

**Life in the cold biosphere:**  
**The ecology of psychrophile communities, genomes, and genes**

Jeff Shovlowsky Bowman

A dissertation  
submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington

2014

Reading Committee:

Jody W. Deming, Chair

John A. Baross

Virginia E. Armbrust

Program Authorized to Offer Degree:

School of Oceanography

© Copyright 2014

Jeff Shovlowsky Bowman

## **Statement of Work**

This thesis includes previously published and submitted work (Chapters 2–4, Appendix 1). The concept for Chapter 3 and Appendix 1 came from a proposal by JWD to NSF PLR (0908724). The remaining chapters and appendices were conceived and designed by JSB. JSB performed the analysis and writing for all chapters with guidance and editing from JWD and co-authors as listed in the citation for each chapter (see individual chapters).

## Acknowledgements

First and foremost I would like to thank Jody Deming for her patience and guidance through the many ups and downs of this dissertation, and all the opportunities for fieldwork and collaboration. The members of my committee, Drs. John Baross, Ginger Armbrust, Bob Morris, Seelye Martin, Julian Sachs, and Dale Winebrenner provided valuable additional guidance. The fieldwork described in Chapters 2, 3, and 4, and Appendices 1 and 2 would not have been possible without the help of dedicated guides and support staff. In particular I would like to thank Nok Asker and Lewis Brower for giving me a sample of their vast knowledge of sea ice and the polar environment, and the crew of the icebreaker *Oden* for a safe and fascinating voyage to the North Pole.

None of this work would have been possible without the support of the NASA Astrobiology Institute, the University of Washington Astrobiology Program (via NSF IGERT), NSF Office of Polar Programs (now Division of Polar Programs), and the EPA STAR fellowship program. Programs and initiatives that directly supported individual chapters are listed in chapter acknowledgements.

I would also like to thank the many co-authors and collaborators who worked directly with me on these projects, facilitated the fieldwork, or provided critical insight at some point. In particular, Hans-Werner Jacobi, Kevin Hand, and Søren Rysgaard graciously provided opportunities to go to the field, enabling much of the work described here. Matthias Wietz, Nikolas Blom, Shelly Carpenter, Jesse Colangelo-Lillis and many others made that fieldwork a lot more enjoyable; thanks to all of you for the assistance, entertainment, and collaboration! Eric Collins, Britney Schmidt, and a host of colleagues, co-authors, and friends, including the whole



crew at the shaka house, provided, and continue to provide, critical insight and great discussion on this and related work.

Last I would like to thank Sarah Purkey, my colleague and partner, for all the ideas, discussion, and help over the last six years. I can't wait to see where life leads us next!

University of Washington

Abstract

**Life in the cold biosphere:**

**The ecology of psychrophile communities, genomes, and genes**

Jeff S. Bowman

Chair of Supervisory Committee:

Prof. Jody W. Deming

The prevalence of low temperature habitats on Earth makes the ecology of organisms adapted to low temperature environments (psychrophiles) an important area of research. Studies of low temperature ecosystems including the deep sea, sea ice, glacial ice, permafrost, and snow have provided a wealth of knowledge on the resilience of psychrophilic microbial ecosystems in the face of anthropogenic and natural disturbance, the history of microbial life on Earth, and the potential distribution of life in extraterrestrial environments. Taking these three knowledge areas as motivation, this dissertation further explores psychrophile ecology.

Chapter 1 introduces the history of research on psychrophiles, and the current state of knowledge regarding sea ice microbial communities, an important ecosystem dominated by psychrophiles. This chapter also explores one method for making the jump from descriptive ecological studies to process-based studies that have predictive power. Chapters 2 and 3 explore the diversity of Bacteria found in two understudied psychrophile habitats; multiyear sea ice and frost flowers. In the first quantitative analysis of sea ice community composition, Chapter 2 describes the multiyear sea ice microbial community as diverse, and in agreement with previous studies, dramatically different from the seawater microbial community. Chapter 3 describes an

unusual community of Bacteria in frost flowers, dominated by members of the order Rhizobiales. The metabolic plasticity of the Rhizobiales could have important implications for elemental cycling within young sea ice, particularly for nitrogen and sulfur species, an idea explored through an analysis of metabolic potential in Chapter 4. Chapter 5 describes a new method for evaluating genomic plasticity within any group of genomes, and applies this method to a comparative analysis of psychrophile and mesophile genomes, finding evidence for greater genome plasticity within psychrophile genomes. Application of a molecular clock to horizontal gene transfer events in these genomes suggests a link between cold periods in the Phanerozoic and the rate of retention of genes acquired by psychrophiles. Chapter 6 examines how psychrophilic enzymes are optimized for low temperatures through amino acid substitutions and introduces a model for further exploration of amino acid preferences. Experiments with this model suggest a role for serine, preferred in some psychrophile proteins, in adaptation to both low temperature and high salinity. Chapter 7 explores the potential for psychrophiles to degrade alkanes, a major component of crude oil, by the presence of genes coding for alkane hydroxylases. Several likely alkane hydroxylase genes in psychrophiles, not currently annotated as alkane hydroxylases, were identified. Analysis of the protein physical parameters coded by these genes suggests some optimization to low temperature. These findings have important implications for the in situ bioremediation of crude oil in cold environments, and for efforts to develop remediation technologies that function at low temperature.

## Table of Contents

LIST OF FIGURES.....	xvi
LIST OF TABLES.....	xviii

### **CHAPTER 1: AN INTRODUCTION TO SEA ICE MICROBIAL ECOLOGY.....1**

1.1 A BRIEF HISTORY OF RESEARCH ON SEA ICE MICROBIAL COMMUNITIES.....	2
1.2 THE PRESENT DAY UNDERSTANDING OF SEA ICE MICROBIAL ECOLOGY.....	7
1.3 THE FUTURE OF SEA ICE MICROBIAL ECOLOGY.....	9
REFERENCES.....	14
FIGURES.....	19

### **CHAPTER 2: MICROBIAL COMMUNITY STRUCTURE OF ARCTIC MULTIYEAR SEA ICE AND SURFACE SEAWATER BY 454 SEQUENCING OF THE 16S RNA GENE.....24**

*Bowman, J. S., Rasmussen, S., Blom, N., Deming, J. W., Rysgaard, S., & Sicheritz-Ponten, T. (2011). Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. The ISME journal, 6(1), 11–20.*

ABSTRACT.....	24
2.1 INTRODUCTION.....	25
2.2 METHODS.....	27
2.2.1 SAMPLE COLLECTION AND PREPARATION.....	27
2.2.2 COMMUNITY STRUCTURE.....	28
2.2.3 COMMUNITY COMPOSITION.....	29
2.2.4 COMMUNITY SIMILARITY.....	29
2.2.4 16S rRNA GENE DIVERGENCE.....	30
2.3 RESULTS.....	31
2.3.1 ENVIRONMENTAL.....	31

2.3.2 COMMUNITY STRUCTURE.....	32
2.3.3 COMMUNITY COMPOSITION.....	33
2.3.4 16S rRNA GENE DIVERGENCE.....	34
2.3.5 COMMUNITY SIMILARITY.....	35
2.4 DISCUSSION.....	36
2.5 SUPPLEMENTARY METHODS.....	39
2.5.1 SAMPLE COLLECTION.....	39
2.5.2 ENVIRONMENTAL PARAMETERS.....	40
2.5.3 DNA EXTRACTION, PURIFICATION, AND SEQUENCING.....	41
REFERENCES.....	43
TABLES AND FIGURES.....	48

**CHAPTER 3: SELECTIVE OCCURRENCE OF RHIZOBIALES IN FROST FLOWERS  
ON THE SURFACE OF YOUNG SEA ICE NEAR BARROW, ALASKA AND  
DISTRIBUTION IN THE POLAR MARINE RARE BIOSPHERE.....71**

*Bowman, J. S., Larose, C., Vogel, T. M., & Deming, J. W. (2013). Selective occurrence of Rhizobiales in frost flowers on the surface of young sea ice near Barrow, Alaska and distribution in the polar marine rare biosphere. Environmental microbiology reports, 5(4), 575–582.*

ABSTRACT.....	71
3.1 INTRODUCTION.....	72
3.2 RESULTS AND DISCUSSION.....	75
3.3 CONCLUSIONS.....	80
3.4 SUPPLEMENTARY METHODS.....	80
3.4.1 SAMPLE COLLECTION AND DNA EXTRACTION.....	80
3.4.2 CLONE LIBRARY.....	82
3.4.3 MICROARRAY ANALYSIS.....	83
3.4.4 T-RFLP.....	84
3.4.5 METAGENOMIC ANALYSIS.....	85
3.4.6 COMPARISON WITH OTHER DATA SETS.....	86

REFERENCES.....	88
TABLES AND FIGURES.....	94

## **CHAPTER 4: THE GENETIC POTENTIAL FOR KEY BIOGEOCHEMICAL PROCESSES IN ARCTIC FROST FLOWERS AND YOUNG SEA ICE REVEALED BY METAGENOMIC ANALYSIS.....101**

*Bowman, J. S., Berthiaume, C. T., Armbrust, E., & Deming, J. W. (2014). The genetic potential for key biogeochemical processes in Arctic frost flowers and young sea ice revealed by metagenomic analysis. FEMS Microbiology Ecology, 89: 376–387.*

ABSTRACT.....	101
4.1 INTRODUCTION.....	102
4.2 MATERIALS AND METHODS.....	103
4.2.1 SAMPLE COLLECTION AND SEQUENCING.....	103
4.2.2 SEQUENCE PROCESSING AND TAXONOMIC ANALYSIS.....	103
4.2.3 METABOLIC PROFILES.....	104
4.3 RESULTS.....	106
4.3.1 METABOLIC PROFILES.....	109
4.4 DISCUSSION.....	110
4.5 CONCLUSIONS.....	115
4.6 SUPPLEMENTAL INFORMATION.....	116
4.6.1 POTENTIAL TRANSPORT MECHANISM METHODS.....	116
4.6.2 POTENTIAL TRANSPORT MECHANISM RESULTS.....	118
4.6.3 DISCUSSION OF POTENTIAL TRANSPORT MECHANISMS.....	119
REFERENCES.....	123
TABLES AND FIGURES.....	128

## **CHAPTER 5: EVIDENCE FOR ENHANCED HORIZONTAL GENE TRANSFER IN PSYCHROPHILE GENOMES AND POSSIBLE LINKS TO PHANEROZOIC CLIMATE.....138**

ABSTRACT.....	138
5.1 INTRODUCTION.....	139
5.2 RESULTS.....	141
5.2.1 GENOME DIVERGENCE.....	141
5.2.2 CDS WITH ANOMALOUS GC CONTENT.....	143
5.3 DISCUSSION.....	145
5.4 MATERIALS AND METHODS.....	150
5.4.1 GENOME SELECTION AND PROTEOME PREDICTION.....	150
5.4.2 GENOME DIVERGENCE.....	151
5.4.3 CDS WITH ANOMALOUS GC CONTENT.....	152
REFERENCES.....	156
TABLES AND FIGURES.....	161

**CHAPTER 6: AMINO ACID PREFERENCE IN THE PROTEOMES OF  
PSYCHROPHILIC BACTERIA: IS SERINE THE ANSWER TO ENZYME  
ADAPTATION TO LOW TEMPERATURE AND HIGH SALINITY?.....177**

ABSTRACT.....	177
6.1 INTRODUCTION.....	179
6.2 METHODS.....	181
6.2.1. GENOME SELECTION AND ANNOTATION.....	174
6.2.2 SECONDARY STRUCTURE PREDICTION.....	182
6.2.3 ANALYSIS OF PROTEIN PARAMETERS.....	183
6.2.4 PEPC.....	184
6.3 RESULTS.....	185
6.3.1 METPALLY AND REDDY DATASET.....	185
6.3.2 EXPANDED DATASET.....	186
6.3.3 PEPC.....	188
6.4 DISCUSSION.....	189
REFERENCES.....	193

TABLES AND FIGURES.....	197
-------------------------	-----

## **CHAPTER 7: ALKANE HYDROXYLASE GENES IN PSYCHROPHILE GENOMES AND THE POTENTIAL FOR COLD ACTIVE CATALYSIS.....207**

ABSTRACT.....	207
7.1 INTRODUCTION.....	208
7.2 METHODS.....	211
7.2.1 IDENTIFYING ALKANE HYDROXYLASES.....	211
7.2.2 ANALYSIS OF PROTEIN PARAMETERS.....	213
7.3 RESULTS.....	214
7.4 DISCUSSION.....	217
REFERENCES.....	222
TABLES AND FIGURES.....	226

## **APPENDIX 1: ELEVATED BACTERIAL ABUNDANCE AND EXOPOLYMERS IN SALINE FROST FLOWERS AND IMPLICATIONS FOR ATMOSPHERIC CHEMISTRY AND MICROBIAL DISPERSAL.....233**

*Bowman, J. S., & Deming, J. W. (2010). Elevated bacterial abundance and exopolymers in saline frost flowers and implications for atmospheric chemistry and microbial dispersal. Geophys. Res. Lett., 37(13), L13501.*

ABSTRACT.....	233
A1.1 INTRODUCTION.....	234
A1.2 METHODS.....	235
A1.3 RESULTS.....	237
A1.4 DISCUSSION.....	238
A1.5 CONCLUSION.....	242
REFERENCES.....	243
FIGURES.....	246



**APPENDIX 2: BACTERIAL PHYLOTYPES SHARED BETWEEN THE YOUNG SEA  
ICE AND SUPRAGLACIAL ENVIRONMENTS.....248**

ABSTRACT.....	248
A2.1 INTRODUCTION.....	249
A2.2 METHODS.....	250
A2.2.1 SAMPLE COLLECTION AND PROCESSING.....	250
A2.2.2 DATA ANALYSIS.....	251
A2.3 RESULTS.....	253
A2.3.1 ENVIRONMENTAL.....	253
A2.3.2 COMMUNITY COMPOSITION.....	254
A2.4 DISCUSSION.....	256
REFERENCES.....	260

**APPENDIX 3: THE CRYOSPHERE FROST FLOWER REACTOR FOR ORGANIC  
GEOCHEMISTRY (CRYO-FROG).....266**

A3.1 INTRODUCTION.....	266
A3.2 METHODS.....	266
A3.3 RESULTS AND DISCUSSION.....	267
REFERENCES.....	270

**APPENDIX 4: CURRICULUM VITAE.....297**

## List of Figures

Figure 1.1. The theoretical distribution of habitable space.....	19
Figure 1.2. Occurrence in the peer-reviewed literature of the terms “psychrophile” and “psychrophile+astrobiology”.....	20
Figure 1.3. Factor-fold concentration of solutes in water ice relative to starting solute concentration.....	21
Figure 1.4. Bacterial abundance and chlorophyll a from land-fast first-year sea ice during the Austral spring of 2011, McMurdo Sound, Antarctica.....	22
Figure 1.5. The Taxonomy-Metabolic Potential-Metabolism-Biogeochemistry (TMMB) model framework.....	23
Figure 2.1. Distribution of genetic distances for MYI and seawater communities.....	51
Figure 2.2. Numerically dominant clades in the MYI and seawater communities.....	52
Figure 2.3. Divergence of the 16S rRNA gene using a taxonomy-based method.....	53
Figure 2.4. Divergence of the 16S rRNA gene using a taxonomy-independent analysis.....	54
Figure S2.1. Sampling locations in the Arctic Ocean.....	65
Figure S2.2. Correlation between Simpson’s index of diversity and the non-parametric implementation of Shannon’s index.....	66
Figure S2.3. Fraction of reads unclassified at taxonomic level.....	67
Figure S2.4. Primer set exclusions.....	68
Figure S2.5. Ice core profiles.....	69
Figure S2.6. Rarefaction curves.....	70
Figure 3.1. Frost flowers and young sea ice at the study site.....	95
Figure 3.2. Relative fluorescence of genera evaluated by microarray.....	96
Figure 3.3. Placement of Barrow frost flower (FF) and young ice (YI) 16S rRNA gene clones on Rhizobiales reference tree.....	97
Figure 3.4. Placement of 16S rRNA gene reads from available deep sequence libraries (Table 3.1).....	98
Figure S3.1. Heatmap of T-RFLP fragment relative abundance.....	100
Figure 4.1. Coverage of chromosomes and plasmids in the frost flower metagenome.....	132

Figure 4.2. Coverage for chromosomes and plasmids for strains with well covered plasmids.....	133
Figure 4.3. Coverage across the chromosome of <i>Candidatus Pelagibacter ubique</i> (top) and <i>Sinorhizobium fredii</i> HH103 (bottom).....	134
Figure 4.4. Phylogenetic placement of betaine methyltransferase read translations from the frost flower metagenome on a reference tree of UniProt sequences .....	135
Figure 4.5. Phylogenetic placement of NodG and FabG read translations from the frost flower metagenome on a reference tree of UniProt sequences .....	136
Figure S4.1. Wind magnitude at the Barrow NOAA observatory for April 14 – 21, 2010.....	137
Figure 5.1. Proteome vs. 16S rRNA gene divergence.....	162
Figure 5.2. Measures of divergence between groups.....	163
Figure 5.3. Div <sub>G</sub> values between strains in the psychrophile and mesophile groups.....	164
Figure 5.4. Frequency of HGT events across geologic time for the psychrophile and mesophile groups.....	165
Figure S5.1. Sensitivity of div <sub>G</sub> to gene acquisition.....	169
Figure S5.2. Normalized compositional vector distance between strains in the psychrophile and mesophile groups.....	170
Figure S5.3. 16S rRNA gene distance between strains in the psychrophile and mesophile groups.....	171
Figure S5.4. Synteny plot of <i>Octadecabacter antarcticus</i> 307 and <i>Octadecabacter arcticus</i> 238.....	172
Figure S5.5. GC anomalies of coding sequences in the psychrophile genomes.....	173
Figure S5.6. GC anomalies of coding sequences in the mesophile genomes.....	174
Figure S5.7. Phylogenetic tree of strains used in this study.....	175
Figure S5.8. Change of HGT time since acquisition probability window size with time.....	176
Figure 6.1. Maximum-likelihood tree of 16S rRNA genes from Bacteria and Archaea in the expanded dataset.....	199
Figure 6.2. Correlation between parameter means (amino acid composition) in Metpally and Reddy (Metpally & Reddy, 2009) and in our reanalysis.....	200
Figure 6.3. Flexibility vs. isoelectric point for all proteins in the expanded dataset.....	201

Figure 6.4. Flexibility vs. isoelectric point for proteins in the high flexibility/low isoelectric point quadrant (Fig. 6.3).....	202
Figure 6.5. Mutation collector's curves generated with the PEPC model.....	203
Figure 6.6. Location in parameters space defined by flexibility and isoelectric point during PEPC model runs.....	204
Figure 6.7. Mutations preferred in each PEPC model run.....	205
Figure 7.1. Euclidean distance in 2S NMDS space as a function of bit score and phylogenetic distance.....	229
Figure 7.2. NDMS plots of genetic distance within four protein families (pfams).....	230
Figure 7.3. Alignment of the flexibility parameter between putative alkane hydroxylases in psychrophiles and mesophiles.....	231
Figure 7.4. Predicted 3D structure for psychrophile representatives of the four clusters with high confidence predictions.....	232
Figure A1.1. Correlations between bacterial abundance and salinity (of melted samples) for laboratory grown (A) and naturally occurring FF from Barrow (B) and the central Arctic Ocean (C).....	246
Figure A1.2. Concentrations of pEPS, brine skim, and sea ice at Barrow, AK.....	247
Figure A2.1. Sample locations near McMurdo Sound, Antarctica.....	264
Figure A2.2. Abundance of most dominant OTUs across samples, subsampled to a depth of 3000 contigs .....	265
Figure A3.1. CRYO-FROG schematic.....	272
Figure A3.2. Frost flower growth conditions.....	273

## List of Tables

Table 2.1. Means values of richness and diversity for multiple OTU definitions for the MYI and seawater communities. ....	48
Table 2.2. Estimates of richness and diversity for OTU definition of 100 % similarity for each MYI and seawater sample.....	49
Table 2.3. Pairwise comparison of sample similarity for OTU definition of 100 % similarity...	50
Table S2.1. Locations and environmental parameters for the MYI and seawater samples.....	55
Table S2.2. Pairwise comparison of sample similarity for 97, 93, and 90 % similarity definitions of OTU using richness and abundance-based Sørensen indices .....	56
Table S2.3. Classification of reads at each taxonomic level for the MYI community (I13 and I14 combined).....	57
Table S2.4. Classification of reads at each taxonomic level for the seawater community (S15, S16, and S17 combined).....	60
Table S2.5. Quantities from DNA collection, isolation, and processing for each MYI and seawater sample.....	64
Table 3.1. Datasets used to evaluate the occurrence of marine Rhizobiales in polar regions ....	94
Table 4.1. Number of reads, coverage, and normalized coverage by protein groups (left column).....	128
Table 4.2. Most abundant orders according to classification of 16S rRNA gene reads .....	129
Table 4.3. Blastn results for assembled contigs longer than 10,000 bp, searched against nt as single queries .....	130
Table 4.4. Putative peptides placed on edges corresponding to proteins from the Rhizobiales.....	131
Table 5.1. Strains used in this study and key values.....	161
Table S5.1. Isolation environment, $\bar{x}div_G$ , and the number of anomalous CDS for each evaluated strain.....	167
Table S5.2. Occurrence of Pfams as likely HGT events in the psychrophile and mesophile groups.....	168
Table 6.1. Psychrophile amino acid preferences as reported by Metpally and Reddy (2009), the reanalysis, and the expanded dataset.....	197

Table 6.2. Psychrophile protein parameter preferences as reported by Metpally and Reddy (2009), the reanalysis, and the expanded dataset.....	198
Table S6.1. Psychrophile protein parameter preferences for the Metpally and Reddy dataset determined using the by-strain method as reported by Metpally and Reddy (bulk protein only) and as determined in this analysis (by secondary structure element and by bulk protein).....	see suppl. file table_s6.1.tsv
Table S6.2. Psychrophile protein parameter preferences for the Metpally and Reddy dataset determined using the by-class method as determined in this analysis (by secondary structure element and by bulk protein).....	see suppl. file table_s6.2.tsv
Table S6.3. Psychrophile protein parameter preferences for the extended dataset using the by-strain method .....	see suppl. file table_s6.3.tsv
Table S6.4. Psychrophile and mesophile protein parameter preferences for the extended dataset using the by-class method .....	see suppl. file table_s6.4.tsv
Table 7.1. Occurrence in each dataset of conserved protein family (pfam) domains linked to alkane hydroxylases (AH).....	226
Table 7.2. Number of candidate alkane hydroxylases observed in each of the psychrophile and mesophile genomes examined .....	227
Table 7.3. Pairwise parameters for candidate alkane hydroxylase with the p450 and Pyr_redox_3 conserved domains. Region refers to: whole; entire protein, C; coil, E; $\beta$ -sheet, H; $\alpha$ -helix ...	228
Table S7.1. Genbank accession and annotation for the protein records corresponding to the gene products evaluated in this study .....	see suppl. file table_s7.1.pdf

## Chapter 1

### **An introduction to sea ice microbial ecology**

Our universe is a cold place, with a background temperature of only 2.73 degrees above absolute zero (de Bernardis et al., 2000). Yet scattered throughout this universe are pinpoints of warmth, the result of nuclear fusion, as in the case of our sun, and gravitational interactions, as in the case of the moon Europa. It is much too hot for life to exist near the energetic centers of these pinpoints, but around each is a thin veneer of conditions that enable liquid water, chemical interactions, and reaction rates familiar to those that permit life on Earth. If we consider each of these pinpoints as a sphere, we find that the habitable space around each increases exponentially with a linear decrease in the temperature minima of life (Fig. 1). Thus these spheres succinctly demonstrate that the lower the temperature that life can tolerate, the more space there is for habitation. With this in mind, the function of life at low temperature becomes a central question to the ecology of life in our universe and on Earth.

