red lines show the loss of potential methanol.*

1.2. OM43/LD28 methylotrophs

1.2.1. Marine methylotrophy

Although methylotrophy has been studied in the oceans for decades, a combination of cultivation, genomic, and proteomic studies recently demonstrated that the Beta-proteobacteria OM43 clade of marine *Methylophilaceae* bacteria represent a previously overlooked component of the methylotrophic marine community (Morris, Longnecker and Giovannoni 2006; Giovannoni *et al.* 2008; Sowell *et al.* 2011; Halsey, Carter and Giovannoni 2012; Georges *et al.* 2014; Jimenez-Infante *et al.* 2015; Taubert *et al.* 2015). Among the marine methylotrophs, the OM43 clade belonging to the Beta-proteobacteria occurs abundantly in productive aquatic environments from coastal waters to brackish and freshwater ecosystems (Giovannoni *et al.* 2008; Huggett, Hayakawa and Rappé 2012; Salcher *et al.* 2015, 2019; Sosa *et al.* 2015). Surveys in the Western Atlantic

Ocean determined that OM43 clade members represented 5% of the bacterial clones retrieved [27]. Furthermore, studies showed that about 4% of stained cells could be attributed to the OM43 clade in the North Sea (Sekar *et al.* 2004). Based on 16S rRNA surveys, the OM43 clade is more abundant and common than other previously described marine methylotrophs (Sekar *et al.* 2004; Morris, Longnecker and Giovannoni 2006; Song, Oh and Cho 2009; Taubert *et al.* 2015). Previous studies have also indicated that the abundance and distribution of this group seem to be associated with phytoplankton blooms(Sowell *et al.* 2011; Georges *et al.* 2014). Genome analysis of OM43 isolates have demonstrated that they have a highly reduced genome, are specialists for C1 compounds, and possess a unique set of methanol oxidation genes compared to previously described methylotrophs (Giovannoni *et al.* 2008; Halsey, Carter and Giovannoni 2012). Environmental proteomic studies have shown the prevalence of these OM43 methanol oxidation genes, particularly methanol dehydrogenase (XoxF). These are commonly expressed in coastal surface waters, where XoxF accounted for up to 3% of all identified peptides, suggesting that the OM43 clade may be responsible for a considerable amount of methanol turnover in the coastal ocean (Sowell *et al.* 2011; Georges *et al.* 2014).

A powerful method to study the physiology of microbial cells is to isolate and cultivate strains from any given community to better study their metabolic capabilities. Cultivated strains can provide complete genome sequences offering insights into their potential metabolism and are useful for interpreting other genomic datasets (Giovannoni and Stingl 2007). Marine OM43 bacteria were first isolated from their respective aquatic environments using dilution to extinction cultivation techniques (Connon and Giovannoni 2002) with sterilized ocean water serving as the growth medium (Connon and Giovannoni 2002; Salcher *et al.* 2015; Sosa *et al.* 2015). These

studies demonstrated that natural marine environments contain adequate growth substrates and nutrients to sustain these methylotrophic bacteria.

Methylotrophic Beta-proteobacteria are globally distributed and express high ecotypic diversity across habitats. This means that there are genetically distinct geographic populations of methylotrophic Beta-proteobacteria that are adapted to specific environmental conditions. To date strains from the marine OM43 clade have been isolated from a variety of habitats: strain HTCC2181 from the Oregon coast (Giovannoni *et al.* 2008), strain HIMB624 (or KB13) from the coast of Hawaii (Huggett, Hayakawa and Rappé 2012), strain MBRS-H7 from the Red Sea (Jimenez-Infante *et al.* 2015), and strains NB0016 and NB0046 from Nahant Bay (Sosa *et al.* 2015). Even though these strains contain some of the smallest genomes among the free-living bacteria, they share about two-thirds of their genomic content with each other (Jimenez-Infante *et al.* 2015; Salcher *et al.* 2015). There is a very high level of genome and, most likely, functional conservation seen in OM43 genotypes from contrasting oceanic habitats giving rise to potential niche adaptations with differences in their pangenome (Jimenez-Infante *et al.* 2015). A study conducted on the Red Sea by Jimenez-Infante et al. (2015) revealed ecotypic separation of the marine OM43 clade members (HTCC2181, HIMB624, MBRS-H7) through phylogenetic analysis of the 16S-23S internal transcribed spacer (ITS) and metagenomic fragment recruitment. These isolates can be distinguished as open ocean (HIMB624 and MBRS-H7) and coastal ocean (HTCC2181 and NB strains) ecotypes. A clear dominance of members of the open ocean clade was seen in higher temperatures and less productive environments whereas the coastal ocean clade showed the opposite trend (Jimenez-Infante *et al.* 2015). At the genome level, several differences have been found among the OM43/LD28 clade, which can account for the ecological success of different subclades in divergent aquatic habitats.

1.2.2. Freshwater methylotrophy

Less is known about methanol oxidation in the surface waters of lakes and rivers than there is in oceans, although a large body of work exists on benthic methylotrophy in freshwater sediments (Kalyuzhnaya *et al.* 2012; Chistoserdova and Lidstrom 2013). Given that terrestrial plants and phytoplankton are sources of methanol and methylated compounds in the ocean (Milne et al. 1995; Heikes 2002b; Millet et al. 2008b; Mincer and Aicher 2016b), methylotrophy is likely important in the surface waters of lakes as well. The occurrence of close relatives of the marine OM43 clade in freshwater habitats, which form a closely related, but separate, lineage known as the LD28 clade (Giovannoni *et al.* 2008; Salcher *et al.* 2015, 2019), also implies that OM43-like organisms have a broad biogeographical distribution. In fact, 16S rRNA gene sequences affiliated with OM43 are common in lakes of Europe and North America (Zwart *et al.* 1998, 2003; Salcher *et al.* 2008; Jezbera *et al.* 2012). LD28 and OM43 are both members of the broader betIV clade (Newton *et al.* 2011), which also contains *Methylophilus* and *Methylotenera* strains found in freshwater sediments and soils (Chistoserdova 2011a). Both *Methylophilus* and *Methylotenera* can use a range of one carbon compounds (Chistoserdova 2011a) and some may also be able to subsist on phenol and humic matter (Hutalle-Schmelzer *et al.* 2010). Isolates from two planktonic freshwater methylotrophic lineages (LD28 and closely affiliated PRD01a001B) were cultured from Lake Zurich, Switzerland (Salcher *et al.* 2015). Subsequent genome analysis confirmed a methylotrophic lifestyle and a reduced set of genes for C1 metabolism (Salcher *et al.* 2015).

1.2.3. Diversity of Beta-proteobacterial methylotrophs across ecosystems

The methylotrophic lineages, LD28 and OM43, that are common to both freshwater and marine ecosystems are a good model family to study diversity, distribution, and activity across a range of aquatic ecosystems. A common feature of the OM43/LD28 methylotrophs is that they

possess a unique methanol dehydrogenase gene, *xoxF4*, that can be used as a specific functional gene marker for environmental studies (Chistoserdova *et al.* 2007; Giovannoni *et al.* 2008; Salcher *et al.* 2015, 2019; Sosa *et al.* 2015). A study was conducted on the basic physiology and genomic traits of an OM43 isolate recovered from the ultraoligotrophic environment of the Red Sea (Jimenez-Infante *et al.* 2015). The authors found biogeographical distribution patterns that are co-related to physiochemical conditions of specific environments, opening the door for further research questions about variability between environments and conditions (Jimenez-Infante *et al.* 2015). These results demonstrated a geographical distribution of OM43/LD28-related methylotrophs across both marine and freshwaters. This suggests that methylotrophy occurring in the water column is an important component of lake, estuary, and marine carbon cycling and biogeochemistry.

1.2.4. Metabolic versatility in OM43/LD28 methylotrophs

The OM43/LD28 clade of bacteria were believed to be obligate methylotrophs, but now it is known that they are also commonly enriched for under a variety of conditions. For example, a study conducted by Sosa et al. (2015) (Sosa *et al.* 2015) expanded on the dilution to extinction technique by enrichment with high molecular weight (HMW) dissolved organic matter (DOM) collected directly from ocean surface waters. They discovered that the most abundant isolates belonged to the OM43 clade revealing previously unsuspected relationships between naturally occurring organic polymers and the microorganisms that consume them (Sosa *et al.* 2015). A follow up study conducted by Gifford et al. (2016) (Gifford *et al.* 2016) revealed that abiotic release of methanol from HMW DOM is negligible and therefore it is likely that the growth using this organic polymer is due to enzymatic cleavage of the carbon substrate directly by the methylotrophs. Similarly, enrichment and cultivation of pelagic freshwater bacteria was performed

by Hutalle-Schmelzer et al. (2010) (Hutalle-Schmelzer *et al.* 2010). They enriched bacterial populations, including LD28 methylotrophs, from a humic lake using phenol and humic matter. These studies demonstrate that organic polymers in both marine and freshwater environments can serve as important growth substrates for methylotrophic bacteria. However, a comprehensive mechanistic understanding to tease apart the complex factors allowing methylotrophs to access these growth substrates remains to be determined.

A study using quantitative transcriptomics showed that OM43 may upregulate the Ribulose monophosphate (RuMP) pathway when nutrient supplies allow methanol to be used for both energy generation and biosynthesis (Gifford *et al.* 2016). In this study, it was also determined that the metabolic range of the OM43 strain (NB0046) has potentially expanded to include other carbon substrates. However, this only holds true for one marine isolate grown and tested in the lab. There is no evidence that the same upregulation would occur for other OM43/LD28 strains. Freshwater habitats are also more spatially and temporally dynamic than marine water bodies. There is far more terrestrial input and that changes the source of the organic matter found in the surface waters and changes the potential for these organisms to use terrestrial organic polymers.

While the OM43, HTCC2181 strain, can grow on methanol as a sole carbon and energy source, this strain also exhibits synergistic growth when methanol is supplemented with additional methylated compounds (i.e. Dimethylsulfoniopropionate (DMSP) and Trimethylamine N-oxide (TMAO)) (Halsey, Carter and Giovannoni 2012). It has also been shown that the addition of methylamine increased cell yield and led to significantly higher cell numbers in freshwater methylotrophs found in Lake Zurich (Salcher *et al.* 2015). In addition, Gifford et al. (2016) revealed that in marine methylotrophs high expression levels of the *xoxF* gene was correlated with nitrogen availability, therefore suggesting that favourable growth would be seen using a

methylamine amended media. The study illustrated that the OM43, NB0046 strain, can grow on HMW DOM amended media without the addition of methanol (Gifford *et al.* 2016). There are many unanswered metabolic and physiological questions that we can begin to answer using cultivated strains. Physiological and metabolic interpretations of individual populations of OM43/LD28 methylotrophs will provide a knowledge base for predicting the carbon-processing activities of marine and freshwater populations.

1.5. Northern aquatic ecosystems

1.5.1. The Saint-Lawrence Estuary

The Saint Lawrence Estuary, located in in Québec, Canada is an interesting ecosystem that supports a diverse range of microbial life. The estuary has a gradient of salinity, with freshwater from the river mixing with saltwater from the ocean, creating a distinct brackish environment. Nutrients from both the river and ocean influence the bacterial community composition across the estuary. Previous studies have reported on the distribution of specific bacterial taxa across the estuarine surface salinity gradient and on the distribution of metabolic pathways across the stratified waters using meta-omics approaches (Colatriano *et al.* 2015; Cui *et al.* 2020). Climate change is also having a profound impact on the environmental conditions of the Saint Lawrence Estuary. Increased surface temperatures, changes in precipitation patterns, and increased river runoff have been observed, all of which could be causing changes to the salinity gradient across the estuary. This makes the Saint Lawrence Estuary an important ecosystem to study.

1.5.2. The Bedford Basin

The Bedford Basin is a large northwest Atlantic inlet in Eastern Canada and is considered a good indicator of what is happening in coastal waters along the Atlantic continental shelf. As a fjord with a long narrow entrance, Bedford Basin has limited freshwater input mostly localized to

the surface waters. As we increasingly see the effects of climate change on ecosystems, the Bedford Basin is an important study site due to continued time-series monitoring by the Department of Fisheries and Oceans. Previous studies have identified changes in microbial community composition in relation to seasonality and changing environmental conditions (Georges *et al.* 2014; El-Swais *et al.* 2015). In fact, a recent study shows changes in phytoplankton community composition in response to warmer temperatures (Robicheau *et al.* 2022). These changes in phytoplankton may also affect the abundance and diversity of methylotrophs that use the methanol released by phytoplankton as a carbon and energy source.

1.5.3. The Arctic Ocean

The waters of the Western Arctic Ocean are strongly stratified by a gradient in salinity where fresher water with a higher nutrient content overlies more saline water. The high nutrient concentration is due to inflow directly from the relatively fresh and nutrient-rich Pacific Ocean (Carmack *et al.* 2011). The halocline also receives a significant volume of freshwater from the Mackenzie River, contributing to the low salinity surface waters. The Canada Basin, one of the 4 basins of the Arctic Ocean, also receives inflow from the relatively warm and saline Atlantic Ocean. The sea ice extent and seasonality play an important role in shaping the structure of the Arctic Ocean. Sea ice cover has significantly diminished in the past decade, with one of the lowest ever recorded sea ice minima starting in the summer of 2007. Additionally, the oldest and thickest ice (five years or older) disappeared almost entirely from the Arctic in 2010 (McLaughlin *et al.* 2011; IPCC 2019; ICCI 2022). Questions about how this increase in freshening will affect the water column of the Canada Basin have risen from these drastic changes.

The majority of life in the Arctic is microbial but the consequences of climate-mediated changes on the microbial food web and in turn biogeochemical cycles are not well known. Global

climate change is impacting the oceans by increasing sea surface temperatures and leading to melting sea ice, which together alter water stratification and circulation patterns. These changes can in turn affect the marine biome by changing marine food web dynamics (Doney 2010; Carmack and McLaughlin 2011). Evidence for such a change has already been reported from the upper 150m of the Canada Basin where smaller phytoplankton size-classes are becoming now more prevalent (Li *et al.* 2013). The increase of freshwater in the Arctic Ocean leads to questions about whether there are bacterial species, with freshwater and marine lineages, that can adapt and evolve in response to these new conditions. It is also important to understand the mechanisms of freshwater-marine transition zones. In fact, scientists and many international reports have highlighted the role of Arctic microbes as first responders to climate change and the resulting need for increased microbial genomics research. Additionally, international reports have highlighted the lack of microbial time-series data showing the effects of climate change at the base of the marine food web (CAFF 2017; IPCC 2019, 2021). Long-term observations of the state of the bacterial community will allow us to link the changing physical and chemical pressures to the impact on marine ecosystems. With climate change greatly affecting the Arctic Ocean, it is important to assess genetic diversity and increased adaptability to decreased salinity and low nitrate concentrations in the surface waters. Investigating the changes in bacterial community composition in association to various changes in environmental factors. This will allow us to determine how changes in the Arctic Ocean are directly affecting the microbes living there.

1.6. Rationale

Given what we know about the importance of aquatic environments and methylotrophs, the overarching objective of this thesis was to further understand the metabolic diversity and evolution of bacteria caused by changes in environmental variables, specifically salinity. In the

first study, we investigated the diversity of methylotrophic bacteria in pelagic marine and freshwater ecosystems. The second set of studies focus on the Arctic Ocean ecosystem to 1) investigate the distribution, diversity, and metabolism of OM43/LD28 clades in the Western Arctic Ocean, and 2) to determine how decadal changes in microbial community composition in the Western Arctic Ocean is linked to shifts in environmental variables.

2. **Investigation of XoxF methanol dehydrogenases reveals new methylotrophic bacteria in pelagic marine and freshwater ecosystems**

Ramachandran A, Walsh DA. FEMS Microbiol Ecol 2015;91:fiv105.

2.1. Abstract

The diversity and distribution of methylotrophic bacteria have been investigated in the oceans and lakes using the methanol dehydrogenase *mxaF* gene as a functional marker. However, pelagic marine (OM43) and freshwater (LD28 and PRD01a001B) methylotrophs within the *Betaproteobacteria* lack *mxaF*, instead possessing a related *xoxF4*-encoded methanol dehydrogenase. Here, we developed and employed *xoxF4* as a complementary functional gene marker to *mxaF* for studying methylotrophs in aquatic environment. Using *xoxF4,* we detected OM43-related and LD28-related methylotrophs in the ocean and freshwaters of North America, respectively, and showed the coexistence of these two lineages in a large estuarine system (St Lawrence Estuary). Gene expression patterns of *xoxF4* supported a positive relationship between *xoxF4*-containing methylotroph activity and spring time productivity, suggesting phytoplankton blooms are a source of methylotrophic substrates. Further investigation of methanol dehydrogenase diversity in pelagic ecosystems using comparative metagenomics provided strong support for a widespread distribution of *xoxF4* (as well as several distinct *xoxF5*) containing methylotrophs in marine and freshwater surface waters. In total, these results demonstrate a geographical distribution of OM43/LD28-related methylotrophs that includes marine and freshwaters and suggest that methylotrophy occurring in the water column is an important component of lake and estuary carbon cycling and biogeochemistry.

2.2. Introduction

Methanol is an abundant organic compound in the ocean where it can serve as a carbon and energy source for methylotrophic microorganisms. Marine methanol can originate from terrestrial inputs as well as production during phytoplankton growth and decay (Milne *et al.* 1995; Heikes 2002). In addition to high oceanic concentrations (50–400 nM), the turnover time of methanol in seawater is short, on the order of one to a few days, suggesting that methanol biogeochemistry is a significant com- ponent of the marine carbon cycle (Dixon, Beale and Nightingale 2011; Dixon and Nightingale 2012). To elucidate the role of microorganisms in marine methanol cycling, it is important to identify and quantify the diversity and distribution of methylotrophic bacteria in the ocean. To this end, (Mcdonald and Murrell 1997) first proposed the use of the *mxaF* gene, which encodes the alpha subunit of methanol dehydrogenase (MDH), as a functional gene marker for methylotrophs. MDH is an enzyme containing a pyrroloquinoline quinone (PQQ) cofactor that oxidizes methanol to formaldehyde, and over the years the *mxaF* gene has been used in numerous cultivation-independent assessments of marine methylotroph diversity (Mcdonald and Murrell 1997; Neufeld *et al.* 2008b; Lau *et al.* 2013).

Although *mxaF* as a marker has proved useful, an important limitation is that marine methylotrophs that lack the *mxaF* gene have recently been reported. Cultivation of marine methylotrophs of the OM43 clade of *Betaproteobacteria* and subsequent genome analysis showed that these methylotrophs lack *mxaF*, yet possess a homologous *xoxF4* gene that is implicated in methanol oxidation (Giovannoni *et al.* 2008; Chistoserdova 2011b). In fact, previous phylogenetic analyses of MxaF and XoxF proteins revealed that this MDH protein family comprises at least six different clades (MxaF and XoxF1–5) and that methylotrophs may possess multiple MDH

homologs (Chistoserdova 2011b; Keltjens *et al.* 2014). However, representative genomes of the OM43 clade possess only a single *xoxF4* gene.

Based on 16S rRNA surveys, the OM43 clade is more abundant and common than other previously described proteobacterial methylotrophs (Sekar *et al.* 2004; Morris, Longnecker and Giovannoni 2006; Song, Oh and Cho 2009; Taubert *et al.* 2015). Environmental proteomic studies have shown that the XoxF4 protein is commonly expressed in coastal surface waters, suggesting the OM43 clade may be responsible for a considerable amount of methanol turnover in the coastal ocean (Sowell *et al.* 2011; Georges *et al.* 2014). Based on these recent observations, it is therefore likely that the current approaches that rely on the *mxaF* gene marker do not provide an accurate account of the abundance and diversity of marine methylotrophic bacteria.

Compared to oceans, less is known about methanol oxidation in the surface waters of lakes and rivers, although a large body of work exists on benthic methylotrophy in sediments (Kalyuzhnaya *et al.* 2012; Chistoserdova and Lidstrom 2013). Given that terrestrial plants and phytoplankton are sources of methanol and methylated compounds in the ocean (Milne *et al.* 1995; Heikes 2002; Millet *et al.* 2008), methylotrophy is likely important in the surface waters of lakes as well. Interestingly, 16S rRNA gene sequences affiliated with OM43 are common in lakes of Europe and North America where they are referred to as the LD28 tribe (Zwart *et al.* 1998; Salcher *et al.* 2008; Jezbera *et al.* 2012). LD28 and OM43 are both members of the broader betIV clade (Newton *et al.* 2011), which also contains *Methylophilus* and *Methylotenera* isolates from freshwater sediments and soils (Chistoserdova 2011a). These isolates can use a range of one carbon compounds (Chistoserdova 2011a) and some may also be able to subsist on phenol and humic matter (Hutalle-Schmelzer *et al.* 2010). Very recently, isolates from two planktonic freshwater lineages (LD28 and closely affiliated PRD01a001B) were cultured from Lake Zurich, Switzerland

(Salcher *et al.* 2015). Subsequent genome analysis confirmed a methylotrophic lifestyle and a reduced set of genes for one carbon (C1) metabolism. Similar to OM43, these freshwater lineages lack *mxaF*, but possess a single *xoxF4* homolog. As such, we hypothesized that *xoxF4* genes present in lakes and rivers can also serve as a functional gene marker for freshwater LD28/PRD01a001B methylotrophy.

The objective of this study was to first develop the *xoxF4* gene as a complementary functional gene marker to the canonical *mxaF* gene for studying methylotroph diversity in aquatic environments. We then investigated the distribution, diversity, and *xoxF4* expression levels of methylotrophic bacteria across marine, estuarine, and freshwater environments using a combination of PCR-based analyses of *xoxF4* genes and comparative metagenomics of PQQ-dependent methanol dehydrogenases present in publically available marine and freshwater metagenomes.

2.3. Materials and Methods

2.3.1. Sampling and environmental DNA and RNA extraction

Microbial biomass for DNA and RNA extraction was collected from the surface water (5 m depth) at the compass buoy station in Bedford Basin, Nova Scotia, and along a transect of the St Lawrence Estuary, Quebec (**Figure 2.1**). Sampling of the St Lawrence Estuary occurred during 16–20 May 2011. It should be noted that no biological replicates of DNA or RNA were collected. Seawater was prefiltered (2 L from Bedford Basin and 2.5 L from the St Lawrence Estuary) through a Whatman GF/D filter (2.7 µm cut- off) and cells were collected on a 0.22µm Sterivex filter. After filtration, 1.8 mL of sucrose-based lysis buffer was added and filters were stored at –80°C. DNA was extracted following the protocol described in (Zaikova *et al.* 2010). RNA was

extracted using the mirVana miRNA Isolation kit (Life Technologies, Burlington, Ontario, Canada) following the modified methodology in (Stewart *et al.* 2012).

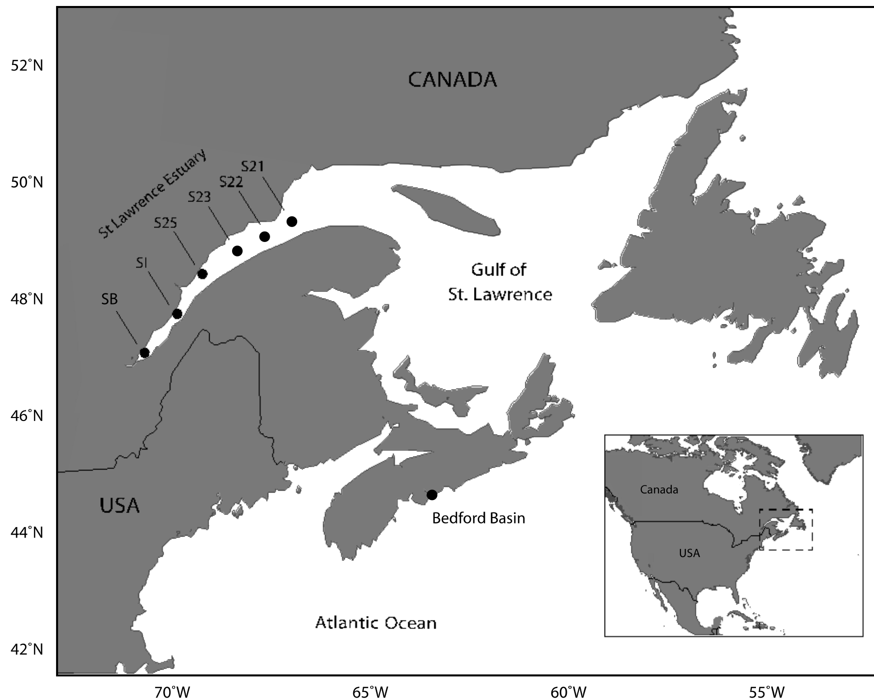


Figure 2.1 *Map of the coastal Atlantic Ocean and St Lawrence Estuary showing station locations (circles) used in this study. The inset map at the lower left shows the study region (dashed box) in the broader context of North America.*

Temperature and salinity were measured by CTD. Bacterial production was measured using the [^{3}H]leucine incorporation method detailed in (Smith and Azam 1992). Bacterial abundance was determined through flow cytometry in which the cells were stained with SYBR Gold as detailed in (Li and Dickie 2001). Chlorophyll a measurements were determined *in vitro* by fluorescence of 90% acetone extracts of plankton as detailed in (Li and Dickie 2001).

2.3.2. *xoxF4* PCR primer design and amplification

Multiple sequence alignments of PQQ-dependent dehydrogenase genes (*xoxF1–5* and *mxaF*) from whole genomes were generated using MUSCLE (in MEGA5). Primer sets were designed in order to amplify the *xoxF4* fragment specifically by maximizing primer matching to

marine *xoxF4* genes and minimizing primer matching to genes from the other five *xoxF*/*mxaF* clades. The multiple sequence alignments were created using the nucleotide sequences of genes encoding PQQ-dependent dehydrogenases acquired from the NCBI database. The primers used for the ~500 bp fragment were 974F (5′-AACCCWTCHGTNTGGAAYCC-3′) and 1441R (5′-GCRCCMACCCAWGGYTGACC-3′), which were subsequently used for cloning and sequencing. The primers for the ~230 bp fragments were 1004F (5′-CCWGGYGAYAAYAARTGGTC-3′) and 1232R (5′-GCCCAGTTNAYRAAKGGRTG-3′), which were subsequently used for qPCR and RT-qPCR assays (**Supplementary Figure 2.1**). Genomic DNA from *Methylotenera mobilis* strain #13 (Chistoserdova 2011a) was used as a PCR positive control as this strain possesses both the *mxaF* and the *xoxF4* homologs.

Template genomic DNA from environmental samples was used in a 25 µL reactions containing 16 µL MilliQ, 5 µL 1X Reaction Buffer, 1.25 µL F primer (0.5 µM), 1.25 µL R primer (0.5 µM), 0.5 µL dNTPs (0.2 µM), 0.5 µL Phire Polymerase, 1 µL DNA template (~10 ng). PCR conditions were as follows: 98°C for 30 s for the initial denaturing step; 30 cycles of 98° C for 5 s, 56.5° C (60° C for *mxaF*) for 5 s, 72°C for 10 s; and 72°C for 1 min for the final elongation step. PCR products were visualized using a 1% (w/v) agarose gel via electrophoresis.

2.3.3. Cloning and sequencing of *xoxF4* fragment

The larger fragment (~500 bp) of *xoxF4* was PCR-amplified and gel-purified using the QIAGEN QIAquick Gel Extraction Kit. The samples used for the construction of the clone libraries were from January 2011 and May 2011 from Bedford Basin, and Station B and Station 22 from the St Lawrence Estuary. The purified PCR product was cloned into a pJET vector using the CloneJET PCR Cloning Kit. The transformants were grown on LB+Ampicillin plates and the colonies, 24 for each sample, were screened for the insert by colony PCR using the pJET primers.

Restriction fragment length polymorphism (RFLP) analysis was performed using the following reaction mix: 7.75 µL $H_2 0$, 2 µL NE buffer 4, 0.25 µL HhaI, 10µL Colony PCR product. The 16 clones chosen for sequencing were those that showed variation in fragment size after the RFLP. Plasmid DNA was isolated using the GeneJet Plasmid Miniprep Kit (Thermo Scientific, Ottawa, Ontario, Canada) and sent for paired end Sanger sequencing (OPERON, Louisville, KY, USA). Phylogenetic analysis was performed using maximum likelihood and a GTR + Gamma distribution (four categories) model of evolution.

2.3.4. Quantification of *xoxF4* gene and mRNA abundances

Plasmids harboring the *xoxF4* fragment were linearized overnight at 37°C using the Eco31I restriction enzyme. The linearized plasmid was gel purified and used as the qPCR and RT-qPCR standards. The absolute quantification was measured through serial dilutions of 10^7 to 10^3 gene copies per microlitre. Template genomic DNA from Bedford Basin and St Lawrence Estuary samples were used in 20 µL reactions containing 14.6 µL MilliQ, 4µL 5X EvaGreen qPCR mix, 0.4µL F primer (0.1µM), 0.4 µL R primer (0.1 µM), 1 µL DNA template. The thermal profile used was 15 min polymerase activation at 95°C followed by the PCR cycling stage with 40 cycles (95°C for 45 s, 55°C for 30 s, 72°C for 1 min) and ending with a melting curve (95°C for 15 s, 55°C for 15 s, 95°C for 15 s). The dilution factor used for the unknowns was between 10^0 and 10^{-3} and the reactions were performed in triplicate. In order to calculate the copy number of the *xoxF4* gene in 1 mL of seawater, the copy number given by the qPCR was divided by the amount of seawater (mL) filtered for each sample and corrected for the dilution factor.

The RNA from the Bedford Basin and the St Lawrence Estuary was used as the template to synthesize cDNA using the M-MLV Reverse Transcriptase kit (Invitrogen) using the reverse primer 1441R. The results obtained from the samples (Bedford Basin and St Lawrence Estuary)

were compared to the standard curve and a quantity of copy number was given. Calculation of mRNA transcript levels in the RT-qPCR analysis was similar to that of gene copy number. The efficiency of the qPCR was ≥81% and the R^2 value was ≥93% for all runs.

2.3.5. Metagenomic analysis of PQQ-dependent methanol dehydrogenases

XoxF4 from the Betaproteobacterial strain HTCC2181 was used as the query protein sequence to search for homologs in metagenomic datasets from 35 marine and two freshwater ecosystems from IMG/M using BLASTp and a 10^{-5} E-value cutoff (**Supplementary Table 2.1**). Retrieved protein sequences were clustered at 99% identity using the program cd-hit (Fu *et al.* 2012) to reduce redundancy, resulting in 808 clusters for which a single sequence was selected as a representative. Sequences greater than 300 amino acids in length were added to a multiple sequence alignment, containing PQQ-dependent dehydrogenase reference sequences including sequences from (Lau *et al.* 2013; Keltjens *et al.* 2014). The preliminary tree included metagenomic sequences that were closely related to other PQQ-dependent dehydrogenases (e.g. alcohol and glucose de- hydrogenases) and these were trimmed. After clustering and trimming, 44 protein sequences remained for the final phylogenetic analysis using the MEGA6 software. The following parameters were used: maximum-likelihood method, JTT substitution model, gamma distribution model for the rate variation with four discrete gamma categories, and the nearest-neighbor interchange (NNI) heuristic search method.

2.4. Results

2.4.1. *xoxF4* PCR primer design and specificity

Previous phylogenetic analysis of MxaF and XoxF proteins revealed that these MDH subunits are members of six different clades (MxaF and XoxF1–5) (Chistoserdova 2011b; Keltjens *et al.* 2014). While the genomes of certain bacteria may encode multiple *xoxF* homologs (Keltjens

et al. 2014), the publicly available marine OM43 genomes (strains HTCC2181 and HIMB624/KB13) and freshwater LD28 ('*Candidatus* Methylopumilus planktonicus') and PRD01a001B ('*Candidatus* Methylopumilus turicensis') genomes only contain a single homolog encoding an XoxF4 protein. Therefore, we designed PCR primer sets (**Supplementary Figure 2.1**) that would broadly and specifically target the *xoxF4* gene in order to maximize the detectable diversity of *xoxF4*- containing bacteria in the environment. One primer combination (974F/1441R) targeted a ~500 bp fragment of the *xoxF4* gene, while a second primer combination (1004F/1232R), targeting an internal ~230 bp fragment, was designed for qPCR.

2.4.2. *xoxF4* diversity in aquatic environments

We surveyed the methylotrophic community in seasonal surface waters of a coastal inlet (Bedford Basin, Nova Scotia), as well as along a 200 km estuarine salinity gradient (St Lawrence Estuary, Quebec) (**Figure 2.1, Table 2.1**). Using the 974F/1441R primers, we detected *xoxF4* genes in all samples, while no *mxaF* amplification was detected using the updated primers reported in Lau *et al.* (2013). Cloning and sequencing of the *xoxF4* fragments followed by phylogenetic analysis verified the specificity of the 974F/1441R primer combination and identified three distinct and well-supported *xoxF4* subclades (**Figure 2.2**). We detected *xoxF4* sequences nearly identical to strain HTCC2181 in both Bedford Basin (BB) and the marine end of the St Lawrence Estuary (SLE). The second subclade (*xoxF4–2*) was composed solely of sequences recovered from BB. The third clade was composed of sequences recovered from the SLE, including sequences recovered from a freshwater location (*xoxF4–3*), as well as the *xoxF4* gene from the LD28 isolate '*Candidatus* Methylopumilus planktonicus'. These results expand the known diversity and distribution of *xoxF4*-harboring lineages in both marine and freshwater habitats of North America.

Table 2.1 *Environmental data from sample locations in BB and SLE*

	Temperature (°C)	Salinity (PSU)	Chlorophyll (mg/m^3)	Bacterial abundance (x10^5 ml^{-1})	Bacterial production (ug C l^{-1} day^{-1})
Bedford Basin					
December 15 2010	7.9	29.6	0.97	8.3	n.d.
January 11 2011	5.1	29.9	0.84	2.8	n.d
April 20 2011	4.5	29.9	7.64	5.7	n.d
May 25 2011	7.9	28.4	12.69	8.4	n.d
Saint Lawrence Estuary					
Station B	10.6	0.08	n.d	6.5	21.64
Station SI	4.3	18.5	n.d	2.8	2.94
Station 25	5.5	21.2	n.d	4.2	15.90
Station 23	4.7	24.3	n.d	5.7	42.41
Station 22	5.4	22.2	n.d	6.9	59.12
Station 21	6.0	27.1	n.d	3.7	26.03

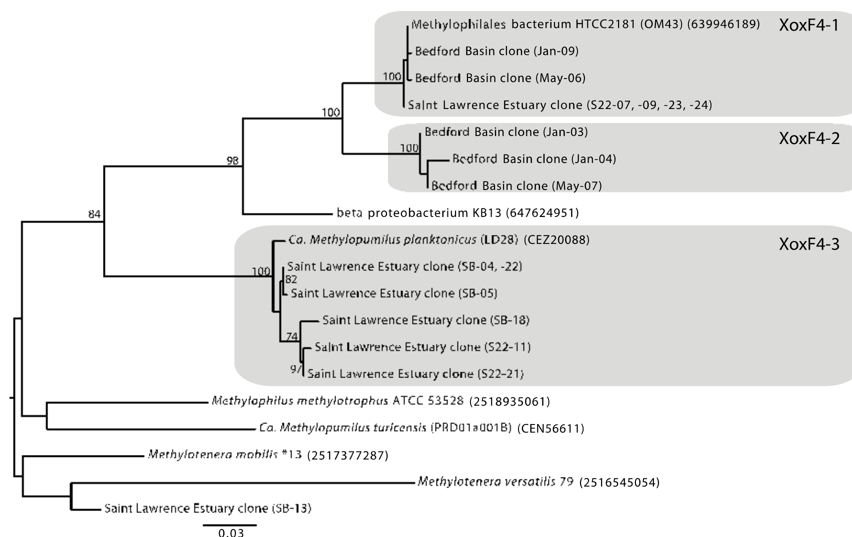


Figure 2.3 *Phylogenetic analysis of partial (~500 bp) xoxF4 gene sequences PCR-amplified from Bedford Basin and Stations B and 22 in the St Lawrence Estuary using primer pair 974F/1441R. The tree was inferred using maximum likelihood and a GTR + Gamma distribution (four categories) model of evolution. Clone identifiers for sequences produced in this study are presented in parentheses.*

2.4.3. Abundance and expression of *xoxF4* genes

Quantification of *xoxF4* gene abundance and mRNA transcripts revealed spatiotemporal variation in methylotroph abundance and expression levels in BB and the SLE (**Figure 2.3**). In BB, we com- pared the *xoxF4* community during the winter (December and January) and during the spring phytoplankton bloom (April and May). Over this time interval, *xoxF4* gene abundance increased twofold (from an average of 1747 mL^{-1} to 3499 mL^{-1}), while mRNA expression levels increased 12-fold (from 419 mL^{-1} to 4964 mL^{-1}). These results demonstrated that *xoxF4*-harboring bacteria exhibit higher abundance and *xoxF4* expression levels in association with productive spring conditions compared to the winter.