In seven chapters this dissertation explores the adaptation of microbial life on Earth to low temperature with a particular emphasis on sea ice environments. Chapter 1 reviews the current state of our understanding of sea ice microbial ecology. Chapter 2 compares the microbial community structure of Arctic multiyear sea ice to that of surface seawater — two proximate but ecologically disparate low temperature environments. Chapter 3 explores the microbial community in frost flowers, highly saline structures on the surface of newly formed sea ice, and the underlying newly formed or young sea ice. Chapter 4 uses metagenomics to probe the functional diversity within this same environment. Chapter 5 evaluates the role that horizontal gene transfer may play in microbial adaptation to low temperature environments

through quantification of genomic plasticity in psychrophile genomes and taxonomically related mesophile genomes. This chapter also explores the rate of horizontal gene acquisition and retention among psychrophiles and mesophiles throughout the Phanerozoic Eon. Chapter 6 explores the adaptation of putatively cold active enzymes to low temperature and describes a simple model, the Protein Evolution Parameter Calculator (PEPC), to evaluate the difficulty of producing enzymes optimized to multiple environmental challenges. Chapter 7 explores the distribution of genes coding for alkane hydroxylase enzymes in psychrophile genomes and the degree to which these enzymes might be optimized to low temperatures. Three appendices also accompany this dissertation, each addressing a different aspect of the young sea ice surface as a low temperature environment. Appendix 1 reports a strong correlation between bacterial abundance and salt in frost flowers and other components of the young sea ice environment, and discusses the potential implications for microbial transport and chemical reactions at the sea ice surface. Appendix 2 explores the potential microbial exchange between two geographically separate ice environments linked by aerial transport: the supraglacial environment and the young sea ice environment. Appendix 3 describes the Cryosphere Frost Flower Reactor for Organic Geochemistry (CRYO-FROG), an experimental apparatus for exploring photochemical reactions in the surface ice environment

## **1.1 A brief history of research on sea ice microbial communities**

Cold-adapted microbes first appeared in the scientific literature as early as 1887, when Forster isolated bioluminescent bacteria capable of growth at 0 °C from cold-stored flounder (Forster, 1887). Likewise, the scientific exploration of sea ice microbial communities dates back to at least that same year, when Nansen described diatoms that adhered to the bottom of Arctic sea ice (see reference in Nansen, 1906). McLlan extended these observations during the



Australian Antarctic Expedition between 1911 and 1914, describing “practically the whole of low life which exists” in Antarctic sea ice— including protists, rotifers, and bacteria — and in fact, was able to culture bacteria from preparations of sea ice algae (McLlan, 1918). Despite these tantalizing early observations, the study of life in low temperature environments proceeded slowly until after World War II, when increased exploration of the Arctic and Antarctic provided new opportunities to study cold microbial habitats. Prior to 1960, such work was motivated primarily by the role bacteria played in food spoilage and focused on a relatively narrow range of isolates (e.g. *Pseudomonas* spp.). Despite these limitations, important advances in our overall understanding of psychrophiles were made during this period, including adoption of the reaction rate terminologies  $Q_{10}$  and  $\mu$  to describe the temperature dependence of enzymes (Ingraham & Bailey, 1959) and an overall appreciation that psychrophilic microbes are ecologically distinct from mesophiles (Ingraham, 1958).

The International Geophysical Year (IGY) of 1957-1958 heralded an era of increased scientific activity in the Arctic and Antarctic. While we cannot link any microbiological studies directly to IGY, it serves as a useful boundary between studies that preceded it, which were primarily limited to laboratory work, and those that followed it, which would be augmented by environmental observations. The first focused study of sea ice bacteria proceeded from the fifth Japanese Research Expedition of 1961. Researchers obtained several isolates from surface seawater and from melted sea ice samples stored at  $-5^{\circ}\text{C}$  for several months (Iizuka, Tanabe, & Meguro, 1966). Isolation of these samples, however, took place at  $25^{\circ}\text{C}$ , a temperature that is deleterious to most sea ice bacteria; thus the dominant members of the community were missed. The authors, however, did note a co-occurrence of bacteria and ice algae in the sampled “plankton ice” and speculated on possible ecological interactions, a recurring theme in sea ice

microbial ecology (e.g. Sullivan & Palmisano, 1984, Feng *et al.*, 2014). While the isolation and characterization of marine psychrophiles from seawater continued throughout the 1960s (e.g. Colwell *et al.*, 1964), studies of sea ice microbial ecology focused almost exclusively on ice algae. The most visible component of the sea ice microbial ecosystem, ice algae were observed to reach densities in excess of  $670 \mu\text{g}$  chlorophyll *a*  $\text{L}^{-1}$  in spring and summer (H Meguro, 1962). Descriptive work of this element of the ecosystem (Bunt, 1963b; Iizuka *et al.*, 1966) rapidly gave way to more quantitative analyses (Bunt, 1963a; Apollonio, 1965; Burkholder & Mandelli, 1965; Hiroshi Meguro, Ito, & Fukushima, 1967; Horner & Alexander, 1972) regarding chlorophyll concentrations, primary production, and the specific challenges of life in ice. These studies clarified that ice algae are wide-spread, important to the polar carbon cycle, and uniquely adapted to the sea ice environment.

During the 1970s and early 1980s, the works of Pomeroy (1974), Azam *et al.* (1983), and others brought forward the role that heterotrophic bacteria play in the marine carbon cycle. The “microbial loop,” wherein dissolved organic carbon (DOC) is recycled via bacterial assimilation and predation by protozoa, became recognized as an important component of the marine food web. Sullivan and a host of co-authors transferred this concept to the sea ice ecosystem in a pivotal series of papers in the 1980s (Grossi, Kottmeier, & Sullivan, 1984; Sullivan & Palmisano, 1984; Kottmeier, Grossi, & Sullivan, 1987; Kottmeier & Sullivan, 1988). Their work confirmed that sea ice bacteria are not only abundant and active within sea ice but also closely coupled to the occurrence of ice algae. These observations, all on mature land-fast ice within McMurdo Sound, Antarctica, were extended to more variable ice types by Helmke and Weyland (1994) and Grossmann and Diekmann (1994). Working on newly formed pelagic sea ice, Grossmann and Diekmann (1994) observed significant bacterial growth rates even in relatively

oligotrophic sea ice, as well as bacterial production rates far in excess of those observed in seawater. Helmke and Weyland (1994) extended these observations to pelagic winter sea ice via the observation of high rates of activity and bacterial biomass relative to the underlying water column; at times, the ATP concentration, a measure of metabolic activity, in a single meter of sea ice exceeded the 100 m depth-integrated value for the underlying water column. These high levels of activity, however, were limited to the bottom-most warmest zone of sea ice.

By the mid-1990s, it was clear that sea ice bacterial communities were composed of physiologically unique psychrophilic bacteria, capable of survival under conditions of severe environmental stress, and had developed a fast response to new inputs of carbon. The taxonomic and functional diversity of this community, however, was almost entirely unknown, with the exception of the phenotypic and morphology-based classifications of a few isolates for the former (Iizuka et al., 1966) and the limited observations of extracellular enzyme activity for the latter (Helmke & Weyland, 1994). Concurrent with the growing appreciation of sea ice bacteria as a unique and potentially important component of the polar marine ecosystem came major advances in understanding taxonomic diversity within microbial communities. In a groundbreaking 1977 paper, Woese and Fox used 16S and 18S rRNA gene sequences to classify life into three broad domains (Woese & Fox, 1977). Improvements in sequence technology nearly a decade later (Smith et al., 1986) opened the door for more wide-spread sequencing of 16S rRNA genes from environmental samples and a rapid shift in the existing paradigms of microbial diversity (e.g. Giovannoni, *et al.*, 1990; Ward, *et al.*, 1990). These methods were further utilized to identify isolates from sea ice (J. P. Bowman, McCammon, Brown, Nichols, & McMeekin, 1997; John P Bowman et al., 1998) and, in a novel application of the technique, to identify sequences from an environmental clone library (Brown & Bowman, 2001). These and

later studies established that while most genera observed in sea ice have members common to other environments, there are specific strains associated with sea ice.

Studies during this time also introduced astrobiology as a new motivation for studying sea ice microbial communities (Deming & Huston, 2000) (Fig. 2). Broadly defined as the study of life in a universal context, astrobiology could be considered the purest form of ecology. A common approach is to study Earth environments that are analogous to potential extraterrestrial habitats. Since, as discussed previously, the universe is a cold place, the study of sea ice, glacial ice, and permafrost are central, but certainly not exclusive of, this approach.

Reports of the first extra-solar planet orbiting a main sequence star in 1995 (Mayor & Queloz, 1995) and the thousands of candidate and confirmed extrasolar planets since then have catalyzed the research on Earth analogues, as has the hotly disputed putative microbial fossils in the Martian meteor ALH84001 in 1996 (McKay et al., 1996) and the now widely accepted liquid-water ocean beneath Europa's icy exterior in 1998 (Carr et al., 1998).

Although the studies of the 1990s began to elucidate the composition of the sea ice microbial community, they did not address its functional role. In some cases, function could be inferred from specific experiments; e.g. Gerdes et al. (2005) and Brakstad et al. (2008) used diesel and crude oil perturbation experiments to explore the ability of the sea ice microbial community to respond to these carbon sources. Analyses of specific functional genes within sea ice, however, have been surprisingly limited, even though they are the most high-throughput measure of community metabolic capability. Koh et al. identified proteorhodopsin genes (Koh et al., 2010) and genes for anoxygenic photosynthesis (Koh, Phua, & Ryan, 2011) within Antarctic sea ice, which suggests that bacterial energy acquisition in sea ice is not limited to the oxidation of ice algal photosynthates, and Møller et al. (2013) found mercury resistance genes in Arctic sea

ice brines. These insights into sea ice microbial community function have been extended by a small but growing number of sequenced genomes from sea ice isolates. Although inferences of function from genomes are restricted to the geography and ecology of the original isolate, commonalities between isolates can provide a broader picture of adaptation to the sea ice environment.

*Colwellia psychrerythraea* 34H was the first bacterium represented in the sea ice environment to have its genome sequenced (Méthe et al., 2005). Although this strain was isolated from sediment, 16S rRNA gene sequences associated with the *C. psychrerythraea* phylotype have been observed repeatedly in sea ice (J. P. Bowman et al., 1997; John P Bowman et al., 1998; K. Junge, Imhoff, Staley, & Deming, 2002), even if the phylotype does not appear to be ubiquitous in the sea ice environment. Since the genome of *C. psychrerythraea* 34H was published, genomes have been sequenced from the sea ice bacteria *Glaciecola psychrophila* 170 (Yin et al., 2013), *Octadecabacter arcticus* 238 and *Octadecabacter antarcticus* 307 (Vollmers et al., 2013), and *Psychroflexus torquis* ATCC 700755 (Feng, Powell, Wilson, & Bowman, 2014), among others. Considering that in July of 2014 there were nearly 12,000 completed and draft genomes in Genbank, it is clear that the genetic diversity of sea ice has been under sampled compared to most other environments. The tiny glimpse into the function of the sea ice microbial community afforded by these genome sequences suggests not only that the community may be exceptionally genetically and physiologically plastic, but also highly adapted to the environmental constraints imposed by sea ice.

## **1.2 The present-day understanding of sea ice microbial ecology**

Like many microbial habitats, ice is a porous medium. It consists of a solid phase composed of ice crystals and a liquid phase composed of water and solutes excluded from the

crystals during growth. Aside from temperature alone, water ice is distinct from other porous media due to its dynamic nature over a temperature range that is relevant to microbial life.

While tap water begins to freeze at 0 °C, seawater, with a salinity of roughly 35 ppt, begins to freeze at -1.8 °C and does not complete the process until at least -36 °C (G. Marion, Farren, & Komrowski, 1999). The higher the salinity of the starting solution, the lower the temperature required to initiate freezing. Once freezing is initiated for a solution with the brine composition of seawater, however, the salinity of the interstitial brines is almost entirely a function of temperature, although organic matter content and other factors do have some affect. At -5 °C the brine salinity of sea ice is approximately 87 ppt; at -20 °C it will have reached 209 ppt.

The concentration of solutes by freezing is most evident with salt but applies to other components of seawater as well. The lower the salinity and the colder the temperature of the starting solution, the more concentrated the solutes are in the brine phase relative to their starting concentrations. This concentration effect extends even to bacteria and viral particles as demonstrated by Junge et al. (Karen Junge, Krembs, Deming, Stierle, & Eicken, 2001) and Wells and Deming (Wells & Deming, 2006). For any solute, the degree to which it is concentrated in ice relative to its concentration in the source material is a function of temperature and the bulk salinity of the ice, the total quantity of salt contained in a volume of melted ice (Cox & Weeks, 1983). Thus the degree of concentration varies widely between ice types (Fig. 3).

Because primary production is most concentrated at the sea ice/seawater interface and because DOC, as with other solutes, is concentrated in sea ice brines (Giannelli et al., 2001), bacterial sea ice specialists are optimized to high concentrations of organic carbon. This trait may aid growth at the coldest temperatures and allow enzymes to maintain uptake rates sufficient to support growth at temperatures well below their optimum if substrate concentrations are high

(Nedwell, 1999; Pomeroy & Wiebe, 2001). Likewise, this discovery has aided the laboratory study of sea ice bacteria, as sea ice specialists tend to also be specialists for common organic-rich culture conditions (K. Junge et al., 2002).

Observations of sea ice microbial community structure during the winter suggest that, while metabolic activity is present (Karen Junge, Eicken, & Deming, 2004), bacterial growth is extremely limited. Observing Arctic sea ice throughout the winter, Collins *et al.* (2010) reported little change to microbial community structure, while that of the underlying seawater changed considerably in the same time period. The authors hypothesized that exopolymers, produced by algae and bacteria and known to act as cryoprotectants (Krembs, Eicken, Junge, & Deming, 2002), may enable the survival of even non-ice associated genera within sea ice. The overwintering community observed by Collins et al. (2010) was similar to a typical seawater community and distinct from the community observed in late-spring and summer sea ice. To date, the transition from winter to summer has not been observed with molecular methods, a surprising deficiency given the relevance of this transition to the polar marine ecosystem and carbon cycle. By tracking bacterial abundance and chlorophyll *a* concentrations, the sea ice microbial community can be seen rapidly responding to the initiation of the spring algal bloom, though the response may be slower than that observed for seawater (Fig. 4). This response is presumed to reflect the rapid growth of the psychrophilic sea ice microbial community as soon as DOC concentrations are sufficient to overcome the temperature inhibition of enzymes.

### **1.3 The future of sea ice microbial ecology**

The over-arching research objectives in microbial ecology follow the classic set of questions: who, what, where, when, why, and how. When these questions are known in sufficient detail, it is possible to make predictions about the ecosystem in question. A goal of sea

ice microbial ecology, for example, is to predict how the microbial ecosystem will respond to environmental perturbations, including slow perturbations, like changing climate, or immediate perturbations, such as a release of crude oil. The latter issue has particular significance in the Arctic, where exploration and extraction continues on significant marine petroleum reserves (Gautier et al., 2009). The Macondo Well disaster in the Gulf of Mexico demonstrated that indigenous deep sea bacteria can degrade a considerable quantity of crude oil despite high pressure and low temperatures (Redmond & Valentine, 2012). Interestingly, one of the dominant groups of bacteria observed responding to this input of crude oil was *Colwellia*, a genus known to sea ice as discussed previously. *Colwellia psychrerythraea* 34H is one of the few sequenced psychrophiles that is commonly used in laboratory studies of cold adaptation. Neither *C. psychrerythraea* 34H nor a close relative sequenced from water in the vicinity of the Macondo wellhead (Mason, Han, Woyke, & Jansson, 2014), however, have recognizable genes for catabolizing low molecular weight alkanes, even though the latter organism is implicated in that process. One explanation is that this bacterium is using a gene without close homology to known alkane degradation genes, an exciting and realistic possibility; alternatively it may be responding to secondary metabolites produced by *Oceanospirales*, the primary responding clade when oil was released into these deep waters.

Either of these scenarios has important implications for the bioremediation of oil that may be released in the proximity of sea ice. Because many sea ice bacteria are optimized to high carbon concentrations, they may be resistant to crude oil toxicity and better prepared to catabolize hydrocarbons and achieve a more rapid bioremediation. This idea is supported by the work of Gerdes et al. (Gerdes et al., 2005) and Brakstad et al. (Brakstad et al., 2008), who observed that the indigenous sea ice microbial community is capable of crude oil degradation.



The rate of crude oil catabolism under the environmental conditions imposed by sea ice, however, remains an unknown. At low temperatures, many crude oil components have reduced bioavailability (Colwell, Walker, & Cooney, 1977) and, at low temperature, the rate of bacterial production is generally below the rate of primary production (Kirchman, Moran, & Ducklow, 2009). This discrepancy indicates the limit to which increased substrate concentration can compensate for reduced substrate affinity. While sea ice bacteria produce enzymes optimized for low temperatures (Adrienne L. Huston, Krieger-Brockett, & Deming, 2000; Adrienne L Huston, Methé, & Deming, 2004; Méthe et al., 2005), this optimization may be insufficient to keep pace with either carbon fixation or a rapid input of crude oil. Predicting the fate of crude oil, or any other perturbation to the sea ice ecosystem, will require a much more complete understanding of microbial functional diversity and physiology, and their combined impact on biogeochemistry.

One pathway to develop a predictive biogeochemical model is the metabolic flux model, an idea that was explored for crude oil degradation in a recent review by Röling and Bodegom (Röling & van Bodegom, 2014). This model makes use of community metabolic potential and gene expression data, information derived from environmental and isolate sequencing experiments, to predict the flow of energy and material between the biotic and abiotic components of the biosphere as well as between members of the microbial community. Coupled to a traditional biogeochemical model, the metabolic model becomes predictive at the ecosystem level; as the flow of energy, carbon, or nutrients into the system change, or as members of the community change in presence or abundance, the impact on biogeochemistry can be quantified. This idea is conceptualized in what I term the Taxonomy-Metabolic potential-Metabolism-Biogeochemistry (TMMB) model framework (Fig. 5). As in any model, however, the predictive value of the TMMB framework depends on the level of detail built into the model itself.

Despite decades of research on sea ice microbial ecology, our grasp of the details of sea ice ecosystem function trails far behind our grasp for the marine microbial ecosystem in general. As outlined in the section *A brief history of research on sea ice microbial communities*, our understanding of sea ice microbial ecology has generally lagged the broader field of marine microbial ecology by a decade or more. The first measurement of primary production by the  $^{14}\text{C}$  method, for example, came in 1952 for seawater (Steeman-Nielsen, 1952) and 1965 for sea ice (Burkholder & Mandelli, 1965). The first study of *in situ* gene expression in seawater came in 1990, while gene transcripts were not studied in sea ice until 2010 (Koh et al., 2010). Reasons for this delay may include the novelty of the environment, which reduces the need for new or sophisticated techniques to warrant funding or publication, technical challenges regarding the application of these techniques to sea ice, the limited number of researchers in the field, and the significant logistical challenge of accessing the sea ice environment and conducting research there. While these are valid reasons, an alternate paradigm for the future may be to view the challenges of sea ice microbial ecology as a strong motivator for greater innovation in research. The field of sea ice microbial ecology is well placed to lead the development of a new framework for understanding microbial ecology.

This preferential placement is because, compared to many other microbial ecosystems, the sea ice microbial ecosystem may be relatively simple. A large number of its dominant members can be cultured (K. Junge et al., 2002) and thus sequenced and subjected to detailed physiological evaluation. Due to its fixed location in the water column, the flux of nutrients and materials into the system is easier to constrain than for many other marine environments (it is not necessary to commit to either a Eulerian or Lagrangian view). A strong seasonality not only defines the sea ice environment, but also provides a predictable annual cycle of perturbation and

community succession that makes it ideal for testing hypotheses regarding the biogeochemical impact of community structure. Sea ice represents an optimal environment to develop and test integrated models connecting community structure, metabolic potential, biological activity, biogeochemistry, and the resulting feedback loop on community structure. Such an undertaking, however, will require a coordinated and long-term research effort involving both modelers and observationalists, with both groups including specialists in physiology, genetics, and biogeochemistry.

While this dissertation does not solve the problem of implementing a TMMB model for the sea ice microbial ecosystem, it seeks to clarify fundamental aspects of sea ice microbial ecology through a better understanding of microbial community structure within several under-explored ice types, the genomic plasticity and metabolic function of psychrophiles, and the evolution of cold-adapted communities. In time, a deeper appreciation of these aspects of the microbial community will become part of the foundation for a more complete understanding of the sea ice ecosystem as a whole.

## References

- Apollonio, S. (1965). Chlorophyll in arctic sea ice. *Arctic*, 118–122.
- Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., & Thingstad, F. (1983). The ecological role of water-column microbes in the sea. *Mar. Ecol. Prog. Ser.*, 10(3), 257–263.
- Bowman, J. P., Gosnik, J. K., McCammon, S. A., Lewis, T. E., Nichols, D. S., Nichols, P. D., . . . Staley, J. T. (1998). *Colwellia demingiae* sp. nov., *Colwellia hornerae* sp. nov., *Colwellia rossensis* sp. nov. and *Colwellia psychrotropica* sp. nov.: psychrophilic Antarctic species with the ability to synthesize docosaheptaenoic acid (22: ω63). *Int. J. Syst. Bacteriol.*, 48(4), 1171–1180.
- Bowman, J. P., McCammon, S. A., Brown, M. V., Nichols, D. S., & McMeekin, T. A. (1997). Diversity and association of psychrophilic bacteria in Antarctic sea ice. *Appl. Environ. Microbiol.*, 63, 3068–3078.
- Brakstad, O., Nonstad, I., Faksness, L.-G., & Brandvik, P. (2008). Responses of microbial communities in Arctic Sea Ice after contamination by crude petroleum oil. *Microb. Ecol.*, 55(3), 540–552. doi: 10.1007/s00248-007-9299-x
- Brown, M. V., & Bowman, J. P. (2001). A molecular phylogenetic survey of sea-ice microbial communities. *FEMS Microbiol. Ecol.*, 35, 267–275.
- Bunt, J. (1963a). Microbiology of Antarctic sea-ice: Diatoms of Antarctic sea-ice as agents of primary production. *Nature*, 199, 1255–1257.
- Bunt, J. (1963b). Microbiology of Antarctic sea-ice: Microalgae and Antarctic sea-ice. *Nature*, 199, 1254–1255.
- Burkholder, P. R., & Mandelli, E. F. (1965). Productivity of microalgae in Antarctic sea ice. *Science*, 149(3686), 872–874.
- Carr, M. H., Belton, M. J. S., Chapman, C. R., Davies, M. E., Geissler, P., Greenberg, R., . . . Veverka, J. (1998). Evidence for a subsurface ocean on Europa. *Nature*, 391(6665), 363–365.
- Collins, R. E., Rocap, G., & Deming, J. W. (2010). Persistence of bacterial and archaeal communities in sea ice through an Arctic winter. *Environ. Microbiol.*, 12(7), 1828–1841.
- Colwell, R. R., & Morita, R. Y. (1964). Reisolation and emendation of description of *Vibrio marinus* (Russell) Ford. *J. Bacteriol.*, 88(4), 831–837.
- Colwell, R. R., Walker, J. D., & Cooney, J. J. (1977). Ecological aspects of microbial degradation of petroleum in the marine environment. *CRC Cr. Rev. Microbiol.*, 5(4), 423–445.

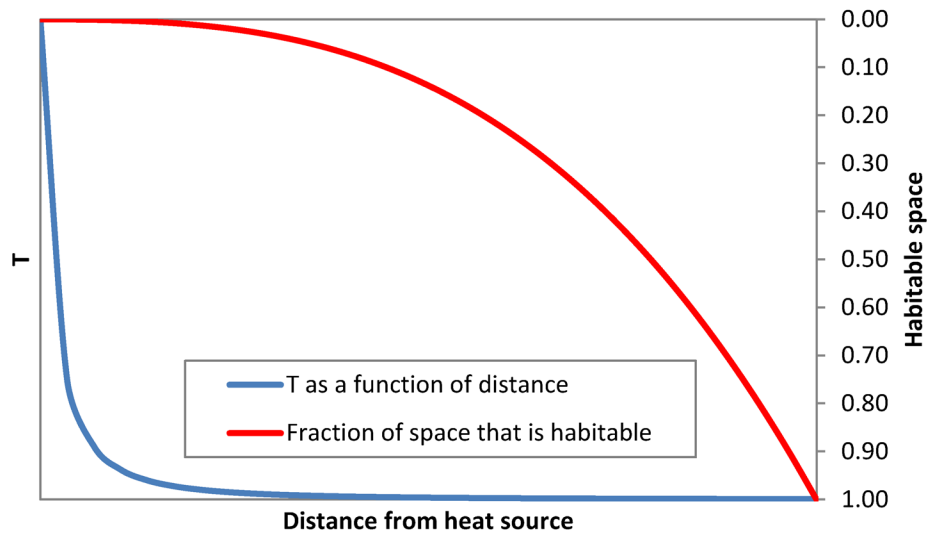
- Cox, G. F. N., & Weeks, W. F. (1983). Equations for determining the gas and brine volumes in sea-ice samples. *J. Glaciol.*, 29(102), 306–316.
- de Bernardis, P., Ade, P., Bock, J., Bond, J., Borrill, J., Boscaleri, A., . . . Farese, P. (2000). A flat Universe from high-resolution maps of the cosmic microwave background radiation. *Nature*, 404(6781), 955–959.
- Deming, J., & Huston, A. (2000). An oceanographic perspective on microbial life at low temperatures with implications for polar ecology, biotechnology and astrobiology *Cellular Origins and Life in Extreme Habitats* (pp. 149–160).
- Feng, S., Powell, S. M., Wilson, R., & Bowman, J. P. (2014). Extensive gene acquisition in the extremely psychrophilic bacterial species *Psychroflexus torquis* and the link to sea-ice ecosystem specialism. *Gen. Bio. Evol.*, 6(1), 133–148.
- Forster, J. (1887). Über einige eigenschaften leuchtender bakterien. *Centr. Bakteriolog. Parasitenk.*, 2, 337–340.
- Gautier, D. L., Bird, K. J., Charpentier, R. R., Grantz, A., Houseknecht, D. W., Klett, T. R., . . . Wandrey, C. J. (2009). Assessment of undiscovered oil and gas in the Arctic. *Science*, 324(5931), 1175–1179. doi: 10.1126/science.1169467
- Gerdes, B., Brinkmeyer, R., Dieckmann, G., & Helmke, E. (2005). Influence of crude oil on changes of bacterial communities in Arctic sea-ice. *FEMS Microb. Ecol.*, 53(1), 129–139. doi: 10.1016/j.femsec.2004.11.010
- Giannelli, V., Thomas, D. N., Haas, C., Kattner, G., Kennedy, H., & Dieckmann, G. S. (2001). Behaviour of dissolved organic matter and inorganic nutrients during experimental sea-ice formation. *Ann. of Glaciol.*, 33(1), 317–321.
- Giovannoni, S. J., Britschgi, T. B., Moyer, C. L., & Field, K. G. (1990). Genetic diversity in Sargasso Sea bacterioplankton. *Nature*, 345, 60–63.
- Grossi, S. M., Kottmeier, S. T., & Sullivan, C. (1984). Sea ice microbial communities. III. Seasonal abundance of microalgae and associated bacteria, McMurdo Sound, Antarctica. *Microb. Ecol.*, 10(3), 231–242.
- Grossmann, S., & Dieckmann, G. S. (1994). Bacterial standing stock, activity, and carbon production during formation and growth of sea ice in the Weddell Sea, Antarctica. *Appl. Environ. Microbiol.*, 60(8), 2746–2753.
- Helmke, E., & Weyland, H. (1994). Bacteria in sea ice and underlying water of the eastern Weddell Sea. *Mar. Ecol. Prog. Ser.*, 117, 269–287.
- Horner, R., & Alexander, V. (1972). Algal populations in arctic sea ice: An investigation of heterotrophy. *Limnol. Oceanogr.*, 17(3), 454–458.

- Huston, A. L., Krieger-Brockett, B. B., & Deming, J. W. (2000). Remarkably low temperature optima for extracellular enzyme activity from Arctic bacteria and sea ice. *Environ. Microbiol.*, 2(4), 383–388. doi: 10.1046/j.1462-2920.2000.00118.x
- Huston, A. L., Methé, B., & Deming, J. W. (2004). Purification, characterization, and sequencing of an extracellular cold-active aminopeptidase produced by marine psychrophile *Colwellia psychrerythraea* strain 34H. *Appl. Environ. Microbiol.*, 70(6), 3321–3328.
- Iizuka, H., Tanabe, I., & Meguro, H. (1966). Microorganisms in plankton-ice of the Antarctic Ocean. *J. Gen. Appl. Microbiol.*, 12(1), 101–102. doi: 10.2323/jgam.12.101
- Ingraham, J. (1958). Growth of psychrophilic bacteria. *J. Bacteriol.*, 76(1), 75.
- Ingraham, J., & Bailey, G. (1959). Comparative study of effect of temperature on metabolism of psychrophilic and mesophilic bacteria. *J. Bacteriol.*, 77(5), 609.
- Junge, K., Eicken, H., & Deming, J. W. (2004). Bacterial activity at –2 to –20°C in Arctic wintertime sea ice. *Appl. Environ. Microbiol.*, 70(1), 550–557. doi: 10.1128/aem.70.1.550-557.2004
- Junge, K., Imhoff, J., Staley, J., & Deming, J. (2002). Phylogenetic diversity of numerically important Arctic sea-ice bacteria cultured at subzero temperature. *Microb. Ecol.*, 43(3), 315–328.
- Junge, K., Krembs, C., Deming, J., Stierle, A., & Eicken, H. (2001). A microscopic approach to investigate bacteria under in situ conditions in sea-ice samples. *Ann. of Glaciol.*, 33, 304–310.
- Kirchman, D. L., Moran, X. A. G., & Ducklow, H. (2009). Microbial growth in the polar oceans: role of temperature and potential impact of climate change. *Nat. Rev. Micro.*, 7(6), 451–459.
- Koh, E. Y., Atamna-Ismaeel, N., Martin, A., Cowie, R. O. M., Beja, O., Davy, S. K., . . . Ryan, K. G. (2010). Proteorhodopsin-bearing bacteria in Antarctic sea ice. *Appl. Environ. Microbiol.*, 76(17), 5918–5925. doi: 10.1128/aem.00562-10
- Koh, E. Y., Phua, W., & Ryan, K. G. (2011). Aerobic anoxygenic phototrophic bacteria in Antarctic sea ice and seawater. *Environ. Microbiol. Rep.*, 3(6), 710–716. doi: 10.1111/j.1758-2229.2011.00286.x
- Kottmeier, S. T., Grossi, S., & Sullivan, C. W. (1987). Sea ice microbial communities. VIII. Bacterial production in annual sea ice of McMurdo Sound, Antarctica. *Mar. Ecol. Prog. Ser.*, 35, 175–186.
- Kottmeier, S. T., & Sullivan, C. W. (1988). Sea ice microbial communities (SIMCO). *Polar Biology*, 8(4), 293–304.

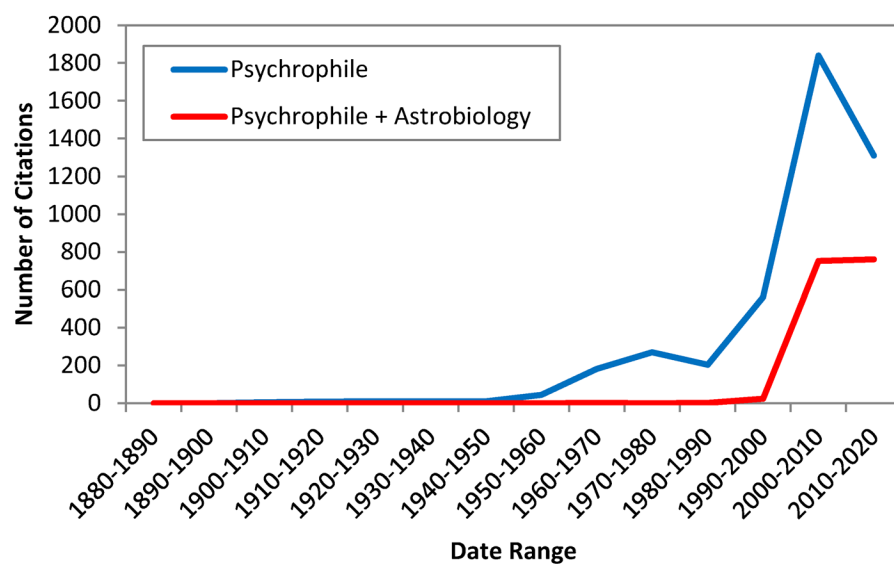
- Krembs, C., Eicken, H., Junge, K., & Deming, J. W. (2002). High concentrations of exopolymeric substances in Arctic winter sea ice: implications for the polar ocean carbon cycle and cryoprotection of diatoms. *Deep Sea Res. Part I*, 49(12), 2163–2181.
- Marion, G., Farren, R., & Komrowski, A. (1999). Alternative pathways for seawater freezing. *Cold Reg. Sci. Technol.*, 29(3), 259–266.
- Marion, G. M., & Farren, R. E. (1999). Mineral solubilities in the Na-K-Mg-Ca-Cl-SO<sub>4</sub>-H<sub>2</sub>O system: a re-evaluation of the sulfate chemistry in the Spencer-Møller-Weare model. *Geochimica et Cosmochimica Acta*, 63(9), 1305-1318.
- Mason, O., Han, J., Woyke, T., & Jansson, J. (2014). Single-cell genomics reveals features of a *Colwellia* species that was dominant during the Deepwater Horizon oil spill. *Frontiers in Microbiology*, 5, 332.
- Mayor, M., & Queloz, D. (1995). A Jupiter-mass companion to a solar-type star. [10.1038/378355a0]. *Nature*, 378(6555), 355–359.
- McKay, D. S., Gibson, E. K., Thomas-Keprta, K. L., Vali, H., Romanek, C. S., Clemett, S. J., . . . Zare, R. N. (1996). Search for Past Life on Mars: Possible Relic Biogenic Activity in Martian Meteorite ALH84001. *Science*, 273(5277), 924–930. doi: 10.1126/science.273.5277.924
- McLlan, A. L. (1918). Bacteria of ice and snow in Antarctica. *Nature*, 102(2550), 35-39.
- Meguro, H. (1962). Plankton ice in the Antarctic Ocean. *Antarct. Rec.*, 14, 1192–1199.
- Meguro, H., Ito, K., & Fukushima, H. (1967). Ice flora (bottom type): a mechanism of primary production in polar seas and the growth of diatoms in sea ice. *Arctic*, 114–133.
- Méthe, B. A., Nelson, K. E., Deming, J. W., Momen, B., Melamud, E., Zhang, X., . . . Fraser, C. M. (2005). The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *P. Natl. Acad. Sci.*, 102(31), 10913–10918. doi: 10.1073/pnas.0504766102
- Møller, A., Barkay, T., Hansen, M., Norman, A., Hansen, L., Sørensen, S., . . . Kroer, N. (2013). Mercuric reductase genes (merA) and mercury resistance plasmids in high Arctic snow, freshwater and sea-ice brine. *FEMS Microb. Ecol.*, 73, 2230–2238.
- Nansen, F. (1906). Protozoa on the ice floes of the North Polar Sea. *Scient. Results Norw. N. Polar Exped.*, 5(16), 1–22.
- Nedwell, D. B. (1999). Effect of low temperature on microbial growth: lowered affinity for substrates limits growth at low temperature. *FEMS Microbiol. Ecol.*, 30(2), 101–111. doi: 10.1111/j.1574-6941.1999.tb00639.x
- Pomeroy, L. R. (1974). The ocean's food web, a changing paradigm. *BioScience*, 499–504.

- Pomeroy, L. R., & Wiebe, W. J. (2001). Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. *Aquat. Microb. Ecol.*, 23(2), 187–204.
- Redmond, M. C., & Valentine, D. L. (2012). Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. *P. Natl. Acad. Sci.*, 109(50), 20292–20297.
- Röling, W. F., & van Bodegom, P. M. (2014). Toward quantitative understanding on microbial community structure and functioning: a modeling-centered approach using degradation of marine oil spills as example. *Frontiers in Microbiology*, 5(125). doi: 10.3389/fmicb.2014.00125
- Smith, L. M., Sanders, J. Z., Kaiser, R. J., Hughes, P., Dodd, C., Connell, C. R., . . . Hood, L. E. (1986). Fluorescence detection in automated DNA sequence analysis. *Nature*, 321, 674–679.
- Steeman-Nielsen, E. (1952). The use of radioactive carbon ( $C^{14}$ ) for measuring organic production in the sea. *J. Conseil*, 18(2), 117–140.
- Sullivan, C. W., & Palmisano, A. C. (1984). Sea ice microbial communities: Distribution, abundance, and diversity of ice bacteria in McMurdo Sound, Antarctica, in 1980. *Appl. Environ. Microbiol.*, 47(4), 788–795.
- Vollmers, J., Voget, S., Dietrich, S., Gollnow, K., Smits, M., Meyer, K., . . . Daniel, R. (2013). Poles apart: arctic and antarctic *Octadecabacter* strains share high genome plasticity and a new type of xanthorhodopsin. *PLoS ONE*, 8(5), e63422.
- Ward, D. M., Weller, R., & Bateson, M. M. (1990). 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. *Nature*, 345, 63–65.
- Wells, L. E., & Deming, J. W. (2006). Modelled and measured dynamics of viruses in Arctic winter sea-ice brines. *Environ. Microbiol.*, 8(6), 1115–1121.
- Woese, C. R., & Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *P. Natl. Acad. Sci.*, 74(11), 5088–5090.
- Yin, J., Chen, J., Liu, G., Yu, Y., Song, L., Wang, X., & Qu, X. (2013). Complete genome sequence of *Glaciecola psychrophila* strain 170T. *Genome announcements*, 1(3), e00199–00113.