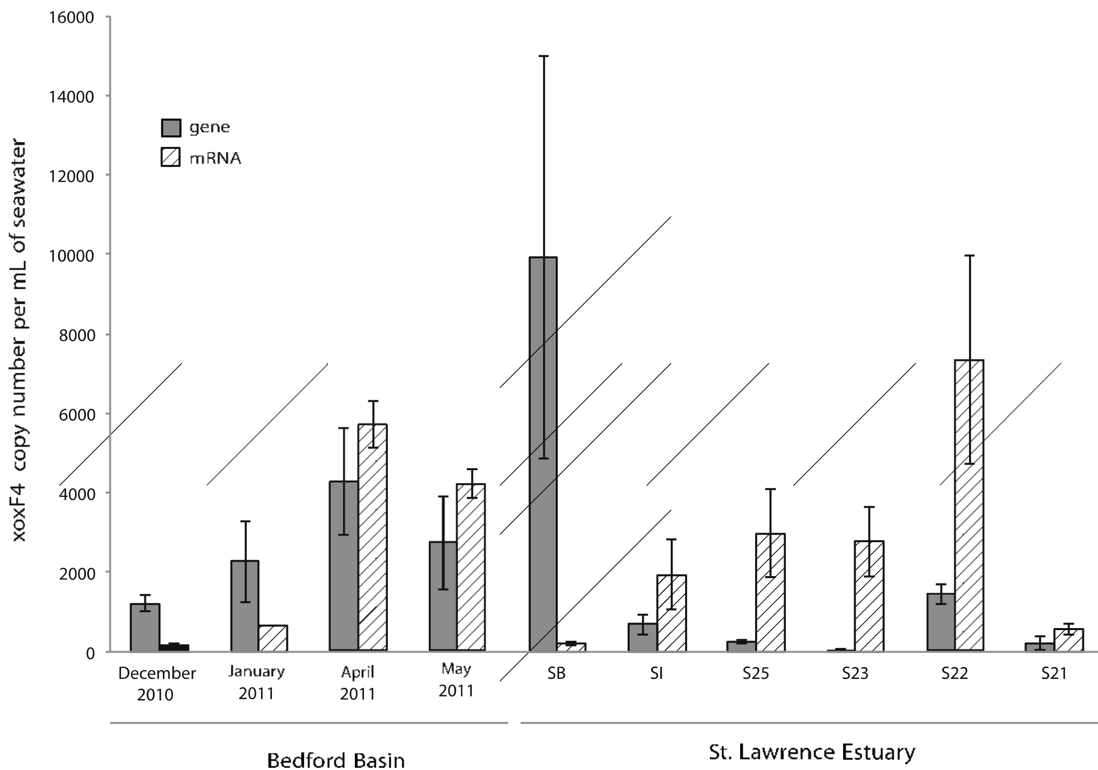
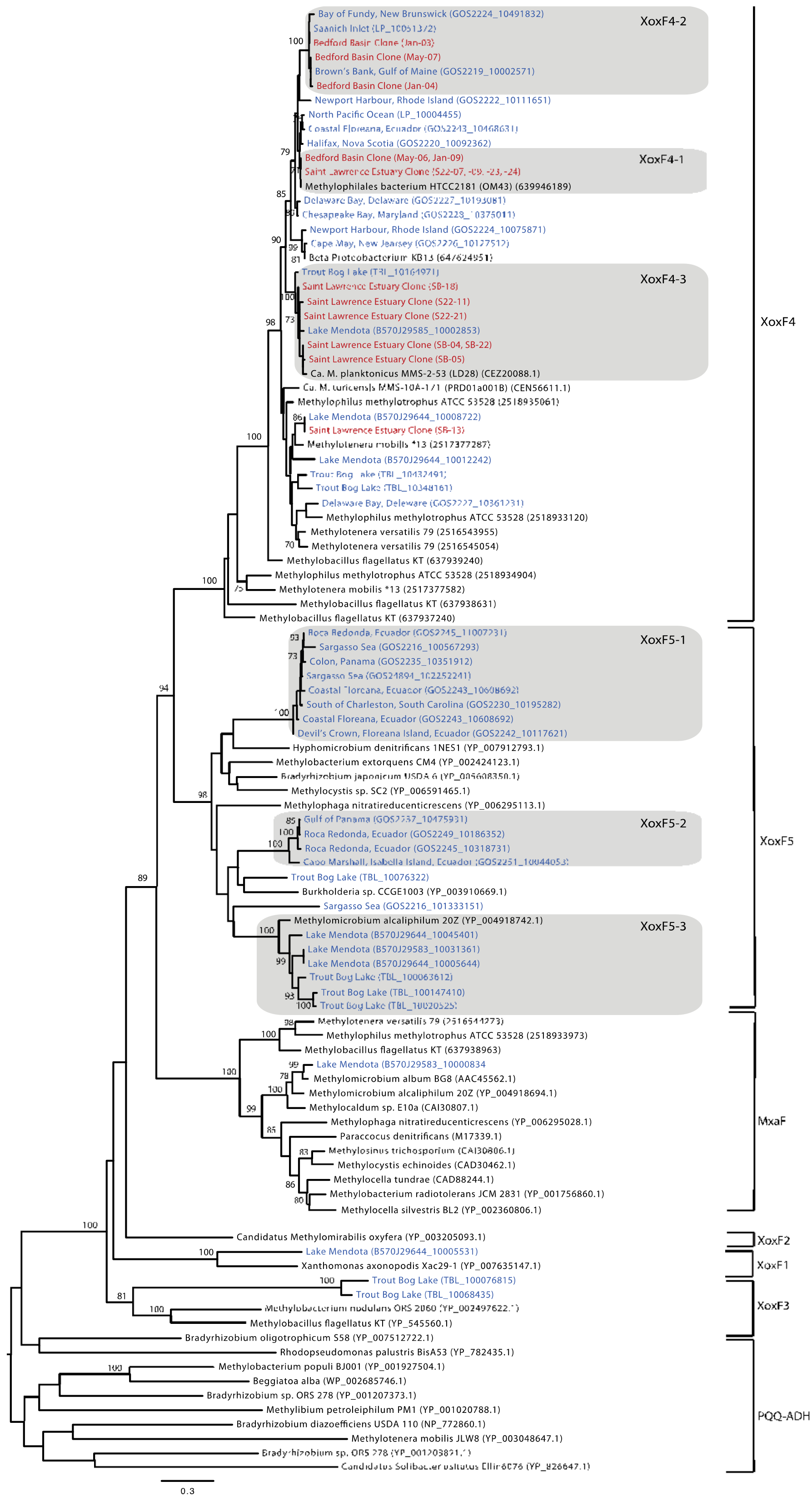


Figure 2.4 *Abundance of xoxF4 genes and mRNA transcripts from seasonal samples collected in Bedford Basin and along the salinity gradient of the St Lawrence Estuary. Error bars indicate standard deviation.*

In the SLE, the *xoxF4* genes were present and expressed across a salinity gradient ranging from freshwater (Station B) to brackish-marine waters (Station 21). However, unlike in BB, we did not observe a positive correlation between abundance and expression. While *xoxF4* abundance was maximal at the freshwater end of the SLE (Station B, 9931 mL^{-1}), *xoxF4* expression was at its lowest (198 mL^{-1}), suggesting the presence of an abundant, yet inactive freshwater assemblage of methylotrophs. In contrast, *xoxF4* gene expression was elevated within the brackish estuary and exhibited a maximum at the most productive location near the marine end of the estuary (Station 22, 7345 copies mL^{-1}).

2.4.4. Comparative metagenomics of *mxaF* and *xoxF* diversity and biogeography

Based on our observation of *xoxF4* genes in the freshwater end of the SLE and the recent report of freshwater methylotrophs by (Salcher *et al.* 2015), we were motivated to investigate the distribution of PQQ-dependent methanol dehydrogenase across a diversity of marine and freshwater habitats using metagenomics. To do so, we searched for XoxF/MxaF in publically available metagenomic datasets, including those from surface waters of the oceans and freshwater lakes in North America (**Supplementary Table 2.1**). After clustering and trimming, we identified 44 representative sequences. Phylogenetic analysis placed the metagenome-derived protein sequences into five separate clades, which were previously reported in Keltjens *et al.* (2014) (XoxF1, 3, 4, 5 and MxaF) (**Figure 2.4**). To visualize the geographical distribution of methylotrophs, we then mapped the distribution of the different clades to the metagenome sample locations (**Figure 2.5**).



Bay of Fundy, New Brunswick (GOS2224_10491832)
Saanich Inlet (LP_10051372)
Bedford Basin Clone (Jan-03)
Bedford Basin Clone (May-07)
Brown's Bank, Gulf of Maine (GOS2219_10002571)
Bedford Basin Clone (Jan-04)
Newport Harbour, Rhode Island (GOS2222_10111651)
North Pacific Ocean (LP_10004455)
Coastal Floreana, Ecuador (GOS2243_10468631)
Halifax, Nova Scotia (GOS2220_10092362)
Bedford Basin Clone (May-06, Jan-09)
Saint Lawrence Estuary Clone (S22-07, -09, -23, -24)
Methylophilales bacterium HTCC2181 (OM43) (639946189)
Delaware Bay, Delaware (GOS2227_10193081)
Chesapeake Bay, Maryland (GOS2228_10375011)
Newport Harbour, Rhode Island (GOS2224_10075871)
Cape May, New Jearsey (GOS2226_10127512)
Beta Proteobacterium KB13 (647624951)
Trout Bog Lake (TBL_10161971)
Saint Lawrence Estuary Clone (SB-18)
Saint Lawrence Estuary Clone (S22-11)
Saint Lawrence Estuary Clone (S22-21)
Lake Mendota (B570J29585_10002853)
Saint Lawrence Estuary Clone (SB-04, SB-22)
Saint Lawrence Estuary Clone (SB-05)
Ca. M. planktonicus MMS-2-53 (LD28) (CEZ20088.1)
Ca. M. turicensis MMS-10A-171 (PRD01a001B) (CEN56611.1)
Methylophilus methylotrophus ATCC 53528 (2518935061)
Lake Mendota (B570J29644_10008722)
Saint Lawrence Estuary Clone (SB-13)
Methylotenera mobilis *13 (2517377287)
Lake Mendota (B570J29644_10012242)
Trout Bog Lake (TBL_10432491)
Trout Bog Lake (TBL_10348161)
Delaware Bay, Delaware (GOS2227_10361231)
Methylophilus methylotrophus ATCC 53528 (2518933120)
Methylotenera versatilis 79 (2516543955)
Methylotenera versatilis 79 (2516545054)
Methylobacillus flagellatus KT (637939240)
Methylophilus methylotrophus ATCC 53528 (2518934904)
Methylotenera mobilis *13 (2517377582)
Methylobacillus flagellatus KT (637938631)
Methylobacillus flagellatus KT (637937240)
Roca Redonda, Ecuador (GOS2245_11007231)
Sargasso Sea (GOS2216_100567293)
Colon, Panama (GOS2235_10351912)
Sargasso Sea (GOS21894_102252241)
Coastal Floreana, Ecuador (GOS2243_10608692)
South of Charleston, South Carolina (GOS2230_10195282)
Coastal Floreana, Ecuador (GOS2243_10608692)
Devil's Crown, Floreana Island, Ecuador (GOS2242_10117621)
Hyphomicrobium denitrificans 1NES1 (YP_007912793.1)
Methylobacterium extorquens CM4 (YP_002424123.1)
Bradyrhizobium japonicum USDA 6 (YP_005608350.1)
Methylocystis sp. SC2 (YP_006591465.1)
Methylophaga nitratireducenticrescens (YP_006295113.1)
Gulf of Panama (GOS2237_10475931)
Roca Redonda, Ecuador (GOS2249_10186352)
Roca Redonda, Ecuador (GOS2245_10318731)
Cabo Marshall, Isabella Island, Ecuador (GOS2251_10044053)
Trout Bog Lake (TBL_10076322)
Burkholderia sp. CCGE1003 (YP_003910669.1)
Sargasso Sea (GOS2216_101333151)
Methylomicrobium alcaliphilum 20Z (YP_004918742.1)
Lake Mendota (B570J29644_10045401)
Lake Mendota (B570J29583_10031361)
Lake Mendota (B570J29644_10005644)
Trout Bog Lake (TBL_100063612)
Trout Bog Lake (TBL_100147410)
Trout Bog Lake (TBL_10020525)
Methylotenera versatilis 79 (2516544273)
Methylophilus methylotrophus ATCC 53528 (2518933973)
Methylobacillus flagellatus KT (637938963)
Lake Mendota (B570J29583_10000834)
Methylomicrobium album BG8 (AAC45562.1)
Methylomicrobium alcaliphilum 20Z (YP_004918694.1)
Methylocaldum sp. E10a (CAI30807.1)
Methylophaga nitratireducenticrescens (YP_006295028.1)
Paracoccus denitrificans (M17339.1)
Methylosinus trichosporium (CAI30806.1)
Methylocystis echinoides (CAD30462.1)
Methylocella tundrae (CAD88244.1)
Methylobacterium radiotolerans JCM 2831 (YP_001756860.1)
Methylocella silvestris BL2 (YP_002360806.1)
Candidatus Methylomirabilis oxyfera (YP_003205093.1)
Lake Mendota (B570J29644_10005531)
Xanthomonas axonopodis Xac29-1 (YP_007635147.1)
Trout Bog Lake (TBL_100076815)
Trout Bog Lake (TBL_10068435)
Methylobacterium nodulans ORS 2060 (YP_002497622.1)
Methylobacillus flagellatus KT (YP_545560.1)
Bradyrhizobium oligotrophicum S58 (YP_007512722.1)
Rhodopseudomonas palustris BisA53 (YP_782435.1)
Methylobacterium populi BJ001 (YP_001927504.1)
Beggiatoa alba (WP_002685746.1)
Bradyrhizobium sp. ORS 278 (YP_001207373.1)
Methylibium petroleiphilum PM1 (YP_001020788.1)
Bradyrhizobium diazoefficiens USDA 110 (NP_772860.1)
Methylotenera mobilis JLW8 (YP_003048647.1)
Bradyrhizobium sp. ORS 278 (YP_001203821.1)
Candidatus Solibacter usitatus Ellin6076 (YP_826647.1)
XoxF4-2
XoxF4-1
XoxF4-3
XoxF4
XoxF5-1
XoxF5-2
XoxF5-3
XoxF5
MxaF
XoxF2
XoxF1
XoxF3
PQQ-ADH
0.3

Figure 2.5 *Phylogenetic analysis of PQQ-dependent methanol dehydrogenase MxaF/XoxF proteins. Protein sequences generated from our study are in red, while those from lake and ocean metagenomes are in blue. Some methylotrophs have multiple homologs of XoxF that can branch into different clades. The following parameters were used: maximum-likelihood method, 100 bootstrap iterations, JTT substitution model, gamma distribution model for the rate variation with four discrete gamma categories.*

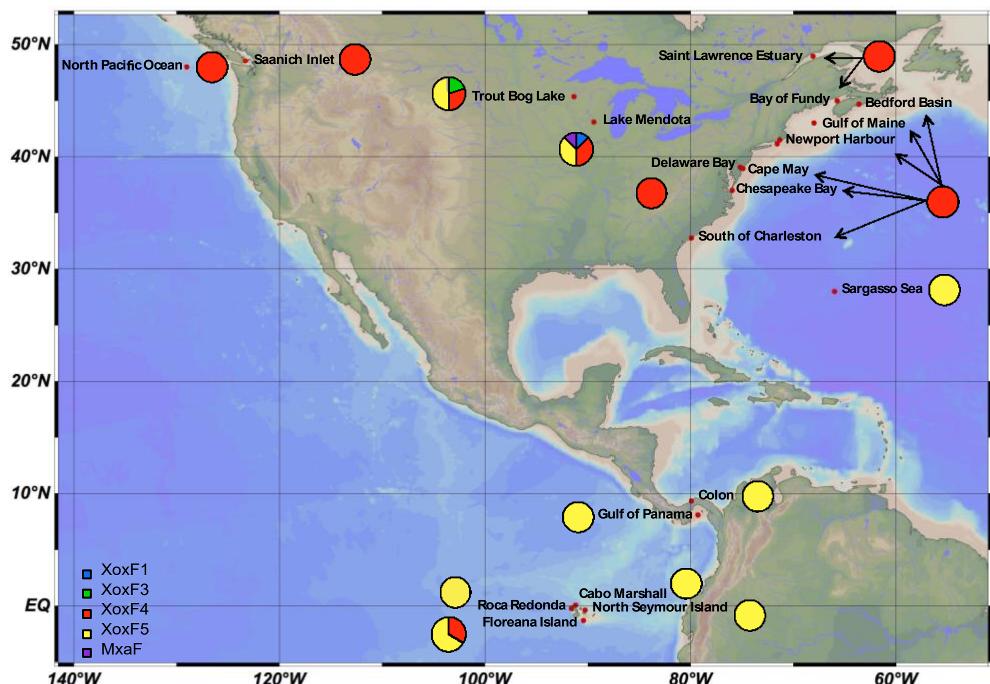


Figure 2.6 *Survey of PQQ-dependent methanol dehydrogenase genes across aquatic metagenomes.*

Based on our metagenomic analysis, the detection of MxaF was restricted to lakes, while XoxF4 was more broadly distributed across lake and ocean locations. XoxF4 proteins affiliated with OM43 were common in coastal habitats of the North Pacific and North Atlantic Oceans, yet not detectable offshore (Stations 12, 16 and 26 of Line P in the north-eastern Pacific and a single station in the Indian Ocean) (**Supplementary Table S1**). In support of our observation of freshwater *xoxF4*-containing methylotrophs in the St Lawrence River, we also identified closely related homologs in the epilimnion of two freshwater lakes (Trout Bog and Lake Mendota). Phylogenetic analyses placed these freshwater XoxF4 homologs into two distinct groups. One

clade contained cultured isolates (e.g. *Methylophilus* and *Methylotenera*) from freshwater sediments (Hernandez *et al.* 2015), while the second clade (XoxF4–3) was represented by environmental sequences recovered from the water column and the *xoxF4* homolog from '*Candidatus* Methylopumilus planktonicus' (Salcher *et al.* 2015). These observations support the presence of phylogenetically distinct benthic and pelagic lineages of methylotrophic *Betaproteobacteria* in lakes.

Metagenomics also revealed the presence and broad distribution of the XoxF5 clade in freshwater and marine environments. In the ocean, the XoxF5 showed a contrasting latitudinal distribution compared to XoxF4: XoxF4 dominated in the north- ern hemisphere, while XoxF5 dominated environments near the equator. Moreover, there were 3 distinct XoxF5 subclades, including two that were only identified in the ocean (XoxF5–1, XoxF5–2), and a third that was restricted to lakes (XoxF5–3). While the freshwater clade was closely affiliated with *Methylomicrobium alcaliphilum*, a halophilic methanotrophic bacterium containing the *mxaF* gene (Vuilleumier *et al.* 2012) the marine subclades represented uncultivated and poorly characterized taxa that may represent novel one carbon compound oxidizers in the sea.

2.5. Discussion

2.5.1. OM43 methylotroph abundance and activity in coastal habitats

Although methylotrophy has been studied in the oceans for decades, a combination of cultivation, genomic, and proteomic studies recently demonstrated that the *Betaproteobacteria* OM43 clade represent a previously overlooked component of the methylotrophic marine community (Morris, Longnecker and Giovannoni 2006; Giovannoni et al. 2008; Sowell et al. 2011; Halsey, Carter and Giovannoni 2012; Georges et al. 2014; Taubert et al. 2015b). Using PCR-based assays specific to *xoxF4*, we detected the presence and expression of *xoxF4* genes across a set of

samples from the coastal north-west Atlantic Ocean and along the St Lawrence Estuary salinity gradient. The universal detection of the *xoxF4* gene and an absence of detectable *mxaF* demonstrate that OM43 are numerically dominant over *mxaF*-harboring methylotrophs in the ocean. A recent study of *xoxF* diversity in the western English Channel also identified *xoxF4* genes closely related to the OM43 clade (Taubert *et al.* 2015). Moreover, our comparative metagenomics analysis showed a common presence of *xoxF4* in coastal habitats, and a lack of detectable *mxaF*, supporting a previous report by Chistoserdova (2011b). In the Atlantic, *xoxF4* gene abundance ranged from 1747 to 3499 copies mL^{-1} suggesting that about 1% of the bacterioplankton contain a *xoxF4* homolog. Although it is possible that some OM43 bacteria may possess multiple copies of xoxF4, if we assume *xoxF4* is a single copy gene in OM43, then this value agrees well with previous studies showing that OM43 constitute ~2% of bacterial cells in coastal microbial communities (Rappé, Vergin and Giovannoni 2000; Sekar *et al.* 2004; Galand *et al.* 2008; Song, Oh and Cho 2009). We observed highest *xoxF4* gene expression during the spring phytoplankton bloom. This observation is in agreement with previous work showing a high OM43 abundance associated with diatoms (Morris, Longnecker and Giovannoni 2006), suggesting that diatom-derived methanol and methylated substances such as dimethylsulfoniopropionate (DMSP) and trimethylamine-*N*-oxide (TMAO) may serve as methylotrophic growth substrates.

The widespread distribution and expression of *xoxF4* suggests that OM43 could be playing a role in marine methylotrophy. However, these findings are at odds with earlier studies that implicate *Gammaproteobacteria* with marine methanol metabolism (Neufeld *et al.* 2007, 2008b). In a series of elegant studies, Neufeld *et al.* (2007, 2008b, 2008a) used DNA-SIP with [^{13}C]methanol to link methylotrophy with *Methylophaga* and other *Gammaproteobacteria*, but not OM43, irrespective of season. One possible explanation is that OM43 are adapted to very low

methanol concentrations, and it has been suggested that XoxF-type MDHs oxidize methanol with higher rates and affinities than MxaF-MDHs (Keltjens *et al.* 2014). However, even in incubations at near *in situ* methanol concentration (1µM), *Methylophaga* are detected among the predominant methylotrophs (Neufeld *et al.* 2008a). Perhaps OM43 methylotrophs were overlooked in these studies since low resolution 16S rRNA DGGE analysis and/or the *mxaF* gene marker were used to assess the identity of methylotrophs. On the other hand, it may be that OM43 does not readily assimilate methanol under the incubation conditions as it may be adapted to take advantage of multiple growth substrates simultaneously when available in the environment. In fact, although OM43 (strain HTCC2181) can grow on methanol as a sole carbon and energy source, this strain exhibits synergistic growth when methanol is amended with additional methylated compounds including DMSP and TMAO, which are compatible osmolytes produced by phytoplankton (Halsey, Carter and Giovannoni 2012). In any case, the application of *xoxF4* as a functional gene marker for OM43 methylotrophy should be valuable in further deciphering the ecological niches and metabolic preferences of methylotrophs in the ocean.

2.5.2. Expanding the diversity of putative marine methylotrophs

In addition to the XoxF4 family, our metagenomic survey revealed that the XoxF5 family is common in the ocean. Although XoxF5 function remains undescribed, it is large family of proteins implicated in one carbon metabolism and found in methylotrophs of *Alpha*-, *Beta*- and *Gammaproteobacteria* (Keltjens *et al.* 2014). For the most part, detection of XoxF5 was restricted to locations near the equator, while XoxF4 was more prevalent at northern latitudes. Indeed, Salcher *et al.* (2015, in press) reported that *xoxF4*-containing freshwater strains from Lake Zurich, Switzerland were psychrophilic, growing more rapidly at 6° C compared to 22° C. Hence, the biogeographical divide we observed in the oceans may be driven by environmental conditions.

However further work is required to link these *xoxF5* genes to particular phylogenetic lineages and assess their role in marine methylotrophy.

2.5.3. Rare-earth elements and their relationship to methylotrophs

Rare-earth elements (REEs) are used as the cofactor for the XoxF methanol dehydrogenase. These elements are not as rare as their name implies as they are found in every type of soil, sand and sediment (Keltjens *et al.* 2014). Previous studies indicate that *xoxF* gene expression is induced by REEs found in the lanthanide group such as cerium (III) and lanthanum (III) (Keltjens *et al.* 2014). A study conducted by (Hara *et al.* 2009) showed that during phytoplankton growth, the concentration of dissolved REEs decreased in the surface waters suggesting that REEs can be concentrated by phytoplankton. Phytoplankton decay may therefore not only serve as a source of methylotrophic growth substrates, but also as a source of REE cofactors; however, further experimentation is required to test this hypothesis.

2.5.4. Freshwater and estuarine pelagic methanol-oxidizing bacteria

Methylotrophs have been extensively studied in freshwater sediments (Chistoserdova 2011a), yet much less is known about methylotroph distribution and activities in the surface waters of lakes and rivers. Potential sources of methanol and methylated compounds in freshwater include production in anoxic bottom waters and sediments, degradation of methylated sugars, direct production by phytoplankton and trans- port from terrestrial sources. Using our *xoxF4* primer set we identified a group of *xoxF4* sequences (xoxF4–3) in the St Lawrence River, and in metagenomes from the epilimnion of Lake Mendota and Trout Bog in North America. These results expand on a recent study by Salcher *et al.* (2015, in press), who isolated previously uncultivated members of two fresh- water lineages (LD28 and PRD01a001) affiliated with the marine OM43 clade. The *xoxF4* sequences from our study were members of the LD28 group, and

we were unable to identify *xoxF4* sequences from the PRD01a001 group. Although primer bias cannot be completely ruled out, in silico analysis would suggest that this is unlikely, as our *xoxF4* primers will amplify *xoxF4* sequences from both groups. This observation is in line with the findings of Salcher *et al.* (2015, in press) who demonstrated that LD28 were generally more abundant than PRD01a001 in Lake Zurich and this may be a general pattern that extends to freshwaters of North America.

In addition to the identification of LD28-derived *xoxF4* in freshwaters, we also identified a mixture of LD28- and OM43- derived *xoxF4* genes in brackish waters of the SLE. In a recent study by Taubert *et al.* (2015, in press), a diversity of *xoxF4* genes were identified in a salt marsh demonstrating their common presence in brackish environments. In the SLE, a peak in *xoxF4* gene expression was also concomitant with the most productive location (S22) in the estuary. Hence, as shown for marine (Milne *et al.* 1995; Heikes 2002; Sinha *et al.* 2007; Halsey, Carter and Giovannoni 2012) and freshwater systems ((Millet *et al.* 2008; Salcher *et al.* 2015), it appears that methylotrophs are associated with phytoplankton abundance in estuaries as well.

Now that it is possible to differentiate between distinct populations of *xoxF4* containing methylotrophs, it will be interesting to assess the activity and distribution of LD28 and OM43 methylotrophs along salinity gradients. Moreover, the phylogenetic relationships between the marine and freshwater lineages show that this group of methylotrophs likely has a freshwater ancestry and that the OM43 arose due to colonization of the marine environment. This is supported by the observation that marine OM43 genomes have acquired ‘marine genes’ such as Na^{+}-pumping respiratory proteins by lateral gene transfer from marine *Alphaproteobacteria* (Walsh, Lafontaine and Grossart 2013) during their evolution. Hence the OM43/LD28 methylotrophs may

serve as a useful evolutionary model to investigate habitat transitions, ecological specialization and speciation in the microbial world.

2.6. Acknowledgements

We wish to thank the captains and crews of the R/V Coriolis II for providing essential assistance during field operations in the SLE and Bill Li, Kevin Pauley and Jeff Anning for logistical support in collecting samples from Bedford Basin. We thank Paul del Giorgio for providing bacterial abundance data and Roxane Maranger for providing bacterial production data. We are also grateful to Ludmila Chistoserdova for providing the *M. mobilis* culture. We thank Ludmila Chistoserdova and Hans-Peter Grossart for comments on an earlier version of the manuscript and three anonymous reviewers for their helpful input on the submitted manuscript.

2.7. Funding

This work was supported by National Science and Engineering Research Council of Canada (402214–2011) and Canada Research Chair Program (950–221184) research grants. A.R. was supported by the Concordia Institute for Water, Energy, and Sustainable Systems.

Conflict of interest. None declared.

2.8. Supplementary Figures and Tables

Supplementary Figure 2.1 *Primer design technique*

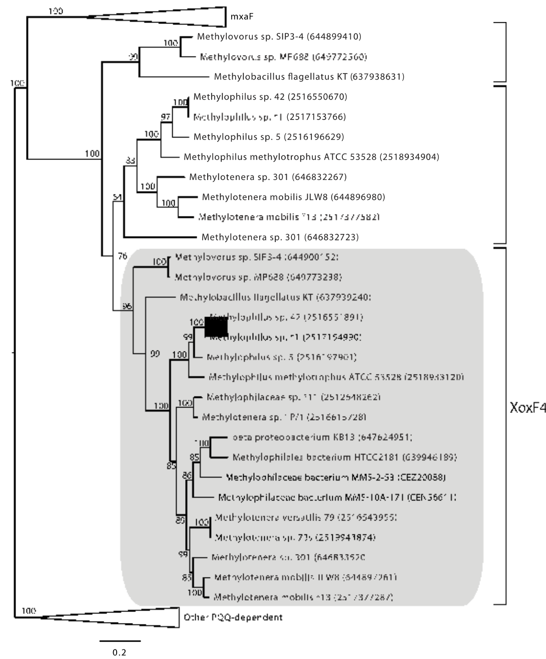


Table 1. Differences in nucleotide sequences

F974	**5'**	**A**	**A**	**C**	**C**	**C**	**W**	**T**	**C**	**H**	**G**	**T**	**N**	**T**	**G**	**G**	**A**	**A**	**Y**	**C**	**C**	**3'**
xoxF4-1		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
xoxF4-2		-	-	-	-	-	-	R	G	-	A	C	-	-	-	-	-	-	-	-	-	
xoxF4-3		-	T	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	S	-	
mxaF		-	-	Y	-	-	N	G	C	N	C	C	N	-	-	-	-	-	-	G	A	
Other PQQ		-	-	Y	-	-	N	D	S	N	A	C	H	-	-	-	-	-	-	-	-	
F1004	**5'**	**C**	**C**	**W**	**G**	**G**	**Y**	**G**	**A**	**Y**	**A**	**A**	**Y**	**A**	**A**	**R**	**T**	**G**	**G**	**T**	**C**	**3'**
xoxF4-1		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
xoxF4-2		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
xoxF4-3		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	-	
mxaF		-	-	N	-	-	B	-	-	-	-	-	-	-	-	-	-	-	-	A	-	
Other PQQ		-	-	N	-	-	S	-	V	C	-	-	-	M	R	V	-	-	-	D	-	
R1232	**5'**	**G**	**C**	**C**	**C**	**A**	**G**	**T**	**T**	**N**	**A**	**Y**	**R**	**A**	**A**	**K**	**G**	**G**	**R**	**T**	**G**	**3'**
xoxF4-1		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
xoxF4-2		-	-	-	-	-	-	-	-	-	-	-	-	-	-	N	-	-	-	-	K	
xoxF4-3		-	M	-	-	-	R	-	-	-	-	-	-	-	-	R	-	-	-	-	-	
mxaF		A	M	B	M	M	R	-	-	V	-	C	B	G	Y	N	K	B	-	-	C	
Other PQQ		R	S	-	-	-	-	-	-	V	-	C	N	V	H	B	K	K	-	-	C	
R1441	**5'**	**G**	**C**	**R**	**C**	**C**	**M**	**A**	**C**	**C**	**C**	**A**	**W**	**G**	**G**	**Y**	**T**	**G**	**A**	**C**	**C**	**3'**
xoxF4-1		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
xoxF4-2		-	M	-	-	-	H	R	-	-	-	-	-	-	-	-	-	-	V	-	-	
xoxF4-3		-	-	-	-	-	N	-	-	-	-	-	-	-	-	-	-	-	M	-	-	
mxaF		-	-	D	-	-	V	-	-	R	A	-	R	A	A	-	-	-	R	-	-	
Other PQQ		N	G	D	R	C	C	N	A	-	R	W	A	N	-	G	Y	W	K	D	-	

Supplementary Table 2.1 *Metagenomic sample sites*

Sampling Sites	Depths	Dates	Description
Antarctic Peninsula	5 m	06-02-28	Marine bacterioplankton communities from Palmer Station B, Arthur Harbor, Antarctica
Bay of Fundy, Nova Scotia, Canada - GS006	1 m	03-08-23	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific

Bedford Basin, Nova Scotia, Canada - GS005	1 m	03-08-22	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Bermuda, Atlantic Ocean - GS001	5 m	03-05-15	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Block Island, New York, USA - GS009	1 m	03-11-18	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Browns Bank, Gulf of Maine - GS003	1 m	03-08-21	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Cabo Marshall, Isabella Island, Equador - GS036	2.1 m	04-03-02	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Cape May, New Jersey, USA - GS010	1 m	03-11-18	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Chesapeake Bay, Maryland, USA - GS012	2.07 m	04-10-26	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Coastal Floreana, Equador - GS028	2 m	04-02-04	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Cocos Island, Costa Rica - GS023	2 m	04-01-21	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Colon, Panama - GS019	1.7 m	04-01-12	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Delaware Bay, New Jersey, USA - GS011	1 m	03-11-18	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Devil's Crown, Floreana Island, Equador - GS027	2.2 m	04-02-04	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Dirty Rock, Cocos Island, Costa Rica - GS025	1.1 m	04-01-28	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Equatorial Pacific Ocean - GS037	1.8 m	04-03-17	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Galapagos, Equador - GS026	2 m	04-02-01	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Gulf of Mexico, USA - GS016	2 m	04-01-08	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Gulf of Panama, Panama - GS021	1.6 m	04-01-19	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Halifax, Nova Scotia, Canada - GS004	2 m	03-08-22	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific

Indian Ocean - GS112	1.8 m	05-08-08	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Key West, Florida, USA - GS015	1.7 m	04-01-08	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Lake Gatun, Panama - GS020	2 m	04-01-15	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Lake Mendota	Epilimnion	11-06-01	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	11-11-01	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-04-02	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	10-06-02	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-06-02	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-08-03	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	11-10-03	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	11-09-04	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	10-08-05	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-03-05	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	10-05-05	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-05-05	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-11-05	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	10-07-06	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	10-07-06	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	07-07-07	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	09-10-07	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-09-07	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-06-08	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-10-08	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	11-08-09	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	09-06-09	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-11-09	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA

Lake Mendota	Epilimnion	09-08-10	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	11-07-12	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-10-12	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	08-09-12	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-07-13	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	10-09-13	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	12-08-13	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	09-11-14	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	08-10-17	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	10-05-20	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Lake Mendota	Epilimnion	11-07-25	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA
Line P, North Pacific Ocean - P04	10 m	Aug-09	Marine microbial communities from expanding oxygen minimum zones in the northeastern subarctic Pacific Ocean
Line P, North Pacific Ocean - P12	10 m	Jun-09	Marine microbial communities from expanding oxygen minimum zones in the northeastern subarctic Pacific Ocean
Line P, North Pacific Ocean - P16	10 m	Aug-09	Marine microbial communities from expanding oxygen minimum zones in the northeastern subarctic Pacific Ocean
Line P, North Pacific Ocean - P26	10 m	Aug-08	Marine microbial communities from expanding oxygen minimum zones in the northeastern subarctic Pacific Ocean
Moorea, Cooks Bay, Polynesia Archipelagos - GS049	1.4 m	04-05-19	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Nags Head, North Carolina, USA - GS013	2.1 m	03-12-19	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Newport Harbor, Rhode Island, USA - GS008	1 m	03-11-16	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
North James Bay, Santigo Island, Equador	2.1 m	04-02-08	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
North Seamore Island, Equador - GS034	2.1 m	04-02-19	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Northern Gulf of Maine, Canada - GS007	1 m	03-08-25	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Panama City, Panama - GS022	2 m	04-01-19	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific

Polynesia - GS047	30 m	04-03-28	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Rangirora Atoll, Polynesia Archipelagos - GS051	1 m	04-05-22	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Roca Redonda, Equador - GS030	19 m	04-02-09	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Rosario Bank, Honduras - GS018	1.7 m	04-01-10	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Sargasso Sea - GS000a&b	5 m	03-02-26	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Sargasso Sea - GS000c	5 m	03-02-25	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
South of Charleston, South Carolina, USA - GS014	1 m	03-12-20	The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific
Trout Bog Lake	Epilimnion	07-06-07	Freshwater microbial communities from Lake Mendota and Trout Bog Lake, Wisconsin, USA

Bridging text

The first chapter focused on being able to differentiate between distinct populations of *xoxF4* containing methylotrophs. We designed specific *xoxF4* primers and detected OM43-related and LD28-related methylotrophs in the ocean and freshwaters of North America and showed the coexistence of these two lineages in a large estuarine system. Also, the phylogenetic relationships between the marine and freshwater lineages show that this group of methylotrophs likely has freshwater ancestry and that the OM43 clade arose due to colonization of the marine environment. Due to the diversity and distribution of OM43/LD28 methylotrophs as well as their ability to cross the freshwater-marine barrier, we are interested in using this group as a model to study mechanisms of adaptation and evolution. By further studying the Methylophilaceae family in a region where freshwater input is significantly increasing, we can determine the processes of salinity-driven change. Adaptation to decreased salinity in marine environments is becoming more important, especially in the Arctic Ocean, where freshening of the ocean is rapidly increasing with climate change. Therefore, we chose to investigate the diversity, distribution, and metabolism of the Methylophilaceae family in the Western Arctic Ocean.

Chapter 3. A novel freshwater to marine evolutionary transition revealed within *Methylophilaceae* bacteria from the Arctic Ocean

Ramachandran A, McLatchie S, Walsh DA. mBio 2021;12:e0130621.

3.1. Abstract

Bacteria inhabiting polar oceans, particularly the Arctic Ocean, are less studied than those at lower latitudes. Discovering bacterial adaptations to Arctic Ocean conditions is essential for understanding responses to the accelerated environmental changes occurring in the North. The *Methylophilaceae* are emerging as a model for investigating the genomic basis of habitat adaptation because related lineages are widely distributed across both freshwater and marine ecosystems. Here we investigated *Methylophilaceae* diversity in the salinity-stratified surface waters of the Canada Basin, Arctic Ocean. In addition to a diversity of marine OM43 lineages, we report on the genomic characteristics and evolution of a previously undescribed *Methylophilaceae* clade (BS01) common to polar surface waters, yet related to freshwater sediment *Methylotenera* species. BS01 is restricted to the lower salinity surface waters, while OM43 is found throughout the halocline. An acidic proteome supports a marine lifestyle for BS01, but gene content shows increased metabolic versatility compared to OM43 and evidence for ongoing genome-streamlining. Phylogenetic reconstruction shows that BS01 colonized the pelagic ocean independently of OM43 via convergent evolution. Salinity adaptation and differences in one-carbon and nitrogen metabolism may play a role in niche differentiation between BS01 and OM43. In particular, urea utilization by BS01 is predicted to provide an ecological advantage over OM43 given the limited amount of inorganic nitrogen in the Canada Basin. These observations provide further evidence that the Arctic Ocean is inhabited by distinct bacterial groups, and that at least one group (BS01) evolved via a freshwater to marine environmental transition.

3.2. Importance

Global warming is profoundly influencing the Arctic Ocean. Rapid ice melt and increased freshwater input is increasing ocean stratification, driving shifts in nutrient availability and the primary production that supports marine food webs. Determining bacterial responses to Arctic Ocean change is challenging because of limited knowledge on the specific adaptations of Arctic Ocean bacteria. In this study, we investigated the diversity and genomic adaptations of a global distributed group of marine bacteria, the *Methylophilaceae*, in the surface waters of the Arctic Ocean. We discovered a novel lineage of marine *Methylophilaceae* inhabiting the Arctic Ocean whose evolutionary origin involved a freshwater to marine environmental transition. Crossing the salinity barrier is thought to rarely occur in bacterial evolution. However, given the ongoing freshening of the Arctic Ocean, our results suggest that these relative newcomers to the ocean microbiome may increase in abundance and therefore ecological significance in a near-future Arctic Ocean.

3.3. Introduction

Studies in low latitude oceans have provided numerous insights into the eco-evolutionary processes that underlie patterns of marine bacterial biogeography (Biller *et al.* 2014; Sunagawa *et al.* 2015; Delmont *et al.* 2019). However, bacterial communities inhabiting polar oceans, particularly the Arctic Ocean, are far less studied. There are a number of characteristics of the Arctic Ocean that make this region informative to include in studies on bacterial biogeography and evolutionary diversification in the global ocean. The Arctic Ocean is uniquely surrounded by Eurasian and North American land masses, which restrict water exchange with other oceans and influences ocean hydrology through freshwater input via large river systems (Guéguen *et al.* 2012; Krishfield *et al.* 2014; Carmack *et al.* 2016; Proshutinsky *et al.* 2019). A second motivation is that

increasing global temperatures are having a profound influence on the Arctic. In addition to warming (Jackson, Williams and Carmack 2012; Timmermans, Toole and Krishfield 2018), the Arctic Ocean is freshening due to a dramatic loss of sea ice, increased precipitation, and river discharge (Carmack *et al.* 2016; Proshutinsky *et al.* 2019). The assembly of distinct bacterial communities in the Arctic Ocean in response to this unique and changing environmental setting is evident (Ghiglione *et al.* 2012; Boeuf *et al.* 2013; Swan *et al.* 2013; Royo-llonch *et al.* 2020), as is the existence of Arctic-adapted ecotypes within globally prevalent marine bacteria such as SAR11 and SAR202 (Colatriano *et al.* 2018; Kraemer *et al.* 2019). Moreover, time-series studies in the Canada Basin (Western Arctic) have shown the subsequent increase in stratification due to surface freshening, which affects nutrient transport and primary production in the photic zone (Li *et al.* 2009, 2013). There is evidence that the physicochemical changes are influencing microbial community structure. For example, a study comparing microbial community structure in the Beaufort Sea before and after the 2007 record sea ice minimum demonstrated significant differences in all three domains of life (Comeau *et al.* 2011). In combination, such studies are beginning to reveal the unique community composition and genomic adaptations within Arctic marine microbiomes. Further studies that compare Arctic populations with their lower latitude relatives should broaden our understanding of how bacterial taxa are adapted to life in the Arctic Ocean, and provide insights into how these communities may respond to the rapid environmental changes currently underway.

The *Methylophilaceae* are emerging as a model for investigating evolutionary diversification and habitat adaptation in aquatic ecosystems as closely related lineages are distributed across marine and freshwater ecosystems. *Methylophilaceae* are methylotrophs specialized to use one-carbon (C1) compounds, particularly methanol (Giovannoni *et al.* 2008;

Chistoserdova 2011b; Chistoserdova and Lidstrom 2013). Evolutionary studies based on comparative genomics suggests that ancestral *Methylophilaceae* inhabited sediments and subsequently colonized freshwater pelagic habitats (origin of LD28 and PRD01a001B clades) before further diversifying into marine pelagic habitats (OM43 clade) (Walsh, Lafontaine and Grossart 2013; Salcher et al. 2019a). The transition from a sediment to a pelagic lifestyle involved extensive genome reduction, while the transition from freshwater to marine habitats involved metabolic innovation via lateral gene transfer (LGT) (Salcher *et al.* 2019).

Within the *Methylophilaceae*, the marine OM43 lineage is among the most successful bacterial groups in the ocean, inhabiting diverse environments from tropic to polar seas (Morris, Longnecker and Giovannoni 2006; Giovannoni *et al.* 2008; Huggett, Hayakawa and Rappé 2012; Jimenez-Infante *et al.* 2015; Ramachandran and Walsh 2015; Sosa *et al.* 2015; Taubert *et al.* 2015; Salcher *et al.* 2019). OM43 is commonly found in coastal waters and brackish environments (Giovannoni *et al.* 2008; Huggett, Hayakawa and Rappé 2012; Jimenez-Infante *et al.* 2015; Ramachandran and Walsh 2015; Sosa *et al.* 2015) and is often associated with phytoplankton blooms (Georges *et al.* 2014; Taubert *et al.* 2015). Phylogenetic analyses using 16S-23S internal transcribed spacer (ITS) sequences shows that OM43 is broadly divided into two ecotypic clusters, OM43-A (represented by strain HTCC2181) and OM43-B (also referred to as Hawaii-Red Sea [H-RS] cluster) (Jimenez-Infante *et al.* 2015). OM43-B is associated with low-chlorophyll a and/or warm oceans, whereas OM43-A is more prevalent in colder, higher productivity waters (Jimenez-Infante *et al.* 2015). Additional OM43 microdiversity exists (e.g. OM43-A1 and OM43-A2) which may reflect further niche specialization. Given the broad distribution of *Methylophilaceae* in freshwater to marine habitats and their diversification linked to differences in salinity, temperature,

and primary productivity, these methylotrophs may be an informative group for investigating bacterial adaptation in the rapidly changing Arctic Ocean.

In this study, we characterized the phylogenetic and genomic diversity of *Methylophilaceae* in the Canada Basin, Arctic Ocean. A major feature of the Canada Basin is the Beaufort Gyre. As of 2018, the freshwater content of the Canada Basin has increased approximately 40% relative to the 1970's because of increased sea ice melt and river water accumulation driven by an anticyclonic Beaufort Gyre (Proshutinsky *et al.* 2019). Stratification is increasing, and nutrient availability and primary production are shifting as a result of this freshening (Li *et al.* 2009, 2013; Lovejoy, Galand and Kirchman 2011; Jackson, Williams and Carmack 2012; Peralta-ferriz and Woodgate 2014). Here we provide a snapshot of *Methylophilaceae* diversity in vertically stratified metagenomes located along a latitudinal gradient of the Canada Basin. In doing so, we report on the discovery, genomic characteristics, and evolutionary origin of a previously undescribed lineage of marine *Methylophilaceae* that appears to be common in polar oceans.

3.4. Results

3.4.1. Environmental context

Methylophilaceae diversity was investigated along a four station (CB2, CB4, CB8, CB11) latitudinal transect (~73°-77° N) at 150 °W in the Canada Basin during late summer-autumn of 2015 **(Table 3.1, Supplementary Figure 3.1)**. The summer mixed layer depth ranged between 10-30 m. Surface (5-7 m) salinity ranged from 25.7 to 27.3 PSU and nitrate concentrations were below the detection limit. The deep chlorophyll maximum (DCM) was located between 25 to 79 m, where salinity ranged from 29.7 to 31.5 PSU. In the deeper Pacific winter waters (PWW; defined as a salinity of 33.1 PSU) nitrate concentration was approximately 16 $mmol/m^3$.

Table 3.1 *Location and environmental characteristics of samples collected for metagenomic analyses*

Station	Latitude (N)	Longitude (W)	SML[a] depth (m)	Sample[b] feature	Depth (m)	Temp (°C)	Salinity (PSU)	Fluorescence (mg/m^3)	Nitrate (mmol/m^3)	Silicate (mmol/m^3)	Phosphate (mmol/m^3)
CB2	72°59′	149°59′	10	Surface	6.3	−1.26	25.7	0.15	bdl[c]	2.55	0.52
				DCM	70.7	−0.89	31.5	0.33	4.45	10.05	1.07
				PWW	181.3	−1.45	33.2	0.05	15.98	32.9	1.84
CB4	75°00′	150°00′	30	Surface	4.5	−1.389	26.1	0.12	bdl	2.43	0.51
				DCM	80	−0.03	31.2	0.23	4.65	10.8	1.11
				PWW	212.3	−1.47	33.1	0.05	16.13	33.7	1.87
CB8	76°59′	149°58′	16	Surface	6	−1.46	27.2	0.19	bdl	2.73	0.54
				DCM	61.7	−0.15	31	0.3	0.35	5.13	0.78
				PWW	217.8	−1.45	33.1	0.05	16.24	35.02	1.9
CB11	78°59′	149°59′	17	Surface	7.5	−1.48	27.3	0.24	bdl	2.83	0.56
				DCM	27.6	−1.04	29.7	0.25	bdl	3.01	0.64
				PWW	194.6	−1.46	33.2	0.05	15.82	35.04	1.91

[a]SML, surface mixed layer.
[b]DCM, deep chlorophyll maximum; PWW, Pacific winter water.
[c]bdl, below detection limit.

3.4.2. *Methylophilaceae* in the Canada Basin

Methylophilaceae 16S rRNA sequences were analysed in Canada Basin metagenome assemblies from surface, DCM, and PWW samples. Within OM43, 16S rRNA sequences from OM43-A1 and OM43-A2 were detected, while OM43-B was not. **(Figure 3.1)**. We also identified 16S rRNA sequences distantly related to previously described marine (OM43) or freshwater (LD28 and PRD001a001B) *Methylophilaceae* in all surface water metagenomes. These sequences formed a clade (herein referred to as BS01) with sequences previously recovered from Arctic and Antarctic surface seawater and bottom waters of the Gulf of Mexico **(Figure 3.1)**.

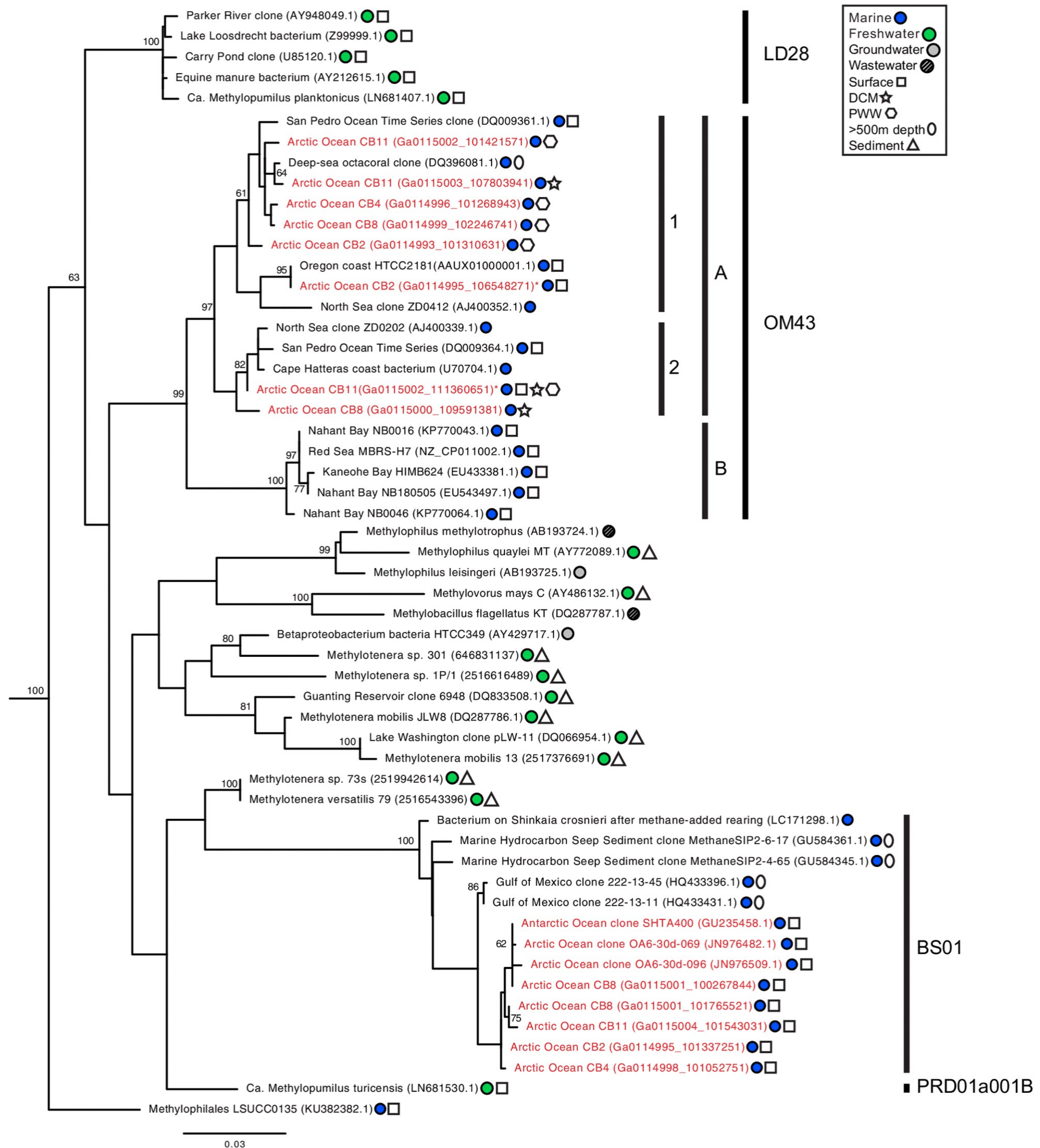


Figure 3.1 *Phylogenetic analysis of 16S rRNA genes from Methylophilaceae from Canada Basin metagenomes and a diversity of aquatic ecosystems. The tree was inferred using maximum likelihood (500 bootstraps) and GTR + gamma distribution (four categories) with invariant site model of evolution and the nearest-neighbor interchange heuristic search method. The tree was rooted using Methylibium as an outgroup to the Methylophilaceae. Sequences from the current study are highlighted in red. Only bootstrap values of >60 are included in the tree.*

We analyzed ITS diversity to provide finer phylogenetic resolution of *Methylophilaceae*. Thirty-nine ITS variants formed six ITS subclades (we designated these clades as OM43-A1a to OM43-A1d and OM43-A2a to OM43-A2b) **(Figure 3.2a)**. Similar to 16S rRNA diversity, we detected ITS sequences from OM43-A throughout the water column, as well as a distantly related group that likely represent BS01. OM43-A2b was detected in all Canada Basin metagenomes, irrespective of water layer, while OM43-A1 ITS clades (A1a, A1b, and A1d) were more restricted to the surface and DCM layers (**Figure 3.2b)**. A Principal Component Analysis showed that most of the variation in OM43 diversity was in ITS subclades contribution to DCM assemblages, although the pattern was not related to any clear differences in environmental conditions or nutrient availability **(Figure 3.2c)**. Overall, 16S rRNA and ITS diversity demonstrated a diverse assemblage of OM43 bacteria in the upper layers of the Arctic Ocean and identified a previously undescribed lineage of *Methylophilaceae* in the ocean.

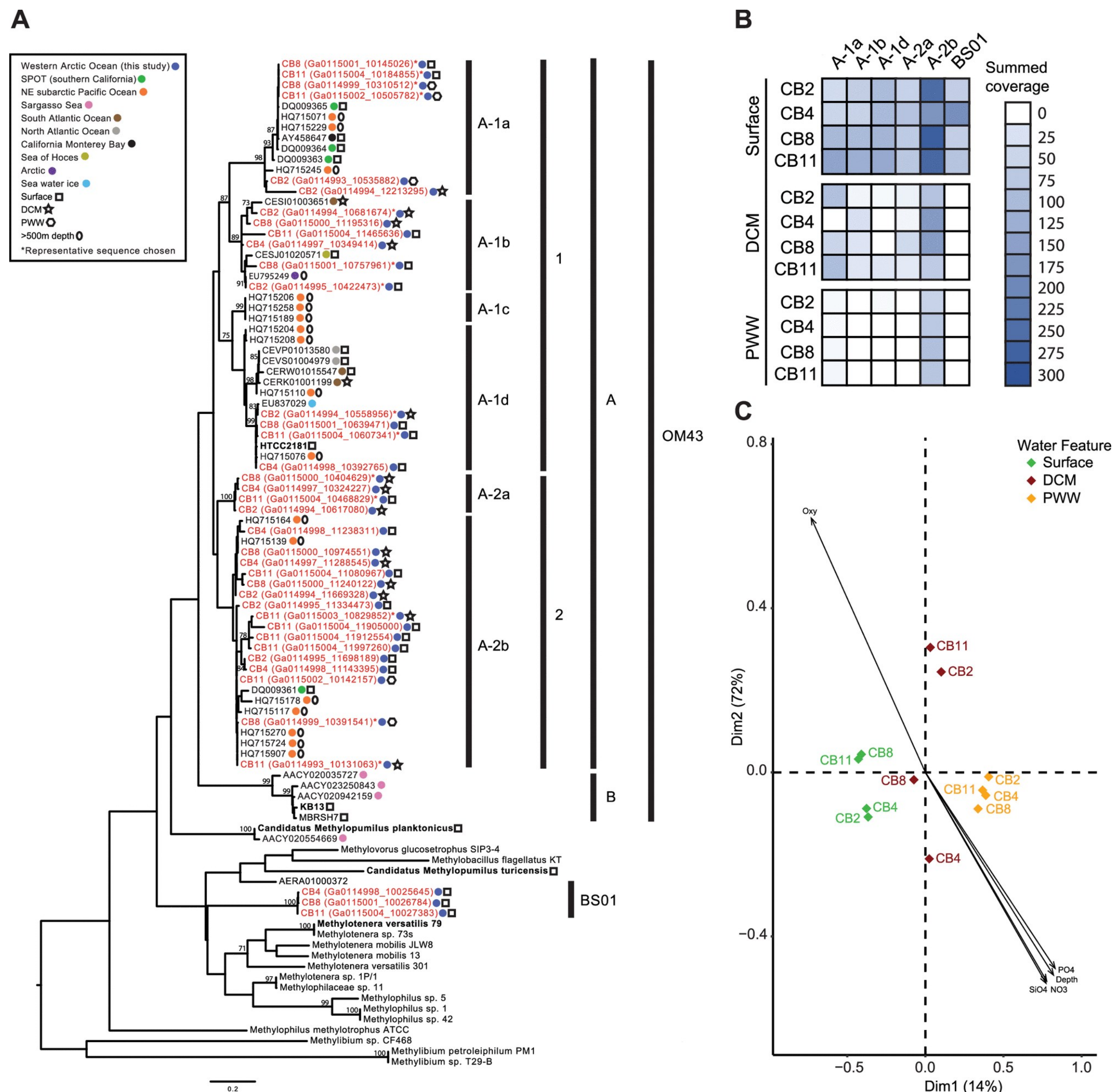


Figure 3.2 *Diversity and biogeography of Methylophilaceae based on ITS variants recovered from Canada Basin metagenomes. (A) Phylogenetic analysis of the Methylophilaceae group across various aquatic regions and depths using the ITS region. The tree was inferred using maximum likelihood (500 bootstraps) and a GTR + gamma distribution (four categories) with invariants sites model of evolution and the nearest-neighbor interchange heuristic search method. Sequences from the current study are highlighted in red. Only bootstrap values of >60 are included in the tree. (B) Abundance of six ITS variants based on summed coverage in metagenome assemblies. (C) Principal coordinate analysis ordination of Bray-Curtis dissimilarities of Arctic samples based on summed coverage of six ITS variants.*

3.4.3. Genomic characteristics of BS01

To further characterize BS01, we reconstructed a representative MAG from a CB2 surface water metagenome. The BS01 MAG (Met-BS01-1) was 1.48 Mb in length and 92% complete. Concatenated protein phylogeny showed Met-BS01-1 was more closely related to sediment-derived *Methylotenera* species than pelagic marine (OM43) or freshwater (LD28 and PRD01a001B) *Methylophilaceae*. **(Figure 3.3a)**. In agreement with the phylogeny, Met-BS01-1 exhibited higher average amino acid identity (%) with genomes from freshwater *Methylotenera* and *Ca.* "M. turicensis" (62-68 %) compared to marine OM43 genomes (53-55 %) **(Supplementary Table 3.1).** The Met-BS01-01 genome was more similar in size to those of the genome-streamlined pelagic OM43 and LD28 compared to *Methylotenera* strains **(Supplementary Table 3.1; Figure 3.3b)**. However, GC content of Met-BS01-01 (43% G+C) showed the opposite trend, exhibiting higher similarity to *Methylotenera* than OM43/LD28 genomes (**Figure 3.3b**). Previous studies on genome-streamlining have reported on shifts in amino acid usage as a response to nitrogen limitation, measured as increases in the lysine to arginine ratio of the proteome (Giovannoni 2005; Giovannoni, Cameron Thrash and Temperton 2014; Salcher *et al.* 2019). The lysine (6%) and arginine (4.5%) content of the Met-BS-01 proteome is more similar to *Methylotenera* than OM43/LD28 genomes **(Supplementary Table 3.1)**. In total, these observations suggest that genome streamlining has occurred during the evolution of BS01, but that the commonly associated shift towards lower GC content and reduced nitrogen amino acid usage were not apparent.

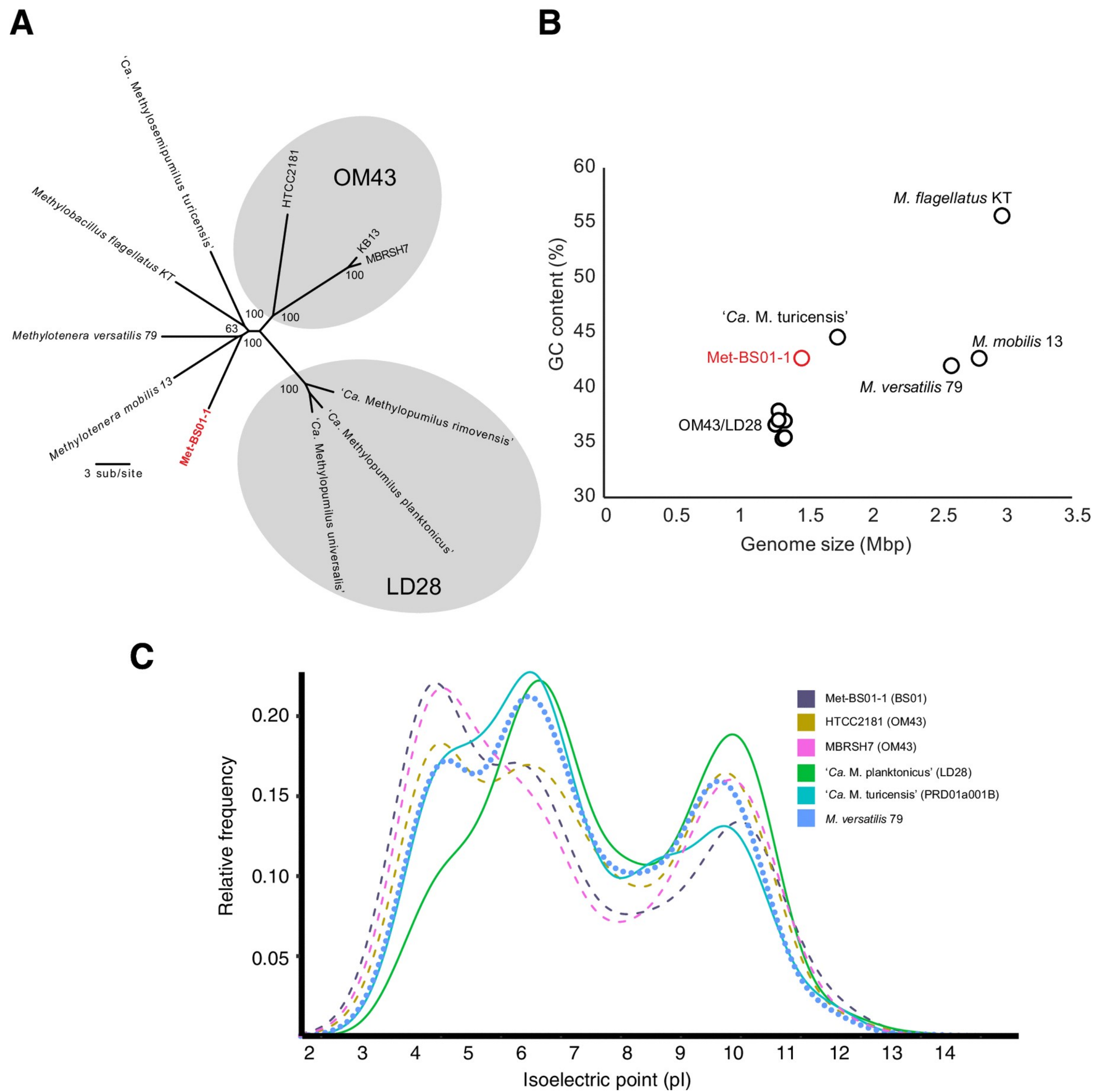


Figure 3.3 *Phylogenomic comparison of BS01 with representative Methylophilaceae genomes. (A) Maximum likelihood phylogenetic analysis of a concatenated alignment of 48 orthologs shared between all Methylophilaceae genomes. Values at the nodes are bootstrap values (100 pseudoreplicates). (B) Scatterplot comparing G+C content and genome size. (C) Whole-proteome pI values versus relative frequency in select Methylophilaceae genomes from freshwater and marine habitats.*

Proteome amino acid content is a valuable trait for predicting the preferred habitat of an organism, since marine bacteria exhibit more acidic values of protein isoelectric points (pI) than

freshwater bacteria (Cabello-Yeves and Rodriguez-Valera 2019). To elucidate habitat adaptation in BS01, we compared global pI plots between Met-BS01-1 and related *Methylophilaceae*. Met-BS01-1 exhibited the highest peak at an acidic pI (~4.5), which is similar to marine OM43 genomes (HTCC2181 and MBRSH7) (**Figure 3.3c**). In contrast, freshwater *Methylophilaceae* exhibit a peak at ~6.5. Interestingly, the Met-BS01-1 pI plot exhibits a frequency of highly acidic proteins (~4.5) that was more similar to the OM43-B representative (MBRSH7) compared to OM43-A (HTCC2181). Overall, the acidic skew of proteins in Met-BS01-1 provide strong support that BS01 is a lineage of marine *Methylophilaceae* rather than a freshwater lineage introduced to the Canada Basin by river discharge.

<u>3.4.4. Gene content variation and marine adaptation in BS01</u>

Gene content was compared between MetBS01-1 and a set of freshwater and marine *Methylophilaceae* genomes **(Supplementary Figure 3.2)**. In total, 503 Met-BS01-1 genes were conserved among the analyzed *Methylophilaceae* genomes, while 349 genes were unique to Met-BS01-1. In agreement with the phylogenetic affiliation of BS01 with sediment *Methylotenera*, an additional 259 genes were shared between Met-BS01-1, *M. versatilis* and *M. mobilis*. Genes shared exclusively between MetBS01-1 and one or more marine OM43 genomes were not detected.

We next interrogated Met-BS01-1 for genes that may be associated with a marine lifestyle, such as osmoregulation and ion metabolism. A H^+-translocating NADH dehydrogenase (NDH) was present in Met-BS01-1 rather than the Na^+-translocating NADH:quinone oxidoreductase (NQR) that is often associated with marine bacteria (**Supplementary Figure 3.3**) (Walsh, Lafontaine and Grossart 2013). We identified 196 genes that exhibit highest similarity with homologs from outside the *Methylophilaceae* family. Of these, none recognizably originated from typical marine bacteria. However, several were associated with sodium transport, including a

Na^+/melibiose symporter related to Alphaproteobacteria and a small-conductance mechanosensitive channel and a calcium/sodium antiporter related to Gammaproteobacteria (**Supplementary Figure 3.3**). An additional set of genes associated with Na^+ metabolism were shared between Met-BS01-01 and *Methylotenera*, including Na^+/proline (*putP*), Na^+/H^+-dicarboxylate (*gltT*), alanine/glycine:cation (*agcS*), and neurotransmitter/Na^+ symporters, as well as Na^+/H^+ (*nhaA*) and monovalent cation/H^+ antiporters (**Supplementary Figure 3.3**).

3.4.5. BS01 energy and nutrient metabolism

Metabolic reconstruction of Met-BS01-1 indicated the ability to grow on methanol as a sole source of carbon and energy **(Figure 3.4a)**. Similar to other pelagic *Methylophilaceae*, a single lanthanide-dependent methanol dehydrogenase (*xoxF4*) was present, while the calcium-dependent methanol dehydrogenase (*mxaF*) was not detected. Methylotrophic activity of BS01 in Arctic Ocean surface water was supported by an abundance of *xoxF4* transcripts in Canada Basin metatranscriptomes **(Figure 3.4b; Supplementary Table 3.3)**. Quantitative PCR targeting either BS01 or OM43 xoxF4 genes specifically verified that BS01 is restricted to the surface waters of the Canada Basin, while OM43 is more broadly present in the water column (**Figure 3.4c**). Similar to other pelagic methylotrophs, Met-BS01-1 encoded all genes for the tetrahydrofolate (H_4F) pathway for formaldehyde oxidation, the ribulose monophosphate (RuMP) cycle for formaldehyde assimilation/oxidation, and formate oxidation via formate dehydrogenase **(Figure 3.4a)**. In addition, Met-BS01-1 possessed the tetrahydromethanopterin (H_4MPT) pathways for formaldehyde oxidation, which was thought to be restricted to sediment methylotrophs (Chistoserdova 2011b) but recently identified in *Ca.* 'M. turicensis' (Salcher *et al.* 2019). Interestingly, we only detected Met-BS01-1 transcripts from the H_4MPT pathway, and not the H_4F pathway in Canada Basin metatranscriptomes (**Figure 3.4b**). Known genes involved in the

processing of other C1 and C1-related compounds including DMSP, glycine betaine, methylated amines were not present in Met-BS01-1.

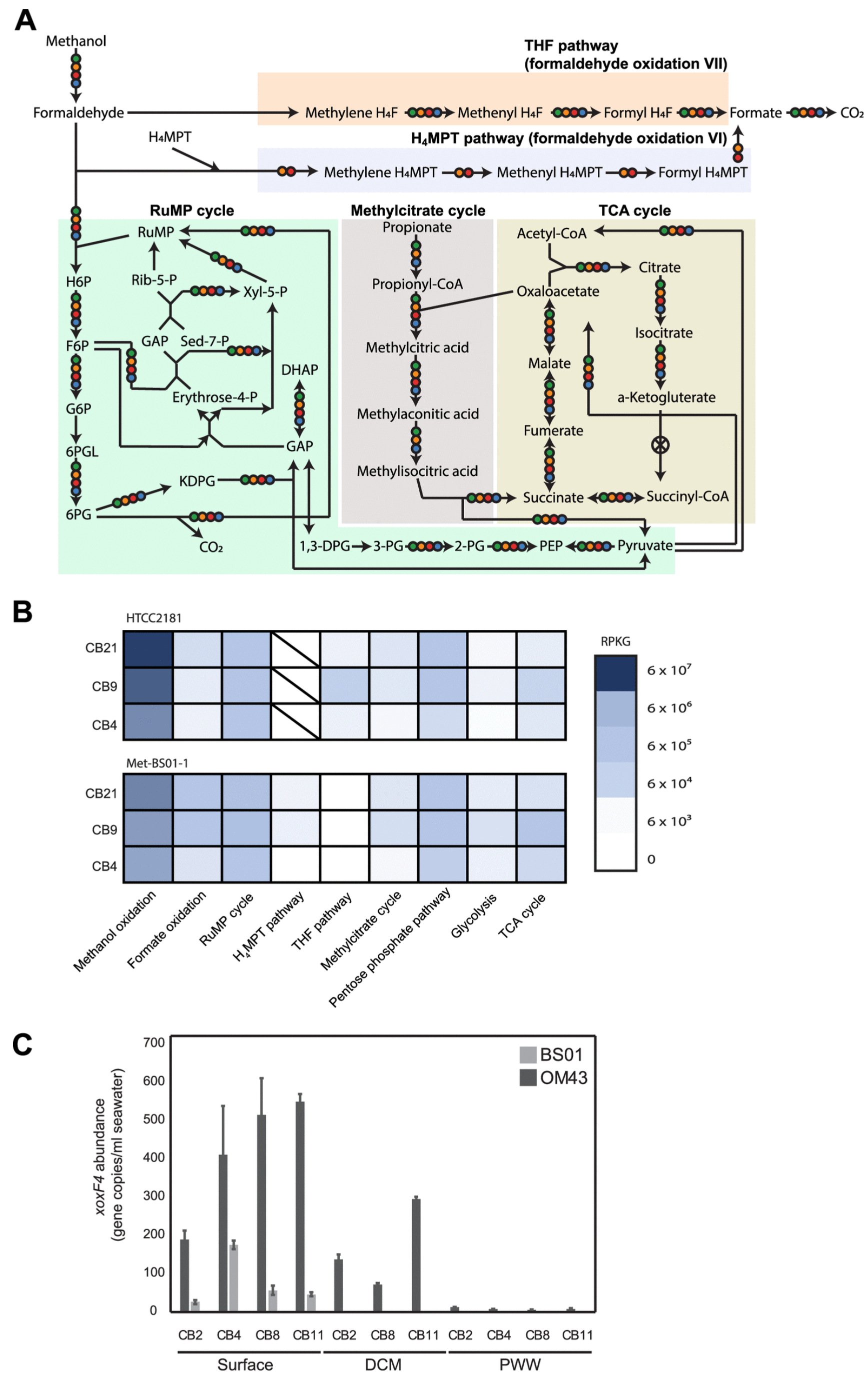
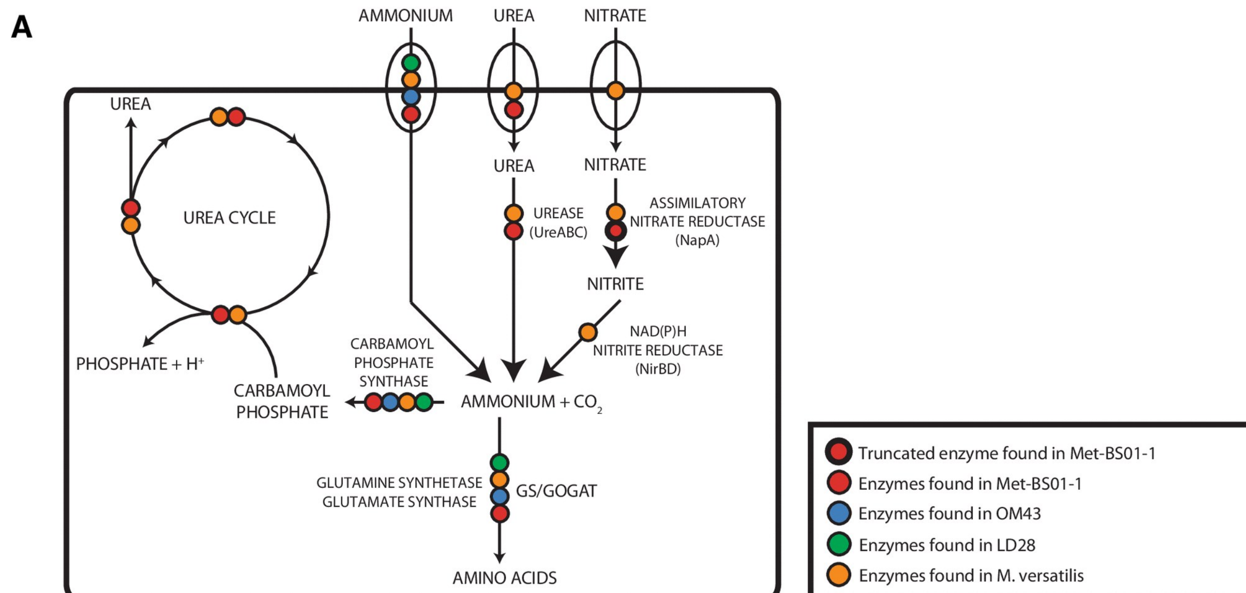


Figure 3.4 *Reconstruction of methylotrophic metabolism in BS01 and comparison to other Methylophilaceae. (A) Distribution of central one-carbon metabolism genes. (B) Gene expression pattern for central carbon metabolism pathways in Canada Basin surface waters revealed through fragment recruitment of metatranscriptomics against Met-BS01-1 and*

HTCC2181 genomes. (C) Quantification of BS01 and OM43 abundances in the Canada Basin using qPCR analysis of xoxF4 gene abundance. Error bars indicate standard deviation. DCM, deep chlorophyll maximum; PWW, Pacific winter water.

With respect to nitrogen acquisition, Met-BS01-1 encoded an ammonium transporter (*amtB*) and the glutamine synthetase/glutamine oxoglutarate aminotransferase (GS/GOGAT) assimilation pathway similar to all known *Methylophilaceae* **(Figure 3.5a)**. The complete set of genes required for assimilatory nitrate reduction was not identified (missing nitrate and nitrite transporters and *nirB*), and *napA* was truncated, encoding only the last 385 amino acids of the typical 800^+ amino acids, suggesting a non-functional pseudogene **(Figure 3.5a)**. In contrast to previously described pelagic methylotrophs, Met-BS01-1 shared the ability for urea utilization with *Methylotenera*. A urea ABC-type transporter encoded by the *urtABCDE* operon and an operon encoding the full urease enzyme and accessory proteins (*ureABCDEFG*) were present in Met-BS01-1 and exhibited highest similarity to orthologs from the freshwater *Methylotenera* sediment isolates (*e.g.* 80-93% for the UreA-UreC protein subunits). Urea use by BS01 was evident as transcripts for urea transport and assimilation and the GS/GOGAT pathway were detected Canada Basin metatranscriptomics. **(Figure 3.5b; Supplementary Table 4)**. Met-BS01-1 also shared an incomplete urea cycle with the sediment methylotrophs. Genes annotated as amino acid transporters were not identified in Met-BS01-1. Interestingly, Met-BS01-1 encodes an ABC-type phosphate transport system as well as polyphosphate kinase and exopolyphosphatase, suggesting an ability to store phosphorus under nitrogen limited conditions, and mobilize the stored phosphorus when enough nitrogen is available for growth.



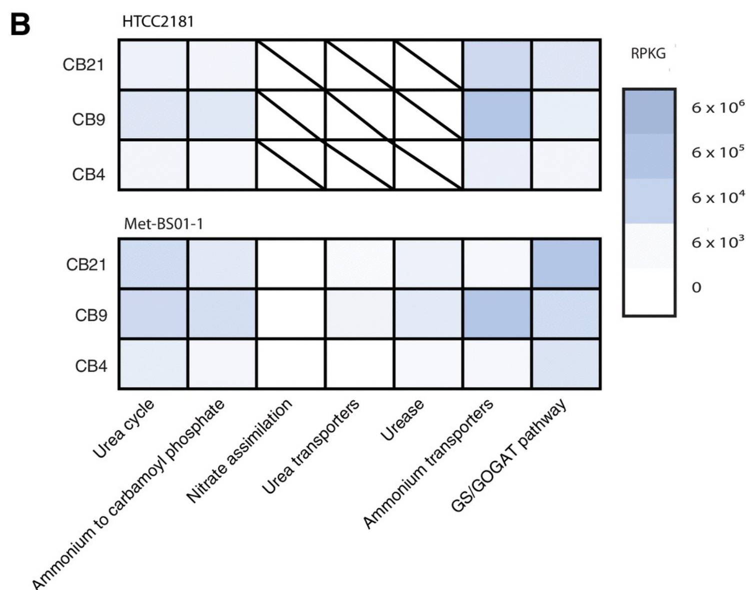


Figure 3.5 *Reconstruction of nitrogen metabolism in BS01 and comparison to other Methylophilaceae. (A) Distribution of central nitrogen metabolism genes. (B) Gene expression pattern for central nitrogen metabolism pathways in Canada Basin surface waters revealed through fragment recruitment of metatranscriptomics against Met-BS01-1 and HTCC2181 genomes.*

3.4.6. BS01 biogeography across aquatic ecosystems

The presence of BS01 in ecosystems outside of the Arctic Ocean was investigated by applying a combination of phylogenetic marker (*xoxF4*) and fragment recruitment analyses to

diverse aquatic metagenomes. We identified several BS01 *xoxF4* genes in metagenomes from Antarctic seawater and a broad *xoxF4* diversity in eastern North American estuary (Chesapeake and Delaware Bays) samples that ranged in salinity from 15 to 30 PSU **(Figure 3.6a)**. Since methanol dehydrogenase genes are prone to lateral gene transfer (Kolb and Stacheter 2013; Kang, Dunfield and Semrau 2019), we verified the presence of BS01 in these samples using metagenomic fragment recruitment. Fragment recruitment only detected Met-BS01-1 in polar surface waters, including the Southern Ocean (Scotia Sea) and estuaries **(Figure 3.6b)**. In contrast to the restricted detection of Met-BS01-1, fragment recruitment against the HTCC2181 genome was observed for all marine biomes (coastal, polar, trades, westerlies) analyzed as well as estuaries **(Figure 3.6b)**.