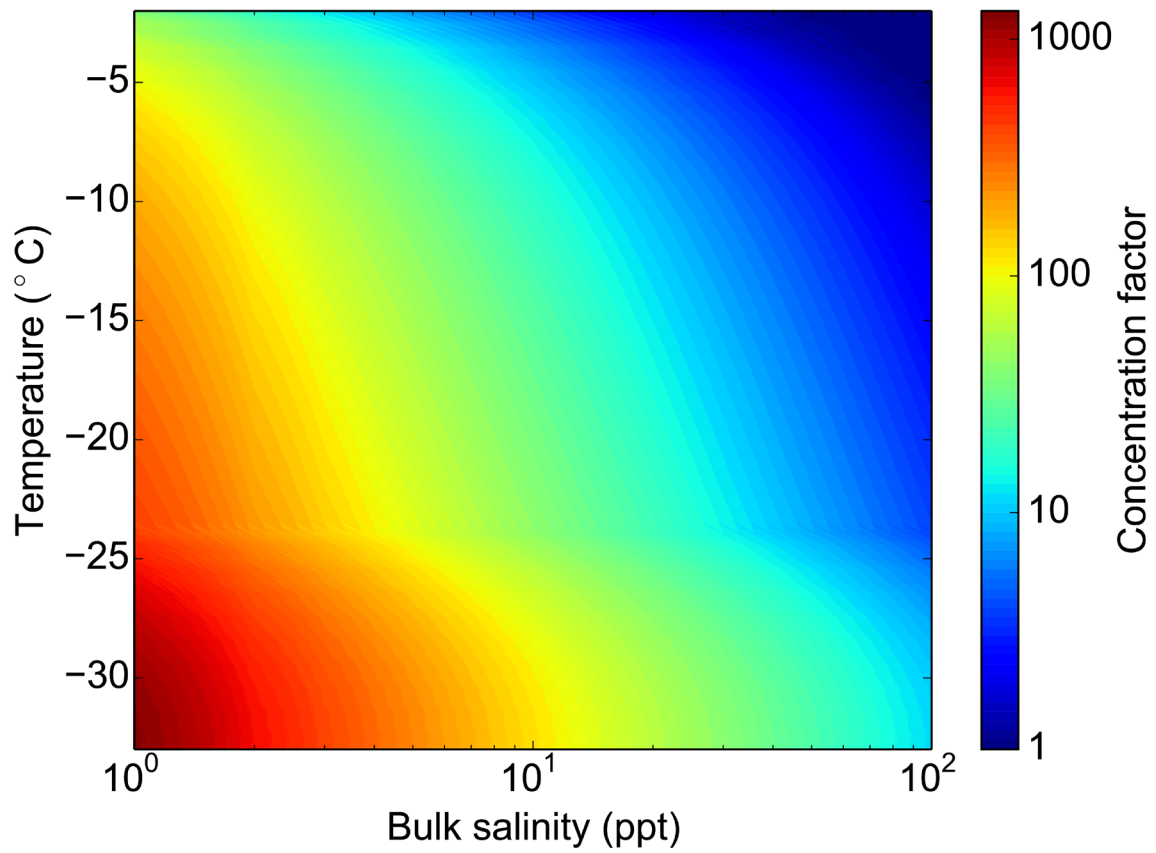




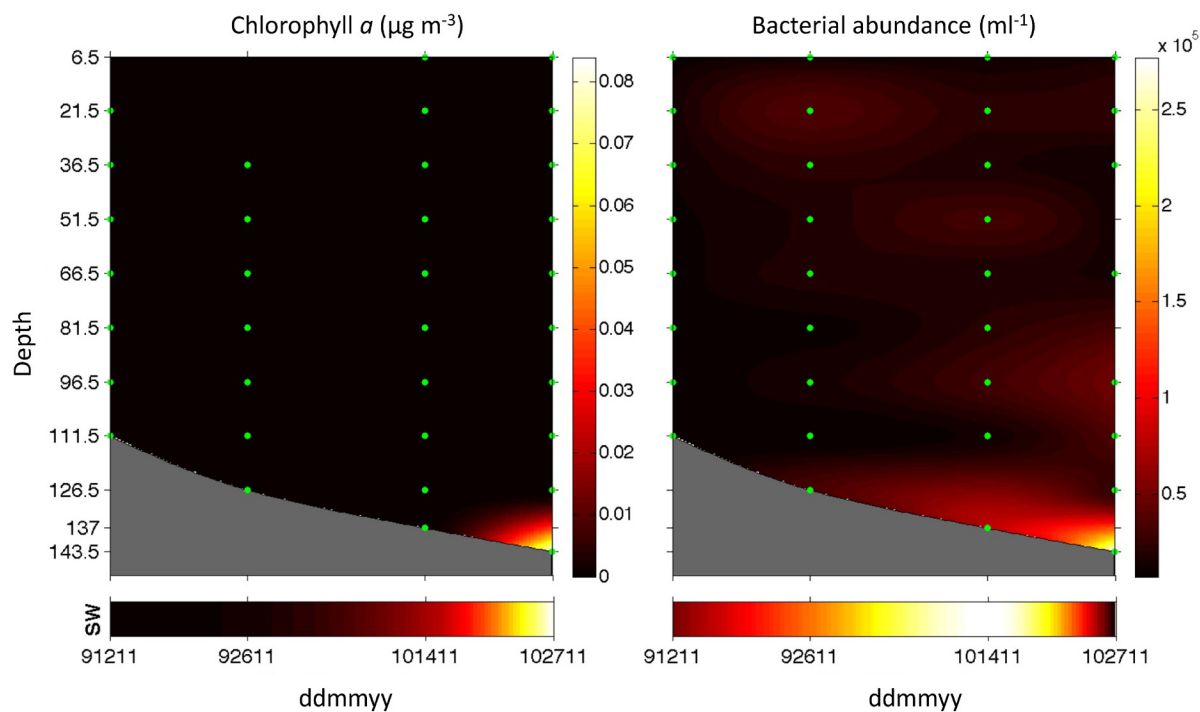
**Fig. 1. The theoretical distribution of habitable space.** Given a point source of heat, the amount of energy (here temperature is used as a proxy for energy) that reaches a point, distance  $d$ , from the source can be estimated from the inverse square law. The amount of space contained in a sphere with radius  $d$  is given by the volume of the sphere ( $v$ ). If the temperature ( $T$ ) calculated for the distance  $d$  represents the lowest temperature permissible for life, the fraction of habitable space is  $v$  divided by the volume of a sphere defined by the maximum size of the system (e.g., the size of a solar system).



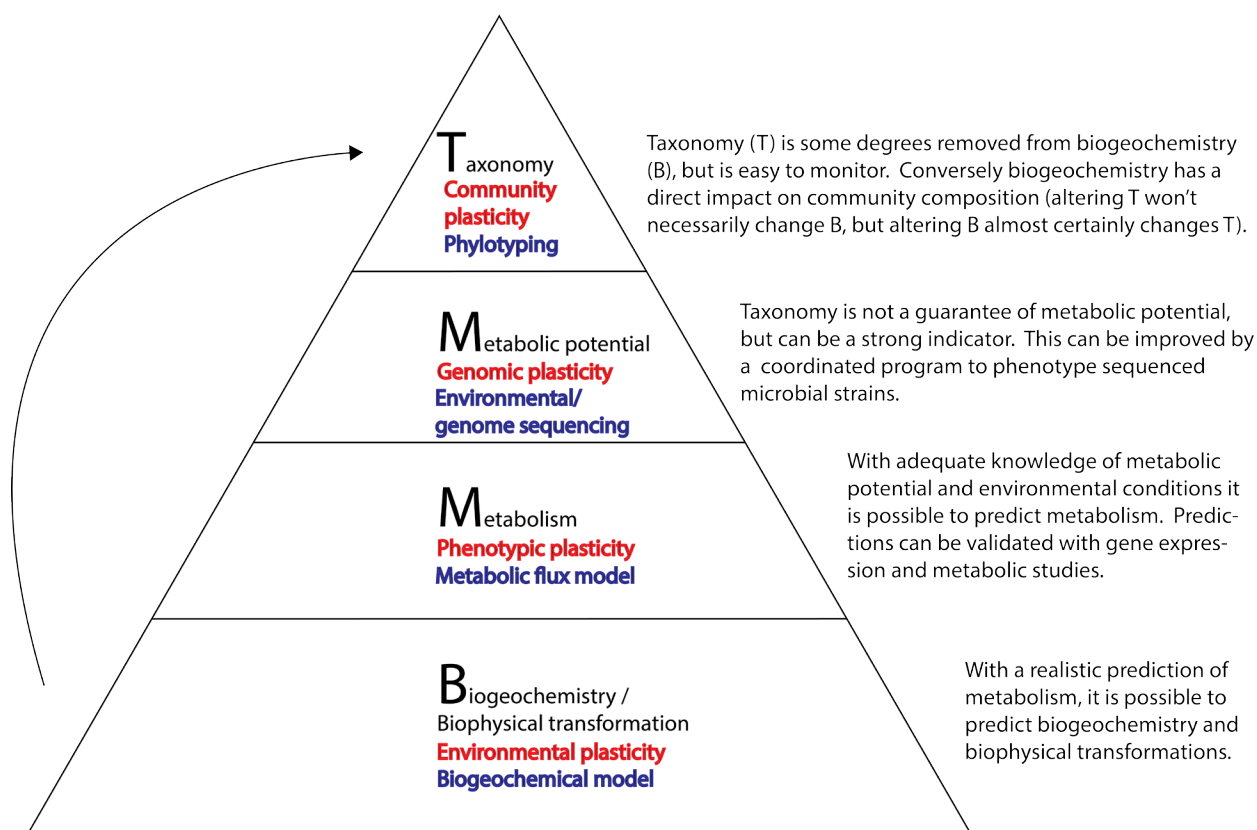
**Fig. 2. Occurrence in the peer-reviewed literature of the word “psychrophile” and “psychrophile+astrobiology.”** Data were taken from Google Scholar searches for 10 year intervals starting with 1880. The final bin represents the period 2010 to 2014. Patents and citations were excluded from the search.



**Fig. 3. Factor-fold concentration of solutes in water ice relative to starting solute concentration.** The relative solute concentration is the inverse of the brine volume fraction, calculated along a gradient of temperature and salinity from the equations of Cox and Weeks (1983).



**Fig. 4. Bacterial abundance and chlorophyll *a* from land-fast first-year sea ice during the Austral spring of 2011, McMurdo Sound, Antarctica.** Grey area indicates seawater below the advancing ice front. Actual values for seawater are given in the box below each primary frame. Bacterial abundance was determined by epifluorescence microscopy while chlorophyll *a* was determined by standard methods, as in Chapter 2 of this work.



**Fig. 5. The Taxonomy-Metabolic potential-Metabolism-Biogeochemistry (TMMB) model framework.** The pyramid represents a conceptual framework linking community composition (taxonomy), easily monitored in the environment and a direct function of environmental conditions, with metabolic potential, metabolism, and biogeochemistry/biophysical transformation. The dynamic nature of the system is commonly referred to as “plasticity”; four different types of plasticity are shown in red. The specific analytical techniques relevant to the development of a model are shown in blue. Given adequate knowledge of M and M from experiments and environmental observations, it should be possible to predict B from T via an analysis of predicted metabolisms, and T from B via a predicted microbial community response.

## Chapter 2

### **Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene**

(Citation: Bowman, J. S., Rasmussen, S., Blom, N., Deming, J. W., Rysgaard, S., & Sicheritz-Ponten, T. (2011). Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. *The ISME journal*, 6(1), 11–20.)

#### ABSTRACT

Dramatic decreases in the extent of Arctic multiyear ice (MYI) suggest this environment may disappear as early as 2100, replaced by ecologically different first-year ice (FYI). To better understand the implications of this loss on microbial biodiversity we undertook a detailed census of the microbial community in MYI at two sites near the geographic North Pole using parallel tag sequencing of the 16S rRNA gene. While the composition of the MYI microbial community has been characterized by previous studies microbial community structure has not. Although richness was lower in MYI than in underlying surface water we found diversity to be comparable using the Simpson and Shannon's indices (for Simpson  $t = 0.65$ ,  $p = 0.56$ ; for Shannon  $t = 0.25$ ,  $p = 0.84$  for a Student's T-test of mean values). Cyanobacteria, comprising 6.8 % of reads obtained from MYI, were observed for the first time in Arctic sea ice. In addition several low-abundance clades not previously reported in sea ice were present, including the phylum TM7 and the classes Spartobacteria and Opitutae. Members of *Coraliomargarita*, a recently described genus of the class Opitutae, were present in sufficient numbers to suggest niche occupation within MYI.

## 2.1 Introduction

The sea ice environment is currently extensive on Earth, covering some  $27\text{--}36 \times 10^6 \text{ km}^2$  in the high latitudes of both hemispheres (Cavalieri, Gloersen, Parkinson, Comiso, & Zwally, 1997; Stroeve, Holland, Meier, Scambos, & Serreze, 2007), an area approximately the size of Canada. This environment is composed primarily of two ice types; first-year ice (FYI) which forms in the fall and melts in the summer, and multiyear ice (MYI), perennial sea ice that has persisted through at least one complete year. In parallel with observed global climate trends the MYI environment is in a state of sharp decline (Barber et al., 2009; Boe, Hall, & Qu, 2009; Comiso, Parkinson, Gersten, & Stock, 2008; Nghiem et al., 2007). A rapid decrease in minimal sea ice extent suggests that the MYI environment could be nonexistent by 2100 (Boe et al., 2009). This dramatic reduction in MYI is expected to have a significant impact on the polar microbial community as ocean surface temperature, salinity, nutrients, and availability of the unique sea ice environment change (Kirchman, Moran, & Ducklow, 2009; Vincent, 2010). These changes will in turn impact the way carbon and other elements are cycled through the Arctic marine food web, most readily through the coupling of primary and bacterial production (Kirchman et al., 2009).

In 2006 the project *DNA of the Polar Seas* was launched as part of the Danish expedition *Galathea III*, to more fully catalogue microbial diversity within the climatically vulnerable Arctic and Southern Oceans. As participants in a 2009 extension of *DNA of the Polar Seas* onboard the Swedish icebreaker *Oden* we sampled MYI and surface seawater near the geographic North Pole with the aim of achieving a high resolution analysis of microbial diversity in these proximate but disparate environments. Previous microbiological investigations of sea

ice, dating back to at least 1893 (McLean, 1918), have identified a specialized sea ice microbial community (Bowman, McCammon, Brown, Nichols, & McMeekin, 1997; Brown & Bowman, 2001; Junge, Imhoff, Staley, & Deming, 2002). Because of methodological limitations however, these studies have reported mainly on community composition rather than structure. They have focused almost exclusively on FYI. The only previous microbial investigation of Arctic MYI, using fluorescence *in situ* hybridization (FISH) and 16S rRNA clone libraries, reported a summer community dominated by Gammaproteobacteria, followed by Alphaproteobacteria and Cytophaga-Flavobacterium-Bacteriodes (CFB) (Brinkmeyer et al., 2003), results similar to those obtained for a summer sampling of Antarctic MYI (Brown & Bowman, 2001). The dynamical processes that control microbial community structure in sea ice are not well understood, however a growing body of literature suggests that microbial communities in multiyear ice differ substantially from those in FYI (Brinkmeyer et al., 2003; Brown & Bowman, 2001; Jody W. Deming, 2007).

Sea ice is considered selective for psychrophilic microbes (Junge et al., 2002), providing the “seed population” for the microbial community reported to dominate carbon cycling in Arctic coastal waters (Connelly, Tilburg, & Yager, 2006). MYI in particular has been observed to host a strongly cold-active microbial enzyme assemblage compared to Arctic seawater and FYI (Jody W. Deming, 2007). Loss of this seed population can be expected to impact the pelagic fate and downwards flux of carbon from sea surface to seafloor (Kellogg, Renfro, Cochran, & Deming, 2011). If the microbial populations present in MYI are more complex than previously anticipated, then the predicted loss of MYI may result in a greater loss of biodiversity and an as yet unquantified impact on elemental cycling. Here we employed 454 pyrosequencing of the 16S rRNA gene to obtain a snapshot of microbial community structure within MYI at a greater



resolution than previous studies, comparing it to that of surface seawater to test for a greater overlap of phylogenetic groups. This approach, in confirming the unique nature of the sea ice microbial community, also revealed an unexpectedly high degree of diversity in MYI and several new clades not previously found in sea ice.

## **2.2 Methods**

### **2.2.1 Sample collection and preparation**

Sea ice and seawater samples were collected during the LOMROG 2009 expedition on the Swedish icebreaker *Oden*. Two MYI samples (designated I13 and I14) and three surface seawater samples (S15, S16, S17) were collected within 150 km of the geographic North Pole (Fig. S2.1, station coordinates given in Table S2.1). Although the precise age the sampled MYI floes could not be determined, ice thickness (301 and 310 cm), temperature and salinity conditions (Fig. S2.1), location (50–120 km from the geographic North Pole), and time of sampling suggest the ice had persisted through more than one summer. Detailed methods for sample collection and processing shipboard and for subsequent DNA extraction, purification, and sequencing and other analyses are given in Supplementary Information. For each MYI sample two cores spanning the depth of the MYI floe were sectioned and melted together as one sample. Microorganisms in the melted samples were captured on a 0.2  $\mu\text{m}$  filter after a 2.0  $\mu\text{m}$  pre-filtration. The decision to melt an entire core length precluded a study of vertical community structure within the ice, but ensured that enough DNA was recovered for this exploratory effort.

Extraction, purification, and sequencing were conducted at the University of Copenhagen. Following lysis and phenol-chloroform DNA extraction, amplification of a section of the 16S RNA gene including the V3 hypervariable region was conducted with the universal primers PRK341F (CCTAYGGGRBGCAACAG) and PRK806R

(GGACTACNNGGGTATCTAAT) (Yu *et al.*, 2005). Sequencing was performed on a GS FLX pyrosequencer (454 Life Sciences, Branford, CT, USA).

### **2.2.2 Community structure**

Analysis was conducted using the microbial ecology community software program Mothur (Schloss *et al.*, 2009). The reads were processed by removing tags and primer, only accepting reads with an average quality score above 20 and read lengths between 200 and 300 nt. Identical sequences were grouped and a representative aligned against the Greengenes reference set (DeSantis *et al.*, 2006) using the Needleman-Wunsch algorithm (Needleman & Wunsch, 1970). A further screening step (pre.cluster) was applied to reduce sequencing noise by clustering reads differing by only 1 bp (Huse, Welch, Morrison, & Sogin, 2010). Chimeric sequences were detected and removed using the implementation of ChimeraSlayer developed by the Broad Institute. The remaining high quality reads were used to generate a distance matrix and clustering with the furthest-neighbor algorithm. Representative sequences for shared operational taxonomic units (OTUs) as defined by 100 (unique), 97, 93, and 90 % similarity were obtained.

Sequence coverage was calculated as the percent of unique reads observed against the Chao1 estimate of richness (Chao, 1984), with coverage ranging from 30.8 to 35.8 %. Rarefaction curves were generated for each sample at the 100 %, 97 %, 90 %, and 80 % levels of similarity. To describe community structure diversity was calculated using Simpson's index (Simpson, 1949), which estimates the probability that two randomly selected reads will belong to the same OTU. Because coverage was less than 100 %, diversity was also calculated using a non-parametric instance of Shannon's index (Chao & Shen, 2003). The Pearson's correlation coefficient was calculated to determine the level of agreement between these two methods.

To analyze community breadth, the distance between all clades present in each community, a distance matrix of all sequences contained within the ice and seawater communities was created using ClustalX (Larkin et al., 2007), correcting for multiple substitutions (Kimura, 1983) and ignoring sites with gaps. Central tendencies were used to describe the distribution of distances within each community.

### **2.2.3 Community composition**

Community taxonomy at the unique sequence level was determined using Mothur's classification tool with 1,000 iterations and a bootstrap cutoff score of 60. The taxonomic database consisted of 14,956 bacterial and 2,297 archaeal sequences from the SILVA database (Pruesse et al., 2007) classified with the Ribosomal Database Project (RDP) (Cole et al., 2007). Samples were grouped by their respective environment (MYI or seawater) and classified against this database. Maximum-likelihood (ML) trees were constructed of all genera identified as common to both the MYI and seawater samples (present above 0.035 % of the total community in both environments) to evaluate whether lineages within these genera were exclusive to either MYI or seawater. Reference trees were constructed from all high quality, > 1,200 bp sequences available on the RDP for each genus, with the addition of all type strains available for the relevant family, using the GTR model and rapid bootstrapping (100 bootstraps) option in the parallelized version of RAxML vs. 7.0.4 (Stamatakis, 2006). The online version of pplacer (Matsen, 2010) was used to place the reads attributed to each genus on the appropriate reference tree.

### **2.2.4 Community similarity**

Differences between communities were analyzed by first constructing a ML tree of all unique sequences aligned with Mothur using RAxML as previously described. Differences in

tree topology were then assessed using the unweighted and weighted UniFrac algorithms (Lozupone & Knight, 2005). Abundance information was not embedded into the tree files necessary for these algorithms beyond the taxonomic level of unique. Therefore Sørensen's abundance and richness-based indices of similarity were used to compare the communities at different taxonomic levels (unique, 97, 93, and 90 % similarity). All indices were calculated with Mothur. For compatibility with pplacer trees were constructed in RAxML using the GTRGAMMA option, precluding calculating of UniFrac significance scores from predicted branch length. To calculate significance a separate UniFrac analysis was conducted on a neighbor-joining (NJ) tree constructed using ClustalW (Larkin et al., 2007) and bootstrapped 1,000 times.

#### **2.2.5 16S rRNA divergence**

Divergence within the 16S rRNA gene was assessed using both taxonomy and taxonomy-independent approaches. In the taxonomy approach the number of member reads and number of unique members were tallied for each clade at each taxonomic level. The slope of the best fit line for both seawater and sea ice communities was determined using a least-squares analysis. Slopes for the two communities were compared using a Student's T-test. To assess 16S rRNA gene divergence independent of taxonomy, and thus avoid bias due to unequal characterizations of the MYI and seawater communities, the number of OTUs was estimated for various definitions of OTU, from 0 to 20 % divergence, for each community using the Chao1 estimate of richness (Chao, 1984). The number of OTUs calculated for seawater was normalized to that of sea ice by reducing the number estimated at each definition to 36 %, the fraction of maximum seawater richness found within the ice. The relationship between the number of OTUs for both

communities was described using a line of best fit. The difference between the slope of the best fit line and unity was assessed with a Student's T-test.