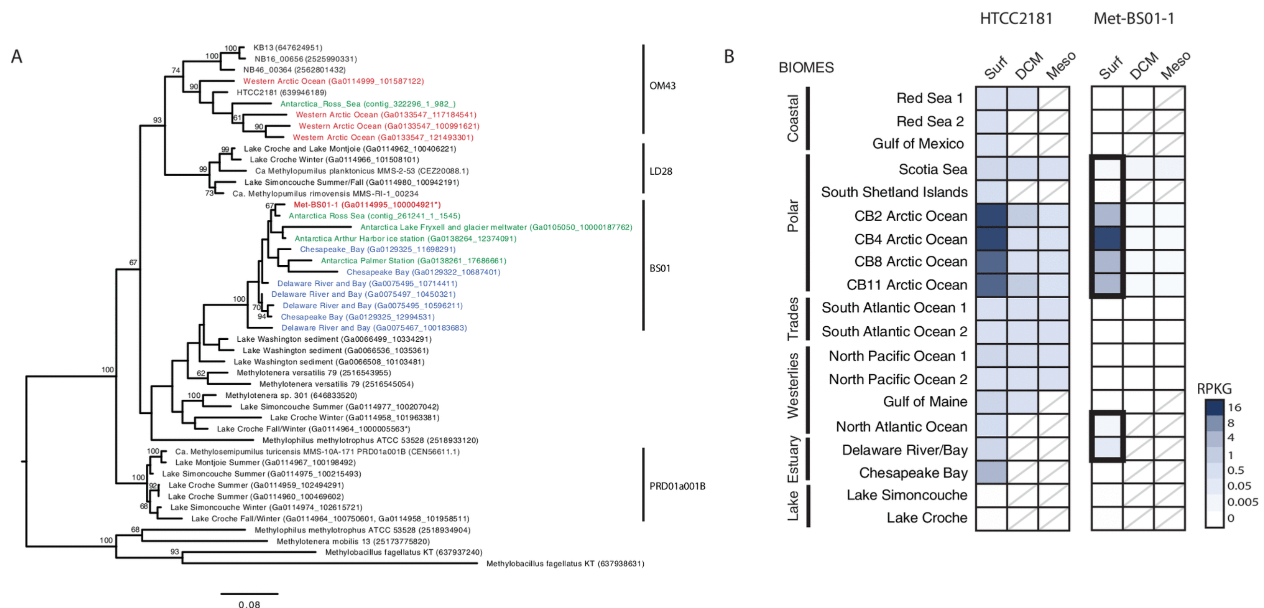


Figure 3.6 *Biogeography of BS01 across the global ocean and estuaries. (A) Phylogenetic analysis of the Methylophilaceae family across multiple aquatic regions and depths using the XoxF4 methanol dehydrogenase protein recovered from metagenomes. The tree was inferred using maximum likelihood (500 bootstraps) and JTT + gamma distributed with invariants (four categories) sites model of evolution, with nearest-neighbor interchange heuristic search method. Colored sequences are those from the Arctic Ocean (red), the Antarctic (green), or estuaries (blue). Only bootstrap values of >60 are included in the tree. (B) Distribution of BS01 and OM43-A revealed through fragment recruitment of aquatic metagenomes against Met-BS01-1 and HTCC2181 genomes reported as reads per kilobase of the MAG per gigabase of metagenome (RPKG). The diagonal lines are to signify no metagenome data for that water column feature.*

3.5. Discussion

3.5.1. *Methylophilaceae* diversity in the salinity-stratified Arctic Ocean

In this study, we discovered a diverse assemblage of *Methylophilaceae* in the salinity-stratified waters of the Canada Basin. Arctic *Methylophilaceae* were comprised of an array of marine OM43 subclades and a distantly related lineage herein termed BS01 (**Figure 3.1**). All OM43 subclades identified in the Arctic Ocean were within OM43-A (**Figure 3.2a**), in agreement with previous findings that OM43-A is generally more common in colder habitats compared to OM43-B (Jimenez-Infante *et al.* 2015). At lower latitudes, OM43-A is strongly associated with phytoplankton blooms (Morris, Longnecker and Giovannoni 2006; Giovannoni *et al.* 2008), which is a likely source of methanol for growth (Mincer and Aicher 2016). In contrast, we find that OM43 was relatively rare in the chlorophyll maximum of the Canada Basin and was in fact more abundant in the extremely oligotrophic surface waters. This vertical structuring of OM43 was observed across all analyses including the metagenome ITS diversity (**Figure 3.2b**) and fragment recruitment analyses (**Figure 3.6b**), as well as OM43 distribution assessed by quantitative PCR (**Figure 3.4c**). One explanation for low abundance in the Canada Basin chlorophyll maximum is that it is comprised mostly of eukaryotic picophytoplankton (Lovejoy, Galand and Kirchman 2011), which may not be as significant a source of methanol as larger bloom-forming phytoplankton. But, what supports the methanol metabolism of the surface water *Methylophilaceae*? In addition to *in situ* production, methanol in the surface waters may originate from atmospheric deposition, which is elevated in the Arctic compared to lower latitude regions due to the colder temperatures (Millet *et al.* 2008; Sjostedt *et al.* 2012). Another possibility is that Arctic *Methylophilaceae* are associated with phytoplankton blooms that periodically occur in the surface waters. In particular, under-ice blooms have been observed throughout the Arctic Ocean

(Yager *et al.* 2001; Fortier *et al.* 2002; Arrigo *et al.* 2012; Mundy *et al.* 2014; Assmy *et al.* 2017) including the Beaufort Sea (Hill *et al.* 2018; Ardyna *et al.* 2020). Betaproteobacteria are present under the ice in the Beaufort Sea during the winter/spring (Galand *et al.* 2008; Pedrós-Alió, Potvin and Lovejoy 2015) and (Collins, Rocap and Deming 2010) identified OM43 under, as well as within sea ice. Overall, these results demonstrate that marine *Methylophilaceae* are common in the stratified waters of the Canada Basin and investigation of their seasonal dynamics would provide a deeper understanding of their ecological role in Arctic Ocean ecosystems.

3.5.2. Freshwater-marine transitions within the *Methylophilaceae*

Insights into the evolutionary adaptations associated with major habitat transitions, such as between marine and freshwater environments, can be revealed through comparison of closely related taxa from different habitats (Walsh, Lafontaine and Grossart 2013; Bižić-Ionescu and Ionescu 2016; Paver *et al.* 2018). Here, we provided strong evidence that BS01 represents a second lineage of marine *Methylophilaceae* that arose independently of marine OM43. The discovery of multiple evolutionary origins of pelagic marine *Methylophilaceae* provides a rare opportunity to compared the pathways of evolutionary adaptation to the ocean. OM43 is thought to have arisen from a sediment to pelagic transition in a freshwater ecosystem, followed by a second transition from fresh to marine waters (Salcher *et al.* 2015). BS01 is related to freshwater sediment *Methylotenera*, as well as pelagic *Ca. "*M. turicensis" and may therefore have originated through a similar sediment to pelagic, then freshwater to marine, series of transitions. However, we propose that pelagic BS01 evolved along a different path, directly from a marine ancestor residing in sediments. Compared to freshwater *Methylotenera*, the BS01 proteome has undergone extensive changes in amino acid compositions (acidic shifts in pI) (**Figure 3.3c**), which requires long evolutionary time (Cabello-Yeves and Rodriguez-Valera 2019). Yet, changes associated with

oligotrophic conditions such as reduced GC content and a shift to less nitrogen-rich amino acids are not as striking. And, although the Met-BS01-01 genome is approaching the small size of other oligotrophs (Giovannoni, Cameron Thrash and Temperton 2014) including LD28 and OM43, it is still considerably larger (**Figure 3.3b**). Moreover, ongoing genome reduction was evident by the identification of partial deletion of the nitrate reductase, with only a *napA* pseudogene remaining. Hence, BS01 appears to have an established set of proteome modifications associated with a marine lifestyle but is at an intermediate stage with respect to pelagic adaptation. These findings further the notion that *Methylophilaceae* serve as a valuable model for "evolution in action" studies (Salcher *et al.* 2015). Additional comparative studies that include new isolates (Lanclos *et al.* 2016) and MAGs from a broader diversity of environments including marine sediments should further advance our understanding of microbial habit transitions in aquatic ecosystems.

Lateral gene transfer plays a significant role in bacterial diversification and the acquisition of genes involved in osmoregulation has been implicated in marine transitions (Penn and Jensen 2012; Walsh, Lafontaine and Grossart 2013). Several sodium transporters were identified in the Met-BS01-1 genomes that may have originated by lateral transfer from Gamma and Alphaproteobacteria. But, overall there wasn't a striking pattern of "marine gene" acquisition in BS01. Given that marine adaptation requires extensive adaptation across the whole proteome (**Figure 3.3c**), we hypothesized that orthologous gene replacement of freshwater-adapted core proteins with more acidic marine homologs may have played a role in BS01 evolution. Phylogenetically related proteins that only differ slightly in their acidic amino acid content would have the most chance of successful incorporation in the recipient genome (Cabello-Yeves and Rodriguez-Valera 2019). With this in mind, we looked for evidence of OM43 serving as a donor of marine orthologs of core genes to BS01. We did not detect any genes exclusive to BS01/OM43.

Moreover, BS01 genes shared with OM43 were consistently more similar to *Methylotenera* homologs, which is evidence against orthologous replacement. These result show that although BS01 and OM43 share the same habitat and are phylogenetically related, they apparently rarely undergo genetic exchange. A similar observation was made for a newly discovered freshwater lineage of SAR11 in ancient Lake Baikal, where Cabello-Yeves et al. (2018) wondered why the lineage had not acquired genetic material from previously established freshwater SAR11 relatives (Cabello-Yeves *et al.* 2018). Perhaps finer resolution studies will identify orthologous replacements or even within-gene recombination events as evolving aquatic bacteria share the same pool of marine or freshwater alleles.

3.5.3. Niche differentiation of BS01 and OM43

The coexistence of BS01 and OM43 in Canada Basin surface waters leads to questions on how these related methylotrophs are ecologically differentiated and how BS01 successfully competes with more established OM43 bacteria. Salinity may play a direct role in niche differentiation of OM43 and BS01, but neither the biogeographic distributions or genome characteristics provide immediate insights about how. Both groups were detected across a broad salinity range in the Canada Basin and estuaries (**Figure 3.6b**) and possessed similarly acidic proteomes (**Figure 3.3c**). Both groups also appear to be highly specialized for methanol oxidation, but differences were apparent that may be ecologically relevant. Genome reduction in both groups has converged on a similar methylotrophic metabolism comprised of the same core metabolic modules identified by Salcher et al. (2019): methanol oxidation via XoxF, formaldehyde oxidation via H_4F and the RuMP cycle, and formate oxidation. However, one metabolic difference between BS01 and OM43 may be related to formaldehyde oxidation rates because BS01 has retained the H_4MPT pathway for formaldehyde oxidation. Evidently, the H_4MPT, but not the H_4F pathway is

expressed in the Canada Basin. Perhaps the H_4MPT pathway can more rapidly remove toxic formaldehyde. Under conditions where growth is limited by inorganic nutrient availability and the reducing power of methanol is being funneled into maintenance energy generation rather than carbon assimilation and growth, perhaps the H_4MPT pathway provides an advantage over the H_4F pathway.

Another ecologically-relevant difference between OM43 and BS01 may be related to nitrogen metabolism. Both groups have independently lost the genes for nitrate transport and assimilatory reduction, while retaining ammonia transporters (**Supplementary Figure 3.3**). However, although BS01 has undergone genome reduction compared to sediment dwellers, the full complement of genes for urea transport and utilization as a nitrogen source has been retained. Urea is present at relatively high, but variable concentrations in the Arctic Ocean (Gratton, Price and Simpson 2008), and often exceeds that of ammonia. In fact, urea has been show to fuel Arctic nitrification (Alonso-Saez *et al.* 2012). Sources of urea include riverine input, excretion and sloppy feeding by zooplankton (Conover and Gustavson 1999) and inputs form the melting of seasonal fast ice (Conover *et al.* 1999). Urea is hypothesized to replace ammonia as a waste product for organisms residing in confided habitats like sea ice because it is less toxic. The input from melting sea ice may be particularly relevant if these methylotrophs are associated with under-ice and ice-associated phytoplankton. The unique ability of BS01 to use urea during these conditions would provide an advantage over OM43 given the limited amounts of inorganic nitrogen in the Canada Basin. In fact, a seasonal study in the Arctic using SIP showed no evidence of ^{15}N urea uptake in summer, whereas winter samples yielded estimates of 30 % isotopic labelling of bacterial populations, including Betaproteobacteria (Connelly *et al.* 2014). Taken together, salinity, differences in C1 metabolism, and nitrogen utilization may all play a role in niche differentiation.

Although genomes can provide some insights into metabolic traits that differentiate ecology, additional physiological traits that cannot be predicted from genomes such as temperature optimum, substrate transport affinities and biological interactions with phage and predators may be additionally relevant.

3.5.4. Bacterial evolution in a changing Arctic Ocean

The Arctic Ocean is a relatively enclosed ocean that is intensely influenced by riverine input. As such, Arctic Ocean surface waters represent a mosaic marine-freshwater interface that could be a global evolutionary hotspot for aquatic bacteria. Along with the introduction of vast amounts of freshwater and terrestrial nutrients (Guéguen *et al.* 2012; Carmack *et al.* 2016; Proshutinsky *et al.* 2019), Arctic rivers may also serve as a conduit for the flow of organisms and genes from terrestrial to marine ecosystems. Indeed, a previous study showed that the ability of Arctic marine bacteria to use aromatic compounds of terrestrial origin evolved, at least in part, by lateral acquisition of genes from terrestrial bacteria (Colatriano *et al.* 2018). In the current study, we have shown that Arctic surface waters may support the evolution of bacteria that relatively recently colonized the oceans. Traditionally, these freshwater-marine transitions were considered to be rare in bacterial evolution (Logares *et al.* 2009). Some years ago, it was suggested that cross colonization may be more common than currently realized and that previously undescribed invaders may be hiding in the "rare biosphere" (Walsh, Lafontaine and Grossart 2013). Indeed, a number of studies have since supported this hypothesis (Bižić-Ionescu and Ionescu 2016; Paver *et al.* 2018) and BS01 certainly fits the criteria for rarity. Hence the relatively fresh Arctic Ocean surface waters that are strongly influenced by terrestrial inputs may support a wider diversity of rare marine bacteria with relatively recent freshwater origins that are awaiting discovery. Finally, given the ongoing freshening of the Arctic Ocean, our results suggest that these relative newcomers

to the ocean microbiome may increase in abundance and therefore ecological significance in a near-future Arctic Ocean.

3.6. Materials and Methods

3.6.1. Sampling and metagenomic data generation

Samples from 4 stations in the Canada Basin (CB2, CB4, CB8, and CB11) were collected aboard the CCGS Louis S. St-Laurent during the Joint Ocean Ice Study (JOIS) research mission in September 2015. Twelve samples were collected and the associated environmental variables were measured for each sample **(Table 3.1)**. Between 4-7 L of seawater was sequentially filtered through a 50 µm pore mesh, followed by a 3 µm pore size polycarbonate filter and a 0.22 µm pore size Sterivex filter (Durapore; Millipore, Billerica, MA, USA). Filters were preserved in RNAlater and stored at −80 °C until processed in the laboratory. DNA was extracted from the Sterivex filters using a SDS lysis protocol as described in Colatriano, *et al*. (2018). DNA sequencing was performed at the Department of Energy Joint Genome Institute (Walnut Creek, CA, USA) on the HiSeq 2500-1TB (Illumina) platform using 150 PE technology. The metagenomic data is deposited in the IMG database under GOLD Project IDs: Gp0134345- Gp0134356.

3.6.2. 16S rRNA gene and ITS analysis

We used the complete 16S rRNA gene from OM43 strain HTCC2181 (Giovannoni *et al.* 2008) to extract the 16S rRNA gene from each of the twelve single sample metagenome assemblies using BLASTn (Altschul *et al.* 1990). The sequences > 500 bp were included in a phylogenetic analysis with reference sequences and 16S rRNA sequences from other biogeographic studies. A multiple sequence alignment was generated using the MUSCLE algorithm as implemented in MEGA v.7 (Kumar, Stecher and Tamura 2016). A maximum likelihood tree was constructed using

the GTR + Gamma distribution (4 categories) model of nucleotide substitution in MEGA v.7 with 100 bootstraps (Kumar, Stecher and Tamura 2016).

We used the complete ITS region from a reference OM43 bacterium (HTCC2181) to extract ITS regions from each metagenome assembly using BLASTn (Altschul *et al.* 1990). The sequences were clustered using CD-hit (Fu *et al.* 2012) at an identity of 100%. To assign sequences to specific clades, the Arctic ITS sequences were analyzed using reference sequences from published genomes and ITS sequences used in a previous biogeographic study (Jimenez-Infante *et al.* 2015). The sequences were aligned using the MUSCLE algorithm as implemented in MEGA v.7 (Kumar, Stecher and Tamura 2016) with poorly aligned sequences removed after visual inspection. A maximum likelihood tree was constructed using the GTR + Gamma distribution (4 categories) model of nucleotide substitution in MEGA v.7 with 100 bootstraps (Kumar, Stecher and Tamura 2016). The distribution of subclades across the stations and depths was determined by summing the average read depth of all ITS sequences within each subclade. A Principal Coordinates Analysis (PCoA) ordination of Bray-Curtis dissimilarities of the Arctic samples was performed based on the read depth of different ITS clades to determine the distribution across samples. The envdist function as implemented in vegan (Oksanen *et al.* 2020) with 999 permutations was used for *Post-Hoc* tests of environmental variables.

3.6.3. MAG generation and analysis

Metagenomic binning was performed on scaffolds > 5 kb in length using MetaWatt (Strous *et al.* 2012). Binning was performed using tetranucleotide frequency and the relative weight of coverage was set to 0.75, with the optimize bins and polish bins options on. The taxonomic identity of MAGs was assessed using a concatenated phylogenetic tree based on 138 single-copy conserved genes as implemented in MetaWatt (Strous *et al.* 2012). Estimation of MAG completeness and

contamination was performed using CheckM (Parks *et al.* 2015) and suspected contamination was manually removed. A single putative BS01 MAG (Met-BS01-1) was identified from the CB2 surface metagenome for further analysis.

3.6.4. Concatenated protein phylogeny

The distribution of orthologous genes was analyzed using ProteinOrtho (Lechner *et al.* 2011). 48 single-copy orthologous genes present in all genomes were identified and selected for concatenated phylogenetic analysis. Each orthologous protein family was aligned using MUSCLE (implemented in MEGA6) and alignment positions were masked using the probabilistic masker ZORRO (Wu, Chatterji and Eisen 2012), masking columns with weights <0.5. Phylogenetic reconstructions were conducted by maximum likelihood using MEGA6-v.0.6 and the following settings: JTT substitution model, gamma distribution with invariant sites model for the rate variation with four discrete gamma categories, and the nearest-neighbor interchange (NNI) heuristic search method with a bootstrap analysis using 100 replicates.

3.6.5. Comparative genomics

Inference of protein function and metabolic reconstruction were based on the IMG annotations provided by the JGI and using the Pathologic software available through Pathway Tools (Karp *et al.* 2010). The pangenomic visualization of *Methylophilaceae* was created using the anvi'o tool (Eren *et al.* 2015; Delmont and Eren 2018). Rodriguez-R and Konstantinidis (2016) toolbox was used to compute the average amino acid identity (AAI) (Rodriguez-R and Konstantinidis 2016). The proteome isoelectric point, was calculated with the software Pepstats from the EMBOSS package (Rice, Longden and Bleasby 2000).

3.6.6. Metatranscriptomic analysis

RNA samples were collected during a JOIS mission in September 2017. RNA was

extracted from the Sterivex® filters (3 to 0.22 µm size fraction) with a modified protocol (Shi, Tyson and DeLong 2009; Stewart, Ottesen and Delong 2010), which employs both the mirVana miRNA isolation kit (Invitrogen) and the RNeasy RNA cleanup kit (Qiagen). cDNA library preparation and sequencing was performed at the Department of Energy Joint Genome Institute (JGI) (Walnut Creek, CA) on the HiSeq 2500-1TB (Illumina) platform using 150 PE technology. To determine the activity and distribution of the Arctic *Methylophilaceae* MAG and reference genome HTCC2181 in the Arctic Ocean, unassembled metatranscriptomic data was recruited against the protein-coding gene sequences using BBMAP (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/), with a minimum identity of 95%. The reads per kilobase of the MAG per gigabase of metatranscriptome (RPKG) was calculated to control for differences in raw reads between samples. The metatranscriptome data is deposited in the IMG database under GOLD Project IDs: Gp0323995, Gp0324000 and Gp0323990.

3.6.7. Methanol dehydrogenase (XoxF4) phylogeny

The full length XoxF4 amino acid sequences from the reference isolate HTCC2181 and from the Met-BS01-1 MAG were used to assess the presence of BS01 in 1,362 metagenome assemblies from aquatic communities available at IMG/M using BLASTp (Altschul *et al.* 1990). The sequences > 300 bp were included in a phylogenetic analysis with reference sequences and sequences from other biogeographic studies. A multiple sequence alignment was generated using the MUSCLE algorithm as implemented in MEGA v.7 (Kumar, Stecher and Tamura 2016). A maximum likelihood tree was constructed using the JTT substitution model, gamma distribution (4 categories) and the nearest-neighbor interchange (NNI) heuristic search method in MEGA v.7 with 100 bootstraps (Kumar, Stecher and Tamura 2016).

3.6.7. Fragment recruitment

The distribution of the Arctic *Methylophilaceae* MAG and HTCC2181 in the multiple aquatic biomes was determined using the best-hit reciprocal blast approach reported in Colatriano et *al.* (2018). Unassembled metagenomics data from 45 samples at 20 sites **(Table S2)** were recruited to the Arctic *Methylophilaceae* MAG and HTCC2181. All hits from the initial blast were then reciprocally queried against the Arctic *Methylophilaceae* MAG and HTCC2181. The best-hit was reported and hits with an alignment length ≥ 100 bp and a percent identity ≥ 95% were counted. To compare the results among the different datasets, the number of recruited reads was normalized to the total number of reads in each sample. The final coverage results were expressed as the number of reads per kilobase of the MAG per gigabase of metagenome (RPKG).

3.6.8. *xoxF4* primer design and qPCR

Primer sets were designed to amplify a ~200 bp fragment of the *xoxF4* gene f from BS01 or OM43 specifically. The primer set specific to BS01 was F1024_BS01 (5'- ATT GCT AAA TGG GGC TAC - 3') and R1161_BS01 (5'- GTT GAA TGT ATA TGC GAA ACC - 3'). The primer set specific to OM43 was F1015_OM43 (5'- GAY TTA GAY ACA GGT ATG GCR - 3') and R1161_OM43 (5'- CCA TGT GTA WGC AAA ACC GTT TCT - 3'). Specificity of primer sets was validated through cloning and sequencing of the DNA inserts **(Supplementary Figure 3.4)**. Cloning, cDNA synthesis, PCR and qPCR were performed as described in Ramachandran and Walsh (2015), with the annealing temperature of 52.3°C for both primer sets.

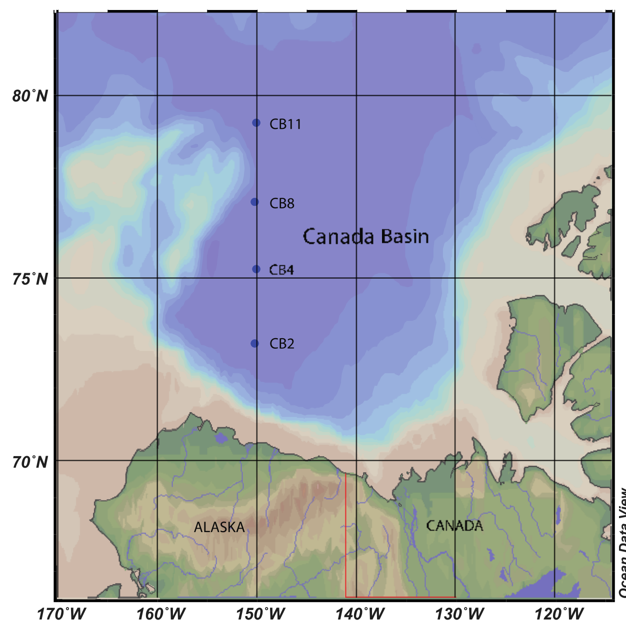
3.7. Acknowledgements

We would like to thank both the Captain and crew of the CCGS Louis S. St-Laurent and the scientific teams aboard. The work was conducted in collaboration with the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, and was supported under

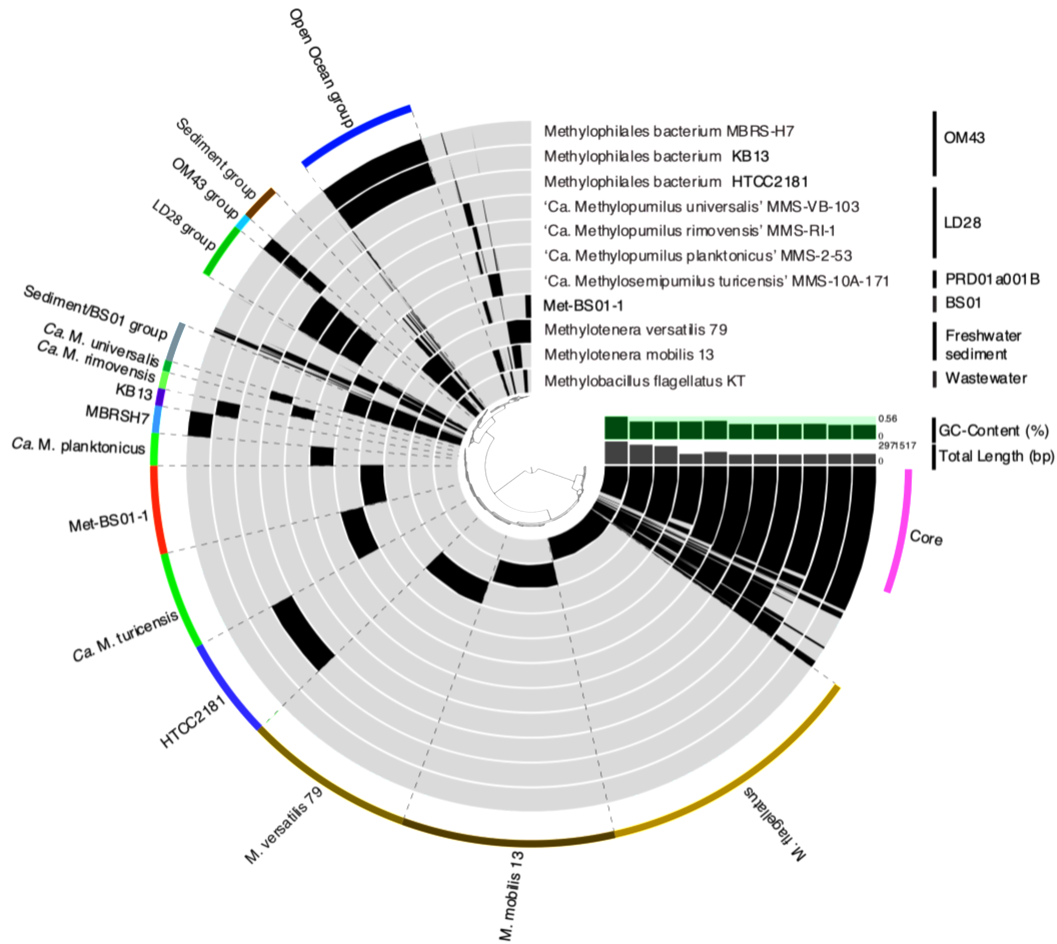
Contract No. DE-AC02-05CH11231. Funding from the Canadian Natural Science and Engineering Research Council (NSERC) Discovery (D.W.) and Canada Research Chair programs (D.W.) are acknowledged. A.R. was supported by Concordia's Institute for Water, Energy and Sustainable Systems. We would like to thank Barbara Campbell for providing access to metagenomes and environmental data associated with the Chesapeake Bay and Delaware Bay samples used in the fragment recruitment analysis.

3.8. Supplementary Figures and Tables

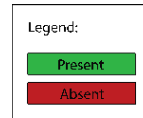
Supplementary Figure 3.1 *Map of the study location in the Canada Basin, Arctic Ocean*



Supplementary Figure 3.2 *Comparative genomics of shared gene content among Methylophilaceae strains visualized using Anvi'o.*



Supplementary Figure 3.3 *Summary of the distribution of metabolism modules across Methylophilaceae genomes.*



		HTCC2181	KB13	MBRS-H7	M. pla	M. tur	**Met-BS01-1**	M. ver 79	M. mob 13	M. fla
Motility, Extracellular structures	flagella	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present	Present
	chemotaxis	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present	Present
	type IV pili	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present	Present
	T4SS	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
Methylotrophy	MOX-PQQ	Present	Present	Present	Present	Present	Present	Present	Present	Present
	XoxF	Present	Present	Present	Present	Present	Present	Present	Present	Present
	MxaF	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present	Present
	MADH	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent	Present
	FaDH	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
	H4F - folD	Present	Present	Present	Present	Present	Present	Present	Present	Present
	H4MPT	Absent	Absent	Absent	Absent	Present	Present	Present	Present	Present
	FOX	Present	Present	Present	Present	Present	Present	Present	Present	Present
	RuMP	Present	Present	Present	Present	Present	Present	Present	Present	Present
Nitrogen	NH4 transporter	Present	Present	Present	Present	Present	Present	Present	Present	Present
	NO3 transporter	Absent	Absent	Absent	Absent	Present	Absent	Present	Present	Present
	ass. NO3 reduction	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present
	diss. NO3 reduction	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent
	NO detoxification	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
	denitrification	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent
	urea utilization	Absent	Absent	Absent	Absent	Absent	Present	Present	Absent	Present
	cyanate utilization	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
Sulfur	SO4 transporter	Present	Present	Present	Absent	Present	Present	Present	Present	Present
	ass. SO4 reduction	Present	Absent	Present	Absent	Present	Absent	Absent	Present	Absent
	sulfonate uptake	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
	DMS degradation	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent	Absent
Salinity	NDH-dehydrogenase	Absent	Present	Present	Present	Present	Present	Present	Present	Present
	NQR-dehydrogenase	Present	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
	ectoine biosynthesis	Absent	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent
	DtpD	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent
	AlsT	Absent	Present	Absent	Absent	Absent	Absent	Absent	Present	Absent
	ActP	Present	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent
	GltT	Present	Present	Present	Absent	Present	Present	Present	Present	Present
	NhaP2	Absent	Present	Present	Absent	Absent	Absent	Absent	Absent	Absent
	NhaE-like	Present	Present	Present	Absent	Absent	Present	Present	Absent	Absent
Assorted Pathways	CRISPR Cas	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present
	AA transporter	Absent	Absent	Absent	Absent	Absent	Absent	Present	Absent	Present
	fructose PTS	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Absent
	MCA	Present	Present	Present	Absent	Present	Present	Present	Present	Absent
	B12 biosynthesis	Absent	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present
	B12 salvage	Absent	Absent	Absent	Absent	Absent	Absent	Present	Present	Present
	B12 tranporter	Present	Present	Present	Present	Present	Present	Present	Present	Present
	proteorhodopsins	Absent	Absent	Absent	Present	Present	Absent	Absent	Absent	Absent
	xantho-like rhodopsin	Present	Present	Present	Present	Absent	Absent	Absent	Absent	Absent

Supplementary Table 3.1 *Characteristics of Methylophilaceae genomes.*

Table S1. Characteristics of Methylophilaceae genomes

	Met-BS01-1	HTCC2181	KB13	MBRS-H7	'Ca. M. rimovensis' MMS-RI-1	'Ca. M. planktonicus' MMS-2-53	'Ca. M. universalis' MMS-VB-103	'Ca. M. turicensis' MMS-10A-171	Methylotenera versatilis 79	Methylotenera mobilis 13	Methylobacillus flagellatus KT
Isolation location	Beaufort Sea	Oregon	Hawaii	Red Sea	Římov reservoir	Zurich	Zurich	Zurich	Washington	Washington	-
Isolation source	Seawater	Seawater	Seawater	Seawater	Lake	Lake	Lake	Lake	Lake sediment	Lake sediment	Activated sludge
Genome size (Mbp)	1.5	1.3	1.3	1.4	1.3	1.4	1.3	1.8	2.6	2.8	3.0
GC content (%)	42.7	37.9	35.3	35.5	36.7	37.0	37.0	44.5	41.9	45.6	55.7
AAI with Met-BS01-1 (%)	-	56	53	54	59	59	59	64	69	66	62
Lysine content (%)	6.0	7.5	8.1	8.1	7.8	7.8	7.9	5.7	5.8	5.6	4.2
Arginine content (%)	4.4	3.9	3.5	3.6	4.0	4.0	4.0	4.7	4.3	4.7	5.9

Supplementary Table 3.2 *List of aquatic metagenomes used in the fragment recruitment analysis.*

Accession Numbers	Name	Latitude (°N)	Longitude (°W)	Sampling Depth (m)	Environmental Feature	Biomes
ERS488569	Red Sea 2	23.36	37.2183	5	SRF	Coastal
ERS488599	Red Sea 2	23.4183	37.2450	80	DCM	Coastal
ERS488621	Red Sea 1	21.9467	38.2517	5	SRF	Coastal
ERS488685	Red Sea 1	18.4417	39.8567	60	DCM	Coastal
ERS490542	South Atlantic Ocean 2	-20.9354	-35.1803	5	SRF	Trades
ERS490597	South Atlantic Ocean 2	-21.0292	-35.3498	120	DCM	Trades
ERS490633	South Atlantic Ocean 2	-20.9315	-35.1794	800	MES	Trades
ERS490659	South Atlantic Ocean 1	-30.1367	-43.2899	5	SRF	Trades
ERS490691	South Atlantic Ocean 1	-30.1484	-43.2705	120	DCM	Trades
ERS490714	South Atlantic Ocean 1	-30.1471	-43.2915	800	MES	Trades
ERS491001	South Shetland Islands	-60.2287	-60.6476	5	SRF	Polar
ERS491044	Scotia Sea	-62.0385	-49.5290	5	SRF	Polar
ERS491095	Scotia Sea	-62.2231	-49.2139	90	DCM	Polar
ERS491110	Scotia Sea	-61.9689	-49.5017	790	MES	Polar
ERS493340	North Pacific Ocean 2	31.5168	-159.0460	115	DCM	Westerlies
ERS493372	North Pacific Ocean 2	31.528	-159.0224	550	MES	Westerlies
ERS493390	North Pacific Ocean 1	35.3671	-127.7422	5	SRF	Westerlies
ERS493431	North Pacific Ocean 1	35.4002	-127.7499	45	DCM	Westerlies
ERS493460	North Pacific Ocean 1	35.2698	-127.7268	650	MES	Westerlies
ERS493938	Gulf of Mexico	25.5264	-88.3940	5	SRF	Coastal
ERS493981	Gulf of Mexico	25.6168	-88.4532	125	DCM	Coastal
ERS494006	Gulf of Mexico	25.6236	-88.4500	640	MES	Coastal
ERS494170	Gulf of Maine	39.2305	-70.0377	5	SRF	Westerlies
ERS494208	Gulf of Maine	39.2392	-70.0343	590	MES	Westerlies
ERS494579	North Atlantic Ocean 1	43.6792	-16.8344	5	SRF	Westerlies
ERS494628	North Atlantic Ocean 1	43.7056	-16.8794	25	MIX	Westerlies
ERS493300	North Pacific Ocean 2	31.5213	-158.9958	5	SRF	Westerlies
Ga0075496	Delaware River/Bay	39.283	-75.3633	1	SRF	Estuary
Ga0129319	Chesapeake Bay	38.0754	-76.7533	1	SRF	Estuary
Ga0114958	Lake Croche	46.4955	-72.3000	1	SRF	Lake
Ga0114974	Lake Simoncouche	48.2311	-71.2508	1	SRF	Lake

Supplementary Table 3.3 *Gene expression pattern for central carbon metabolism genes in Canada Basin surface waters revealed through fragment recruitment of metatranscriptomics against Met-BS01-1 and HTCC2181 genomes.*

Pathway	Genes	Met-BS01-1 RPKG			HTCC2181 RPKG		
		CB4	CB9	CB21	CB4	CB9	CB21
Methanol oxidation	xoxF4	1.39E+07	1.77E+07	2.81E+07	6.21E+07	5.62E+07	6.43E+07
Formate oxidation to CO2	formate dehydrogenase iron-sulfur subunit	3.93E+04	9.94E+04	1.76E+05	1.19E+05	9.66E+04	1.12E+05
	formate dehydrogenase major subunit	9.27E+04	1.64E+05	1.89E+05	5.15E+04	0.00E+00	4.81E+04
	formate dehydrogenase subunit delta	0.00E+00	1.35E+05	0.00E+00	0.00E+00	0.00E+00	1.43E+04
	formate dehydrogenase subunit gamma	0.00E+00	1.38E+05	9.92E+04	5.61E+04	6.72E+04	8.00E+04
	FdhD protein	4.65E+04	1.88E+05	1.73E+05	N/A	N/A	N/A
RuMP cycle	hexulose-6-phosphate synthase	1.04E+06	2.15E+06	2.29E+06	1.24E+06	2.05E+06	1.47E+06
	6-phospho-3-hexuloisomerase	0.00E+00	3.62E+04	2.08E+05	1.21E+05	1.28E+05	5.39E+04
formaldehyde oxidation V (H4MPT pathway)	methylene-tetrahydromethanopterin dehydrogenase	0.00E+00	0.00E+00	0.00E+00	N/A	N/A	N/A
	methenyltetrahydromethanopterin cyclohydrolase	0.00E+00	4.02E+04	3.25E+04	N/A	N/A	N/A
formaldehyde oxidation (H4F pathway)	methenyl tetrahydrofolate cyclohydrolase/dehydrogenase	0.00E+00	0.00E+00	0.00E+00	4.70E+04	8.09E+04	2.12E+04
	NADP-dependent methylenetetrahydrofolate dehydrogenase	0.00E+00	0.00E+00	0.00E+00	1.39E+04	8.64E+04	2.21E+04
2-methylcitrate cycle I	2-methylcitrate synthase	1.98E+04	5.01E+04	4.49E+04	3.93E+04	4.40E+04	4.40E+04
	2-methylcitrate dehydratase	0.00E+00	3.98E+04	1.67E+04	1.36E+04	3.09E+04	3.09E+04
Penthose phosphate pathway	6-phosphogluconate dehydrogenase	2.52E+04	7.65E+04	1.35E+05	8.17E+04	8.61E+04	4.97E+04
	ribulose-phosphate 3-epimerase	4.34E+04	3.57E+05	1.13E+05	9.03E+04	1.17E+05	7.36E+04
	ribose 5-phosphate isomerase A	8.36E+04	2.64E+04	1.52E+05	5.54E+04	4.82E+04	4.92E+04
	transaldolase	1.05E+05	5.31E+05	4.14E+05	2.18E+05	2.85E+05	2.49E+05
Glycolysis	glucose-6-phosphate isomerase	0.00E+00	0.00E+00	0.00E+00	6.04E+03	2.74E+04	9.08E+03
	phosphoglycerate mutase	9.90E+03	3.76E+04	2.25E+04	1.08E+04	8.74E+03	1.50E+04
	enolase	4.77E+04	1.21E+05	6.77E+04	1.67E+04	3.25E+04	9.68E+03
	pyruvate kinase	2.64E+04	0.00E+00	1.20E+04	0.00E+00	0.00E+00	0.00E+00
TCA cycle	citrate synthase	1.18E+04	7.46E+04	1.61E+04	3.40E+04	7.77E+04	2.98E+04
	isocitrate dehydrogenase	0.00E+00	4.70E+04	2.81E+04	3.60E+04	7.00E+04	3.33E+04
	succinyl-CoA synthetase alpha subunit	2.65E+04	4.40E+04	4.74E+04	1.52E+04	4.59E+04	1.82E+04
	succinyl-CoA synthetase beta subunit	0.00E+00	1.67E+04	4.20E+04	2.41E+04	4.30E+04	1.15E+04
	malate dehydrogenase	1.66E+05	2.51E+05	5.52E+04	2.50E+05	1.31E+05	4.42E+04

Supplementary Table 3.4 *Gene expression pattern for central nitrogen metabolism genes in Canada Basin surface waters revealed through fragment recruitment of metatranscriptomics against Met-BS01-1 and HTCC2181 genomes.*

Pathway	Genes	Met-BS01-1 RPKG			HTCC2181 RPKG		
		CB4	CB9	CB21	CB4	CB9	CB21
Urea cycle	argininosuccinate lyase	143.983946	364.271966	163.400154	19223.445	41190.21	14078.715
	acetylornithine aminotransferase	173.219295	0	216.235675	40155.975	47471.715	20386.155
	Ornithine carbamoyltransferase	223.765252	495.350114	558.667765	30871.755	36622.335	18005.835
	argininosuccinate synthase	125.764842	477.267869	323.508259	27284.505	35765.415	31971.855
Ammonium to carbamoyl phosphate	carbamoyl-phosphate synthase small subunit	0	631.700835	302.250447	18898.635	25623.645	17034.87
	carbamoyl-phosphate synthase large subunit	319.514966	525.431548	449.625803	23337.45	49839.06	12933.195
Nitrate assimilation	assimilatory nitrate reductiase catalytic subunit	0	0	0	N/A	N/A	N/A
	response regulator NasT	0	0	0	N/A	N/A	N/A
Urea transporters	urea transport system substrate-binding protein	0	0	0	N/A	N/A	N/A
	urea transport system permease protein	0	0	0	N/A	N/A	N/A
	urea transport system ATP-binding protein	0	0	0	N/A	N/A	N/A
Urease	urease subunit beta	0	403.422654	0	N/A	N/A	N/A
	urease subunit alpha	60.0777935	227.990618	54.5434455	N/A	N/A	N/A
	urease accessory protein (UREF)	0	0	0	N/A	N/A	N/A
	urease accessory protein (UREG)	0	0	101.910054	N/A	N/A	N/A
	urease accessory protein (UREJ)	157.254353	198.922714	392.61216	N/A	N/A	N/A
	urease accessory protein (URED)	180.551316	342.589067	163.918846	N/A	N/A	N/A
	urease accessory protein	56.6847286	71.7046903	128.657233	N/A	N/A	N/A
	urease subunit gamma	0	213.694208	153.36966	N/A	N/A	N/A
	urea transport system ATP-binding protein	0	187.679224	67.3492606	N/A	N/A	N/A
	urea transport system substrate-binding protein	0	104.014986	37.3261102	N/A	N/A	N/A
Ammonium transporter	ammonium transporter, Amt family	72.2970213	685.903946	65.6370488	51146.16	152201.295	55472.535
GS/GOGAT	glutamine synthetase	71.2901216	390.780605	291.252775	0	0	0
	glutamate synthase (NADPH/NADH) small chain	286.757975	317.398689	488.140788	57016.59	36372.3	77265.315
	glutamate synthase (NADPH/NADH) large chain	381.275966	241.152075	908.65095	33807.99	49321.26	32527.08

Bridging text

The third chapter focused on investigating the *Methylophilaceae* family in the continuously freshening surface waters of the Arctic Ocean. In addition to studying the diversity of OM43 lineages, we discovered a previously undescribed clade, BS01, that has also crossed the freshwater-marine barrier through a different evolutionary history when compared to OM43. BS01 is common to lower-salinity polar surface waters yet most closely related to freshwater sediment *Methylotenera* species. Given the ongoing freshening of the Arctic Ocean, the results of this chapter showed that the Arctic Ocean is inhabited by distinct bacterial groups and that at least one group, BS01, evolved via a freshwater to marine environmental transition. This led us to also investigate the changes in microbial community composition and the *Methylophilaceae* family in response to shifts in environmental conditions over time in the Arctic Ocean.

Chapter 4. A 2004-2012 time series of bacterial and archaeal community dynamics in a changing Arctic Ocean

4.1. Abstract

Climate change is profoundly impacting the Arctic, leading to a loss of multi-year sea ice and a warmer, fresher upper Arctic Ocean. The response of microbial communities to these climate-mediated changes is largely unknown. Here, we document the interannual variation in bacterial and archaeal communities across a 9-year time series of the Canada Basin that includes two historic sea ice minima (2007 and 2012). We report an overall loss of bacterial and archaeal community richness and significant shifts in community composition. The magnitude and period of most rapid change differed between the stratified water layers. The most pronounced changes in the upper water layers (surface mixed layer and upper Arctic water) occurred earlier in the time series, while changes in the lower layer (Pacific-origin water) occurred later. Shifts in taxonomic composition across time were subtle, but a decrease in Bacteroidota taxa and increase in Thaumarchaeota and Euryarchaeota taxa were the clearest signatures of change. This time series provides a rare glimpse into the potential influence of climate change on Arctic microbial communities; extension to the present day should contribute to deeper insights into the trajectory of Arctic marine ecosystems in response to warming and freshening.

4.2. Main Text

The Arctic Ocean is rapidly changing. Surface waters are warming due to climate change and sea ice retreat (Stroeve and Notz 2018), and freshening due to melting sea ice and river runoff (Proshutinsky *et al.* 2019; Brown, Holding and Carmack 2020). Warming and freshening are having profound effects on Arctic marine ecosystems (Frainer *et al.* 2017), including reduced nutrients (Zhuang *et al.* 2018) and shifts in primary production (Ardyna and Arrigo 2020).

Although changes in phytoplankton communities have been documented through time-series observations (Li *et al.* 2009; Oziel *et al.* 2020; Freyria, Joli and Lovejoy 2021), and a single study has reported on temporal dynamics of microbial diversity over the 2007 sea ice minimum (Comeau *et al.* 2011), there is a conspicuous lack of information on the change in bacterial and archaeal diversity over interannual time frames relevant to climate change.

In this study, we documented the variation in bacterial and archaeal communities in the Arctic Ocean (Canada Basin) across a 9-year time-series of summer waters that includes two historic sea ice minima (2007 and 2012) (**Figure 4.1a-b**). The dataset encompassed four stations (CB29, CB21, CB15, and CB9) along a latitudinal transect of the Canada Basin (72° N to 78° N) (**Supplementary Table 4.1**). Samples were analyzed from three water layers comprising the relatively fresh summer mixed layer (SML; 3-9 m), the Upper Arctic Waters (UAW; 16-78 m) underlying the SML, and the deeper Pacific-origin water (PW; 49-154 m) bounded by 31-33 PSU. The dataset captured the temporal variation in environmental conditions previously reported from a spatially more extensive time-series of the Canada Basin (Li *et al.* 2013). In particular, surface water freshening and increased stratification associated with sea ice melt was apparent (**Figure 4.1c-d**). Similarly, the previously observed shift towards smaller phytoplankton cell size and a later increase in bacterioplankton abundance was also captured in the dataset (**Figure 4.1e-f**).

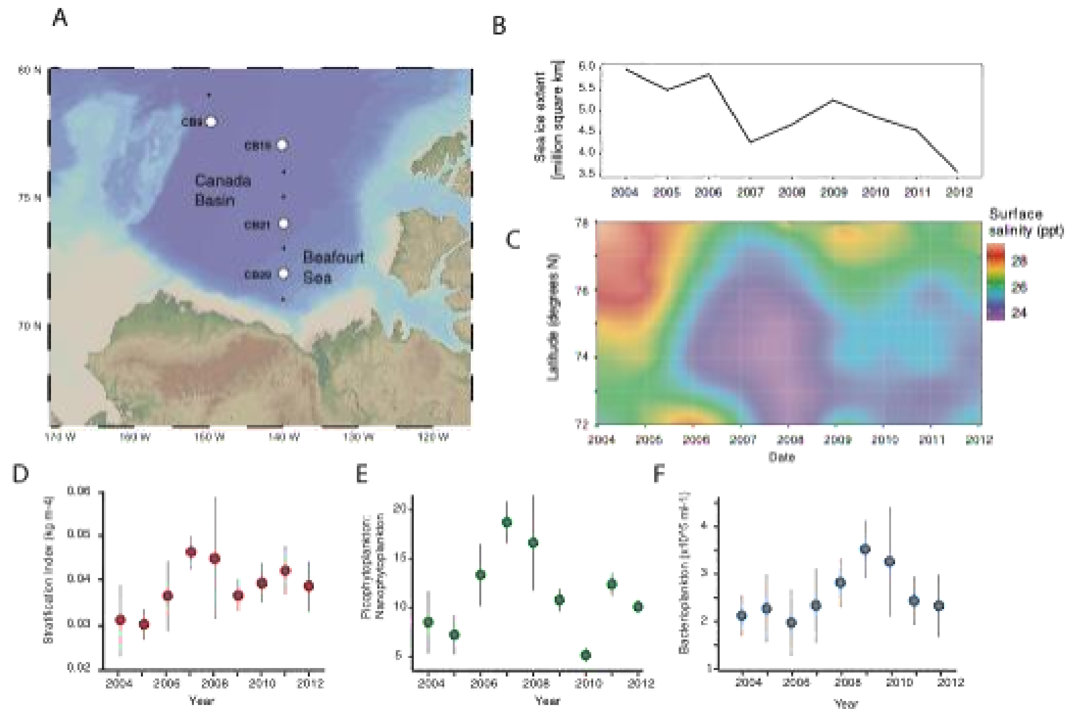


Figure 4.1 *The change in environmental conditions in the Canada Basin over the 2004-2012 time-series: (a) Map of the Arctic Ocean showing station locations (circles) included in this study; (b) Variability in summer sea ice extent in the Arctic Ocean, data is from the National Snow and Ice Data Center; (c) the latitudinal gradient in sea surface salinity over the time series; (d) stratification index, error bars indicate among station standard deviation; (e) ratio of picophytoplankton to nanophytoplankton; (f) bacterioplankton. In panels e-f, values are the average of all measurements >15 m in depth and salinity < 31 PSU, and error bars indicate among depth and among-station standard error.*

Bacterial and archaeal diversity was assessed using a universal filter-PCR approach (Kirchman *et al.* 2001; El-Swais *et al.* 2015) targeting the V4 region of the 16S rRNA gene. From 179 samples, we generated 2,985 amplicon sequence variants (ASVs). Chao-1 estimates of ASV richness were significantly correlated with sampling year, latitude, and depth for the UAW and PW layers, but not for the SML layer (**Supplementary Table 4.2**). We modelled the nonlinear influence of oceanographic variables, including sampling year, on ASV richness using a random forest approach. Random forest models explained 39-54% of the variation in ASV richness (**Supplementary Table 4.3**). Of all variables, year best explained changes in ASV richness in the

UAW (27.7%) and was also significant for the SML (13.3 %) and PW (4.3 %). Visualizing the relationship between year and modelled ASV richness revealed a temporal decline in richness across all water layers (**Figure 4.2a**). Remarkably, the period of most sudden decline varied between water layers. Single declines in ASV richness were observed in 2005-2007 for the SML and in 2010-12 for the UAW, while a two-step decline was observed in the PW.

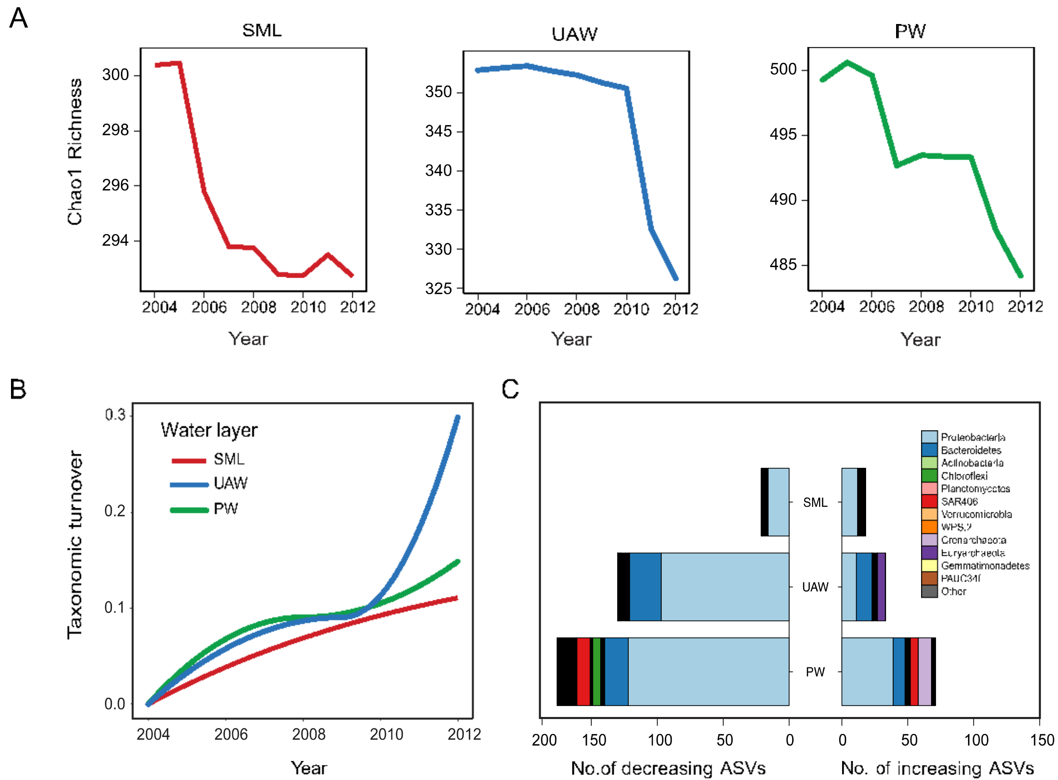


Figure 4.2 *The change in bacterial and archaeal communities in the Canada Basin over the 2004-2012 time series: (a) Partial dependency plots of changes in Chao1 richness predicted across the time series by the random forest approach; (b) The magnitude and rate of change in community composition (taxonomic turnover) across the time series as predicted by generalized dissimilarity modelling; (c) the number and taxonomic composition of bacterial and archaeal ASVs that significantly increased or decreased across the time series. SML, surface mixed layer; UAW, upper Arctic water; PW, Pacific water.*

We next investigated changes in community composition (i.e., turnover) using generalized dissimilarity models (GDMs). GDMs allowed us to model both the rate and magnitude of change in turnover across the time-series and in response to oceanographic variables. Sampling depth and

year were the two most important variables in a GDM constructed from the full dataset (**Supplementary Table 4.4**). When GDMs were constructed for each of the three water layers separately, sampling depth and year were still defined by sampling year was the top predictor of community turnover in the UAW (**Supplementary Table 4.4**). The period of most rapid turnover in the UAW was 2010-2012 (**Figure 4.2b**), coinciding with the previously observed decline in richness (**Figure 4.2a**). A similar overlap was observed for the PW, where community turnover mirrored the decline in richness over the time series, but was not significant.

The combined observation of an overall decrease in diversity and corresponding shift in community composition suggested a pattern of taxon loss across the time series. So, we next used Threshold Indicator Taxa Analysis to identify ASVs that significantly changed in each of the three water layers. As expected, we consistently identified more ASVs that decreased over time compared to ASVs that increased (**Supplementary Table 4.5, Figure 4.2c**). Generally, decreasing ASVs represented a higher taxonomic diversity compared to increasing ASVs, and the majority of decreasing ASVs belonged to Proteobacteria and Bacteroidetes. A significant number of increasing ASVs were from Archaea; Euryarchaeota ASVs increased in the UAW, while Crenarchaeota ASVs increased in the deeper PW.

An earlier study in the Arctic Ocean documented microbial community composition before and after the 2007 sea ice minimum (Comeau *et al.* 2011). Our study, which covers a larger oceanographic region and an extended time series, builds on this seminal study. We observed patterns that were consistent and some that were in contrast to the previous work. For example, we observed a loss in Bacteroidetes diversity as previously reported. Bacteroidetes are involved in the degradation of complex organic matter produced by phytoplankton (Fernández-Gómez *et al.* 2013; Krüger *et al.* 2019) and changes may indicate a response to shifting phytoplankton communities.

In contrast to the previous report of a decrease in ammonia-oxidizing Archaea (AOAs) within the Thaumarchaeota, we detected a signal for an increase in AOA diversity. Given the importance of AOAs in the nitrogen cycle (Santoro, Richter and Dupont 2019) further in-depth investigations on their temporal dynamics are warranted.

Further, we selected the *Methylophilaceae* family for a targeted analysis of change over time **(Supplementary Figure 4.1)** due to their unique evolutionary history and distribution. These Betaproteobacterial methylotrophs are comprised of a more broadly distributed clade, OM43, and more specialized groups, such as the recently discovered BS01, across marine habitats (Salcher *et al.* 2019; Ramachandran, McLatchie and Walsh 2021). The *Methylophilaceae* family is heavily implicated in carbon cycling, with methanol being a key carbon and energy source (Halsey, Carter and Giovannoni 2012; Ramachandran and Walsh 2015; Sargeant *et al.* 2016; Ramachandran, McLatchie and Walsh 2021). In this time-series, 15 *Methylophilaceae* ASVs were identified in all three biomes **(Supplementary Figure 4.2)**. 3 ASVs were present throughout the time series and the relative abundance was not significantly different between the years. On the other hand, six ASVs increased in relative abundance and one decreased after 2007, which could be attributed to the extreme change in environmental conditions after the sea-ice minima. This occurred mainly in the SML where rapid change is most visible **(Supplementary Figure 4.3)**. Due to the diversity of ASVs, which stayed constant or increased or decreased, within the *Methylophilaceae* family, they could be viewed as sentinels of change. Continued monitoring could lead to a deeper understanding of adaptation and evolution in rapidly changing environmental conditions over time. In fact, higher methanol concentrations in the surface waters of the Arctic Ocean have been reported potentially due to air to sea deposition of methanol in ice-free conditions, which could increase as the sea ice extent decreases (Sjostedt *et al.* 2012; Wohl *et al.* 2022). Additionally, methanol production is

linked to phytoplankton growth/decay, therefore this group can also be changing due to shifts in phytoplankton community composition (Mincer and Aicher 2016; Davie-Martin *et al.* 2020). More studies looking at diversity and the metabolic capabilities of the *Methylophilaceae* family will be needed to understand differences in abundance and distribution over time linked to changing environmental conditions.

Our study provides evidence that significant shifts in bacterial and archaeal community composition occurred between 2004-2012, a period that coincided with two historic sea-ice minima. Although changes in community diversity and composition were associated with nutrient concentrations and phytoplankton composition (**Supplementary Table 4.3, Supplementary Table 4.4**), the links should be interpreted cautiously; significant vertical gradients in these variables were present in single water masses, challenging our ability to disentangle depth versus temporal structure driving these relationships. Instead, we conservatively documented the overall interannual changes in the community, rather than the specific environmental drivers of change. Overall, this relatively short time series provides a rare glimpse into the influence of climate change on Arctic Ocean microbial communities. Given the inherent interannual variability in oceanographic conditions, extension of the time series to the present day and into the future will provide deeper insights into the trajectory of Arctic ecosystems and the consequences of a warmer, fresher Arctic Ocean (Vincent 2010; Lannuzel *et al.* 2020).

4.3. Materials and Methods

4.3.1. Environmental data and sample collection

Samples were collected between 2004 and 2012 during the Joint Ocean Ice Study (JOIS) in the Canada Basin on the CCGS Louis S. St-Laurent. Sampling was conducted between July and August, except for 2009 and 2010 when sampling was conducted through September to October.

The collection and processing of all environmental and oceanographic data used in this study are described in detail in McLaughlin *et al.* (2012) (McLaughlin *et al.* 2012) and Li *et al.*, (2013) (Li *et al.* 2013). Seawater samples (2 mL) were collected from 5-7 depths across the upper 200 m of the Canada Basin, fixed in 1% paraformaldehyde for 15 minutes at room temperature, flash frozen in liquid nitrogen, and then stored at -80°C for later analysis.

4.3.2. Filter PCR of 16S rRNA gene amplicons

For each sample, 500 µl of fixed seawater was filtered through a 25 mm polycarbonate filter with a 0.2 µm pore size (GE) using vacuum filtration (100 kPa). 5 ml of sterilized MilliQ ultra-pure water was passed through the filters three times to remove the residual paraformaldehyde preservative. Filters were sectioned and a sixteenth of each filter was transferred to a 200 ml PCR tube. Direct amplification of the V4 region on the 16S rRNA gene was conducted on filters similar to the approach used in El-Swais *et al.* (2015) (El-Swais *et al.* 2015). We used the 515F primer (5'-GTG YCA GCM GCC GCG GTA A-3') from Parada *et al*. (2016) (Parada, Needham and Fuhrman 2016) and the 806R primer (5'-GGA CTA CNV GGG TWT CTA AT-3') from Apprill *et al.* (2015) (Apprill *et al.* 2015). In combination this primer set enhances the detection of SAR11 and Crenarachaeota/Thaumarchaeota. PCR reactions (50 µL) contained 0.5 µM each primer, 1 X Phire Reaction Buffer, 0.2 mM deoxynucleotide triphosphates and 1 µL (1 unit) of Phire Hot Start II DNA Polymerase (Thermo Scientific). Cycling conditions were as follows: initial 3-minute denaturing step, 30 cycles of 5 seconds at 98°C, 5 seconds at 50°C and 10 seconds at 72°C, and a final elongation step of 1 minute at 72°C. The PCR amplicons, with the CS1 (5'-ACACTGACGACATGGTTCTACA-3') and CS2 (5'-TACGGTAGCAGAGACTTGGTCT-3') adapters attached, were purified at Genome Quebec

(McGill University) and subsequently sequenced using the Illumina MiSeq 250bp paired-end platform.

4.3.3. Amplicon data processing

Raw reads were processed using the DADA2 package in R [6]. Quality profiles were created for all reads. Reads were then trimmed, merged and dereplicated, followed by ASV identification and chimera removal. Before further analysis we removed ASVs found in less than three samples, which likely result from amplification errors. Taxonomy of ASVs was using the GreenGenes 16S rRNA gene database (DeSantis *et al.* 2006). Further analysis of the ASV table was performed using the phyloseq (McMurdie and Holmes 2013) package in R version 4.2.2. Graphics were created using the ggplot2 package (Wickham 2016).

4.3.4. Statistical analyses

Chao1 richness was calculated on a rarefied dataset (18,992 sequences; the size of the smallest sample) using phyloseq (McMurdie and Holmes 2013). Analysis of variance (anova) of alpha diversity for each water mass was performed using the explanatory variables latitude, year, and depth using R. We used a RF technique from the R party package (Hothorn, Hornik and Zeileis 2006) based on conditional inference regression trees developed by Ryo and Rillig (2017) (Ryo and Rillig 2017) to model the response of Chao1 richness to environmental variables. 5000 regression trees were used to obtain a stable prediction. A measure of importance was calculated for each predictor variable by cross validating each tree with data not used when the tree was constructed, referred to as the out-of-bag (OOB) data. We generated partial dependency plots (PDPs) illustrating the relationships between Chao1 richness and predictor variable from the RF using the R pdp package (Greenwell, Brandon 2017).

We modeled nonlinear relationships between community turnover (Bray Curtis dissimilarity based on rarified count data) and environmental variables using GDMs as implemented in the gdm R package (Fitzpatrick *et al.* 2021). Explanatory environmental parameters considered in each model were sampling year, longitude, latitude, bacterial counts per mL, phytoplankton counts per mL, nanophytoplankton counts per mL, picophytoplankton counts per mL, temperature, salinity, nitrate, silicate, and phosphate concentrations, as well as the ration of pico- to nanophytoplankton. GDMs were constructed for the full dataset, as well as for subsets corresponding to each water mass. Significant variables were backwards selected with 100 permutations per step. We conducted taxon indicator threshold analysis for each water mass as implemented in the TITAN2 R package (Baker, King and Kahle 2019) to determine thresholds of turnover over a temporal based on relative ASV abundances. For each water mass, highly prevalent (present in all samples of the water mass) ASVs were removed before analysis.

All samples were clustered using hierarchical clustering (hclust function in stats package for R with a Bray-Curtis dissimilarity metric). The relative abundance of taxa that were 3.5% or more of the community composition were plotted next to the clustered samples. Phylogenetic analysis of the *Methylophilaceae* group across various aquatic regions and depths using the 16S rRNA sequence. The tree was inferred using the Kimura 2-parameter model (500 bootstraps) and a GTR gamma distribution (four categories) with invariants sites model of evolution and the nearest-neighbor interchange heuristic search method. Sequences from the current study are highlighted in blue and Arctic Ocean sequences from a previous study conducted by the lab (Ramachandran, McLatchie and Walsh 2021) are highlighted in red. Only bootstrap values of ≥60% were included in the tree. A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic samples was performed based on the relative abundance of

Methylophilaceae ASVs to determine the distribution of samples. The available predictor variables were used to determine which environmental variables are driving the distribution of samples. The relative abundance of *Methylophilaeae* ASVs across the three biomes, nine years, and four sites was graphed to determine patterns of change linked to year.

4.3. Acknowledgments

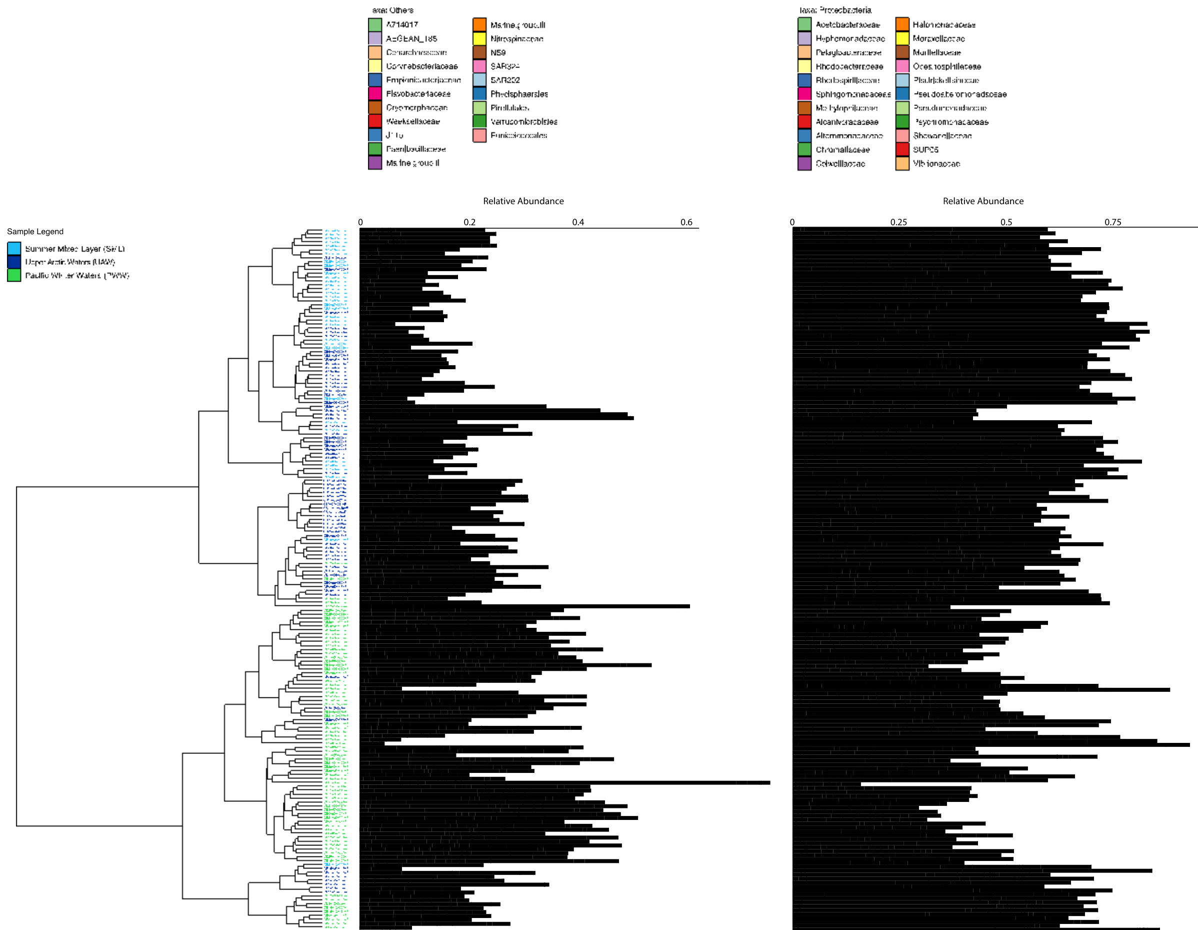
The samples and environmental data were collected aboard the CCGS Louis S. St-Laurent in collaboration with researchers from Fisheries and Oceans Canada at the Institute of Ocean Sciences and Woods Hole Oceanographic Institution's Beaufort Gyre Exploration Program and are available at http://www.whoi.edu/beaufortgyre. We would like to thank both the Captain and crew of the CCGS Louis S. St-Laurent and the scientific teams aboard. We thank Sara Palestini (Concordia University) for providing the surface salinity figure (Figure 1c). Funding from the Canadian Natural Science and Engineering Research Council (NSERC) Discovery program is acknowledged.

4.4. Competing Interests

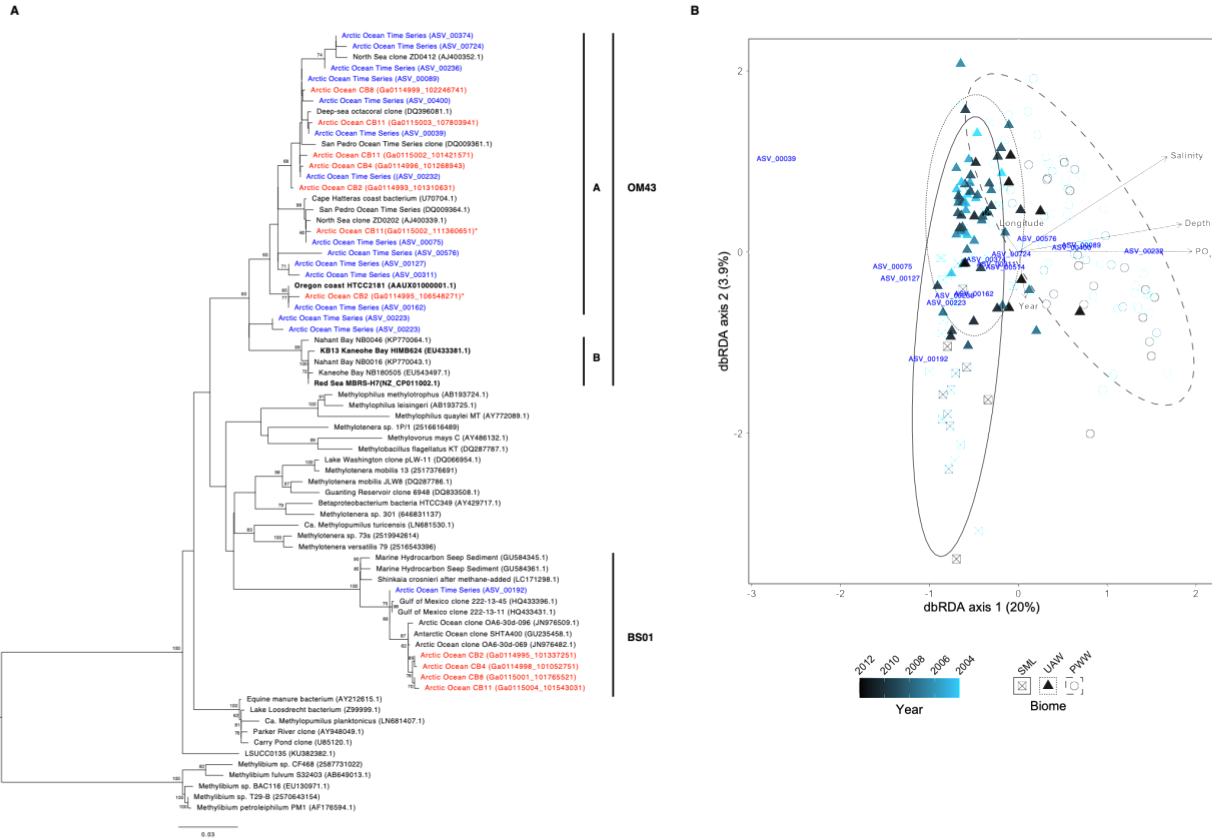
The authors declare there is no competing financial interest in relation to the described work.

4.5. Supplementary Figures and Tables

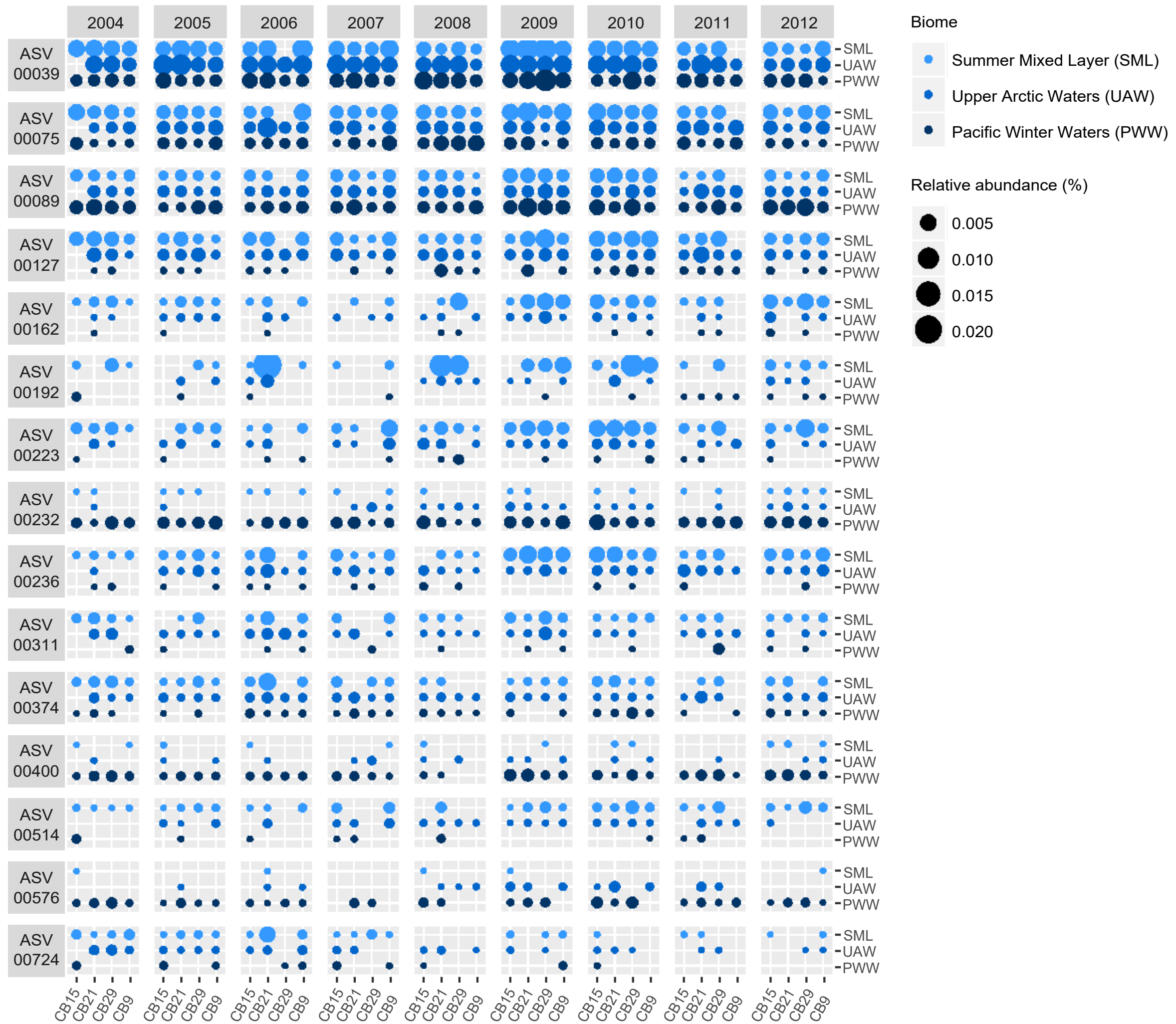
Supplementary Figure 4.1 *A hierarchical clustering analysis of samples based on the ASV distributions using the hclust function in stats package for R with a Bray-Curtis dissimilarity metric. The relative abundance of taxa that were 3.5% or more of the community composition were plotted next to the clustered samples.*



Supplementary Figure 4.2 *(A) Phylogenetic analysis of 16S rRNA genes from Methylophilaceae from Canada Basin time series, metagenomes and a diversity of aquatic ecosystems. (B) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic samples based on the relative abundance of Methylophilaceae ASVs.*



Supplementary Figure 4.3 *The relative abundance of Methylophilaceae ASVs across the three biomes (SML, UAW, PW), nine years (2004-2012), and four sites (CB9, CB15, CB21, CB29).*



Supplementary Table 4.1 *Sample information and environmental data. Values in red were missing from the original dataset and were interpolated as average values from the same depths or water masses.*

Sample	Site	Year	Depth	Sample_ID	Longitude	Latitude
CB15-04-006-	CB15	2004	6	532	-139.92	76.98
CB15-04-027-	CB15	2004	27	531	-139.92	76.98
CB15-04-045-	CB15	2004	45	530	-139.92	76.98
CB15-04-064-	CB15	2004	64	529	-139.92	76.98
CB15-04-096-	CB15	2004	96	527	-139.92	76.98
CB15-05-007-	CB15	2005	7	651	-139.99	76.98
CB15-05-026-	CB15	2005	26	650	-139.99	76.98
CB15-05-057-	CB15	2005	57	649	-139.99	76.98
CB15-05-080-	CB15	2005	80	647	-139.99	76.98
CB15-05-117-	CB15	2005	117	645	-139.99	76.98
CB15-06-007-	CB15	2006	7	1025	-139.88	77.01
CB15-06-023-	CB15	2006	23	1024	-139.88	77.01
CB15-06-055-	CB15	2006	55	1023	-139.88	77.01
CB15-06-097-	CB15	2006	97	1022	-139.88	77.01
CB15-07-007-	CB15	2007	7	1411	-140.25	77.04
CB15-07-022-	CB15	2007	22	1410	-140.25	77.04
CB15-07-046-	CB15	2007	46	1409	-140.25	77.04
CB15-07-064-	CB15	2007	64	1408	-140.25	77.04
CB15-07-106-	CB15	2007	106	1407	-140.25	77.04
CB15-08-006-	CB15	2008	6	1196	-140.01	76.99
CB15-08-022-	CB15	2008	22	1195	-140.01	76.99
CB15-08-053-	CB15	2008	53	1194	-140.01	76.99
CB15-08-084-	CB15	2008	84	1193	-140.01	76.99
CB15-08-113-	CB15	2008	113	1192	-140.01	76.99
CB15-09-007-	CB15	2009	7	906	-139.88	77.02
CB15-09-026-	CB15	2009	26	905	-139.88	77.02
CB15-09-036-	CB15	2009	36	904	-139.88	77.02
CB15-09-051-	CB15	2009	51	903	-139.88	77.02
CB15-09-071-	CB15	2009	71	902	-139.88	77.02
CB15-09-089-	CB15	2009	89	901	-139.88	77.02
CB15-10-007-	CB15	2010	7	1141	-139.73	76.99
CB15-10-028-	CB15	2010	28	1140	-139.73	76.99
CB15-10-048-	CB15	2010	48	1139	-139.73	76.99
CB15-10-086-	CB15	2010	86	1138	-139.73	76.99
CB15-10-104-	CB15	2010	104	1137	-139.73	76.99
CB15-11-007-	CB15	2011	7	613	-139.98	77.08
CB15-11-027-	CB15	2011	27	612	-139.98	77.08
CB15-11-060-	CB15	2011	60	611	-139.98	77.08
CB15-11-084-	CB15	2011	84	608	-139.98	77.08
CB15-11-111-	CB15	2011	111	607	-139.98	77.08

Supplementary Table 4.2 *Sum of squares based on anova of Chao1 richness as a function of sampling year, latitude, and depth for each water mass. Bold values indicate significate covariates.*

	Year	Latitude	Depth
SML	5665	6067	966
UAW	**46653**	**32967**	**77900**
PW	**49398**	**38160**	**64372**

Supplementary Table 4.3 *RF results indicating relative importance (as % of explained variation) of the explanatory variables on bacterioplankton diversity.*

	Chao 1 richness			
	Full	SML	UAW	PW
R^2 fit[1]	0.77	0.34	0.54	0.42
OOB[2]	0.66	0.01	0.20	0.19
Year	1.3	13.8	27.7	4.3
Depth	21.5	0.0	9.8	0.0
Salinity	11.3	20.3	3.8	7.6
Temperature	1.1	0.0	0.18	8.6
Nitrate	14.7	0.0	14.8	14.3
Phosphate	27.2	54.0	6.7	22.4
Silicate	12.1	0.0	7.2	19.4
Latitude	0.7	10.6	2.4	3.0
Longitude	0.1	1.2	1.7	0
Phytoplankton	3.3	0.0	2.2	9.9
Nanophytoplankton	1.9	0.0	22.8	1.2
Picophytoplankton	3.8	0.0	0.3	8.4
Bacterioplankton	1.1	0.0	0.3	0.8

[1]total amount of variation explained
[2]model error rate in out of bag test

Supplementary Table 4.4 *GDM results including the model deviance, deviance explained, and predictor importance. Bold values indicate significant predictors.*

	Full	SML	UAW	PW
Model deviance	607.917	10.014	55.62	113.954
Percent deviance explained	70.413	56.475	55.32	33.053
Year	**1.83**	**4.46**	**23.35**	3.83
Depth	**3.59**	**14.49**	**12.35**	4.4
Salinity	0.022	0.072	0.07	0.13
Temperature	0.052	2.41	1.75	0
Nitrate	0.21	4.13	0.38	0
Phosphate	0.21	4.95	0.67	0
Silicate	0.0045	1.33	1.86	0.45
Latitude	**0.57**	0.67	1.48	**3.42**
Phytoplankton	0.017	0	0	0
Nanophytoplankton	0.022	0.066	1.01	0.0.82
Picophytoplankton	0.000002	0.27	0.4	0.28
Ratio of pico to nano	**1.17**	0.16	0.48	1.76
Bacterioplankton	**1.57**	4.67	1.07	**2.91**

Supplementary Table 4.5 *Summary of threshold indicator analysis with year as the environmental gradient variable.*

Water mass	No. of significantly changing taxa	No. of increasing taxa	No. of decreasing taxa
SML	39	18	21
UAW	163	33	130
PW	314	71	176

Chapter 5. Conclusions, discussions, and future directions

5.1. *Methylophilaceae*: Sentinels of Change?