## **2.3 Results**

### **2.3.1 Environmental**

Averaged vertically the sampled MYI was slightly warmer than the surface seawater ( $-1.1$  and  $-1.2^{\circ}\text{C}$  vs.  $-1.48$  and  $-1.51^{\circ}\text{C}$  respectively) and lower in bulk salinity ( $1.9$  and  $0.7$  compared to  $30.92$  and  $32.57$ ) (Table S2.1). These values are typical for MYI (H. Eicken et al., 1995). Ice thickness, at  $3.0$  m for I13 and  $3.3$  m for I14, was also consistent with values expected for MYI (H. Eicken et al., 1995; Haas, 2004). At the time of sampling the ice was observed to be drifting rapidly ( $5.19\text{ m min}^{-1}$  at  $318.97^{\circ}$  and  $4.84\text{ m min}^{-1}$  at  $198.73^{\circ}$  for I13 and I14, respectively). These variations in drift are common for ice in the central Arctic Ocean and consistent with general transport from the Siberian Sea toward Fram Strait (Pfirman, Colony, Nürnberg, Eicken, & Rigor, 1997). Models from 2002 suggest that ice at the sampled locations should be between 4 and 8 years old, though the same study points to a dramatic decrease in the age of ice across most of the Arctic Ocean (Rigor & Wallace, 2004). As a result the sampled ice could be younger than 4 years.

At  $1.68$  and  $8.40 \times 10^4\text{ cells ml}^{-1}$  melted sea ice, total bacterial and archaeal abundance (averaged vertically throughout the entire ice thickness) was lower in MYI than in the surface seawater ( $1.84 \times 10^5\text{ cells ml}^{-1}$  for sample S17), though concentrations of chlorophyll *a* were similar (sea ice:  $0.025$  and  $0.055\text{ mg m}^{-3}$  melted ice averaged vertically, seawater:  $0.037\text{ mg m}^{-3}$ ). Concentrations of particulate exopolymers (pEPS) were low in both ice samples and undetectable in seawater (sample S17; Table S2.1). Calculating the concentration factor for the sea ice brine as the inverse of the brine volume fraction equation of Cox and Weeks (Cox &

Weeks, 1983; Hajo Eicken, 2009) suggests that the abundance of microbes, salt, pEPS, chlorophyll *a*, and other constituents partitioned into the liquid brine fraction would be an order of magnitude higher than these values (Table S2.1).

### **2.3.2 Community structure**

Richness was higher in seawater than sea ice for all definitions of OTU, reaching average values at the unique sequence level of 1,552 for MYI and 3 042 for seawater with a confidence level of 89 % ( $p = 0.11$ ; Table 2.2). The limited degrees of freedom ( $df = 3$ ) precluded tighter confidence levels. Even though a preclustering step was performed to reduce pyrosequencing noise, the absolute values for richness should be treated cautiously because pyrosequencing noise can artificially inflate sample richness (Quince et al., 2009), as can sequence analysis of the 16S rRNA gene by other techniques (Acinas, Marcelino, Klepac-Ceraj, & Polz, 2004).

Diversity was similar for all definitions of OTU between ice and seawater samples (Table 2.2). Although the MYI community appeared less diverse by both Shannon's index of diversity and Simpson's index, the differences were not statistically significant by a Student's T-test (Table 2.2). Good agreement was observed between the parametric Simpson's index and the non-parametric Shannon's index ( $df = 18$ ,  $p < 10^{-4}$ ; Fig. S2.2), suggesting that despite low coverage using Simpson's definition of abundance when describing these samples is appropriate.

Breadth was significantly different for the ice and seawater communities. A histogram of distances (Fig. 2.1) shows a multimodal distribution for the seawater community (mean = 0.43,  $sd = 0.30$ ,  $n = 4,941,729$ ), compared to a more simplified (bimodal) distribution within the ice (mean = 0.33,  $sd = 0.16$ ,  $n = 912,025$ ). A reanalysis of the seawater distribution using only data from the domain Bacteria revealed a narrower breadth (mean = 0.30,  $sd = 0.12$ ,  $n = 3,988,009$ ).

The difference between this mean and the mean calculated for the MYI community by a Student's T-test was significant ( $p < 10^{-4}$ ,  $df = 4,900\ 036$ ).

### 2.3.3 Community composition

Within the seawater samples, 2,927 reads (represented by 106 unique reads) could not be classified at the domain level using the reference taxonomy (Fig. S2.3). NCBI BLASTN (Altschul et al., 1997) was used to find similar sequences. All scored as highly similar ( $E \leq 10^{-24}$ ) to sequences classified as belonging to isolates or environmental sequences of the domain Archaea. A much smaller number of reads in MYI (13, six unique) could not be classified at the domain level, but they showed a high degree of similarity with sequences classified as Archaea ( $E \leq 10^{-37}$ ). Few archaeal reads were detected in MYI (16, 0.13 % of all reads) compared to seawater (5,485, 15.9 %). Within the seawater a large number of archaeal reads were Euryarchaeota (21.9 %), with the remainder unclassified below the level of domain. *In silico* analysis of the universal prokaryotic primer set indicated a low binding efficiency between primer and marine Crenarchaeota (Fig. S2.4). As a result Crenarchaeota are likely underrepresented in this data set.

Although the bacterial phylum Proteobacteria was numerically dominant in both the MYI and seawater communities, representing 66.3 % and 62.3 % of total reads, respectively, substantial differences in community composition between sample types were observed below the level of phylum (Fig. 2.2). Gammaproteobacteria and Flavobacteria (62.3 % and 19.9 % of total reads) dominated the MYI community, while Alphaproteobacteria dominated the seawater community (48.9 % of total reads). The majority (64.6 %) of seawater Alphaproteobacteria were classified as *Pelagibacter*. At the family level the MYI community was dominated by Moraxellaceae, accounting for 42.8 % of all ice reads. Large numbers of reads were also

assigned to the Flavobacteriaceae family (19.6 %). Although not numerically dominant in either case, cyanobacteria were identified in both the ice and seawater samples, accounting for 6.8 % of total ice reads and 5.4 % of total seawater reads. Cyanobacterial abundance was likely underestimated, as the primer was predicted to have low binding efficiency with the cyanobacterial 16S rRNA gene (Fig. S2.4).

At the level of genus (the lowest level assigned) the reads represented 41 identifiable genera for MYI and 37 for seawater. At this level 32.1 % of ice reads and 64.3 % of seawater reads could not be classified (Fig. S2.3). A complete list of read assignments by clade is given in Tables S2.3 (MYI) and S4 (seawater). Of the clades listed for MYI, 54 have not been reported previously in sea ice. Because the bottom, porous horizons of the MYI cores were included in this analysis we cannot rule out that some of these organisms were contained within entrained seawater. Excluding all clades observed within the two seawater samples leaves 38 clades not previously described in either FYI or MYI.

Four genera were identified as common to both environments; *Polaribacter*, *Rubritalea*, *Psychrobacter*, and *Pelagibacter*. Visual inspection of ML trees constructed of these genera (not shown) suggest that only *Polaribacter* has lineages that have evolved to occupy separate niches within MYI and seawater.

#### **2.3.4 16S rRNA gene divergence**

To assess the level of divergence within the MYI and seawater communities all taxonomies as determined by our classification were tested for a correlation between membership and the number of unique members (Fig. 2.3). For a given taxonomy more members (greater sampling) should produce more unique members. To account for greater coverage of the ice samples (Table 2.1) the number of unique members for ice was reduced by



3.50 % (the difference in mean coverage between the ice and seawater communities). If the unique MYI and seawater communities were subject to similar evolutionary pressures and timescales, they should display a similar relationship between taxon membership and number of unique members. Strong correlations were observed for both the MYI and seawater communities ( $p < 10^{-4}$ ,  $df = 117$  and  $p < 10^{-4}$ ,  $df = 130$ , respectively), but the relationships differed: slopes of the regression lines fit to the data ( $m = 0.0601$  for ice,  $0.0565$  for seawater) were significantly different with a Student's T-test ( $df = 247$ ,  $p < 10^{-4}$ ). This difference suggests that MYI taxa may accumulate more unique reads than seawater for the same amount of sampling.

The non-taxonomic approach to divergence yielded a different result. For a given definition of OTU (% divergence) the seawater samples hosted a greater number of OTUs (Fig. 2.4A) when corrected for greater overall richness. A best fit line for the number of OTUs present in ice for each definition of OTU plotted against the number in seawater yielded a slope significantly different from unity ( $df = 18$ ,  $p < 10^{-4}$ ; Fig. 2.4B) suggesting that seawater hosts more OTUs for an increasingly narrow definition of OTU than does MYI.

### **2.3.5 Community similarity**

Pairwise differences in community composition and structure were assessed for all samples using two richness-based approaches (Sørensen's richness based, unweighted UniFrac) and two abundance-based approaches (Sørensen's abundance based and weighted UniFrac) (Tables 2.3 and S2.2). The abundance based approaches showed substantially lower (more similar) scores between the two sea ice and the three seawater samples than did the richness based approaches. Using a separate, NJ tree to calculate the significance of dissimilarity, only seawater samples S16/S17 were not significantly different ( $p = 0.06$ ) for unweighted UniFrac,

while S16/S17 and I13/I14 were not significantly different using weighted UniFrac ( $p = 0.19$  and  $p = 0.32$ , respectively).

## **2.4 Discussion**

Numerous investigations have characterized the composition and structure of the microbial community within Arctic surface seawater, including some analysis by the 454 method (Galand, Casamayor, Kirchman, & Lovejoy, 2009; Kirchman, Cottrell, & Lovejoy, 2010). Although a further characterization of this environment is not the purpose of this paper, a brief comparison is in order due to spatial and seasonal variation of the sea surface microbial community. As observed in this study Alphaproteobacteria typically dominate Arctic surface water (Alonso-Sáez, Sánchez, Gasol, Balagué, & Pedrós-Alio, 2008; Bano & Hollibaugh, 2002; Galand et al., 2009), though in one coastal study a greater abundance of Gammaproteobacteria were observed (Kirchman et al., 2010). Gammaproteobacteria comprised only a small fraction of the seawater reads in this study ( $< 2\%$ ), which may reflect a regional difference given the mid-ocean location of this study. None of these studies reported cyanobacteria among the major clades present, though they have been found as allochthonous forms in river-impacted, coastal Arctic waters (Waleron, Waleron, Vincent, & Wilmotte, 2007). Further differences were restricted to variations in the relative abundance of the major clades. Deltaproteobacteria, Bacterioidetes, Verrucomicrobiae, and Actinobacteria are widely recognized as abundant Arctic seawater clades within the domain Bacteria.

Previous phylogenetic studies of the microbial communities in spring and summer sea ice (FYI and MYI) have indicated dominance by the clades Gammaproteobacteria, Alphaproteobacteria, and CFB (Bowman et al., 1997; Brinkmeyer et al., 2003; Junge et al., 2002; Petri & Imhoff, 2001), followed by Chlamydiales and Verrucomicrobiales (Brown & Bowman,

2001). With the exception of the order Chlamydiales, which has no representative reads in this data set, our findings confirm the dominance of these clades. Reports of Archaea in sea ice were inconclusive (J. W. Deming, 2009) until Collins et al. (2010) confirmed their presence in winter FYI. They have not been detected in FYI of other seasons or in MYI using less sensitive techniques than 454 pyrosequencing. Although our data are biased by the performance of the universal primer set, our findings of limited archaeal reads in the MYI community suggest that marine archaea, along with many pelagic bacterial clades, are not competitive against the spring/summer sea ice bacterial community. Some Archaea may persist interannually in MYI, or the archaeal reads observed in our MYI samples may be the result of recent seawater infiltration of the lower ice horizons. Similarly, cyanobacteria have not previously been reported in either FYI or MYI, though they are common to freshwater polar environments (Jungblut, Lovejoy, & Vincent, 2009; Waleron et al., 2007). Their presence within MYI may be the result of colonization of surface melt ponds, with subsequent incorporation into MYI through infiltration or melt pond burial. Further study is required to determine whether cyanobacteria may be indigenous to MYI, perhaps occupying upper ice horizons where light and nutrient conditions may provide an open niche.

Several clades not previously reported in sea ice were observed exclusively within MYI, including the Gram-positive phylum TM-7 and family Micrococcineae. These and several other new clades, including the Spartobacteria and Verrucomicrobiales, are commonly found in soil environments, suggesting that they may have reached the MYI environment through dust deposition or inclusion of riverine sediments (Brinkmeyer et al., 2003). Other newly observed clades are likely to be of marine origin, including the genera *Persicivirga*, *Ulvibacter*, *Granulosicoccus*, and *Lewinella*. Representatives from the first three clades have characteristics

that suggest an association with marine algae (Barbeyron et al., 2010; Kurilenko et al., 2010; Nedashkovskaya et al., 2004). The possibility that these clades were present in MYI because of sea ice diatoms warrants further study. Other newly observed clades of biogeochemical interest include *Methylmicrobium*, a dimethylsulfide (DMS) oxidizing genus of the Gammaproteobacteria, *Cycloclasticus*, a PAH degrading clade also of this class, and *Jannaschia*, an anoxygenic phototroph of the Roseobacter clade. Although all of these taxa were present at low levels within MYI (below 0.36 % of all MYI reads), they may represent an important seed population for bloom conditions within the surrounding seawater or sea ice, a hypothesis further supported by their absence from the seawater samples.

A final newly observed genus of note is *Coraliomargarita*, a recently described genus of the class Opitutae (Yoon et al., 2007). *Coraliomargarita* accounted for 197 reads in this dataset (1.6 % of all MYI reads), split among 14 lineages. This genus has no psychrophilic members in culture and, despite a sequenced genome for *C. akajimensis* (Mavromatis et al., 2010), there is little in the natural history of this genus to suggest why it should be associated with MYI. Like many sea ice-associated microbes *C. akajimensis* is copiotrophic, a feature that might favor its persistence within organic-rich horizons of MYI. The frequency of *Coraliomargarita* reads in MYI also raises questions about the role of psychrotolerant bacteria in sea ice.

Past efforts to quantify microbial community structure within sea ice likely underestimated both richness and diversity. Brown and Bowman (2001) were able to measure these parameters on a single MYI sample from the Southern Ocean, identifying it as the most diverse of several sampled ice types. From a clone library they estimated richness at  $24 \pm 12$  and Shannon's index at 1.050, with coverage at 53.8 %. Brinkmeyer et al. (2003) conducted a similar clone-library based analysis on several MYI cores recovered from Fram Strait. This

study reports similar coverage (53.3 to 76 %), richness (15 to 19 OTUs), and diversity (Shannon's index between  $1.065 \pm 0.002$  and  $1.291 \pm 0.002$ ). By comparison with 454 sequencing and only 34.8 and 35.8 % coverage we observed richness an order of magnitude higher, with the Chao1 estimate two orders of magnitude higher. Shannon's index of diversity was calculated to be three-fold higher than these earlier estimates.

The low coverage obtained even with the application of 454 sequencing, the identification of new clades not previously found in sea ice, and continuing questions regarding the dynamical processes of recruitment, succession, and persistence highlight the need for a greater sequencing effort across the sea ice environment. Given the high degree of temporal and spatial heterogeneity inherent to the sea ice environment, and resulting differences in microbial community structure, a three-dimensional sequencing effort across time, space, and sequence depth is needed. Such an effort could provide not only the missing characterization of a vast and changing environment, but also identify where analyses targeting specific metabolic genes and gene expression could be applied to our best advantage.

## **2.5 Supplementary methods**

### **2.5.1 Sample collection**

Samples were collected as part of the *Galathea III* project *DNA of the Polar Seas*, during the LOMROGII expedition of the Swedish icebreaker *Oden*. Two MYI samples were collected (I13 and I14) along with three surface seawater samples (S15, S16, S17). Sampling locations are shown in Fig. S2.1 and coordinates given in Table S2.1.

Ice samples were collected using a 9 cm hand-powered ice corer sterilized with ethanol. The snow was cleared away from the site prior to coring, the ice floe was cored through its full thickness, and the cores were extruded onto autoclaved foil. To generate sufficient material for

downstream processing two adjacent cores were collected. Both cores were then cut into sections using an ethanol-sterilized hand saw and placed together in large double sterile plastic bags in an ethanol-cleaned cooler. All sections from the entire ice thickness were included in the sample. Shipboard each sample was melted at room temperature into 17 L of 0.2  $\mu\text{m}$  filtered brine (salinity of 80) composed from Sigma-Aldrich artificial sea salts (St. Louis, MI, USA). The volume of brine was estimated so that the final melt volume would approximate a salinity of 20, near the salinity of natural brines within the ice matrix as calculated using the equations of Cox and Weeks (1983). Final melt salinities for I13 and I14 were 35 and 22, respectively. This step was taken to prevent cell lysis as a result of osmotic stress. As soon as the cores were melted (while still cold) the meltwater was filtered using a peristaltic pump. One liter of milliQ water and 500 ml of sample were flushed through the tubing prior to sample processing. Volumes filtered are shown in Table S5. Samples were prefiltered through a 2.0  $\mu\text{m}$  filter and collected on 0.2  $\mu\text{m}$  SteriPak filters (Billerica, MA, USA), buffered with 30 mL of sterile sucrose buffer, and frozen at  $-80^{\circ}\text{C}$  until analysis. Seawater samples were collected by a rosette attached to the CTD. Seawater was transferred aseptically to sterile polycarbonate bottles and filtered immediately using a procedure identical to that described for sea ice. Volumes filtered are shown in Table S5.

### **2.5.2 Environmental parameters**

Parallel cores were taken to determine temperature, chlorophyll *a*, cell count, salinity, and pEPS (Fig. S2.5). These cores were sectioned on site into 20-cm sections and placed within sterile Whirlpak bags (Ft. Atkinson, WI, USA). Shipboard the cores were melted into an equal volume of 0.2  $\mu\text{m}$  filtered brine. Salinity was calculated from the concentrations of salt required to produce the final salinity observed with a refractometer from a known volume of melt solution

combining with a known volume of ice melt. Aliquots for the analysis of chlorophyll *a* were filtered onto Whatman GFF filters (0.7  $\mu\text{m}$  retention) and stored in dark tubes at  $-20\text{ }^{\circ}\text{C}$  until analysis by fluorescence (Arar & Collins, 1997). Aliquots for cell counts were fixed with 37 % formaldehyde to a final concentration of 2 % and stored at  $0\text{ }^{\circ}\text{C}$  until staining with the DNA specific stain 4',6-diamidino-2-phenylindole (DAPI) and analysis by epifluorescence microscopy. Fractions for the analysis of pEPS were filtered onto 0.4  $\mu\text{m}$  polycarbonate filters and frozen at  $-20\text{ }^{\circ}\text{C}$  until analysis by the phenol-sulfuric assay as described previously (Marx, Carpenter, & Deming, 2009).

### **2.5.3 DNA extraction, purification, and sequencing**

DNA extraction, purification, and sequencing were performed at the University of Copenhagen, Denmark. Cells were lysed directly on filter and the DNA extracted with phenol-chloroform. One nanogram of DNA was amplified using the universal prokaryotic primers PRK341F (CCTAYGGGRBGCAACAG) and PRK806R (GGACTACNNGGGTATCTAAT) (Yu, Lee, Kim, & Hwang, 2005). PCR product was reamplified using the same primers with different 5' tags. Product from the second PCR was purified by gel electrophoresis. Tagged PCR products were pooled together ( $5 \times 10^5$  DNA copies sample<sup>-1</sup>) and sequenced on a GS FLX pyrosequencer (454 Life Sciences, Branford, CT, USA). Sequencing in this manner produced an average read length of 221 bp. Sample coverage is illustrated by rarefaction curves for various definitions of OTU (Fig. S2.6).

Performance of the PRK341F and PRK806R primers was analyzed using the OligoCheck program. The primer set was determined to have good performance at the domain level for both Bacteria and Archaea (78.7 % and 67.7 % efficiency); however, the primer set does not perform well for specific clades in both domains (Fig. S2.4). Within the Bacteria, Planctomycetes and

Cyanobacteria have low binding efficiency with the primer set. Within the Archaea, the Crenarchaeota, particularly marine Crenarchaeota, have poor binding efficiency.

### **Acknowledgements**

We thank Christian Marcussen (GEUS) for the opportunity to participate in LOMROG 2009, Matthias Wietz and Jens Blom for assistance with sampling, Shelly Carpenter, Pia Friis, and Peter Holmsgaard for assistance with laboratory analysis, and John Hopper for calculating the ice drift velocities. This work was funded by the Villum Kann Rasmussen Foundation and Lundbeckfonden, the Danish Agency for Science, Technology, and Innovation (SR), an award from the Gordon and Betty Moore Foundation (JSB), an NSF IGERT fellowship through the University of Washington Astrobiology Program (JSB), and NSF OPP award 0908724 (JWD).



## References

- Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V., & Polz, M. F. (2004). Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *J. Bacteriol.*, 186(9), 2629–2635. doi: 10.1128/jb.186.9.2629-2635.2004
- Alonso-Sáez, L., Sánchez, O., Gasol, J. M., Balagué, V., & Pedrós-Alio, C. (2008). Winter-to-summer changes in the composition and single-cell activity of near-surface Arctic prokaryotes. *Environ. Microbiol.*, 10(9), 2444–2454. doi: 10.1111/j.1462-2920.2008.01674.x
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nuc. Acids Res.*, 25, 2289–3402.
- Arar, E. J., & Collins, G. B. (1997). Method 445.0: *In vitro* determination of chlorophyll a and pheophytin a in marine and freshwater algae by fluorescence: National Exposure Research Laboratory, Office of Research and Development.
- Bano, N., & Hollibaugh, J. T. (2002). Phylogenetic composition of bacterioplankton assemblages from the Arctic Ocean. *Appl. Environ. Microbiol.*, 68(2), 505–518. doi: 10.1128/aem.68.2.505-518.2002
- Barber, D. G., Galley, R., Asplin, M. G., De Abreu, R., Warner, K.-A., Pu, . . . Julien, S. (2009). Perennial pack ice in the southern Beaufort Sea was not as it appeared in the summer of 2009. *Geophys. Res. Lett.*, 36(24), L24501. doi: 10.1029/2009gl041434
- Barbeyron, T., Lerat, Y., Sassi, J.-F., Le Panse, S., Helbert, W., & Nyvall Collen, P. (2010). *Persicivirga ulvanivorans* sp. nov., a marine Flavobacteriaceae degrading ulvan from green algae. *Int. J. Syst. Evol. Microbiol.*, ijs.0.024489–024480. doi: 10.1099/ij.s.0.024489-0
- Boe, J., Hall, A., & Qu, X. (2009). September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nature Geosci.*, 2(5), 341–343.
- Bowman, J. P., McCammon, S. A., Brown, M. V., Nichols, D. S., & McMeekin, T. A. (1997). Diversity and association of psychrophilic bacteria in Antarctic sea ice. *Appl. Environ. Microbiol.*, 63, 3068–3078.
- Brinkmeyer, R., Knittel, K., Jurgens, J., Weyland, H., Amann, R., & Helmke, E. (2003). Diversity and structure of bacterial communities in Arctic versus Antarctic pack ice. *Appl. Environ. Microbiol.*, 69(11), 6610–6619. doi: 10.1128/aem.69.11.6610-6619.2003
- Brown, M. V., & Bowman, J. P. (2001). A molecular phylogenetic survey of sea-ice microbial communities. *FEMS Microbiol. Ecol.*, 35, 267–275.

- Cavalieri, D. J., Gloersen, P., Parkinson, C. L., Comiso, J. C., & Zwally, H. J. (1997). Observed hemispheric asymmetry in global sea ice changes. *Science*, 278(5340), 1104–1106. doi: 10.1126/science.278.5340.1104
- Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.*, 11, 265–270.
- Chao, A., & Shen, T.-J. (2003). Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environ. Ecol. Stat.*, 10(4), 429–443.
- Cole, J. R., Chai, B., Farris, R. J., Wang, Q., A. S. Kulam-Syed-Mohideen, D. M. M., Bandela, A. M., . . . Tiedje, J. M. (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nuc. Acids Res.*, 35, D169–D172. doi: 10.1093/nar/gkl889
- Collins, R. E., Rocap, G., & Deming, J. W. (2010). Persistence of bacterial and archaeal communities in sea ice through an Arctic winter. *Environ. Microbiol.*, 12(7), 1828–1841.
- Comiso, J. C., Parkinson, C. L., Gersten, R., & Stock, L. (2008). Accelerated decline in the Arctic sea ice cover. *Geophys. Res. Lett.*, 35(1), L01703. doi: 10.1029/2007gl031972
- Connelly, T. L., Tilburg, C. M., & Yager, P. L. (2006). Evidence for psychrophiles outnumbering psychrotolerant marine bacteria in the springtime coastal Arctic. *Limnol. Ocean.*, 51(2), 1205–1210.
- Cox, G. F. N., & Weeks, W. F. (1983). Equations for determining the gas and brine volumes in sea-ice samples. *J. Glaciol.*, 29(102), 306–316.
- Deming, J. W. (2007). Life in ice formations at very low temperatures. In C. Gerday & N. Glansdorff (Eds.), *Physiology and Biochemistry of Extremophiles*. Washington, D.C.: ASM Press.
- Deming, J. W. (2009). Sea ice bacteria and viruses. In D. N. Thomas & G. S. Dieckmann (Eds.), *Sea ice – An Introduction to its Physics, Chemistry, Biology, and Geology*. Oxford: Blackwell Science Ltd.
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., . . . Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, 72(7), 5069–5072. doi: 10.1128/aem.03006-05
- Eicken, H. (2009). Ice Sampling and Basic Sea Ice Core Analysis. In H. Eicken, R. Gradinger, M. Salganek, K. Shirasawa, D. Perovich & M. Lepparanta (Eds.), *Field Techniques for Sea Ice Research*. Fairbanks.

- Eicken, H., Lensu, M., Leppäranta, M., Tucker, W. B., III, Gow, A. J., & Salmela, O. (1995). Thickness, structure, and properties of level summer multiyear ice in the Eurasian sector of the Arctic Ocean. *J. Geophys. Res.*, *100*(C11), 22697–22710. doi: 10.1029/95jc02188
- Galand, P. E., Casamayor, E. O., Kirchman, D. L., & Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *PNAS*, *106*(52), 22427–22432.
- Haas, C. (2004). Late-summer sea ice thickness variability in the Arctic Transpolar Drift 1991–2001 derived from ground-based electromagnetic sounding. *Geophys. Res. Lett.*, *31*(9). doi: 10.1029/2003gl019394
- Huse, S. M., Welch, D. M., Morrison, H. G., & Sogin, M. L. (2010). Ironing out the wrinkles in the rare biosphere through improved OTU clustering. *Environ. Microbiol.*, *12*(7), 1889–1898.
- Jungblut, A. D., Lovejoy, C., & Vincent, W. F. (2009). Global distribution of cyanobacterial ecotypes in the cold biosphere. *ISME J.*, *4*(2), 191–202.
- Junge, K., Imhoff, J., Staley, J., & Deming, J. (2002). Phylogenetic diversity of numerically important Arctic sea-ice bacteria cultured at subzero temperature. *Microb. Ecol.*, *43*(3), 315–328.
- Kellogg, C., Renfro, A., Cochran, J., & Deming, J. (2011). Evidence for microbial attenuation of particle flux in the Amunden Gulf and Beaufort Sea: elevated activity on sinking aggregates. *Polar Biol.*, Available online in advance of print.
- Kimura, M. (1983). *The neutral theory of molecular evolution*: Cambridge University Press.
- Kirchman, D. L., Cottrell, M. T., & Lovejoy, C. (2010). The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ. Microbiol.*, *12*(5), 1132–1143. doi: 10.1111/j.1462-2920.2010.02154.x
- Kirchman, D. L., Moran, X. A. G., & Ducklow, H. (2009). Microbial growth in the polar oceans: role of temperature and potential impact of climate change. *Nat. Rev. Micro.*, *7*(6), 451–459.
- Kurilenko, V. V., Christen, R., Zhukova, N. V., Kalinovskaya, N. I., Mikhailov, V. V., Crawford, R. J., & Ivanova, E. P. (2010). *Granulosicoccus coccoides* sp. nov., isolated from leaves of seagrass (*Zostera marina*). *Int. J. Syst. Evol. Microbiol.*, *60*(4), 972–976. doi: 10.1099/ijs.0.013516-0
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., . . . Higgins, D. G. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics*, *23*(21), 2947–2948. doi: 10.1093/bioinformatics/btm404

- Lozupone, C., & Knight, R. (2005). UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.*, 71(12), 8228–8235.
- Marx, J. G., Carpenter, S. D., & Deming, J. W. (2009). Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium *Colwellia psychrerythraea* strain 34H under extreme conditions. *Can. J. Microbiol.*, 55, 63–72.
- Matsen, F. A. (2010). pplacer. Seattle: Fred Hutchinson Cancer Center. Retrieved from <http://matsen.fhcrc.org/pplacer/index.html>
- Mavromatis, K., Abt, B., Brambilla, E., Lapidus, A., Copeland, A., Desphande, S., . . . Kyrpides, N. C. (2010). Complete genome sequence of *Coralimargarita akajimensis* type strain (04OKA010-24T). *SIGS*, 2(3).
- McLean, A. L. (1918). Bacteria of ice and snow in Antarctica. *Nature*, 102(2550), 35–39.
- Nedashkovskaya, O. I., Kim, S. B., Han, S. K., Rhee, M. S., Lysenko, A. M., Falsen, E., . . . Bae, K. S. (2004). *Ulvibacter litoralis* gen. nov., sp. nov., a novel member of the family Flavobacteriaceae isolated from the green alga *Ulva fenestrata*. *Int. J. Syst. Evol. Microbiol.*, 54(1), 119–123. doi: 10.1099/ijs.0.02757-0
- Needleman, S. B., & Wunsch, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Mol. Biol.*, 48(3), 443–453.
- Nghiem, S. V., Rigor, I. G., Perovich, D. K., Clemente-Colón, P., Weatherly, J. W., & Neumann, G. (2007). Rapid reduction of Arctic perennial sea ice. *Geophys. Res. Lett.*, 34(19), L19504. doi: 10.1029/2007gl031138
- Petri, R., & Imhoff, J. F. (2001). Genetic analysis of sea-ice bacterial communities of the western Baltic Sea using an improved double gradient method. *Polar Biol.*, 24, 252–257.
- Pfirman, S. L., Colony, R., Nürnberg, D., Eicken, H., & Rigor, I. (1997). Reconstructing the origin and trajectory of drifting Arctic sea ice. *J. Geophys. Res.*, 102(C6), 12575–12586. doi: 10.1029/96jc03980
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glockner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl. Acids Res.*, 35(21), 7188–7196. doi: 10.1093/nar/gkm864
- Quince, C., Lanzen, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., . . . Sloan, W. T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Meth.*, 6(9), 639–641.

- Rigor, I. G., & Wallace, J. M. (2004). Variations in the age of Arctic sea-ice and summer sea-ice extent. *Geophys. Res. Lett.*, *31*(9). doi: 10.1029/2004gl019492
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, *75*(23), 7537–7541. doi: 10.1128/aem.01541-09
- Simpson, E. H. (1949). Measurement of diversity. *Nature*, *163*, 688.
- Stamatakis, A. (2006). RAxML-VI-HP: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–2690. doi: 10.1093/bioinformatics/btl446
- Stroeve, J., Holland, M. M., Meier, W., Scambos, T., & Serreze, M. (2007). Arctic sea ice decline: Faster than forecast. *Geophys. Res. Lett.*, *34*(9), L09501. doi: 10.1029/2007gl029703
- Vincent, W. F. (2010). Microbial ecosystem responses to rapid climate change in the Arctic. *ISME J.*, *4*(9), 1087–1090.
- Waleron, M., Waleron, K., Vincent, W. F., & Wilmotte, A. (2007). Allochthonous inputs of riverine picocyanobacteria to coastal waters in the Arctic Ocean. *FEMS Microbiol. Ecol.*, *59*(2), 356–365. doi: 10.1111/j.1574-6941.2006.00236.x
- Yoon, J., Yasumoto-Hirose, M., Katsuta, A., Sekiguchi, H., Matsuda, S., Kasai, H., & Yokota, A. (2007). *Coraliomargarita akajimensis* gen. nov., sp. nov., a novel member of the phylum 'Verrucomicrobia' isolated from seawater in Japan. *Int J Syst Evol Microbiol*, *57*(5), 959–963. doi: 10.1099/ijs.0.64755-0
- Yu, Y., Lee, C., Kim, J., & Hwang, S. (2005). Group-specific primer and probe sets to detect methanogenic communities using quantitative real-time polymerase chain reaction. *Biotechnology and Bioengineering*, *89*(6), 670–679.

**Table 2.1.** Mean values of richness and diversity for multiple OTU definitions for the MYI and seawater communities. Student's T-test results and significance between MYI and seawater samples are given as t and p, respectively

OTU	Chao1				Np-Shannon's index				Simpsons's index			
	Sea ice	Seawater	t	p	Sea ice	Seawater	t	p	Sea ice	Seawater	t	p
<b>unique</b>	1552	3042	5.57	0.11	3.66	3.93	0.25	0.84	0.15	0.11	0.65	0.56
<b>97 %</b>	614	1309	2.34	0.26	3.38	3.72	0.34	0.79	0.16	0.11	0.65	0.56
<b>93 %</b>	305	768	2.38	0.25	3.08	3.51	0.47	0.72	0.17	0.12	0.75	0.51
<b>90 %</b>	230	600	2.15	0.28	2.84	3.35	0.53	0.69	0.20	0.12	0.87	0.45

**Table 2.2.** Estimates of richness and diversity for OTU definition of 100 % similarity for each MYI and seawater sample.

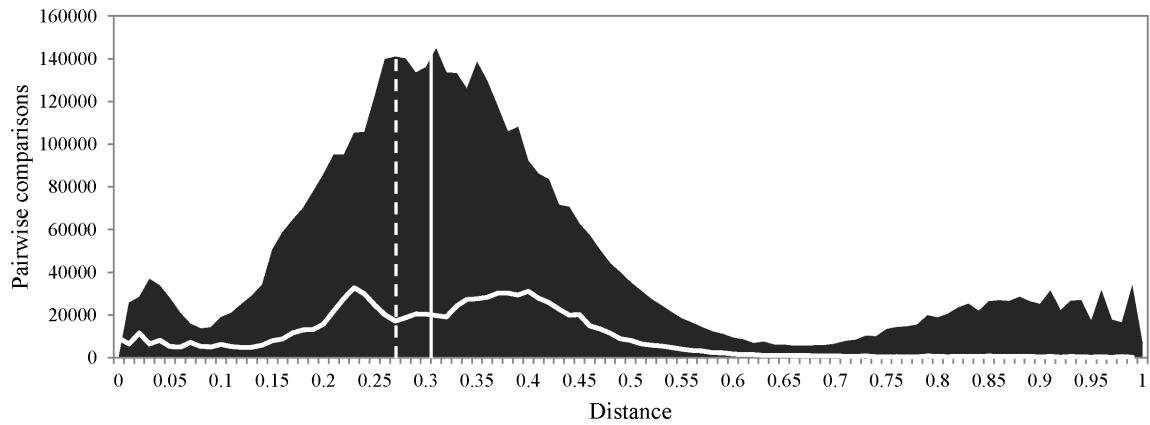
Sample	Unique reads	Chao1	95 % confidence interval	Estimated % coverage	Shannon's index	Simpson's index	95% confidence interval
I13	543	1516	1242-1896	35.8	2.96	0.237	0.23-0.25
I14	552	1588	1298-1990	34.8	4.36	0.071	0.065-0.076
S15	1057	3214	2769-3776	32.9	4.19	0.079	0.076-0.082
S16	853	2770	2320-3358	30.8	3.64	0.137	0.137-0.148
S17	996	3143	2675-3742	31.7	3.95	0.111	0.106-0.115

**Table 2.3.** Pairwise comparison of sample similarity for OTU definition of 100 % similarity.

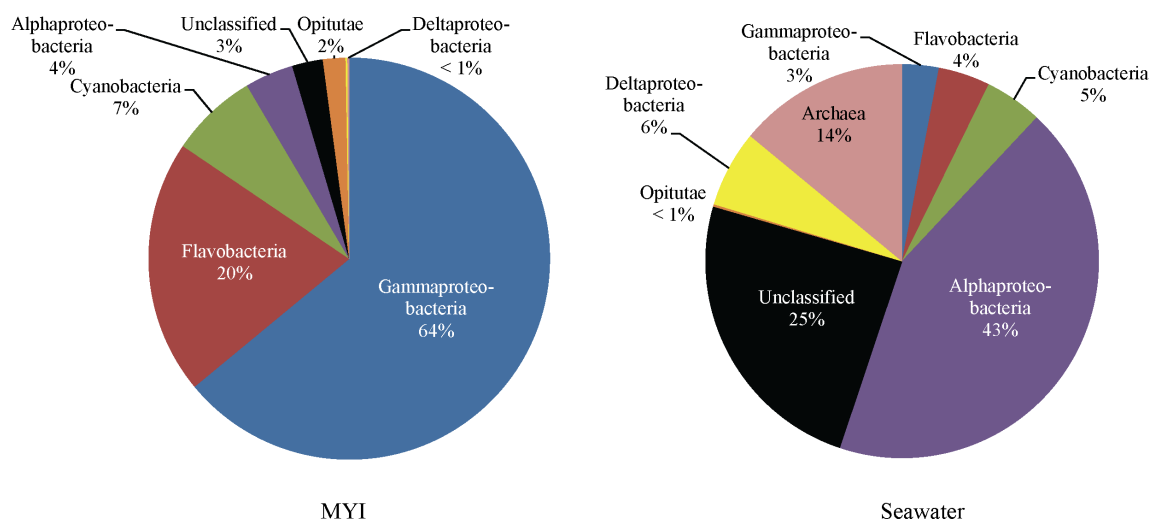
High values for Sørensen's index and UniFrac, indicate a high level of similarity. Value in bold were shown to be significantly similar using a separate weighted UniFrac analysis on a neighbor-joining tree bootstrapped 1000 times.

Sample pair		Richness based Sørensen	Unweighted UniFrac	Abundance based Sørensen	Weighted UniFrac
I13	I14	<b>0.281</b>	<b>0.592</b>	<b>0.914</b>	<b>0.195</b>
I13	S15	0.013	0.841	0.544	0.684
I13	S16	0.012	0.835	0.303	0.635
I13	S17	0.014	0.813	0.548	0.694
I14	S15	0.015	0.813	0.391	0.651
I14	S16	0.027	0.539	0.317	0.226
I14	S17	0.022	0.836	0.362	0.698
S15	S16	0.239	0.824	0.961	0.642
S15	S17	0.212	0.515	0.942	0.164
S16	S17	<b>0.199</b>	<b>0.469</b>	<b>0.953</b>	<b>0.098</b>



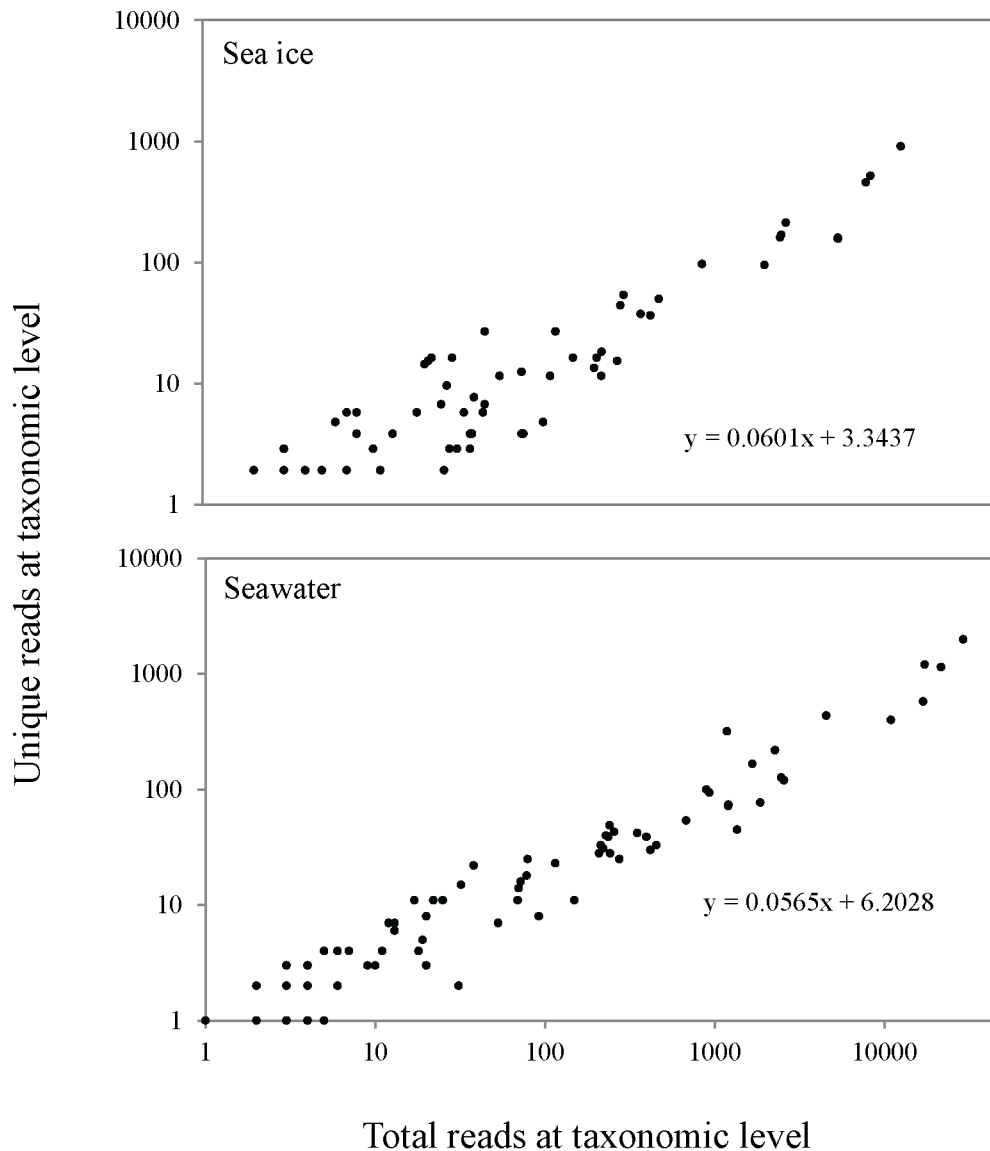


**Fig. 2.1. Distribution of genetic distances for the MYI and seawater communities.** The shaded region indicates the distribution of distances for the seawater community. The multi-modal distribution is a result of the dominance of a single genus (*Pelagibacter*) and the two-domain structure of the community. The white lateral line indicates the distribution of distances for the MYI community and the solid vertical line, the mean distance. The dashed vertical line shows the mean distance for the seawater community, excluding the domain Archaea.