The *Methylophilaceae* family has a unique evolutionary history and distribution, with at least two lineages crossing the freshwater-marine barrier independently through congruent evolution, which makes it an ideal candidate to study change spatially and temporally. This group is ubiquitous in marine, freshwater, and soil environments and is involved in global carbon cycling. That's why it is important to understand their ecological functions, genetic diversity, and genome evolution in response to changing environmental conditions. Furthermore, investigating the *Methylophilaceae* group can provide insights into microbial adaptation and evolutionary mechanisms that are relevant to global ecosystem functioning.

We first studied the diversity and distribution of the *Methylophilaceae* family spatially across a salinity gradient in the Saint-Lawrence Estuary. We were able to detect methylotrophs from the OM43/LD28 clade across the salinity gradient, from freshwater to brackish-marine waters. These distinct populations were related to specific salinity concentrations, which suggests different ecological roles in response to environmental gradients. This led us to also investigate the adaptation and evolution of *Methylophilaceae* in response to changing environmental conditions over time. We conducted our study in an ecosystem that is undergoing rapid and extreme change, the Arctic Ocean. Increased sea ice melt and almost complete loss of multi-year sea ice in the past decades has led to the unprecedented freshening of surface waters over time. We found evidence for endemic Arctic Ocean taxa as well as a novel clade, BS01. Analysis of the Met-BS01-1 MAG showed that it was adapted to the unique surface waters of the Arctic Ocean (i.e. decreased salinity and nitrate, increased urea). We provided strong evidence that BS01 represents a second lineage of marine *Methylophilaceae* that arose independently of marine OM43.

We also found that there was an increase in specific *Methylophilaceae* taxa in 2007 and 2012, which corresponds to years experiencing the highest amount of freshening in our time series.

Ultimately, we can attribute the wide distribution and success of the *Methylophilaceae* family in different habitats to the fact this group is comprised of multiple evolutionarily distinct populations adapted to distinct environmental conditions. The work in this thesis demonstrate that the Methylophilaceae are an ideal candidate to study change spatially and temporally.

Future studies to better understand the adaptive capabilities of the *Methylophilaceae* family include isolation of pure cultures using dilution-to-extinction techniques followed by experimental evolution studies using a gradient of environmental variables (i.e. salinity, nutrients, temperature). Co-occurrence studies using LD28/OM43 and BS01 cultures will also be necessary to understand the dynamics and niche differentiation of these clades.

5.2. Importance of Ecotypes in Understanding Change

Bacterial taxa are comprised of multiple evolutionarily distinct populations that are adapted to distinct environmental conditions, called ecotypes. In this thesis, we endeavoured to further understand how a changing environment affects ecotype diversity and distribution. This added another piece of the puzzle in understanding the larger implications rapid environmental change has on biodiversity and ecosystem functioning. By better understanding the ecotypic diversity of bacterial communities, we gained insight into how these populations may respond to change.

The studies in this thesis expanded our understanding of the various groups in the *Methylophilaceae* family that are adapted to specific environmental conditions and how they can be used as a model to study adaptation to new habitats. The discovery of BS01 in the same habitat

as OM43 revealed that ecotypes can co-exist in the same environment while occupying specialized niches.

We highlight the importance of studying ecotypes in understanding how bacterial populations adapt to changing environmental conditions. By studying ecotypes, we can gain a better understanding of the mechanisms that allow bacterial populations to adapt and persist in changing environments. Overall, studying bacterial evolution and adaptation in the Arctic Ocean is crucial for understanding the impacts of climate change on this vital and unique ecosystem. This understanding can then be applied to predicting how bacterial communities may respond to future environmental changes, and how these changes may affect ecosystem functioning.

Future experiments could include culturing OM43 and BS01 from the Arctic Ocean and testing their co-occurrence patterns and specific ecosystem services. Cultures could also be used for experimental evolution studies to gain a better understanding of how each environmental variable (i.e. salinity, nutrient concentration, temperature) contributes to specific adaptations on the genomic level.

5.3. Microbial Evolution Associated to a Changing Arctic Ocean and the Potential Global Implications

The rapidly changing environmental conditions in the Arctic Ocean have led to many questions about whether there will be an increase in adaptation, evolution, and/or migration of microbial organisms. One outcome could be that bacterial species endemic to the Arctic will adapt and subsequently evolve to the changing environment. Another possible outcome could be that marine bacteria from lower latitudes migrate to the Arctic Ocean in response to, for example, higher temperatures. Alternatively, freshwater bacteria could colonize the fresher Arctic Ocean waters, moving with the influx of freshwater from the rivers emptying into the Arctic Ocean and/or increased sea ice melt and glacier runoff. The research conducted in this thesis suggests that the

movement of lower latitude microbes to the Arctic Ocean is not the pathway driving change in community composition. Using the *Methylophilaceae* family as a model, we provided evidence of endemic species adapting to rapidly changing environments through processes such as genome streamlining, which agrees with previous studies. Further research should be conducted to fully understand the mechanisms and drivers of bacterial evolution and adaptation in the Arctic Ocean, as well as the potential implications for ecosystem functioning and biogeochemical cycling in this rapidly changing environment. This understanding could be critical for predicting and mitigating the impacts of ongoing climate change on the Arctic marine ecosystem.

Changes in the Arctic Ocean are unique. With the Arctic Ocean warming and freshening faster than temperate and tropical oceans, do we observe similar changes reflected in lower latitude oceans? Studies from the Bermuda Atlantic Time Series (BATS), San Pedro Ocean Time Series (SPOTS), Hawaii Ocean Time Series (HOT), and Western Channel that take place over a similar time frame provided interesting comparisons. Unlike our time-series study where community diversity and richness decreased over time, studies conducted at the lower latitude stations showed that the average prokaryotic community was steady over the same time period, showing higher resilience than Arctic Ocean communities (Gilbert *et al.* 2012; Giovannoni and Vergin 2012; Yeh and Fuhrman 2022). This could be attributed to the fact that the Arctic Ocean is experiencing rapid environmental change three to four-fold faster that the global rate. In fact, the Arctic Ocean could even be considered as a potential example of what may occur in these lower latitude oceans as surface temperatures and disruptions to nutrient cycling increases due to climate change.

The study of bacterial evolution and adaptation in the Arctic Ocean is not only relevant for understanding local ecosystem dynamics, but also for gaining insight into the broader impacts of environmental change on global biogeochemical cycling and oceanic processes. We should keep

exploring the complexities of microbial community dynamics in response to changing environmental conditions through continued monitoring. In addition, studies should also investigate the potential for microbial adaptations to affect higher trophic levels in the Arctic food web and the consequences for Arctic biodiversity. Overall, continued research into bacterial evolution and adaptation in the Arctic Ocean will lead to a more comprehensive picture on how to mitigate the impacts of rapid change on this fragile ecosystem.

5.4. Future Directions: Linking Microbial Research in Whole Ecosystem Studies

As environmental change continues to occur at an unprecedented and rapid rate, future studies that have a more holistic view of Arctic ecosystems are needed. Including microbes in multidisciplinary ecosystem studies will allow us to understand whether changes to microbial community composition could have a cascading effect on higher trophic level organisms. After all, we know that in terms of total biomass, most life in the Arctic Ocean is microbial, and that microbes are essential to the health and stability of the ecosystem. Furthermore, considering the interconnectedness and interdependencies within ecosystems, studying microbial communities in isolation may provide an incomplete understanding of their role and function. Integrating the study of microbial communities into ecosystem-focused research will help us better predict and manage future impacts on the Arctic ecosystem.

In conclusion, the Arctic Ocean's microbial communities can be seen as both sentinels of global change. Their responses to climate warming can provide important insights into the overall health and functioning of Arctic marine ecosystems as well as the potential of these systems to influence global biogeochemical cycles. By examining the interactions and dynamics between microbial communities and other biotic and abiotic components of the ecosystem, we can gain a

more comprehensive understanding of the complex processes and feedback mechanisms that drive Arctic ecosystem dynamics in the face of climate change.

References

Alonso-Saez L, Waller AS, Mende DR *et al.* Role for urea in nitrification by polar marine Archaea. *Proceedings of the National Academy of Sciences* 2012;**109**:17989–94.

Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990, DOI: 10.1016/S0022-2836(05)80360-2.

Apprill A, McNally S, Parsons R *et al.* Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology* 2015;**75**:129–37.

Ardyna M, Arrigo KR. Phytoplankton dynamics in a changing Arctic Ocean. *Nat Clim Chang* 2020;**10**:892–903.

Ardyna M, Mundy CJ, Mills MM *et al.* Environmental drivers of under-ice phytoplankton bloom dynamics in the Arctic Ocean. Deming JW (ed.). *Elementa: Science of the Anthropocene* 2020;**8**:30.

Arrigo KR, Perovich DK, Pickart RS *et al.* Massive phytoplankton blooms under arctic sea ice. *Science (1979)* 2012;**336**:1408.

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

Baker M, King R, Kahle D. An Introduction to TITAN2. 2019;**10**:1–26.

Biller SJ, Berube PM, Lindell D *et al.* Prochlorococcus: the structure and function of collective diversity. *Nat Rev Microbiol* 2014;**13**:13–27.

Bižić-Ionescu M, Ionescu D. Crossing the Freshwater/Saline Barrier: A Phylogenetic Analysis of Bacteria Inhabiting Both Freshwater and Marine Ecosystems. In: Glibert PM, Kana TM (eds.). *Aquatic Microbial Ecology and Biogeochemistry: A Dual Perspective*. Cham: Springer International Publishing, 2016, 35–44.

Boeuf D, Cottrell MT, Kirchman DL *et al.* Summer community structure of aerobic anoxygenic phototrophic bacteria in the western Arctic Ocean. *FEMS Microbiol Ecol* 2013;**85**:417–32.

Brown KA, Holding JM, Carmack EC. Understanding Regional and Seasonal Variability Is Key to Gaining a Pan-Arctic Perspective on Arctic Ocean Freshening. *Front Mar Sci* 2020;**7**:1–25.

Brown NJ, Nilsson J, Pemberton P. Arctic Ocean Freshwater Dynamics: Transient Response to Increasing River Runoff and Precipitation. *J Geophys Res Oceans* 2019;**124**:5205–19.

Cabello-Yeves PJ, Rodriguez-Valera F. Marine-freshwater prokaryotic transitions require extensive changes in the predicted proteome. *Microbiome* 2019;**7**:1–12.

Cabello-Yeves PJ, Zemskaya TI, Rosselli R *et al.* Genomes of Novel Microbial Lineages Assembled from the Sub-Ice Waters of Lake Baikal. *Appl Environ Microbiol* 2018;**84**:1–21.

CAFF. *State of the Arctic Marine Biodiversity Report.* Akureyri, Iceland, 2017.

Callahan BJ, McMurdie PJ, Rosen MJ *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016, DOI: 10.1038/nmeth.3869.

Carmack E, McLaughlin F. Towards recognition of physical and geochemical change in Subarctic and Arctic Seas. *Prog Oceanogr* 2011;**90**:90–104.

Carmack E, Proshutinsky A, Krishfield RA *et al.* The Rapid Response of the Canada Basin to Climate Forcing From Bellwether to Alarm Bells. *Oceanography* 2011;**24**:146–59.

Carmack EC, Yamamoto-Kawai M, Haine TWN *et al.* Freshwater and its role in the Arctic Marine System: Sources, disposition, storage, export, and physical and biogeochemical consequences in the Arctic and global oceans. *J Geophys Res Biogeosci* 2016;**121**:675–717.

Chistoserdova L. Methylotrophy in a Lake: from Metagenomics to Single-Organism Physiology. *Appl Environ Microbiol* 2011a;**77**:4705–11.
Chistoserdova L. Modularity of methylotrophy, revisited. *Environ Microbiol* 2011b;**13**:2603–22.
Chistoserdova L. Lanthanides: New life metals? *World J Microbiol Biotechnol* 2016;**32**:138.
Chistoserdova L, Kalyuzhnaya MG, Lidstrom ME. The Expanding World of Methylotrophic Metabolism. *Biological Bulletin* 2009;**223**:477–99.
Chistoserdova L, Lapidus A, Han C *et al.* Genome of Methylobacillus flagellatus, molecular basis for obligate methylotrophy, and polyphyletic origin of methylotrophy. *J Bacteriol* 2007;**189**:4020–7.
Chistoserdova L, Lidstrom ME. Aerobic Methylotrophic Prokaryotes. In: Rosenberg E, DeLong EF, Lory S, et al. (eds.). *The Prokaryotes*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, 267–85.
Chu F, Beck DAC, Lidstrom ME. MxaY regulates the lanthanide-mediated methanol dehydrogenase switch in *Methylomicrobium buryatense*. *PeerJ* 2016;**4**:e2435.
Chu F, Lidstrom ME. XoxF acts as the predominant methanol dehydrogenase in the type I methanotroph Methylomicrobium buryatense. *J Bacteriol* 2016;**198**:1317–25.
Colatriano D, Ramachandran A, Yergeau E *et al.* Metaproteomics of aquatic microbial communities in a deep and stratified estuary. *Proteomics* 2015:n/a-n/a.
Colatriano D, Tran PQ, Guéguen C *et al.* Genomic evidence for the degradation of terrestrial organic matter by pelagic Arctic Ocean Chloroflexi bacteria. *Commun Biol* 2018;**1**:90.
Collins RE, Rocap G, Deming JW. Persistence of bacterial and archaeal communities in sea ice through an Arctic winter. *Environ Microbiol* 2010;**12**:1828–41.
Colwell RK, Futuyma D. On the Measurement of Niche Breadth and Overlap. *Ecology* 1971;**52**:567–76.
Comeau AM, Li WKW, Tremblay J-É *et al.* Arctic Ocean Microbial Community Structure before and after the 2007 Record Sea Ice Minimum. *PLoS One* 2011;**6**:e27492.
Connelly TL, Baer SE, Cooper JT *et al.* Urea uptake and carbon fixation by marine pelagic bacteria and archaea during the Arctic summer and winter seasons. *Appl Environ Microbiol* 2014;**80**:6013–22.
Connon S a., Giovannoni SJ. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl Environ Microbiol* 2002;**68**:3878–85.
Conover R, Gustavson K. Sources of urea in arctic seas:zooplankton metabolism. *Mar Ecol Prog Ser* 1999;**179**:41–54.
Conover R, Mumm N, Bruecker P *et al.* Sources of urea in arctic seas:seasonal fast ice? *Mar Ecol Prog Ser* 1999;**179**:55–69.
Cui TT, Dawson TJ, McLatchie S *et al.* Modelling Free-Living and Particle-Associated Bacterial Assemblages across the Deep and Hypoxic Lower St. Lawrence Estuary. *mSphere* 2020;**5**:1–15.
Davie-Martin CL, Giovannoni SJ, Behrenfeld MJ *et al.* Seasonal and Spatial Variability in the Biogenic Production and Consumption of Volatile Organic Compounds (VOCs) by Marine Plankton in the North Atlantic Ocean. *Front Mar Sci* 2020;**7**:1–15.
Delmont TO, Eren EM. Linking pangenomes and metagenomes: The Prochlorococcus metapangenome. *PeerJ* 2018;**2018**:1–23.
Delmont TO, Kiefl E, Kilinc O *et al.* Single-amino acid variants reveal evolutionary processes that shape the biogeography of a global SAR11 subclade. *Elife* 2019;**8**:1–26.

DeSantis TZ, Hugenholtz P, Larsen N *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;**72**:5069–72.
Dixon JL, Beale R, Nightingale PD. Microbial methanol uptake in northeast Atlantic waters. *ISME J* 2011;**5**:704–16.
Dixon JL, Nightingale PD. Fine-scale variability in methanol uptake and oxidation: From the microlayer to 1000 m. *Biogeosciences* 2012;**9**:2961–72.
Doney SC. The Growing Human Footprint on Coastal and Open-Ocean Biogeochemistry. *Science (1979)* 2010;**328**:1512–6.
Ducklow H. Microbial services: challenges for microbial ecologists in a changing world. *Aquatic Microbial Ecology* 2008;**53**:13–9.
El-Swais H, Dunn K a., Bielawski JP *et al.* Seasonal assemblages and short-lived blooms in coastal north-west Atlantic Ocean bacterioplankton. *Environ Microbiol* 2015;**17**:3642–61.
Eren AM, Esen OC, Quince C *et al.* Anvi'o: An advanced analysis and visualization platformfor 'omics data. *PeerJ* 2015;**2015**:1–29.
Fernández-Gómez B, Richter M, Schüler M *et al.* Ecology of marine bacteroidetes: A comparative genomics approach. *ISME Journal* 2013;**7**:1026–37.
Fitzpatrick MC, Mokany K, Manion G *et al.* gdm: Generalized Dissimilarity Modeling. *R package version 1422* 2021.
Fortier M, Fortier L, Michel C *et al.* Climatic and biological forcing of the vertical flux of biogenic particles under seasonal Arctic sea ice. *Mar Ecol Prog Ser* 2002;**225**:1–16.
Frainer A, Primicerio R, Kortsch S *et al.* Climate-driven changes in functional biogeography of Arctic marine fish communities. *Proc Natl Acad Sci U S A* 2017;**114**:12202–7.
Freyria NJ, Joli N, Lovejoy C. A decadal perspective on north water microbial eukaryotes as Arctic Ocean sentinels. *Sci Rep* 2021;**11**:1–14.
Fu L, Niu B, Zhu Z *et al.* CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;**28**:3150–2.
Fu W, Moore JK, Primeau FW *et al.* A Growing Freshwater Lens in the Arctic Ocean With Sustained Climate Warming Disrupts Marine Ecosystem Function. *J Geophys Res Biogeosci* 2020;**125**:1–14.
Galand PE, Lovejoy C, Pouliot J *et al.* Microbial community diversity and heterotrophic production in a coastal Arctic ecosystem: A stamukhi lake and its source waters. *Limnol Oceanogr* 2008;**53**:813–23.
Galbraith PS, Chassé J, Dumas J *et al.* Physical oceanographic conditions in the Gulf of St. Lawrence during 2021. *DFO Canadian Science Advisory Secretariat Research Document*. Vol 2022/034. 2022, iv-83p.
Georges AA, El-Swais H, Craig SE *et al.* Metaproteomic analysis of a winter to spring succession in coastal northwest Atlantic Ocean microbial plankton. *ISME J* 2014;**8**:1301–13.
Ghiglione J-F, Galand PE, Pommier T *et al.* Pole-to-pole biogeography of surface and deep marine bacterial communities. *Proceedings of the National Academy of Sciences* 2012;**109**:17633–8.
Gifford SM, Becker JW, Sosa O a *et al.* Quantitative Transcriptomics Reveals the Growth- and Nutrient- Dependent Response of a Streamlined Marine Methylotroph to Methanol and Naturally Occurring Dissolved Organic Matter. *mBio* 2016;**7**:e01279-16.
Gilbert JA, Steele JA, Caporaso JG *et al.* Defining seasonal marine microbial community dynamics. *ISME Journal* 2012;**6**:298–308.

Giovannoni S, Stingl U. The importance of culturing bacterioplankton in the "omics" age. *Nat Rev Microbiol* 2007;**5**:820–6.
Giovannoni SJ. Genome Streamlining in a Cosmopolitan Oceanic Bacterium. *Science (1979)* 2005;**309**:1242–5.
Giovannoni SJ, Cameron Thrash J, Temperton B. Implications of streamlining theory for microbial ecology. *ISME Journal* 2014;**8**:1553–65.
Giovannoni SJ, Hayakawa DH, Tripp HJ *et al.* The small genome of an abundant coastal ocean methylotroph. *Environ Microbiol* 2008;**10**:1771–82.
Giovannoni SJ, Vergin KL. Seasonality in ocean microbial communities. *Science (1979)* 2012;**335**:671–6.
Gratton Y, Price NM, Simpson KG. An annual study of inorganic and organic nitrogen and phosphorus and silicic acid in the southeastern Beaufort Sea. 2008;**113**:1–16.
Greenwell, Brandon M. pdp: An R Package for Constructing Partial Dependence Plots. *R J* 2017;**9**:421.
Guéguen C, McLaughlin FA, Carmack EC *et al.* The nature of colored dissolved organic matter in the southern Canada Basin and East Siberian Sea. *Deep Sea Res 2 Top Stud Oceanogr* 2012;**81–84**:102–13.
Gupta A, Gupta R, Singh RL. Microbes and Environment. *Principles and Applications of Environmental Biotechnology for a Sustainable Future*. Singapore: Springer Singapore, 2017, 43–84.
Halsey KH, Carter AE, Giovannoni SJ. Synergistic metabolism of a broad range of C1 compounds in the marine methylotrophic bacterium HTCC2181. *Environ Microbiol* 2012;**14**:630–40.
Hara Y, Obata H, Doi T *et al.* Rare earth elements in seawater during an iron-induced phytoplankton bloom of the western subarctic Pacific (SEEDS-II). *Deep Sea Res 2 Top Stud Oceanogr* 2009;**56**:2839–51.
Heikes BG. Atmospheric methanol budget and ocean implication. *Global Biogeochem Cycles* 2002;**16**:1–13.
Hill VJ, Light B, Steele M *et al.* Light availability and phytoplankton growth beneath arctic sea ice: Integrating observations and modeling. *J Geophys Res Oceans* 2018;**123**:3651–67.
Hothorn T, Hornik K, Zeileis A. Unbiased Recursive Partitioning: A Conditional Inference Framework. *Journal of Computational and Graphical Statistics* 2006;**15**:651–74.
Huggett MJ, Hayakawa DH, Rappé MS. Genome sequence of strain HIMB624, a cultured representative from the OM43 clade of marine Betaproteobacteria. *Stand Genomic Sci* 2012;**6**:11–20.
Hutalle-Schmelzer KML, Zwirnmann E, Krüger A *et al.* Enrichment and cultivation of pelagic bacteria from a humic lake using phenol and humic matter additions. *FEMS Microbiol Ecol* 2010;**72**:58–73.
ICCI. State of the Cryosphere. 2022.
IPCC. *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*. H.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, A. Alegría, M. Nicolai, A. Okem, J. Petzold, B. Rama, N.M. WeyerH.-O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Min NMW (ed.). Cambridge, UK and New York, NY, USA: Cambridge University Press, 2019.
IPCC. *Climate Change 2021 The Physical Science Basis WGI*., 2021.

Jackson JM, Williams WJ, Carmack EC. Winter sea-ice melt in the Canada Basin, Arctic Ocean. *Geophys Res Lett* 2012;**39**, DOI: 10.1029/2011GL050219.

Jezbera J, Jezberová J, Koll U *et al.* Contrasting trends in distribution of four major planktonic betaproteobacterial groups along a pH gradient of epilimnia of 72 freshwater habitats. *FEMS Microbiol Ecol* 2012;**81**:467–79.

Jimenez-Infante F, Ngugi DK, Vinu M *et al.* Comprehensive genomic analyses of the OM43 clade including a novel species from Red Sea indicate ecotype differentiation among marine methylotrophs. *Appl Environ Microbiol* 2015;**966**:AEM.02852-15.

Kalyuzhnaya MG, Beck D a C, Vorobev A *et al.* Novel methylotrophic isolates from lake sediment, description of Methylotenera versatilis sp. nov. and emended description of the genus Methylotenera. *Int J Syst Evol Microbiol* 2012;**62**:106–11.

Kang CS, Dunfield PF, Semrau JD. The origin of aerobic methanotrophy within the Proteobacteria. *FEMS Microbiol Lett* 2019;**366**:1–11.

Karp PD, Paley SM, Krummenacker M *et al.* Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief Bioinform* 2010;**11**:40–79.

Keltjens JT, Pol A, Reimann J *et al.* PQQ-dependent methanol dehydrogenases: rare-earth elements make a difference. *Appl Microbiol Biotechnol* 2014;**98**:6163–83.

Kirchman D, Yu L, Fuchs B *et al.* Structure of bacterial communities in aquatic systems as revealed by filter PCR. *Aquatic Microbial Ecology* 2001;**26**:13–22.

Kolb S, Stacheter A. Prerequisites for amplicon pyrosequencing of microbial methanol utilizers in the environment. *Front Microbiol* 2013;**4**:268.

Kraemer S, Ramachandran A, Colatriano D *et al.* Diversity and biogeography of SAR11 bacteria from the Arctic Ocean. *ISME Journal* 2019, DOI: 10.1038/s41396-019-0499-4.

Krause SMB, Johnson T, Samadhi Karunaratne Y *et al.* Lanthanide-dependent cross-feeding of methane-derived carbon is linked by microbial community interactions. *Proceedings of the National Academy of Sciences* 2016;**114**:201619871.

Krishfield RA, Proshutinsky A, Tateyama K *et al.* Deterioration of perennial sea ice in the Beaufort Gyre from 2003 to 2012 and its impact on the oceanic freshwater cycle. *J Geophys Res Oceans* 2014;**119**:1271–305.

Krüger K, Chafee M, Ben Francis T *et al.* In marine Bacteroidetes the bulk of glycan degradation during algae blooms is mediated by few clades using a restricted set of genes. *ISME Journal* 2019;**13**:2800–16.

Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol* 2016, DOI: 10.1093/molbev/msw054.

Lanclos VC, Henson MW, Pitre DM *et al.* Draft Genome Sequence of Strain LSUCC0135 , an Early Diverging Member of the Order Methylophilales in the Phylum Betaproteobacteria. 2016;**4**:2015–6.

Lannuzel D, Tedesco L, van Leeuwe M *et al.* The future of Arctic sea-ice biogeochemistry and ice-associated ecosystems. *Nat Clim Chang* 2020;**10**:983–92.

Lau E, Fisher MC, Steudler P a *et al.* The methanol dehydrogenase gene, mxaF, as a functional and phylogenetic marker for proteobacterial methanotrophs in natural environments. *PLoS One* 2013;**8**:e56993.

Lechner M, Findeiß S, Steiner L *et al.* Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics* 2011, DOI: 10.1186/1471-2105-12-124.

Levins R. *Evolution in Changing Environments*. Princeton University Press, 1968.

Li WKW, Carmack EC, McLaughlin F A. *et al.* Space-for-time substitution in predicting the state of picoplankton and nanoplankton in a changing Arctic Ocean. *J Geophys Res Oceans* 2013;**118**:5750–9.

Li WKW, Dickie PM. Monitoring Phytoplankton, Bacterioplankton, and Virioplankton in a Coastal Inlet (Bedford Basin) by Flow Cytometry. *Cytometry* 2001;**44**:236–46.

Li WKW, Mclaughlin FA, Lovejoy C *et al.* Smallest Algae Thrive As the Arctic Ocean Freshens. 2009;**326**:2009.

Logares R, Bråte J, Bertilsson S *et al.* Infrequent marine-freshwater transitions in the microbial world. *Trends Microbiol* 2009;**17**:414–22.

Lovejoy C, Galand PE, Kirchman DL. Picoplankton diversity in the Arctic Ocean and surrounding seas. *Marine Biodiversity* 2011;**41**:5–12.

Mcdonald IR, Murrell JC. The methanol dehydrogenase structural gene mxaf and its use as a functional gene probe for methanotrophs and methylotrophs. *Appl Environ Microbiol* 1997;**63**:3218–24.

McLaughlin F, Carmack E, Proshutinsky A *et al.* The Rapid Response of the Canada Basin to Climate Forcing: From Bellwether to Alarm Bells. *Oceanography* 2011;**24**:146–59.

McLaughlin F, Proshutinsky A, Carmack EC *et al.* Physical, chemical, and zooplankton data from the Canada Basin and Canadian Arctic Archipelago, July 20 to September 14, 2006. *Canadian Data Report of Hydrography and Ocean Science* 2012:x + 373 p.

McMurdie PJ, Holmes S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 2013, DOI: 10.1371/journal.pone.0061217.

Millet DB, Jacob DJ, Custer TG *et al.* New constraints on terrestrial and oceanic sources of atmospheric methanol. *Atmospheric Chemistry and Physics Discussions* 2008;**8**:7609–55.

Milne PJ, Riemer DD, Zika RodG *et al.* Measurement of vertical distribution of isoprene in surface seawater, its chemical fate, and its emission from several phytoplankton monocultures. *Mar Chem* 1995;**48**:237–44.

Mincer TJ, Aicher AC. Methanol Production by a Broad Phylogenetic Array of Marine Phytoplankton. 2016:1–17.

Morris RM, Longnecker K, Giovannoni SJ. Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environ Microbiol* 2006;**8**:1361–70.

Mundy C, Gosselin M, Gratton Y *et al.* Role of environmental factors on phytoplankton bloom initiation under landfast sea ice in Resolute Passage, Canada. *Mar Ecol Prog Ser* 2014;**497**:39–49.

Neufeld JD, Boden R, Moussard H *et al.* Substrate-specific clades of active marine methylotrophs associated with a phytoplankton bloom in a temperate coastal environment. *Appl Environ Microbiol* 2008a;**74**:7321–8.

Neufeld JD, Chen Y, Dumont MG *et al.* Marine methylotrophs revealed by stable-isotope probing, multiple displacement amplification and metagenomics. *Environ Microbiol* 2008b;**10**:1526–35.

Neufeld JD, Schäfer H, Cox MJ *et al.* Stable-isotope probing implicates Methylophaga spp and novel Gammaproteobacteria in marine methanol and methylamine metabolism. *ISME J* 2007;**1**:480–91.

Newton RJ, Jones SE, Eiler A *et al.* A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiology and Molecular Biology Reviews* 2011;**75**:14–49.

Oh S, Caro-Quintero A, Tsementzi D *et al.* Metagenomic Insights into the Evolution, Function, and Complexity of the Planktonic Microbial Community of Lake Lanier, a Temperate Freshwater Ecosystem. *Appl Environ Microbiol* 2011;**77**:6000–11.

Oksanen AJ, Blanchet FG, Friendly M *et al.* Package ' vegan .' 2020.

Oziel L, Baudena A, Ardyna M *et al.* Faster Atlantic currents drive poleward expansion of temperate phytoplankton in the Arctic Ocean. *Nat Commun* 2020;**11**:1–8.

Parada AE, Needham DM, Fuhrman JA. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 2016;**18**:1403–14.

Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;**25**:1043–55.

Paver SF, Muratore D, Newton RJ *et al.* Reevaluating the Salty Divide: Phylogenetic Specificity of Transitions between Marine and Freshwater Systems. *mSystems* 2018;**3**:1–15.

Pedrós-Alió C, Potvin M, Lovejoy C. Diversity of planktonic microorganisms in the Arctic Ocean. *Prog Oceanogr* 2015;**139**:233–43.

Penn K, Jensen PR. Comparative genomics reveals evidence of marine adaptation in Salinispora species. *BMC Genomics* 2012;**13**:86.

Peralta-ferriz C, Woodgate RA. Seasonal and interannual variability of pan-Arctic surface mixed layer properties from 1979 to 2012 from hydrographic data, and the dominance of stratification for multiyear mixed layer depth shoaling. *Prog Oceanogr* 2014.

Philippot L, Griffiths BS, Langenheder S. Microbial Community Resilience across Ecosystems and Multiple Disturbances. *Microbiology and Molecular Biology Reviews* 2021;**85**, DOI: 10.1128/mmbr.00026-20.

Proshutinsky A, Krishfield R, Toole JM *et al.* Analysis of the Beaufort Gyre Freshwater Content in 2003–2018. *J Geophys Res Oceans* 2019;**124**:9658–89.

Ramachandran A, McLatchie S, Walsh DA. A Novel Freshwater to Marine Evolutionary Transition Revealed within Methylophilaceae Bacteria from the Arctic Ocean. *mBio* 2021;**12**:e0130621.

Ramachandran A, Walsh DA. Investigation of XoxF methanol dehydrogenases reveals new methylotrophic bacteria in pelagic marine and freshwater ecosystems. *FEMS Microbiol Ecol* 2015;**91**:fiv105.

Rappé MS, Kemp PF, Giovannoni SJ. Phylogenetic diversity of marine coastal picoplankton 16s rRNA genes cloned from the continental shelf off Cape Hatteras, North Carolina. *Limnol Oceanogr* 1997;**42**:811–26.

Rappé MS, Vergin K, Giovannoni SJ. Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. *FEMS Microbiol Ecol* 2000;**33**:219–32.

Rice P, Longden L, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends in Genetics* 2000;**16**:276–7.

Robicheau BM, Tolman J, Bertrand EM *et al.* Highly-resolved interannual phytoplankton community dynamics of the coastal Northwest Atlantic. *ISME Communications* 2022;**2**:1–12.

Rodriguez-R LM, Konstantinidis KT. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. 2016, DOI: 10.7287/peerj.preprints.1900v1.

Royo-llonch M, Sánchez P, Ruiz-gonzález C *et al.* Ecogenomics of key prokaryotes in the Arctic Ocean. *bioRxiv* 2020.
Royo-Llonch M, Sánchez P, Ruiz-González C *et al.* Compendium of 530 metagenome-assembled bacterial and archaeal genomes from the polar Arctic Ocean. *Nat Microbiol* 2021;**6**:1561–74.
Ryo M, Rillig MC. Statistically reinforced machine learning for nonlinear patterns and variable interactions. *Ecosphere* 2017;**8**:e01976.
Salcher MM, Neuenschwander SM, Posch T *et al.* The ecology of pelagic freshwater methylotrophs assessed by a high-resolution monitoring and isolation campaign. *ISME J* 2015:1–12.
Salcher MM, Pernthaler J, Zeder M *et al.* Spatio-temporal niche separation of planktonic Betaproteobacteria in an oligo-mesotrophic lake. *Environ Microbiol* 2008;**10**:2074–86.
Salcher MM, Schaefle D, Kaspar M *et al.* Evolution in action: habitat transition from sediment to the pelagial leads to genome streamlining in Methylophilaceae. *ISME Journal* 2019;**13**:2764–77.
Santoro AE, Richter RA, Dupont CL. Planktonic marine archaea. *Ann Rev Mar Sci* 2019;**11**:131–58.
Sargeant SL, Murrell JC, Nightingale PD *et al.* Seasonal variability in microbial methanol utilisation in coastal waters of the western English Channel. *Mar Ecol Prog Ser* 2016;**550**:53–64.
Scheinin M, Riebesell U, Rynearson T a *et al.* Experimental evolution gone wild. *Journal of the Royal Society, Interface / the Royal Society* 2015;**12**:1–5.
Sekar R, Fuchs BM, Amann R *et al.* Flow Sorting of Marine Bacterioplankton after Fluorescence In Situ Hybridization Flow Sorting of Marine Bacterioplankton after Fluorescence In Situ Hybridization. *Appl Environ Microbiol* 2004;**70**:6210–9.
Shi Y, Tyson GW, DeLong EF. Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* 2009;**459**:266–9.
Sinha V, Williams J, Meyerhöfer M *et al.* Air-sea fluxes of methanol, acetone, acetaldehyde, isoprene and DMS from a Norwegian fjord following a phytoplankton bloom in a mesocosm experiment. *Atmospheric Chemistry and Physics Discussions* 2007;**6**:9907–35.
Sjostedt SJ, Leaitch WR, Levasseur M *et al.* Evidence for the uptake of atmospheric acetone and methanol by the Arctic Ocean during late summer DMS-Emission plumes. *Journal of Geophysical Research: Atmospheres* 2012;**117**, DOI: 10.1029/2011JD017086.
Skovran E, Martinez-Gomez NC. Just add lanthanides. 2015.
Smith DC, Azam F. A simple, economical method for measuring bacterial protein synthesis rates in seawater using tritiated-leucine. *Marine Microbial Food Webs* 1992;**6**:107–14.
Solomon A, Heuzé C, Rabe B *et al.* Freshwater in the Arctic Ocean 2010-2019. *Ocean Science* 2021;**17**:1081–102.
Song J, Oh HM, Cho JC. Improved culturability of SAR11 strains in dilution-to-extinction culturing from the East Sea, West Pacific Ocean. *FEMS Microbiol Lett* 2009;**295**:141–7.
Sosa O a, Gifford SM, Repeta DJ *et al.* High molecular weight dissolved organic matter enrichment selects for methylotrophs in dilution to extinction cultures. *ISME J* 2015:1–15.
Sowell SM, Abraham PE, Shah M *et al.* Environmental proteomics of microbial plankton in a highly productive coastal upwelling system. *ISME J* 2011;**5**:856–65.