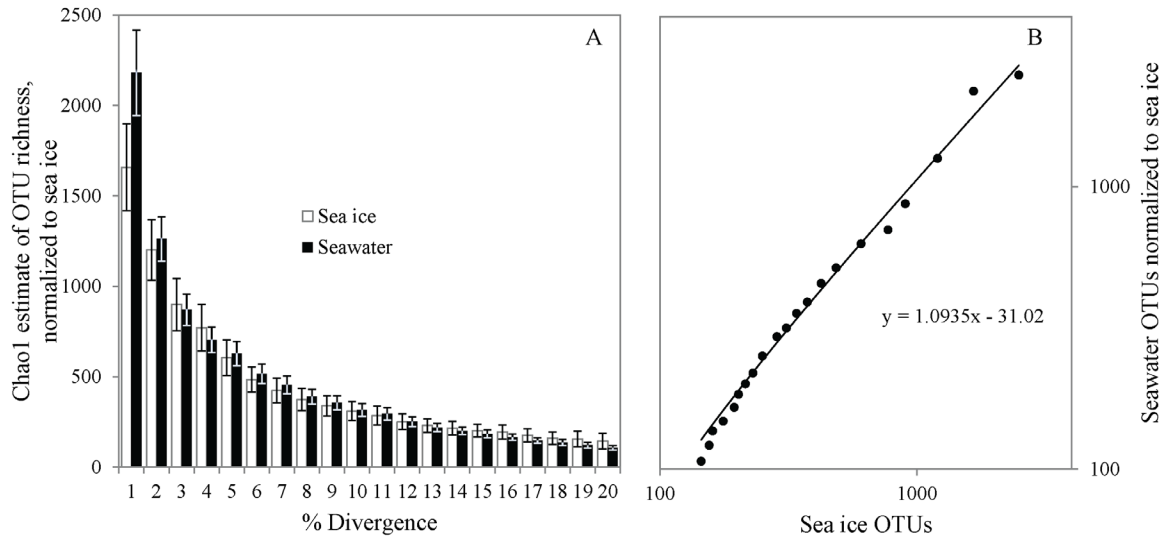


**Fig. 2.2. Numerically dominant clades in the MYI and seawater communities.**

Gammaproteobacteria and Flavobacteria dominated the ice community; Alphaproteobacteria and Archaea, the seawater community. Deltaproteobacteria and Archaea comprised the smallest fractions of the MYI community; Opitutae, of the seawater community.



**Fig. 2.3. Divergence of the 16S rRNA gene using a taxonomy-based method.** Divergence of the 16S rRNA gene (unique reads) with abundance (total reads) for seawater (top):  $r = 0.96$ ,  $p < 10^{-4}$ ,  $df = 130$  and for MYI (bottom):  $r = 0.96$ ,  $p < 10^{-4}$ ,  $df = 117$ . The MYI community appears more divergent than the seawater community, with more new clades formed from increasing membership: the slopes of the lines are significantly different (Student's T-test,  $t = 15.06$ ,  $p < 10^{-4}$ ,  $df = 247$ ).



**Fig. 2.4. Divergence of the 16S rRNA gene using a taxonomy-independent analysis.** The number of OTUs in the seawater community for varying definitions of OTU (0–20 % difference in read sequence) was normalized to the lower richness observed in the MYI community. The accumulation of new OTUs at low percentage divergence appears faster for the seawater than sea ice community (A). Error bars indicate 95 % confidence. The slope of a best fit line (shown with log scale on both axis) for values from both environments (B) is significantly different than one ( $df = 18, p = 10^{-4}$ ).

**Table S2.1.** Locations and environmental parameters for the MYI and seawater samples.

Concentration factor indicates the average fold increase in concentration for constituents partitioned into the brine phase of MYI. Values indicated by a dash are not available, value indicated by nd were not detected.

Sample	Type	Date	Latitude	Longitude	Salinity	T (°C)	Abundance (cells ml <sup>-1</sup> x 10 <sup>4</sup> )	Chlorophyll <i>a</i> (mg m <sup>-3</sup> )	pEPS†	Concentration factor
I13	Ice	8/21/2010	89°27.12' N	129°19.70' W	1.9*	-1.1	1.68	0.025	2.56	12
I14	Ice	8/26/2010	88°42.00' N	69°45.00' W	0.7*	-1.2	8.40	0.055	2.29	35
S15	Water	8/15/2010	88°42.66' N	158°53.42' E	32.57	-1.5	-	-	-	-
S16	Water	8/21/2010	88°27.13' N	129°20.37' W	31.78	-1.5	-	-	-	-
S17	Water	8/26/1010	88°43.74' N	58°27.28' W	30.92	-1.5	18.40	0.037	nd	-

†particulate exopolymers in µg glucose equivalents ml<sup>-1</sup>

\*average value for melted core determined by refractometer, all other salinities determined conductometrically

**Table S2.2.** Pairwise comparison of sample similarity for 97, 93, and 90 % similarity definitions of OTU using richness and abundance-based Sørensen indices.

Sample pair		OTU	Richness based Sørensen	Abundance based Sørensen	OTU	Richness based Sørensen	Abundance based Sørensen	OTU	Richness based Sørensen	Abundance based Sørensen
I13	I14	97	0.469	0.944	93	0.499	0.943	90	0.553	0.947
I13	S15	97	0.044	0.565	93	0.082	0.604	90	0.125	0.622
I13	S16	97	0.072	0.322	93	0.116	0.377	90	0.152	0.419
I13	S17	97	0.074	0.594	93	0.075	0.586	90	0.128	0.624
I14	S15	97	0.051	0.416	93	0.086	0.451	90	0.123	0.474
I14	S16	97	0.092	0.306	93	0.127	0.377	90	0.177	0.394
I14	S17	97	0.065	0.389	93	0.093	0.403	90	0.139	0.439
S15	S16	97	0.428	0.978	93	0.531	0.984	90	0.727	0.990
S15	S17	97	0.445	0.971	93	0.577	0.981	90	0.647	0.987
S16	S17	97	0.539	0.983	93	0.631	0.990	90	0.662	0.994

**Table S2.3.** Classification of reads at each taxonomic level for the MYI community (I13 and I14 combined). Taxa not previously identified in sea ice, and not found within the seawater samples of this study are in bold.

<b>Domain</b>	<b>Abundance</b>	<b>Unique sequences</b>
unclassified	13	6
Bacteria	12336	946
Archaea	3	3
<b>Phylum</b>		
unclassified	161	29
Proteobacteria	8192	540
Bacteroidetes	2622	221
Cyanobacteria	845	101
Verrucomicrobia	369	39
unclassified_Bacteria	88	9
<b>TM7</b>	<b>39</b>	<b>8</b>
Actinobacteria	34	6
Euryarchaeota	2	2
<b>Class</b>		
unclassified	293	56
Gammaproteobacteria	7697	476
Flavobacteria	2462	176
Cyanobacteria	845	101
Alphaproteobacteria	471	52
<b>Opitutae</b>	<b>218</b>	<b>19</b>
Verrucomicrobiae	148	17
Sphingobacteria	117	28
<b>TM7_genera_incertae_sedis</b>	<b>39</b>	<b>8</b>
Actinobacteria	34	6
Deltaproteobacteria	18	6
Betaproteobacteria	2	2
unclassified_Euryarchaeota	2	2
unclassified_Bacteroidetes	1	1
Bacteroidetes_incertae_sedis	1	1
unclassified_Verrucomicrobia	1	1
<b>Spartobacteria</b>	<b>1</b>	<b>1</b>
unclassified_Proteobacteria	1	1
<b>Epsilonproteobacteria</b>	<b>1</b>	<b>1</b>
<b>Order</b>		
unclassified	2680	329
Pseudomonadales	5294	166
Flavobacteriales	2462	176
Chloroplast	845	101
Rhodobacterales	421	38
<b>Puniceococcales</b>	<b>204</b>	<b>17</b>
<b>Verrucomicrobiales</b>	<b>148</b>	<b>17</b>
Sphingobacteriales	117	28
Alteromonadales	45	28
Acidimicrobiales	31	3
Rickettsiales	27	10

Oceanospirillales	22	17
unclassified_Deltaproteobacteria	13	1
<b>Methylococcales</b>	<b>13</b>	<b>4</b>
Sphingomonadales	7	2
<b>Chromatiales</b>	<b>7</b>	<b>6</b>
Thiotrichales	4	1
unclassified_Gammaproteobacteria	3	2
Burkholderiales	2	2
Actinobacteridae	2	2
Rhodospirillales	1	1
<b>Fulvivirga</b>	<b>1</b>	<b>1</b>
<b>Bdellovibrionales</b>	<b>1</b>	<b>1</b>
<b>Spartobacteria_genera_incertae_sedis</b>	<b>1</b>	<b>1</b>
<b>Campylobacteriales</b>	<b>1</b>	<b>1</b>
<b>Family</b>		
unclassified	2910	384
Moraxellaceae	5292	164
Flavobacteriaceae	2422	167
Rhodobacteraceae	421	38
Bacillariophyta	281	46
Chlorophyta	217	12
<b>Puniceicoccaceae</b>	<b>204</b>	<b>17</b>
Cryptomonadaceae	99	5
unclassified_Chloroplast	95	11
Rubritaleaceae	74	4
<b>Verrucomicrobiaceae</b>	<b>74</b>	<b>13</b>
Saprospiraceae	55	12
Cyclobacteriaceae	38	4
Alteromonadaceae	29	17
SAR11	27	10
Acidimicrobidae_incertae_sedis	25	1
Oceanospirillaceae	21	16
<b>Methylococcaceae</b>	<b>13</b>	<b>4</b>
Cryomorphaceae	10	3
Pseudoalteromonadaceae	8	4
Sphingomonadaceae	7	2
<b>Granulosicoccaceae</b>	<b>6</b>	<b>5</b>
Cytophagaceae	5	2
Piscirickettsiaceae	4	1
Sphingobacteriaceae	3	1
Pseudomonadaceae	2	2
Comamonadaceae	2	2
Actinomycetales	2	2
<b>Micrococcineae</b>	<b>2</b>	<b>1</b>
Acidimicrobiales	1	1
Rhodospirillaceae	1	1
Shewanellaceae	1	1
<b>Bacteriovoracaceae</b>	<b>1</b>	<b>1</b>
<b>Campylobacteraceae</b>	<b>1</b>	<b>1</b>
Halomonadaceae	1	1
<b>Genus</b>		
unclassified	3895	532
Psychrobacter	5291	163
Polaribacter	1962	99



Octadecabacter	269	16
<b>Coralimargarita</b>	<b>197</b>	<b>14</b>
Flavobacterium	109	12
Psychroflexus	76	4
Rubritalea	74	4
<b>Persicivirga</b>	<b>45</b>	<b>7</b>
<b>Lewinella</b>	<b>44</b>	<b>6</b>
<b>Jannaschia</b>	<b>37</b>	<b>4</b>
Roseibacillus	37	3
unclassified_Rhodobacteraceae	33	3
<b>Maribacter</b>	<b>28</b>	<b>3</b>
unclassified_Flavobacteriaceae	27	4
Pelagibacter	27	10
<b>Algoriphagus</b>	<b>26</b>	<b>2</b>
<b>Luteolibacter</b>	<b>25</b>	<b>7</b>
Ilumatobacter	25	1
Oleispira	20	15
Loktanella	12	1
<b>Methylobacterium</b>	<b>11</b>	<b>2</b>
Cyclobacterium	11	1
Pseudoalteromonas	8	4
Glaciecola	8	6
unclassified_Saprospiraceae	7	3
unclassified_Sphingomonadaceae	6	1
<b>Granulosicoccus</b>	<b>6</b>	<b>5</b>
Marinobacter	5	1
<b>Cycloclasticus</b>	<b>4</b>	<b>1</b>
Haliea	4	2
<b>Pedobacter</b>	<b>3</b>	<b>1</b>
<b>Ulvibacter</b>	<b>3</b>	<b>2</b>
Pseudomonas	2	2
Sulfitobacter	2	1
<b>Micrococcineae</b>	<b>2</b>	<b>2</b>
Acidimicrobinae	1	1
<b>Rhodobacter</b>	<b>1</b>	<b>1</b>
<b>Thalassobacter</b>	<b>1</b>	<b>1</b>
Sphingomonas	1	1
unclassified_Rhodospirillaceae	1	1
Shewanella	1	1
<b>Leeuwenhoekella</b>	<b>1</b>	<b>1</b>
<b>Peredibacter</b>	<b>1</b>	<b>1</b>
Owenweeksia	1	1
<b>Arcobacter</b>	<b>1</b>	<b>1</b>
Halomonas	1	1

**Table S2.4.** Classification of reads at each taxonomic level for the seawater community (S15, S16, and S17 combined).

<b>Domain</b>	<b>Abundance</b>	<b>Unique sequences</b>
unclassified	2927	106
Bacteria	29032	1997
Archaea	2558	120
<b>Phylum</b>		
unclassified	4541	436
Proteobacteria	21508	1148
Bacteroidetes	2263	219
Cyanobacteria	1855	77
unclassified_Archaea	1353	45
Euryarchaeota	1203	74
Verrucomicrobia	929	94
unclassified_Bacteria	396	39
Actinobacteria	240	49
Acidobacteria	213	33
Firmicutes	13	7
Planctomycetes	3	2
<b>Class</b>		
unclassified	7228	641
Alphaproteobacteria	16877	578
Deltaproteobacteria	2463	127
Cyanobacteria	1855	77
Flavobacteria	1662	167
unclassified_Euryarchaeota	1200	72
Gammaproteobacteria	1180	319
unclassified_Proteobacteria	678	54
unclassified_Verrucomicrobia	418	30
Verrucomicrobiae	350	42
Actinobacteria	240	49
Acidobacteria_Gp6	208	28
Opitutae	78	18
Sphingobacteria	70	14
Clostridia	6	4
Planctomycetacia	3	2
Bacilli	1	1
<b>Order</b>		
unclassified	17205	1207
Rickettsiales	10910	400
Chloroplast	1855	77
Flavobacteriales	1662	167
unclassified_Deltaproteobacteria	1159	53
Verrucomicrobiales	350	42
unclassified_Gammaproteobacteria	276	45
Rhodobacterales	274	25
Acidimicrobiales	229	40
Gp6	208	28
Rhodospirillales	79	25
Puniceicoccales	72	16
Sphingobacteriales	70	14

Alteromonadales	38	22
Chromatiales	32	15
unclassified_Alphaproteobacteria	17	2
Pseudomonadales	17	11
Oceanospirillales	13	6
Kordiimonadales	7	4
Sneathiellales	6	2
Rhizobiales	6	4
Methylococcales	6	4
Opitutales	5	1
Sphingomonadales	5	1
Clostridiales	4	2
Thiotrichales	4	3
Planctomycetales	3	2
Caulobacteriales	2	2
Myxococcales	1	1
Actinobacteridae	1	1
unclassified_Actinobacteria	1	1
<b>Family</b>		
unclassified	19711	1410
SAR11	10910	400
unclassified_Chloroplast	1319	25
Flavobacteriaceae	891	100
Rhodobacteraceae	274	25
Cryomorphaceae	255	43
Rubritaleaceae	242	28
Acidimicrobiales	220	31
Cryptomonadaceae	149	11
Bacillariophyta	115	23
Puniceicoccaceae	72	16
Verrucomicrobiaceae	69	11
Saprospiraceae	53	7
unclassified_Flavobacteriales	42	4
unclassified_Rhodospirillales	29	6
Alteromonadaceae	25	11
Rhodospirillaceae	22	11
Ectothiorhodospiraceae	18	4
Moraxellaceae	12	7
Oceanospirillaceae	11	4
Flammeovirgaceae	10	3
Kordiimonadaceae	7	4
Sneathiellaceae	6	2
Methylococcaceae	6	4
Opitutaceae	5	1
Sphingomonadaceae	5	1
Pseudomonadaceae	5	4
Thiotrichales_incertae_sedis	4	3
Chlorophyta	3	1
Planctomycetaceae	3	2
unclassified_Chromatiales	3	3
Acidimicrobidae_incertae_sedis	3	3
unclassified_Rhizobiales	2	1
Colwelliaceae	2	1
unclassified_Sphingobacteriales	2	1

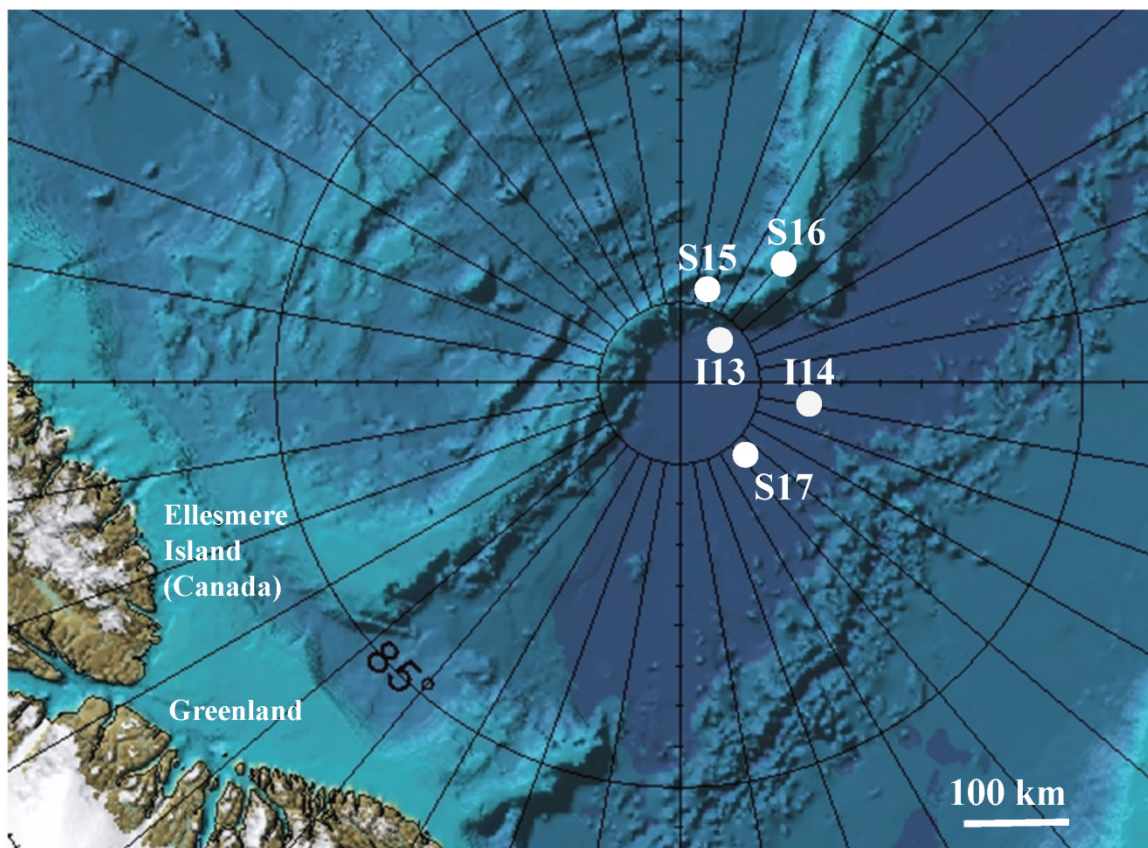
Hyphomonadaceae	2	2
Moritellaceae	2	2
Halomonadaceae	1	1
Aurantimonadaceae	1	1
Idiomarinaceae	1	1
Sorangiineae	1	1
Pseudoalteromonadaceae	1	1
Actinomycetales	1	1
Rhodobiaceae	1	1
Hahellaceae	1	1
<b>Genus</b>		
unclassified	21960	1584
Pelagibacter	10910	400
Polaribacter	453	33
Rubritalea	242	28
Owenweeksia	236	39
Acidimicrobinae	220	31
Tenacibaculum	92	8
unclassified_Flavobacteriaceae	91	13
unclassified_Rhodobacteraceae	59	2
unclassified_Puniceicoccaceae	51	12
Persicirhabdus	31	2
Pelagicoccus	20	3
Haliae	20	8
Roseovarius	19	5
Psychrobacter	12	7
Oleispira	11	4
Roseibacillus	9	3
unclassified_Flammeovirgaceae	8	1
Kordiimonas	7	4
Sneathiella	6	2
unclassified_Ectothiorhodospiraceae	6	2
unclassified_Rhodospirillaceae	5	1
Opitutus	5	1
Sphingomonas	5	1
Sulfitobacter	4	1
Marinobacter	4	2
unclassified_Planctomycetaceae	3	2
Pseudomonas	3	2
Ilumatobacter	3	3
Crocinitomix	2	1
Flavobacterium	2	1
Salagentibacter	2	1
Aureispira	2	2
Moritella	2	2
Psychroserpens	2	2
Halomonas	1	1
Aurantimonas	1	1
Idiomarina	1	1
Polyangiaceae	1	1
Robiginitomaculum	1	1
Pseudoalteromonas	1	1
Glaciecola	1	1
Thalassobius	1	1

Rhodobium	1	1
Endozoicomonas	1	1

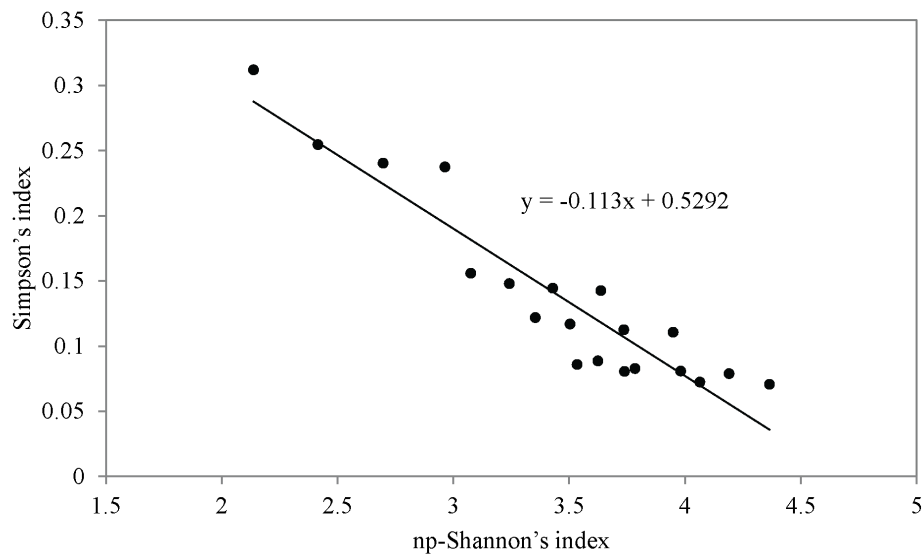
**Table S2.5.** Quantities from DNA collection, isolation, and processing for each MYI and seawater sample.