Stewart FJ, Dalsgaard T, Young CR *et al.* Experimental incubations elicit profound changes in community transcription in OMZ bacterioplankton. *PLoS One* 2012;**7**, DOI: 10.1371/journal.pone.0037118.
Stewart FJ, Ottesen EA, Delong EF. Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *ISME Journal* 2010;**4**:896–907.
Stroeve J, Notz D. Changing state of Arctic sea ice across all seasons. *Environmental Research Letters* 2018;**13**, DOI: 10.1088/1748-9326/aade56.
Strous M, Kraft B, Bisdorf R *et al.* The Binning of Metagenomic Contigs for Microbial Physiology of Mixed Cultures. *Front Microbiol* 2012;**3**:1–11.
Sunagawa S, Coelho LP, Chaffron S *et al.* Structure and function of the global ocean microbiome. *Science (1979)* 2015;**348**:1261359–1261359.
Swan BK, Tupper B, Sczyrba A *et al.* Prevalent genome streamlining and latitudinal divergence of planktonic bacteria in the surface ocean. *Proceedings of the National Academy of Sciences* 2013;**110**:11463–8.
Takemura AF, Chien DM, Polz MF. Associations and dynamics of vibrionaceae in the environment, from the genus to the population level. *Front Microbiol* 2014;**5**:1–26.
Taubert M, Grob C, Howat AM *et al.* XoxF encoding an alternative methanol dehydrogenase is widespread in coastal marine environments. *Environ Microbiol* 2015;**17**:3937–48.
Timmermans ML, Toole J, Krishfield R. Warming of the interior Arctic Ocean linked to sea ice losses at the basin margins. *Sci Adv* 2018;**4**:1–7.
Villaescusa JA, Casamayor EO, Rochera C *et al.* A close link between bacterial community composition and environmental heterogeneity in maritime Antarctic lakes. *Int Microbiol* 2010;**13**:67–77.
Vincent WF. Microbial ecosystem responses to rapid climate change in the Arctic. *ISME Journal* 2010;**4**:1089–91.
Vu HN, Subuyuj GA, Vijayakumar S *et al.* Lanthanide-dependent regulation of methanol oxidation systems in Methylobacterium extorquens AM1 and their contribution to methanol growth. *J Bacteriol* 2016;**198**:1250–9.
Vuilleumier S, Khmelenina VN, Bringel F *et al.* Genome sequence of the haloalkaliphilic methanotrophic bacterium methylomicrobium alcaliphilum 20Z. *J Bacteriol* 2012;**194**:551–2.
Walsh DA, Lafontaine J, Grossart HP. On the Eco-Evolutionary Relationships of Fresh and Salt Water Bacteria and the Role of Gene Transfer in Their Adaptation. In: Gophna U (ed.). *Lateral Gene Transfer in Evolution*. New York: Springer, 2013, 55–77.
Wang Q, Wekerle C, Danilov S *et al.* Arctic Sea Ice Decline Significantly Contributed to the Unprecedented Liquid Freshwater Accumulation in the Beaufort Gyre of the Arctic Ocean. *Geophys Res Lett* 2018;**45**:4956–64.
Wickham H. Package `ggplot2`: Elegant Graphics for Data Analysis. *Springer-Verlag New York* 2016.
Wohl C, Jones AE, Sturges WT *et al.* Sea ice concentration impacts dissolved organic gases in the Canadian Arctic. *Biogeosciences* 2022;**19**:1021–45.
Wu M, Chatterji S, Eisen JA. Accounting For Alignment Uncertainty in Phylogenomics. Salemi M (ed.). *PLoS One* 2012;**7**:e30288.
Yager PL, Connelly TL, Mortazavi B *et al.* Dynamic bacterial and viral response to an algal bloom at subzero temperatures. *Limnol Oceanogr* 2001;**46**:790–801.

Yau S, Lauro FM, Williams TJ *et al.* Metagenomic insights into strategies of carbon conservation and unusual sulfur biogeochemistry in a hypersaline Antarctic lake. *ISME Journal* 2013;**7**:1944–61.
Yeh Y-C, Fuhrman JA. Contrasting diversity patterns of prokaryotes and protists over time and depth at the San-Pedro Ocean Time series. *ISME Communications* 2022;**2**:1–12.
Zaikova E, Walsh DA, Stilwell CP *et al.* Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environ Microbiol* 2010;**12**:172–91.
Zhuang Y, Jin H, Chen J *et al.* Nutrient and phytoplankton dynamics driven by the Beaufort Gyre in the western Arctic Ocean during the period 2008–2014. *Deep Sea Res 1 Oceanogr Res Pap* 2018;**137**:30–7.
Zwart G, Hannen EJ Van, Agterveld MPK *et al.* Rapid Screening for Freshwater Bacterial Groups by Using Reverse Line Blot Hybridization. 2003;**69**:5875–83.
Zwart G, Hiorns WD, Methé B a *et al.* Nearly identical 16S rRNA sequences recovered from lakes in North America and Europe indicate the existence of clades of globally distributed freshwater bacteria. *Syst Appl Microbiol* 1998;**21**:546–56.

Appendix A: Quality Control and Assessment of the filter-PCR method, and Additional Analyses of the Community Composition

In Chapter 4, we used archival samples to document the interannual variation in bacterial and archaeal communities across a 9-year time series of the Canada Basin that includes two historic sea ice minima (2007 and 2012). Here we outline the filter-PCR process as well as additional analyses of the overall community composition and the *Methylophilaceae* family.

A.1. Quality Control and Assessment of the Methodology

A.1.1. Assessment of the filter PCR approach

The use of archival samples allowed us to process over 100 archival samples over a 9-year period. However, there were challenges and caveats that we needed to address prior to analysing the data. We used 500µl samples of seawater that was fixed with 1% paraformaldehyde, which were originally collected for flow-cytometric based analysis. The samples were also freeze-thawed twice leading us to question whether DNA degradation/modification has occurred and how that might decrease our yield and overall diversity. We also needed to determine whether the small volume skewed the diversity of microbes in the water-column, given that bacterioplankton communities can vary remarkably over very small spatial scales.

To address these challenges, we sequenced control samples. We tested the following: (1) Filter PCR versus genomic DNA (gDNA) PCR from the same sampling point, (2) Filter variation between filter pieces from the same sample, and (3) Sequencing run variability by running the same sample on different plates, (4) PCR variation by using the gDNA from the same sample in triplicate PCR reactions.

A.1.2. Community structure of filter vs gDNA replicates at Stations CB9 and CB15 in 2017

We first needed to determine whether we could detect any anomalies in diversity that would be attributed to the filter samples being: (1) fixed in paraformaldehyde, (2) freeze thawed twice, (3) small volume, and (4) not size fractioned (up to 3.7μm). We sequenced filter and gDNA, which was size fractioned at 0.22μm, samples from 2017. To match filter and gDNA samples, we chose three depths at Stations CB9 and CB21, two in the UAW and one in the PW. The samples clustered by depth regardless of the station **(Figure A.1)**. For each depth, the filter PCR samples were more closely related to each other and the gDNA samples were more closely related to each other. Taking a closer look at the family-level taxonomy, we saw that the filters had a higher diversity compared to the gDNA samples and there was a clear distinction in bacterial community composition between the three depths. The first UAW samples at ~20m had a high relative abundance of ASVs identified as part of the Pelagibacteraceae, Halamonadaceae, Oceanospirillaceae, Methylophilaceae, Flavobacteriaceae, Alteromonadaceae, and Rhodobacteraceae families. The deeper UAW samples at the DCM (~60-80m) were enriched in archaeal ASVs belonging to Crenarchaeaceae and Marine group II as well as in the bacterial groups Pelagibacteraceae, Halamonadaceae, and SUP05. The PW samples at 207m has a higher relative abundance of Crenarchaeaceae and Marine group II ASVs as well as Pelagibacteraceae, A714017, and SAR324 ASVs. The first UAW samples were more distinct, likely due to their proximity to the surface, while the PW and the second UAW samples exhibit more similar community structure at the family-level in comparison.

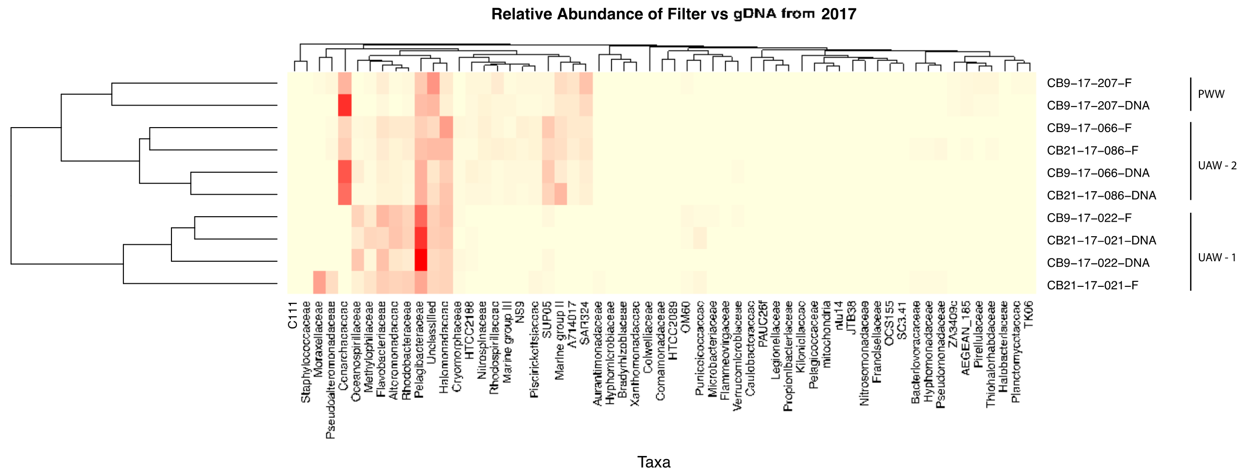


Figure A.1 *A hierarchical clustering analysis of filter-PCR vs gDNA PCR for 3 sampling depths, based on the ASV distributions using the hclust function in stats package for R with a Bray-Curtis dissimilarity metric. The relative abundance of taxa was plotted next to the clustered samples.*

A.1.3. Community structure of filter replicates at Stations CB9 and CB15 in 2004 and 2012

Next, we wanted to determine whether there were any discrepancies between filter pieces from the same sample and/or any variability by running the same sample on different plates. To do so, six replicate filter PCR samples were sequenced and analyzed for CB9 and CB15 SML from 2004 and 2012. The beta diversity analysis showed that the variation between the samples was much higher than within the samples **(Figure A.2a)**. The hierarchical clustering analysis showed the same pattern, with the filter PCR samples that were derived from filter pieces from the sample clustering together (i.e. all CB15−04−006−F samples were clustered together), showing that the results are reproducible **(Figure A.2b).** Samples were also clustering by year, with 2004 samples altogether and 2012 samples altogether. Therefore, it was determined that any discrepancy that may have occurred due to sample variability was negligible.

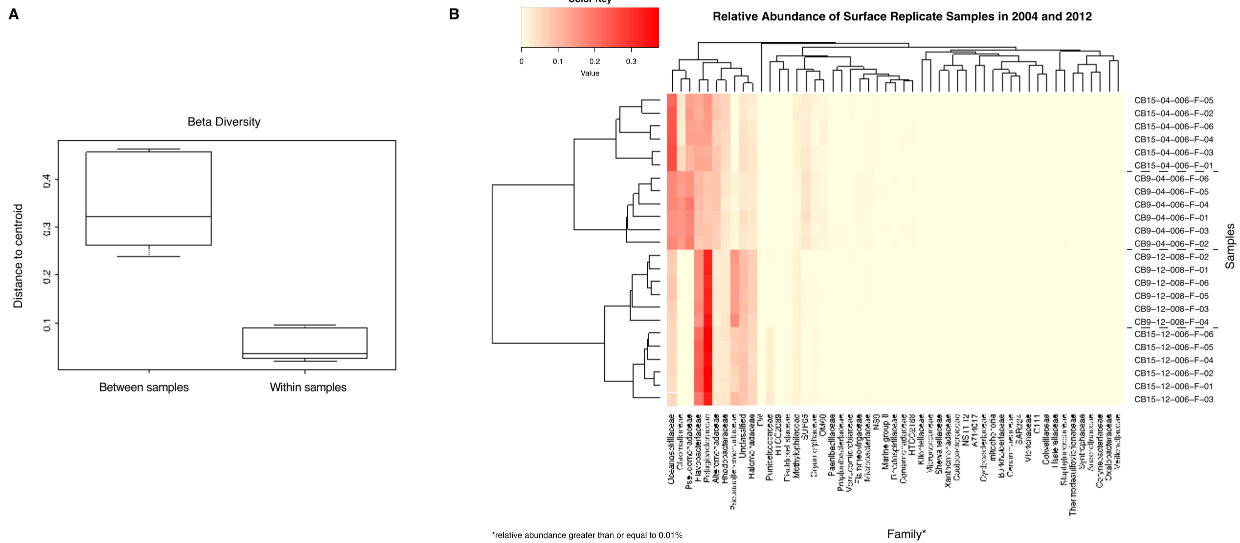


Figure A.2 *(A) A beta diversity analysis showing the variation between the samples and within the samples (B) A hierarchical clustering analysis of surface replicate filter-PCRs for 2 sampling years, based on the ASV distributions using the hclust function in stats package for R with a Bray-Curtis dissimilarity metric. The relative abundance of taxa greater than or equal to 0.01% was plotted next to the clustered samples.*

A.1.4. PCR variation using genomic DNA from 2017 samples

To test whether there is variation between PCRs, we use the gDNA from the same sample in triplicate PCRs. We performed a hierarchical clustering analysis where the heatmap depicts the relative abundance on the family level and the dendrogram shows the relationship on the ASV level **(Figure A.3)**. The relative abundance of taxa is almost identical between each of the six PCR replicates, thereby showing that there is little to no variation introduced by any PCR-related bias.

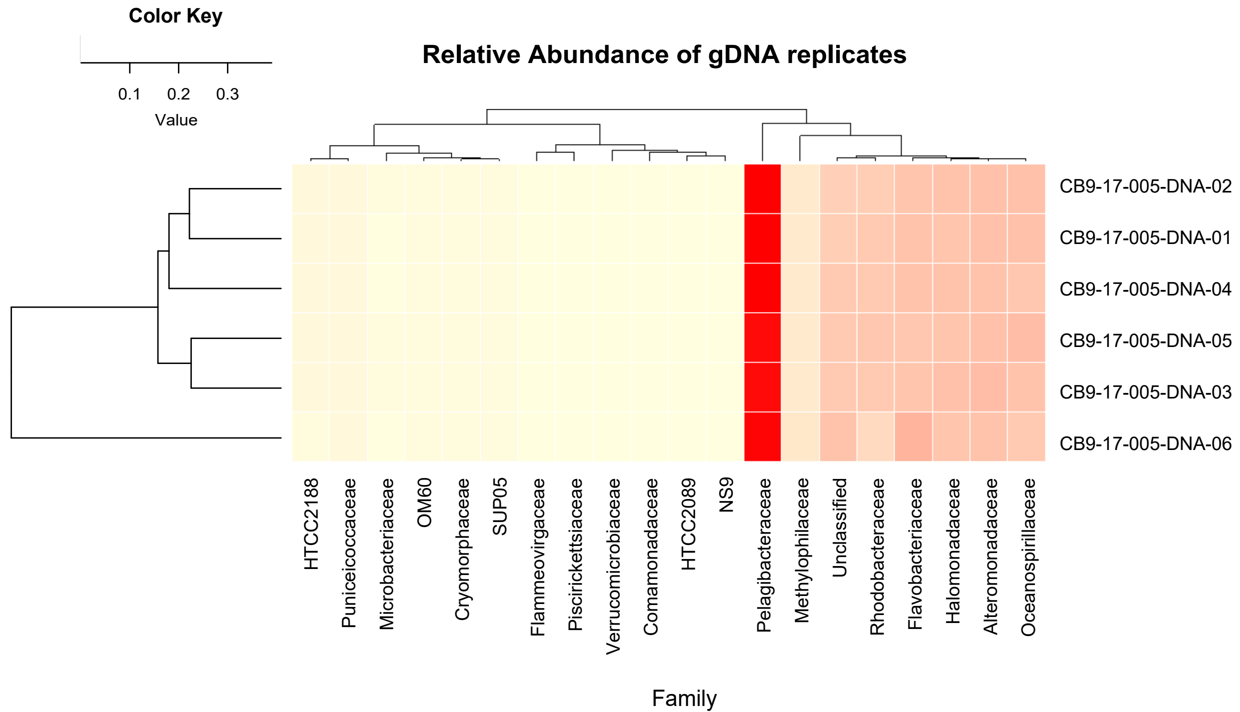


Figure A.3 *A hierarchical clustering analysis of gDNA PCR replicates based on the ASV distributions using the hclust function in stats package for R with a Bray-Curtis dissimilarity metric. The relative abundance of taxa was plotted next to the clustered samples.*

A.2. Additional Analyses of the Community Composition

A.2.1. Overall Community Composition

To further explore the distribution of all samples in relation to environmental conditions of the biomes, we performed dbRDA ordination of the Bray–Curtis distance matrix of the relative abundance values. The samples clustered according to their biome, however, the SML and UAW clusters had some overlap, while PW samples were more distinct in their compositions **(Figure A.4a)**. Along the first dbRDA axis, which accounted for 31.3% of the total variation, samples separated according to a salinity, phosphate, nitrate, and silicate gradient, and depth. The fresher, nutrient poor surface samples were distinguished from saltier, nutrient richer PW samples. Along the second dbRDA axis, which accounted for 5.1% of the variation, samples were separated according to the year. In the SML, silicate, phosphate, temperature and year were the significant

environmental factors that separated samples **(Figure A.4b)**. In the UAW, the explanatory variables were salinity, bacterial count, depth, silicate, longitude and year **(Figure A.4c)**. In the PW, the significant factors were salinity, nitrate, longitude, bacterial count, and year **(Figure A.4d)**. It seemed that year was not an important explanatory variable in the dbRDA of all samples, however, when we break it down to each biome, we found that year drove more of the variation.

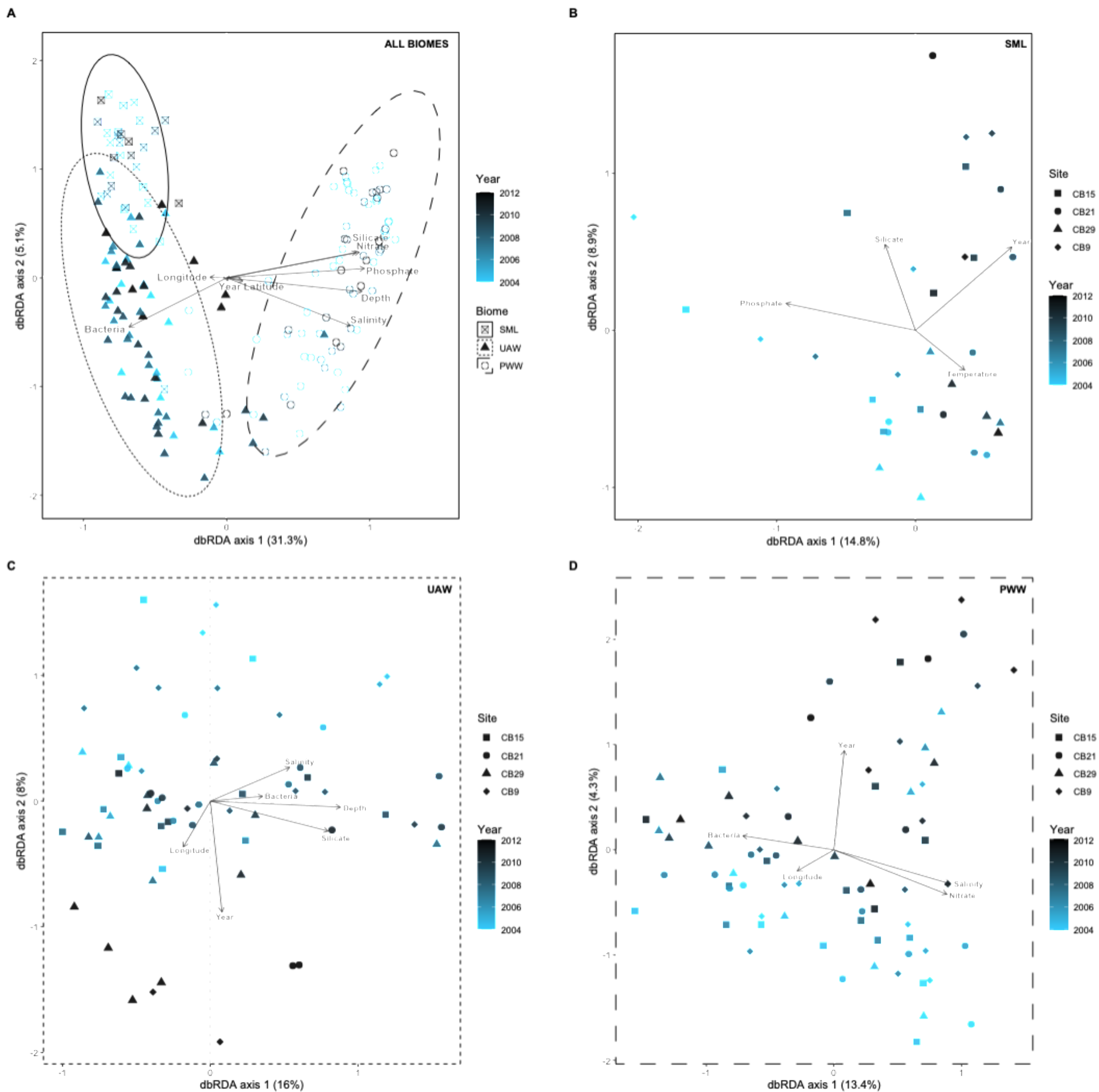


Figure A.4 *(A) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic samples based on the relative abundance of ASVs (B) A distance-based redundancy*

analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic SML samples based on the relative abundance of ASVs (C) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic UAW samples based on the relative abundance of ASVs (D) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic PW samples based on the relative abundance of ASVs

A hierarchical clustering analysis of samples based on the ASV distributions revealed a strong vertical structuring of the bacterioplankton assemblages that largely reflected the three water masses we defined **(Figure A.5)**. In general, PW assemblages were distinct compared to SML and UAW, while the SML and UAW showed some degree of overlap. When taking a closer look at the highly abundant taxa for each water mass, we see that in general, Flavobacteriaceae, Oceanospirillaceae, Halomonadaceae, and Rhodobacteraceae are more abundant in the SML and UAW. The highest abundance of SUP05 is seen in the UAW. The PW is enriched in A714017, Cenarchaeaceae, SAR324, SAR202, and Pseudoalteromonadaceae.

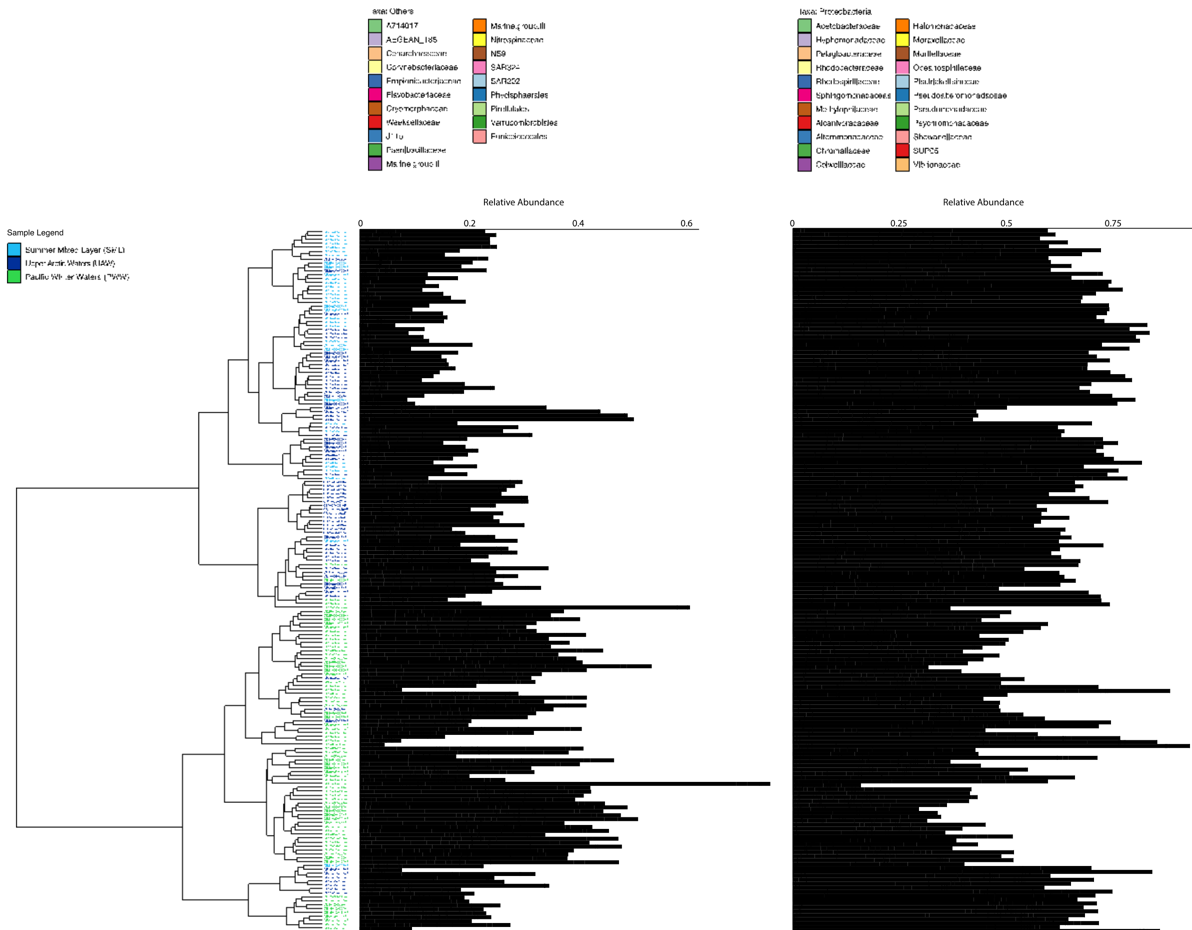


Figure A.5 *A hierarchical clustering analysis of samples based on the ASV distributions using the hclust function in stats package for R with a Bray-Curtis dissimilarity metric. The relative abundance of taxa that were 3.5% or more of the community composition were plotted next to the clustered samples.*

A.2.2. Niche Breadth Analysis of the Whole Community Composition

A.2.2.1. Methodology

Specialist-generalist classification of ASVs was based on Levin's Index (B) (LEVINS 1968; Colwell and Futuyma 1971). The function spec.gen from R package EcolUtils (https://github.com/GuillemSalazar/EcolUtils) was used to calculate B for random permutations of the ASV table (100 for the community composition and 1000 for the Methylophilaceae groups)

and categorize ASVs into generalists if the original B index was larger than its confidence interval (CI 95) or specialists if the original B index was smaller than its confidence interval (CI 95). To standardize this measure of niche breadth on a scale of 0 to 1, Levin's measure of standardized niche breadth was calculated: $B_A = \frac{B-1}{n-1}$ where B is Levin's Index and n is the number of samples (Ecological Methodology, 3rd ed.). The niche breadth analysis was conducted for each biome and for the Methylophilaceae family.

A.2.2.2. Niche breadth analysis and classification as specialists or generalists

In a rapidly changing environment, like the Arctic Ocean, it is important to understand the potential capabilities and niche adaptations of the microbial community to identify different levels of habitat specialization. This can help with understanding where and why select species are more abundant compared to others. We conducted a niche breadth analysis, using Levin's Index (B), to classify the taxa into specialists and generalists **(Figure A.6)**. The species that were most evenly distributed across all samples were classified as generalists, while the unevenly distributed species that usually peak in abundance in fewer samples were specialists (Royo-Llonch *et al.* 2021). We calculated the niche breadth for each water mass. The SML had 336 ASVs assigned **(Figure A.6a)**, the UAW consisted of 480 ASVs **(Figure A.6b)**, and the PW had 913 ASVs **(Figure A.6c)**. There were many ASVs that could not be classified into generalists or specialists: 31% (n=105) in the SML, 28% (n=134) in the UAW, and 31% (n=284) in the PW. In the SML, we classified 60% (n=202) specialists and 9% (n=29) generalists. In the UAW, 62% (n=297) were specialists and 10% (n=47) were generalists. In the PW, 65% (n=597) were specialists and 4% (n=32) were generalists.

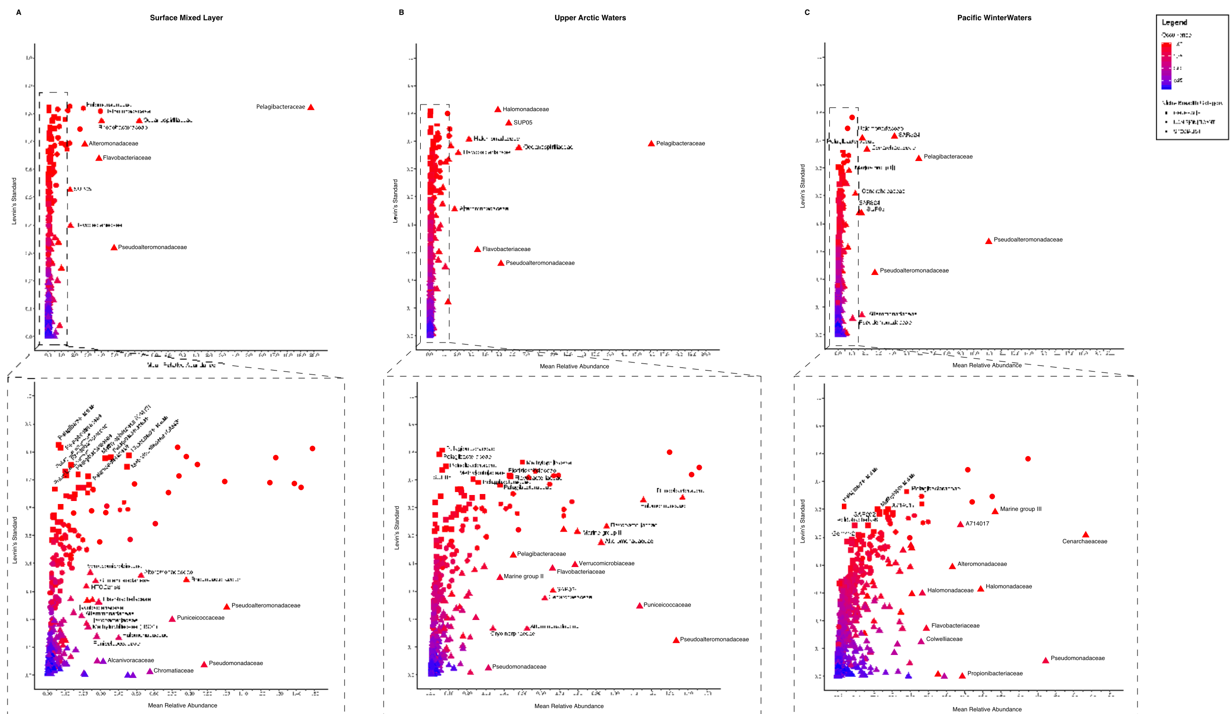


Figure A.6 *Specialist-generalist classification of ASVs was based on Levin's Index (B). The function spec.gen from R package EcolUtils was used to calculate B for random 100 permutations of the ASV table and categorized ASVs into generalists or specialists. (A) Niche-breadth analysis of the SML (B) Niche-breadth analysis of the UAW (C) Niche-breadth analysis of the PW.*

Across all water masses, the highly abundant ASVs that were classified as generalists belonged to Pelagibacteraceae followed by Methylophilaceae and Flavobacteriales. The generalists unique to the UAW were assigned to ASVs belonging to the following groups: SUP05, Oceanospirillaceae, and Actinomycetales. In the PWW, the unique generalists were ASVs that were assigned to SAR202, A714017, AEGEAN_185, Acidobacteria, Chloroflexi, Pseudoalteromonadaceae, and SAR324. Interestingly, there were 17 generalists that were found in both the SML and the UAW, which were dominated mostly by ASVs from Pelagibacteraceae, followed by Methylophilaceae, Flavobacteraceae, Rhodobacteraceae, Halomonadaceae, Microbacteriaceae. We found only 1 shared generalist between the UAW and the PW, which belonged to and ASV from the Arctic97B-4 family. We identified 3 generalists that were found in

all 3 biomes, two Pelagibacteraceae (SAR11) ASVs and one Methylophilaceae (OM43) ASV. On the other hand, there was a high prevalence of habitat specialists with a much larger diversity across all 3 biomes. In the SML, the ASVs with the highest mean relative abundance were from the Pelagibacteraceae, Oceanospirillaceae, Rhodobacteraceae, Flavobacteraceae, Alteromonadaceae, Pseudoalteromonadaceae, and SUP05 families. The remaining specialists had a mean relative abundance of ≤ 1. Some of the families, such as the Methylophilaceae, had ASVs that were generalists and specialists. In the UAW, archaeal specialists emerge, such as Cenarchaeaceae and Marine Group II. In the PW, additional archaeal groups are identified, like Marine Group III, and new bacterial groups, such as A714017 and Propionibacteriaceae.

Methylophilaceae were one of the three families that were found across all biomes and had both specialist and generalist categories; therefore we further investigated the ASVs to understand their unique distribution and diversity. 15 ASVs were assigned to the Methylophilaceae group, of which 3 were generalists and 11 were specialists **(Figure A.7)**. ASV89 belongs to a generalist found in all three biomes and belonged to the OM43 clade. NCBI Blast results showed that this ASV is 100% identical to samples collected in the Arctic and Southern oceans. Interestingly, ASV192, belonging to the BS01 clade, has the most specialized niche and distribution restricted to the SML.

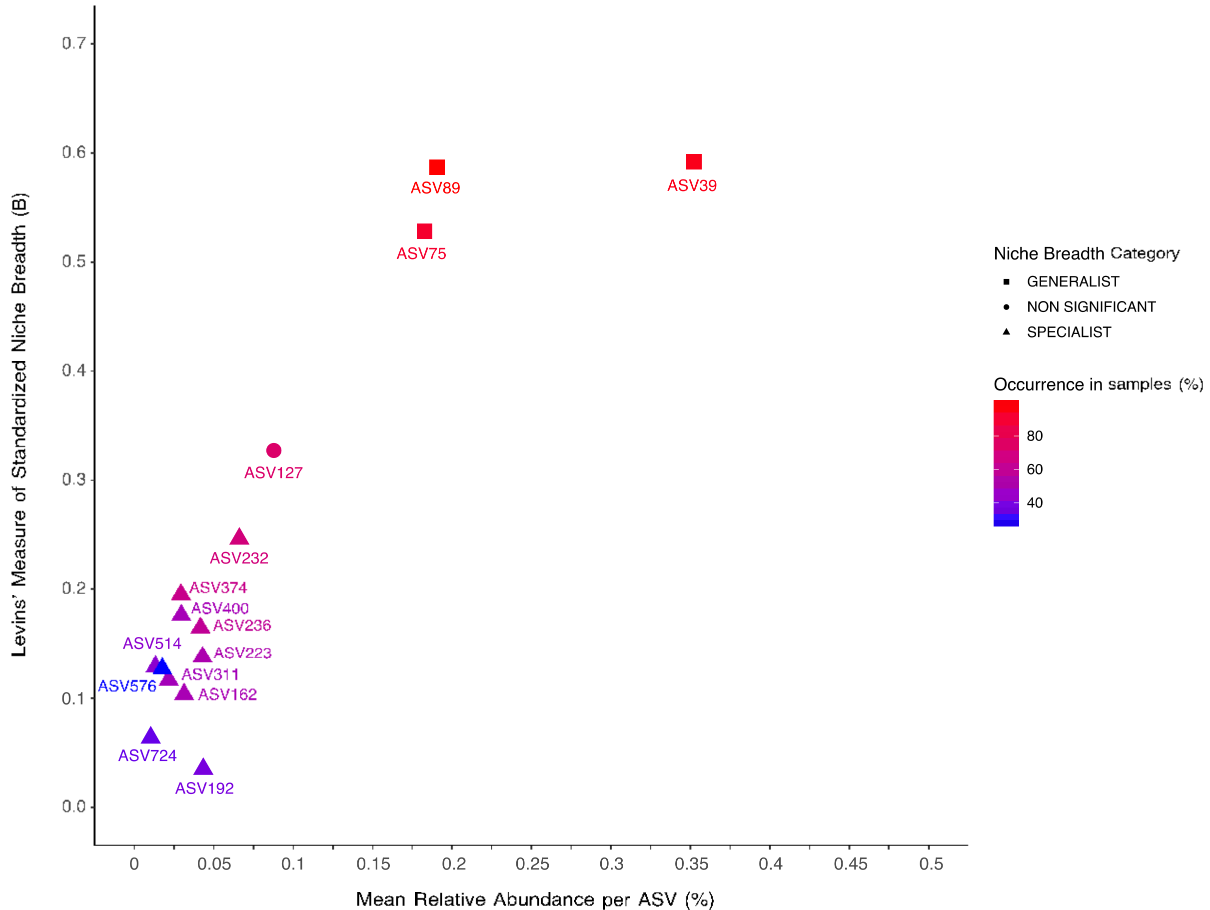


Figure A.7 *Specialist-generalist classification of ASVs was based on Levin's Index (B). The function spec.gen from R package EcolUtils was used to calculate B for random 1000 permutations of the ASV table and categorized ASVs into generalists or specialists.*

A.2.3. Diversity and distribution of the Methylophilaceae family

To assess the diversity and distribution of ASVs within the known phylogeny of the Methylophilaceae family, a phylogenetic analysis of the ASVs was conducted. In total, 15 ASVs were assigned to the Methylophilaceae family **(Figure A.8a)**. Within OM43, 12 ASVs were assigned to the OM43-A clade and none to OM43-B. We also identified 1 ASV belonging to the recently described BS01 clade. Interestingly, we also detected 2 ASVs fell outside of these previously defined clades.

To further explore Methylophilaceae distribution in relation to environmental conditions of Arctic Ocean biomes we defined, we performed dbRDA ordination of the Bray–Curtis distance matrix of the relative abundance values **(Figure A.8b)**. Overall, the samples clustered according to their biome, however, the SML and UAW clusters overlapped, while PW samples were more distinct in their compositions. Along the first dbRDA axis, which accounted for 20% of the total variation, samples separated according to a salinity, depth, and phosphate gradient, with fresher, more oligotrophic surface samples distinguished from saltier, less oligotrophic PW samples. Along the second dbRDA axis, which accounted for 3.9% of the variation, samples were separated according to longitude and year. ASV_00192, which belongs to the BS01 clade, had a different distribution compared to the OM43 clade. It was only found in the SML, while OM43 ASVs were associated with SML, UAW, and PW samples.

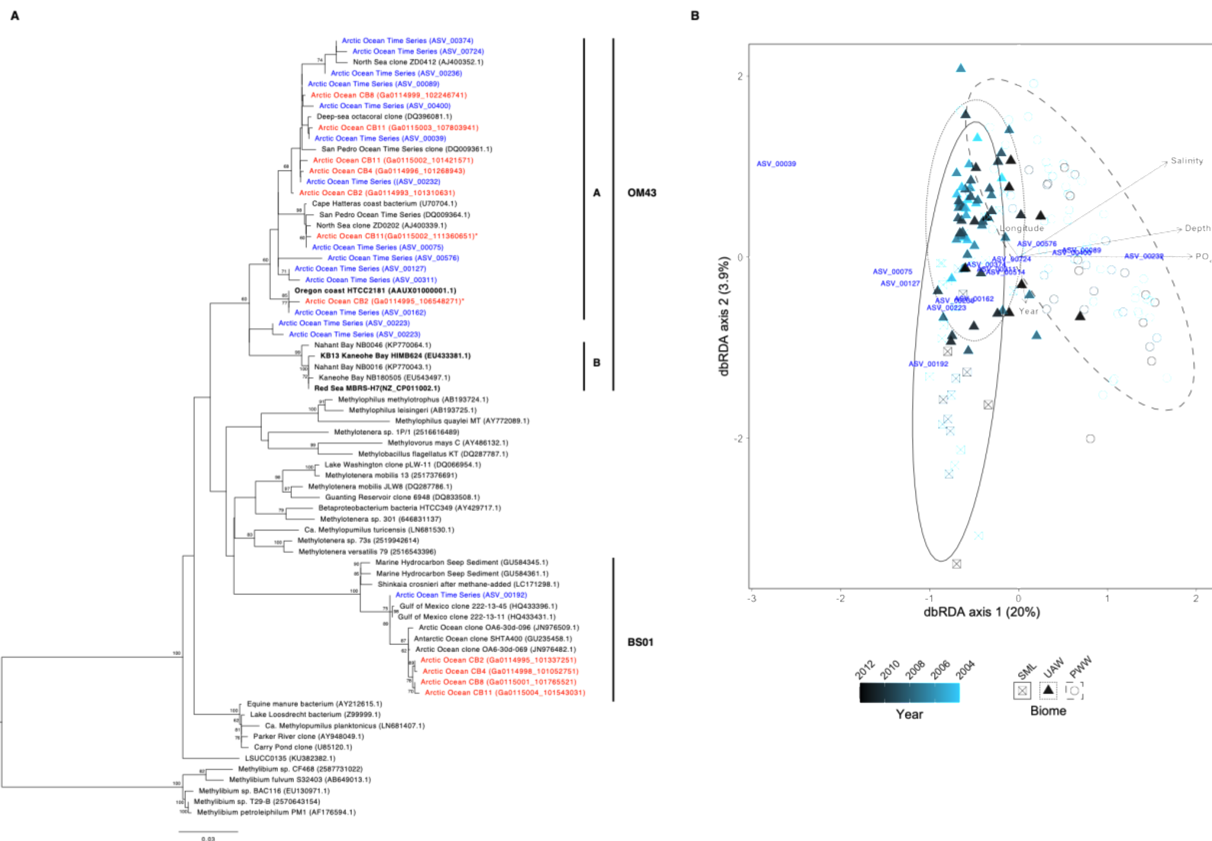


Figure A.8 *(A) Phylogenetic analysis of 16S rRNA genes from* Methylophilaceae *from Canada Basin time series, metagenomes and a diversity of aquatic ecosystems. (B) A distance-based redundancy analysis (dbRDA) of Bray-Curtis dissimilarities of the Arctic samples based on the relative abundance of* Methylophilaceae *ASVs.*

We took a closer look at the relative abundance of Methylophilaceae ASVs over the 9 year time series to decipher any patterns associated with time **(Figure A.9)**. The dataset was split into the 3 biomes and the average was taken when multiple samples were present in each biome. The three ASVs (ASV_00039, ASV_00075, ASV_00089) classified as generalists were present throughout the time series and the relative abundance was not significantly different between the years. For six specialist ASVs (ASV_00162, ASV_00192, ASV_00223, ASV_00236, ASV_00514, ASV_00576), we observed an increase in relative abundance after 2007, which could be attributed to the extreme change in environmental conditions after the sea-ice minima. In contrast, we see a decrease in relative abundance for ASV_00724 after the 2007 sea-ice minima.

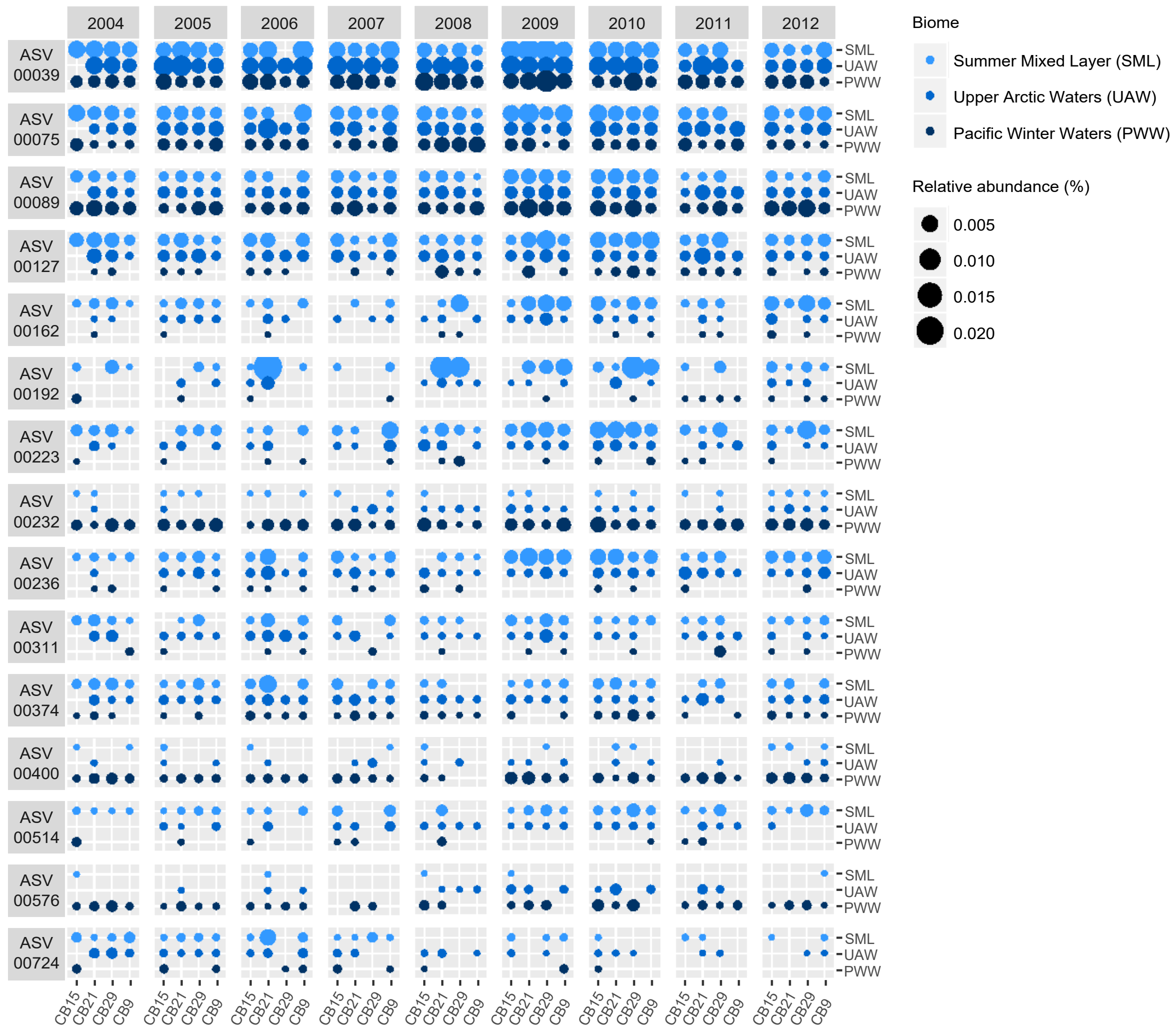


Figure A.9 *The relative abundance of* Methylophilaeae *ASVs across the three biomes (SML, UAW, PW), nine years (2004-2012), and four sites (CB9, CB15, CB21, CB29).*

A.3. Discussion

Quality assessment of Filter-PCR methodology reveals it can be used for future studies, with the appropriate caveats addressed. One of the main goals of this study was to determine the quality and quantity of the 16S rRNA sequence data, and the resulting biases from using small volume samples (500 µl) that have been fixed using 1% paraformaldehyde. We determined that this method can be used to assess community diversity if the caveats are properly addressed and discussed. The ability to use archival samples that describe the community

composition has larger implications for research because of the ability (1) to store more multi-use samples for a longer period of time, and (2) to have multi-use samples (i.e. cell counts, DNA sequencing). There is potential for long-term and large-scale studies to be undertaken. In this study, we were able to sequence the V4 region of the 16S rRNA using a filter PCR approach on samples that have been freeze-thawed multiple times. Paraformaldehyde is known to contribute to DNA degradation and modification, thereby potentially having the ability to decrease PCR yield and increase the failure to amplify, causing error-prone PCRs. Looking at the control reactions, we do not believe that DNA degradation/modification occurred. The controls of this study also revealed that samples were not likely biased due to the multiple freeze-thaws or the small volume. We also addressed bias associated to whether a small volume of seawater is representative of the water-column, given the observation that bacterioplankton communities can vary remarkably over very small spatial scales, using sequenced genomic DNA samples extracted and the corresponding filter samples. Overall, we believe that we have provided evidence that filter PCR on fixed archival samples, although not the most optimal method, is possible as long as biases and caveats are considered and addressed.

Specialists represent most of the diversity in the Arctic Ocean. The Arctic Ocean is undergoing rapid change and our study has been able to capture moments in time for nine years to determine how these changes are affecting the bacterial community composition. We explored the time series data by splitting it up into two groups: generalists and specialists. This was done to understand the potential capabilities of the microbial community and the associated niche adaptations to identify different levels of habitat specialization. Generalists are thought to be less reactive to changes in environmental conditions due to their ability to inhabit a wide range of habitats. On the other hand, specialists tend to respond to change more quickly and have a more

limited distribution. We found that the majority of the ASVs were categorized as specialists using Levin's niche breadth index. A previous metagenomic study in polar regions showed high abundance of specialists as well (Royo-Llonch *et al.* 2021), which could be attributed to rapidly changing environmental variables creating more niche conditions that may support a wider diversity of rare marine bacteria. This could also be driving bacterial evolution in such a way that being a generalist is less favourable than being a specialist. One reason could be the increased energy supply needed to maintain a larger genome size with a variety of metabolic capabilities. Another could be that as nutrients become limited, specialists occupying specific niches can reduce competition for those limiting nutrients, creating interesting interdependencies for nutrient cycling. Further research looking at the co-occurrence patterns and how that affects biogeochemical cycles in rapidly changing environments could provide more insight into why specialists are more abundant. While we were able to taxonomically describe the diversity of bacteria in this time-series, we are still missing the links at the functional and metabolic level. Future studies using whole genome sequencing of the dataset will allow us to decipher whether changes are occurring metabolically and functionally, and if that has caused cascading effects on the environment.

Appendix B: Characterization of novel methylotrophs isolated from the Western Arctic Ocean

In this thesis, I investigated evolution of methylotrophs using metagenomics and discovered that there is a novel clade, BS01, which is potentially adapted to the Arctic Ocean conditions. Given changes in salinity and the fact that there seem to be Arctic Ocean genotypes, we plan on investigating physiology, phenotypic plasticity, and evolutionary change using methylotrophs cultured from the Arctic Ocean as model organisms.

B.1. Introduction

The identification of freshwater relatives of OM43 has led to research questions on the evolutionary diversification of this group. Endemic freshwater bacterial clades have been identified with molecular techniques and significant gene content differences have been described for these populations relative to their marine counterparts (Villaescusa *et al.* 2010; Newton *et al.* 2011; Oh *et al.* 2011; Yau *et al.* 2013). To further explore the relationship between freshwater bacterial clades and their marine counterparts, we will be looking into transitions between a freshwater and marine lifestyle. Transitions like this have occurred rarely over the course of bacterial evolution (Penn and Jensen 2012; Walsh, Lafontaine and Grossart 2013; Takemura, Chien and Polz 2014). While freshwater-marine transitions may be rare, there are a few well-characterized aquatic bacterial groups that have both marine and freshwater lineages (Walsh, Lafontaine and Grossart 2013). A model example of such a transition is the OM43/LD28 clade of methylotrophs. The phylogenetic relationships between the marine and freshwater lineages show that this group of methylotrophs likely has a freshwater ancestry and that the OM43 arose due to colonization of the marine environment. This is supported by the observation that the marine OM43 HTCC2181 genome has acquired 'marine genes' such as Na^{+}-pumping respiratory proteins by lateral gene transfer from marine Alphaproteobacteria (Walsh, Lafontaine and Grossart 2013)

during its evolution. Hence the OM43/LD28 methylotrophs may serve as a useful evolutionary model to investigate habitat transitions, ecological specialization and speciation in the microbial world.

The Arctic Ocean is a transition zone between freshwater and marine habitats owing to its increasingly high input of freshwater and unique stratification (Brown, Nilsson and Pemberton 2019; Proshutinsky *et al.* 2019; Solomon *et al.* 2021). Climate change is affecting the Arctic Ocean at a much faster rate than other ecosystems and provides the opportunity to study the impact of freshening surface waters on microbial communities. As such, it is an ideal study system to investigate the evolutionary history of aquatic methylotrophs that have crossed the freshwater-marine barrier. Identifying the genomic traits that differentiate marine and freshwater lineages will reveal the mechanisms allowing for diversification across habitats.

Beta-proteobacterial methylotrophs are known to use methanol as their sole carbon and energy source. Bacteria from the *Methylophilaceae* family have been isolated using dilution-to-extinction techniques. The media was enriched with methanol and maximum cell concentration was achieved at methanol concentrations between 10 μM to 100 μM (Giovannoni *et al.* 2008; Halsey, Carter and Giovannoni 2012; Salcher *et al.* 2015; Sosa *et al.* 2015). Enriching the media with methanol will increase the probability of isolating methylotrophic bacteria due to the selective pressure.

A study conducted in the Arctic Ocean off the coast of Barrow, Alaska used isotopically labeled [^{15}N]urea to reveal that there are Beta-proteobacterial lineages that have the ability to uptake urea (Connelly *et al.* 2014). In fact, the percentage incorporation of isotopes labeled with [^{15}N]urea was highest in Beta-proteobacteria. In the previous study, the ability of BS01 to potentially use urea as a nitrogen source was discovered. However, OM43/LD28 bacteria do not

possess the genes to transport urea across their cell membranes nor the ability to breakdown urea. The polar surface waters have relatively high concentrations of urea and the metagenomic data from Chapter 1 revealed that the only marine pelagic group potentially capable of using urea as a nitrogen source is BS01. We can therefore use urea as a selective pressure in the enrichment cultures.

Lanthanides are Rare Earth Elements (REEs) that were thought to be biologically inert but have now emerged as essential metals for activity and expression of XoxF (Chistoserdova 2016). XoxF are equipped to use REEs as their cofactor and are currently the only known lanthanide-dependent enzyme (Keltjens *et al.* 2014; Chistoserdova 2016). In fact, the methylotrophic bacteria from soil and sediments have grown on medium enriched with lanthanides (Skovran and Martinez-Gomez 2015; Chistoserdova 2016; Chu, Beck and Lidstrom 2016; Krause *et al.* 2016; Vu *et al.* 2016). Studies have shown that growth on medium enriched with Lanthanides, specifically Lanthanum and Cerium, increases the abundance of bacteria harbouring the *xoxF* gene as well as increasing the expression levels (Chu and Lidstrom 2016; Vu *et al.* 2016). The methylotrophic Beta-proteobacteria in the Arctic Ocean harbour the *xoxF4* methanol dehydrogenase gene and should therefore be enriched when grown in media amended with Lanthanides.

Bacterial communities have high potential to evolve in response to global change. One way to measure and observe the rate of evolution in real time in response to variation in environmental variables is through experimental evolution. Experimental evolution is a method that uses replicate populations, in controlled environments, to measure evolution in real time (Scheinin *et al.* 2015). One advantage to this method is that it produces generalizable results that can further our understanding of how natural selection and evolution work (Scheinin *et al.* 2015). This type of experiment allows manipulation of the environmental factors in replicate populations, allowing

the environmental changes to be linked causally to evolutionary response (Scheinin *et al.* 2015). One limitation of experimental evolution is that control and tractability can usually only be achieved under laboratory conditions, leading to a trade-off between uncovering general evolutionary mechanisms and understanding how they apply in complex natural environments, which in turn limits our understanding of how natural populations evolve in response to environmental drivers (Scheinin *et al.* 2015). This will however give us a knowledgebase to then test, using mesocosm experiments for example, in the natural environment.

The objective is to isolate pure cultures of methylotrophic bacteria. This will be achieved using dilution-to-extinction techniques and enrichments using methanol, urea, and lanthanides.

B.2. Enrichment and Cultivation

Our cultivation protocol was a combination of approaches from Sosa et al. (2015), Salcher et al. (2015), and Vu et al. (2016). The media prep, incubation details, and the dilution and screening of the cultures are taken from all three studies. We employed a dilution-to-extinction approach that has proven to be useful for obtaining abundant but difficult to cultivate bacteria (**Figure B.1**). It allowed for separation of slower growing cells from faster, more easily cultured microorganisms (Sosa *et al.* 2015). Rapid changes are occurring in the surface waters and the subsurface chlorophyll maximum waters (**Figure B.2a-b**). The Arctic Ocean surface and subsurface chlorophyll maximum waters were diluted 10^{-3}, 10^{-4}, and, 10^{-5} using six treatments: 1) Seawater amended with 100 μM MeOH, 2) Seawater amended with 100 μM MeOH and 100 μM Urea, 3) Seawater amended 100 μM MeOH and 30 μM Lanthanum, 4) Seawater amended 100 μM MeOH and 30 μM Cerium, 5) Seawater amended with 100 μM MeOH, 100 μM Urea, and 30 μM Lanthanum, and 6) Seawater amended with 100 μM MeOH, 100 μM Urea, and 30 μM Cerium (**Figure B.2c**). All filtered seawater media also contained a cocktail of vitamins. The liquid

cultures were incubated at 4°C. In order to screen for growth (cell abundance), flow cytometry was used (**Figure B.2d**). Samples that had a concentration of 10^5 cells/mL or greater were scored as positive for cell growth (Sosa *et al.* 2015). The samples that scored positively for cell growth were processed for whole-genome sequencing. These genomes were used in the comparative genomics study to determine the phylogenetic identity and diversity of methylotrophs in the Arctic Ocean.

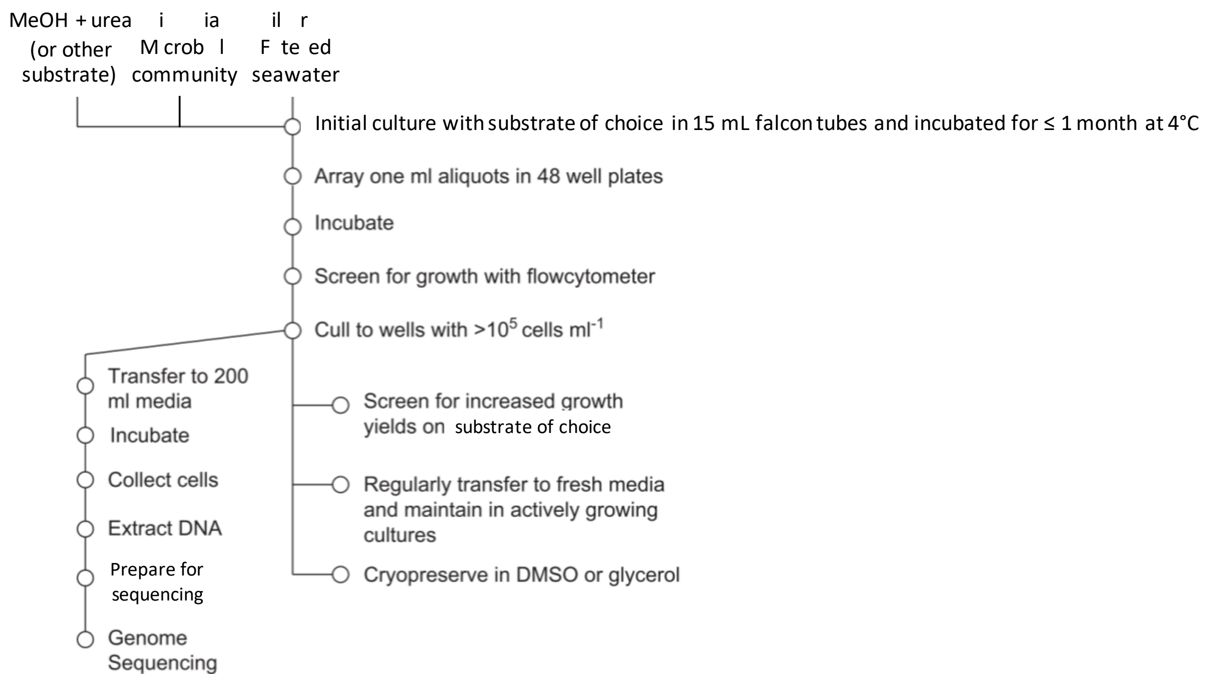


Figure B.1 *General workflow for isolating and sequencing methylotroph strains.*

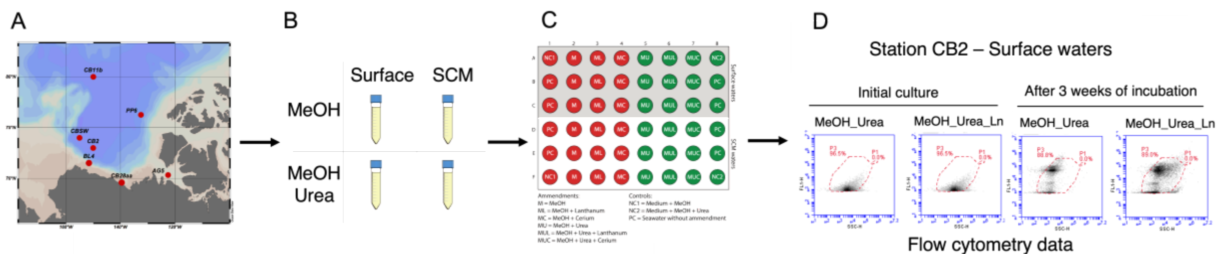


Figure B.2 *General workflow: (A) Sampling sites (B) Incubation of seawater from the surface waters and the subsurface chlorophyll maximum (SCM), supplemented with methanol or methanol*

and urea. (C) 1mL of supplemented seawater with the addition of a rare earth metal plated on 48-well plates and incubated at 4°C. (D) Growth of cultures is measured using flow cytometry.

B.3. Sequencing and Exploring the Metagenome Assembled Genomes (MAGs)

16S rRNA gene analysis of the culture after growth was performed and revealed that the cultures contain organisms that are most closely related to the Rhodobacterales (Sulfitobacter) and Rhodospirillales (Thalassospira) families (**Figure B.3** and **Figure B.4**).

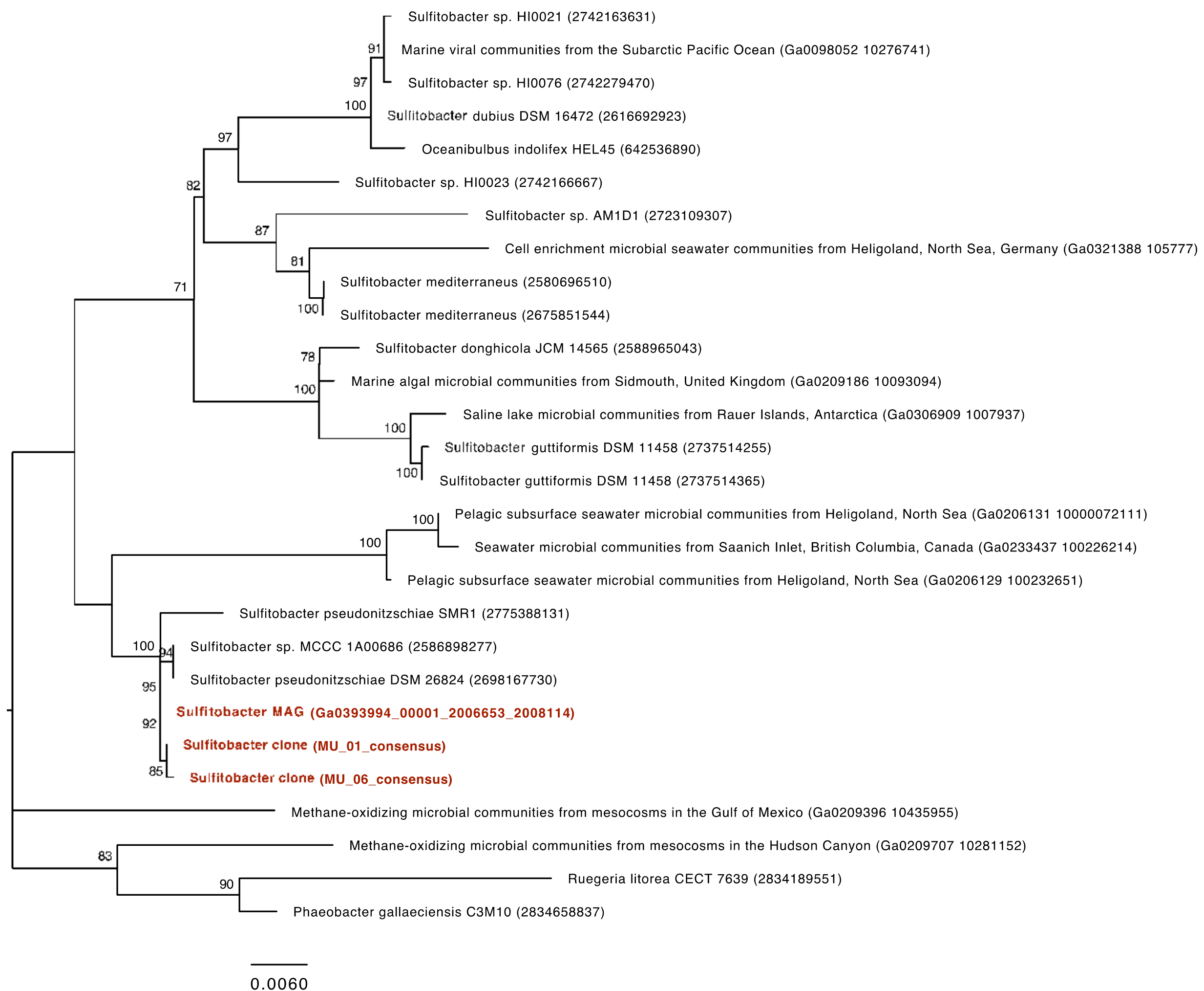


Figure B.3 *Phylogenetic analysis of 16S rRNA genes from the Sulfitobacter MAG from Canada Basin and a diversity of aquatic ecosystems. The tree was inferred using maximum likelihood (500 bootstraps) and GTR 1 gamma distribution (four categories) with invariant site model of evolution and the nearest-neighbor interchange heuristic search method. Sequences from the current study are highlighted in red. Only bootstrap values of .60 are included in the tree.*

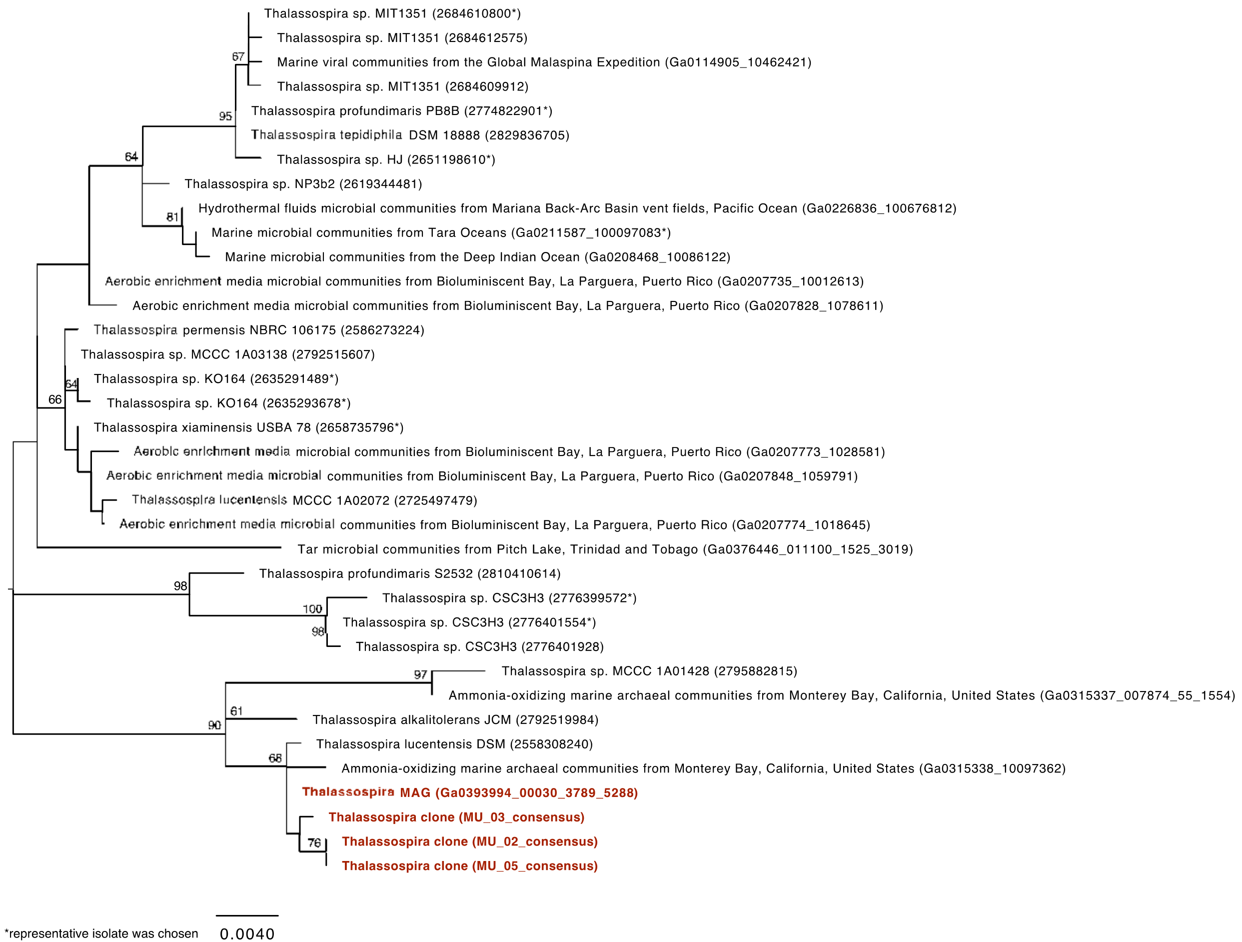


Figure B.4 *Phylogenetic analysis of 16S rRNA genes from the Thalassospira MAG from Canada Basin and a diversity of aquatic ecosystems. The tree was inferred using maximum likelihood (500 bootstraps) and GTR 1 gamma distribution (four categories) with invariant site model of evolution and the nearest-neighbor interchange heuristic search method. Sequences from the current study are highlighted in red. Only bootstrap values of .60 are included in the tree.*

Metagenomic sequencing led to the reconstruction of two Metagenome-Assembled Genomes (MAGs) (**Table B.1**). We were able to recover the complete genomes of two organisms, Sulfitobacter sp. and Thalossospira sp. A comprehensive review of these Sulfitobacter and Thalassospira revealed that neither of these organisms are known to use methanol as a carbon or energy source and have not been shown to possess methanol dehydrogenase genes. Therefore, it was very interesting that these organisms were enriched in the presence of methanol. A comprehensive analysis of the MAGs revealed that they do not have the xoxF4 methanol

dehydrogenase gene not any other PQQ-dependent methanol dehydrogenases. This opened the door to investigate other types of methanol oxidation pathways.

Table B.1 *Assembled and binned metagenome assembled genomes (MAGs) from Arctic Ocean cultures*

MAGs	Genome size (Mbp)	GC content (%)	Completeness (%)
Sulfitobacter sp.	3.66	62	100
Thalossospira sp.	4.62	53.5	100

B.3.1. NAD-dependent alcohol dehydrogenase: Potential for methanol oxidation?

Alcohol dehydrogenases (ADHs) can be grouped into three different categories: (1) nicotinamide adenine dinucleotide- (NAD-) or nicotinamide adenine dinucleotide phosphate- (NADP-) dependent ADHs, (2) pyrroloquinolinequinone-, haem-group and F_{420}-dependent enzymes and (3) flavin adenine dinucleotide-dependent isozymes. The two sequenced MAGs, Sulfitobacter and Thalassospira, both contain NAD-dependent alcohol dehydrogenases. Within this NAD-dependent ADH, there are three families, and the one I will be focusing on is the iron-containing one. Bacteria members of this family includes Bacillus methanolicus, which contains an NAD-dependent methanol dehydrogenase (EC 1.1.1.244).

In order to determine whether these two MAGs have the genes capable of metabolizing methanol, I used BLAST to pull out all sequences closely related to the methanol dehydrogenase gene in Bacillus methanolicus along with all the sequences that were annotated as "NAD-dependent alcohol dehydrogenase". we also pulled out several "iron-containing alcohol dehydrogenase" sequences from both Sulfitobacter strains and Thalassospira strains to ensure that we correctly identify the gene. After trimming, we found that it is likely that some of "NAD-dependent alcohol dehydrogenase" were actually "NAD-dependent methanol dehydrogenase" specifically (**Figure B.5**).

After a thorough literature search, we were unable to find any evidence for Thalassospira species using methanol or containing a methanol dehydrogenase gene. However, we were able to find one Sulfitobacter strain that contains the methanol dehydrogenase gene, however, no studies describing the use of this gene were found. Therefore, we would be showing that there are novel marine bacteria capable of using methanol as a carbon/energy source and specifically have the NAD-dependent methanol dehydrogenase gene rather than the more abundant PQQ-dependent methanol dehydrogenase genes.

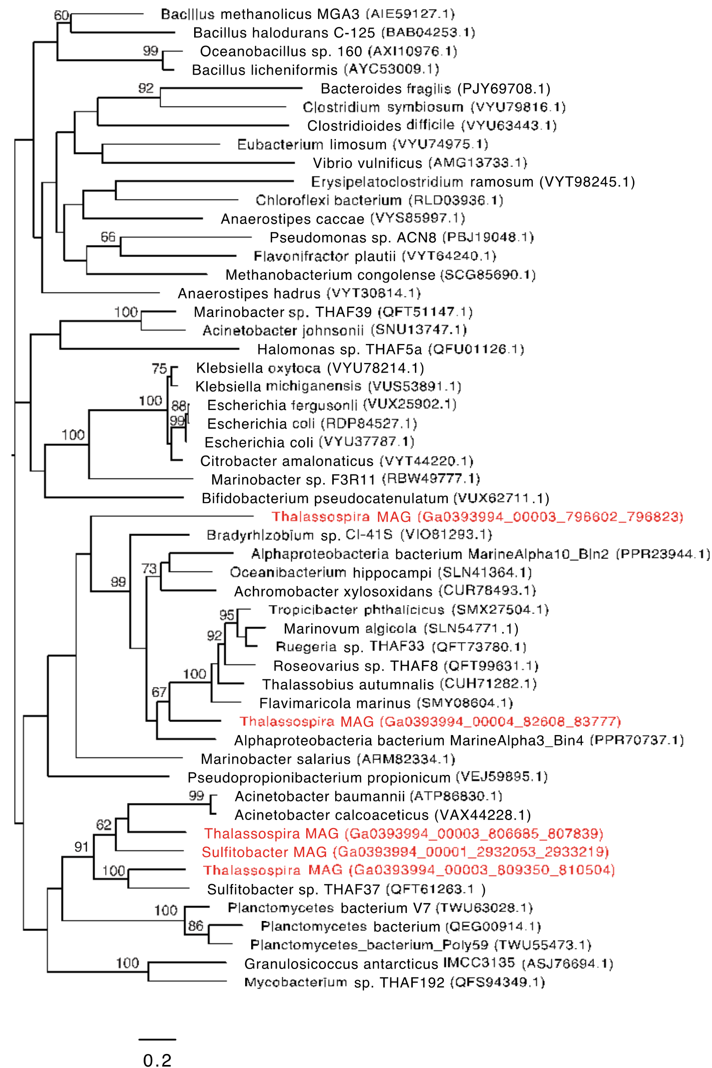


Figure B.5 *Phylogenetic analysis of the NAD-dependent methanol dehydrogenase protein recovered from the MAGs. The tree was inferred using maximum likelihood (500 bootstraps) and JTT 1 gamma distributed with invariants (four categories) sites model of evolution, with nearest-*

neighbor interchange heuristic search method. Colored sequences are those from this study. Only bootstrap values of .60 are included in the tree.

B.4. Conclusion and Future Studies

In this study, we endeavoured to enrich and culture methylotrophs from the Methylophilaceae family using media supplemented with methanol, urea, and rare-earth elements. While we were unable to isolate pure cultures of organisms from the OM34 and BS01 clades, we were still able to expand our understanding of marine methylotrophs. Interestingly, we enriched and characterized novel Alphaproteobacterial methylotrophs, which have not previously been implicated in methanol oxidation. Future studies should include a more comprehensive analysis of the MAGs to further understand the metabolic potential of these novel Alphaproteobacteria. We should also further explore and decipher the physiology of these cultured organisms by observing growth patterns on methylotrophic substrates, and organic and inorganic nitrogen sources. As we know, the surface waters of the Arctic Ocean are rapidly freshening leading to questions about how bacteria will adapt and evolve to this decrease in salinity. By conducting experimental evolution studies using a gradient of salinity, we can try to understand how these changes are affecting the physiology and genomic content of these organisms.