<b>Sample</b>	<b>Volume sampled (L)</b>	<b>DNA quantity (ng <math>\mu\text{L}^{-1}</math>)</b>	<b>Reads</b>	<b>High quality reads</b>
I13	40†	4.89	8 324	8 300
I14	61†	12.5	4 073	4 052
S15	49	5.18	11 777	11 649
S16	84	16.7	11 387	11 321
S17	28	6.33	11 600	11 547

† Volume includes ice melt and 17 L of added brine solution

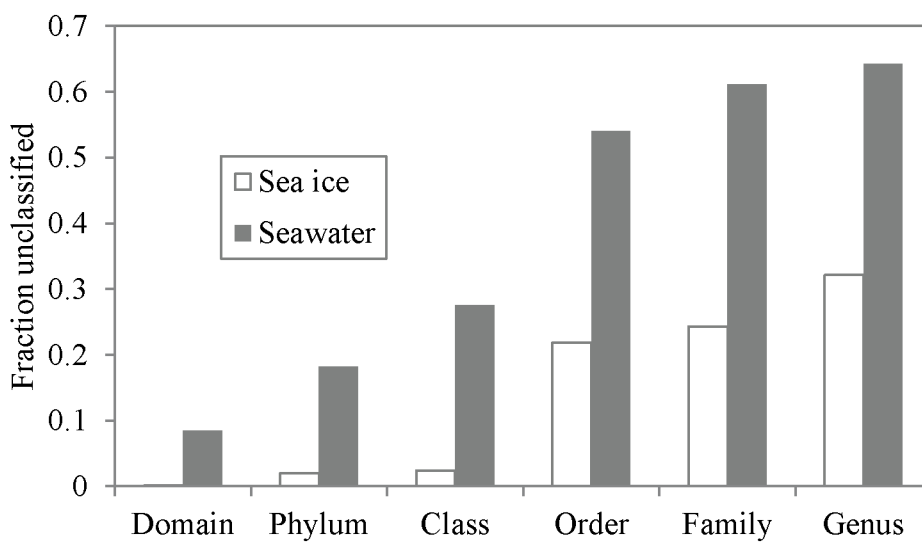


**Fig. S2.1. Sampling locations in the Arctic Ocean.** Two MYI samples (I13 and I1) and three surface seawater samples (S15, S16, S17) were collected.

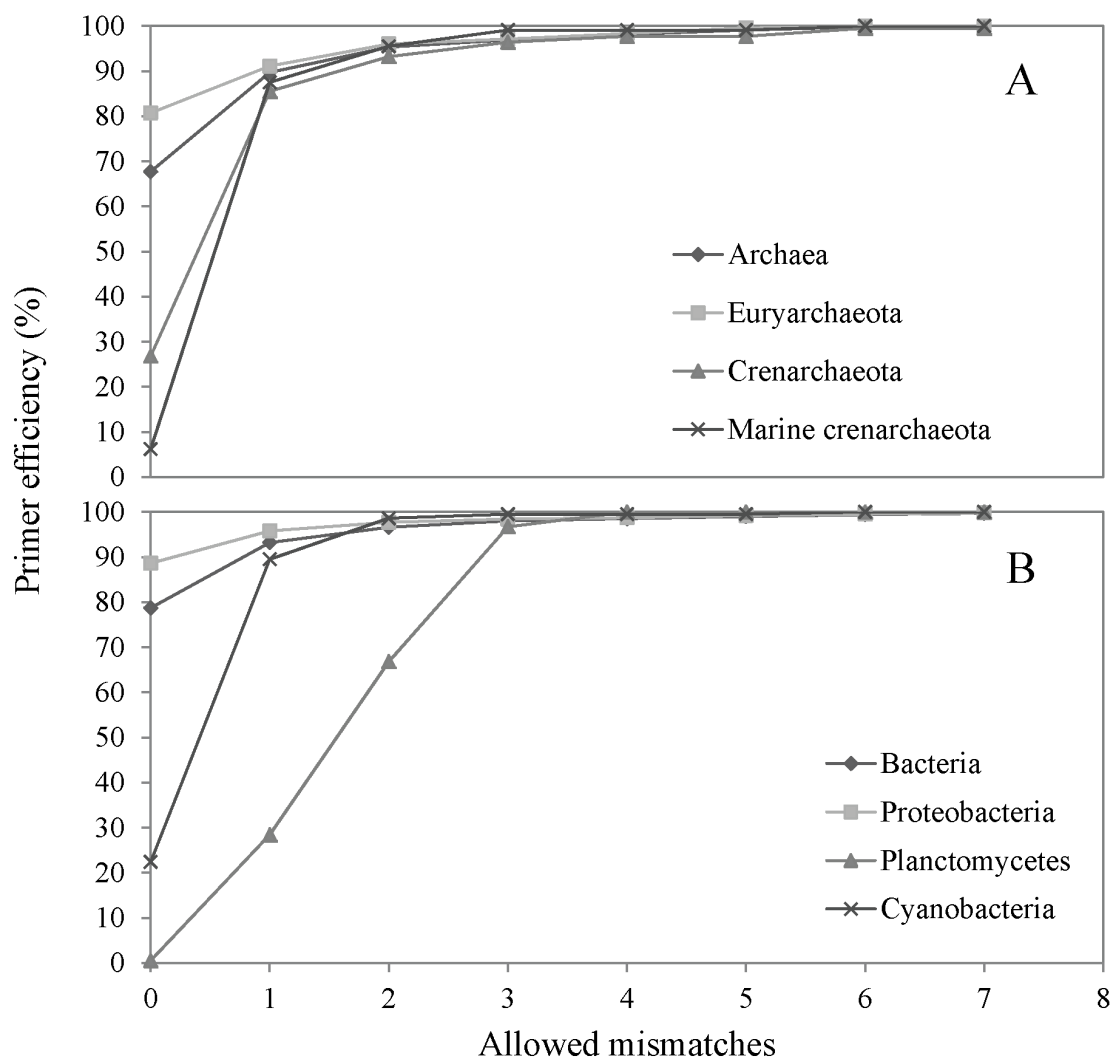


**Fig. S2.2. Correlation between Simpson's index of diversity and the non-parametric implementation of Shannon's index.** Good agreement was observed for all samples at all definitions of OTU ( $r = 0.94$ ,  $df = 18$ ,  $p < 10^{-4}$ ).

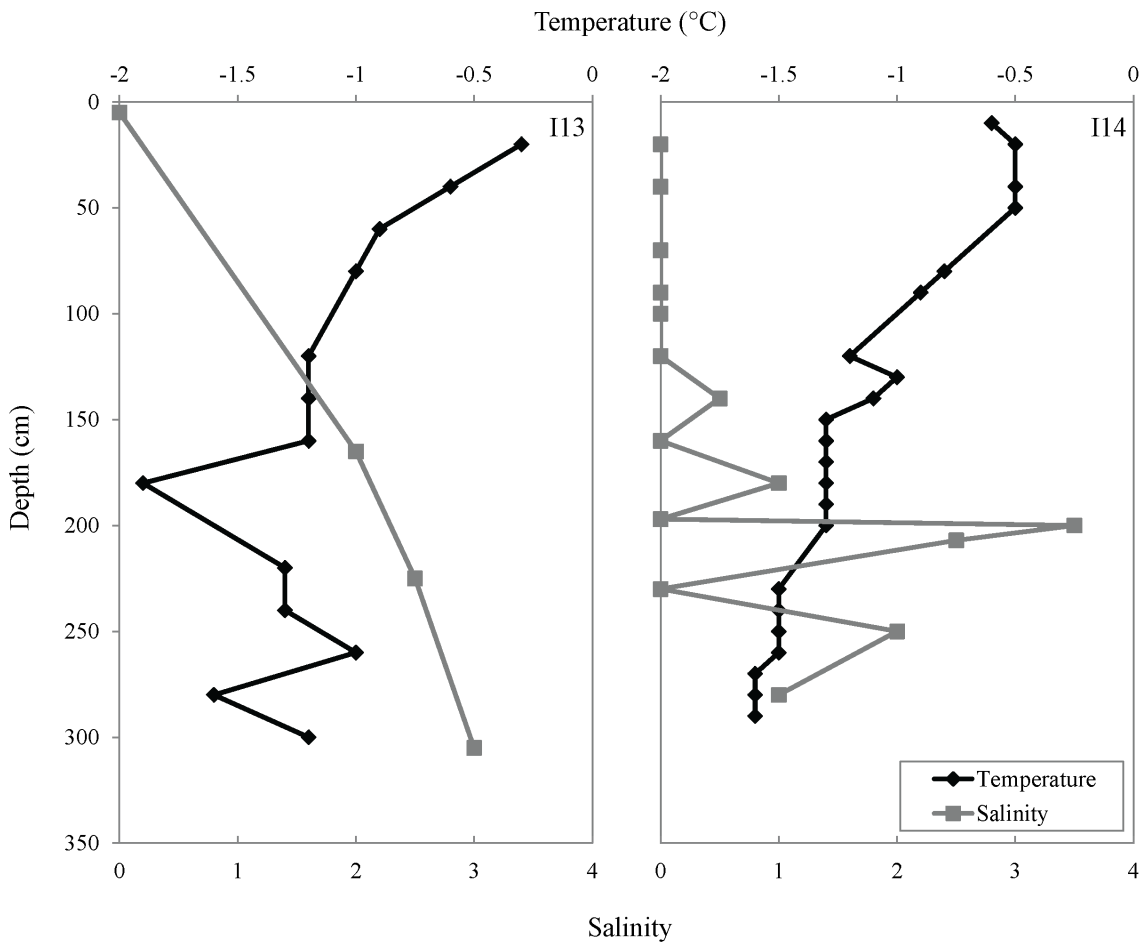




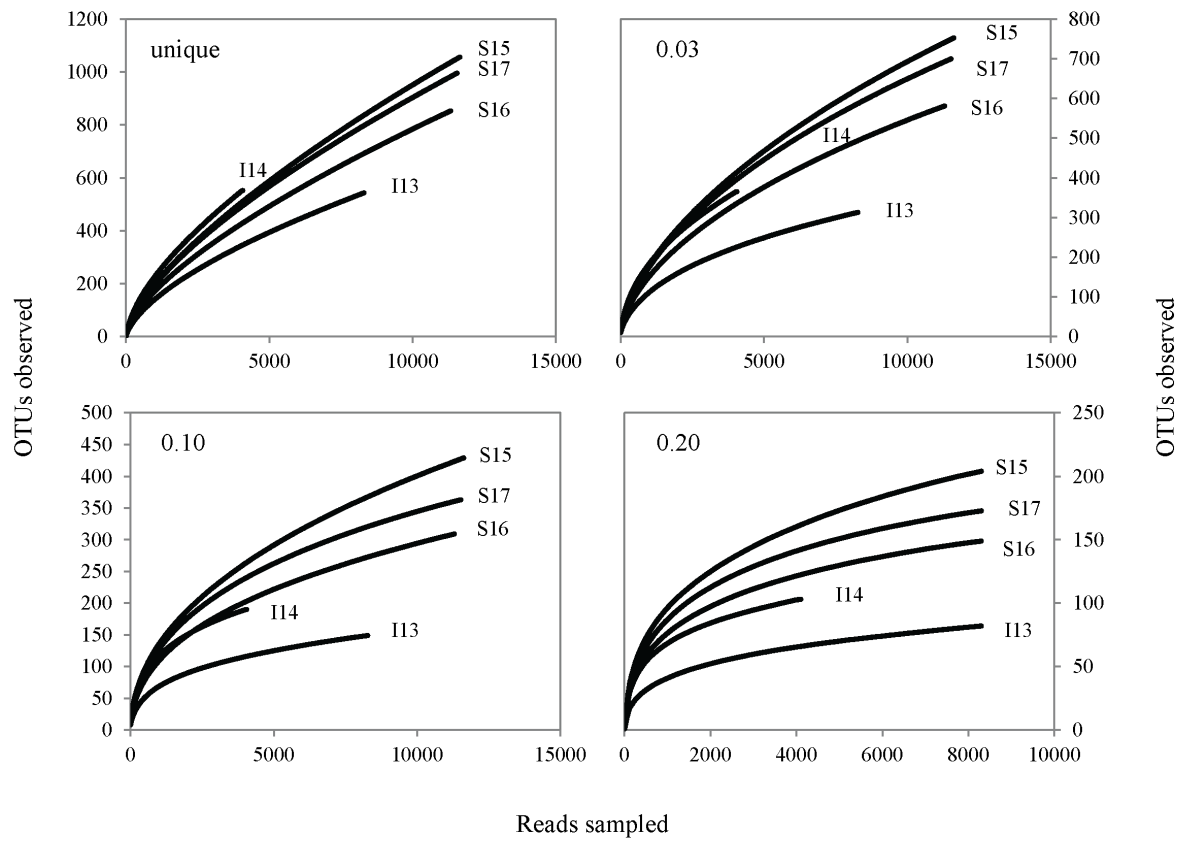
**Fig. S2.3. Fraction of reads unclassified at taxonomic level.** The percentage of reads unclassified at each taxonomic level was determined using Mothur's classification tool and a custom database of 14,956 high quality bacterial and 2,297 high quality archaeal sequences.



**Fig. S2.4. Primer set exclusions.** *In silico* analysis of the primer set suggested good coverage of the domains Bacteria, though specific clades were likely underreported in the dataset as a result of exclusion by the primer set. In the domain Archaea, Crenarchaeota are likely highly underreported as a result of primer set exclusion.



**Fig. S2.5. Ice core profiles.** Temperature and salinity profiles for I13 and I14 are shown.



**Fig. S2.6. Rarefaction curves.** Rarefaction curves at 100, 97, 90, and 80 % similarity for MYI (I13 and I14) and seawater (S15, S16, S17) communities were created using Mothur.

## Chapter 3

### **Selective occurrence of Rhizobiales in frost flowers on the surface of young sea ice near**

### **Barrow, Alaska and distribution in the polar marine rare biosphere**

(Citation: Bowman, J. S., Larose, C., Vogel, T. M., & Deming, J. W. (2013). Selective occurrence of Rhizobiales in frost flowers on the surface of young sea ice near Barrow, Alaska and distribution in the polar marine rare biosphere. *Environmental microbiology reports*, 5(4), 575–582.)

#### ABSTRACT

Frost flowers are highly saline ice structures that grow on the surface of young sea ice, a spatially extensive environment of increasing importance in the Arctic Ocean. In a previous study, we reported organic components of frost flowers in the form of elevated levels of bacteria and exopolymers relative to underlying ice. Here, DNA was extracted from frost flowers and young sea ice, collected in springtime from a frozen lead offshore of Barrow, Alaska, to identify bacteria in these understudied environments. Evaluation of the distribution of 16S rRNA genes via four methods (microarray analysis, T-RFLP, clone library and shotgun metagenomic sequencing) indicated distinctive bacterial assemblages between the two environments, with frost flowers appearing to select for Rhizobiales. A phylogenetic placement approach, used to evaluate the distribution of similar Rhizobiales sequences in other polar marine studies, indicated that some of the observed strains represent widely distributed members of the marine rare biosphere in both the Arctic and Antarctic.

### 3.1 Introduction

Newly formed sea ice, or young ice, is an environment of increasing areal importance, given recent losses in multiyear ice across the Arctic Ocean (Nghiem et al., 2012). It also represents an extensive, extreme, and understudied microbial habitat. In the Arctic Ocean alone, approximately  $8 \times 10^6 \text{ km}^2$  of sea ice is formed every year (Perovich & Richter-Menge, 2009), all of which transitions through a young ice phase. The formation of young ice is not limited to the autumn freeze-up, as substantial quantities of ice are formed throughout the winter and spring over open leads and polynyas. Young ice can be difficult to access in any season, which has limited microbial (and other) studies that require direct sampling or large sample volumes.

In early work examining the temporal progression of bacteria in young ice in austral autumn, Grossman and Dieckmann (1994) observed an initial reduction in activity (leucine turnover) followed by rapid recovery (higher rates) and higher bacterial abundances in the interior of the ice as it aged. Other signs of stress and recovery were reported by Collins and Deming (2012a, 2012b), who observed high concentrations of virus-like particles and bacteria in young ice in Arctic winter. Most of the Bacteria (and Archaea) that recover from, or are unaffected by, the freeze-in process appear to overwinter successfully in the interior of very cold (down to  $-27^\circ\text{C}$ ) Arctic sea ice (Collins and Deming, 2010). In warmer ( $-7^\circ\text{C}$ ) sub-Arctic ice, some bacteria may even grow within the ice during winter (Aslam et al., 2012). The surface of bare (snow-free) young ice, however, presents an environment more extreme than interior or bottom ice due to exposure to atmospheric conditions. In a previous study of young ice sampled near Barrow, Alaska, and the North Pole, we observed an enrichment of bacterial cells on the ice surface that could be attributed to upward expulsion of brine during ice formation (J. S. Bowman & Deming, 2010). When atmospheric conditions at the ice surface allow (temperature below –

8°C and sufficient humidity), centimeter-scale ice structures known as frost flowers (FF) grow abundantly on the young ice, wicking up brine in the process (Perovich & Richter-Menge, 1994). The highly saline FF that we examined, including freezer-grown FF at –21°C, hosted bacterial concentrations higher than those in the expelled brines (absent FF) or the ice and water below (Bowman and Deming, 2010). In an ice-tank study based on sub-Arctic water and lower air temperature (–13°C), Aslam et al. (2012) also observed higher bacterial numbers in FF that grew on the young ice surface, compared to those in the starting source water (though bacterial growth in the tank and ice complicates this analysis).

Young ice in general is expected to contain particulate and dissolved organic material, because the process of sea ice formation entrains a wide assortment of microorganisms, exudates, and detrital particles present in surface water (Riedel, Michel, Gosselin, & LeBlanc, 2007). Exudates may then increase in concentration within the ice as a result of initial stress imposed during freeze-in, which can trigger diatoms and bacteria to produce extracellular polymeric substances (EPS) (R. E. Collins, Rocap, & Deming, 2010; C. Krembs, Eicken, Junge, & Deming, 2002; Riedel et al., 2007). Brine expelled to the ice surface will thus carry organic compounds as well as bacteria. In both natural frost flowers and surface-expelled brines, we observed examples of high concentrations of particulate EPS (Bowman and Deming, 2010); in a study of natural frost flowers, Beine et al. (2012) estimated DOC concentrations to be nearly three-fold higher in FF than seawater. In their tank study under sub-arctic conditions, Aslam et al. (2012) reported elevated levels of POC, DOC, and dissolved EPS (dissolved carbohydrates) in frost flowers over those in interior ice brines and underlying source seawater. The potential biological or enzymatic fate of organic compounds delivered to the ice surface is not known.

Abiotic reactions at the young ice surface have been considered for their significance to polar atmospheric chemistry. Unexpectedly high concentrations of aldehydes measured in surface-ice samples, including frost flowers, have been attributed to photolysis of higher molecular weight organics such as EPS delivered to the ice surface by expelled brine (Thomas A. Douglas et al., 2012). Arctic frost flowers also contain high concentrations of mercury, relative to surrounding snow and ice environments, a sign that they are likely linked to atmospheric mercury depletion events (T. A. Douglas et al., 2005). The young ice surface and nearby open leads have also been implicated in the formation of halide radicals that cause significant reductions in tropospheric ozone during the polar spring (Kaleschke et al., 2004; Nghiem et al., 2012).

The opportunistic sampling that enabled an initial investigation of bacteria on the surface of young ice (Bowman and Deming, 2010) did not yield enough material for us to characterize the communities involved. In April 2010, follow-on sampling at an offshore lead (Fig. 3.1) yielded sufficient frost flowers and underlying young ice to extract DNA for phylogenetic analyses of the bacterial 16S rRNA gene by microarray, terminal restriction fragment length polymorphism (T-RFLP), clone library sequencing, and metagenomics. Our objectives were to determine whether this surface-ice environment harbored a microbial community similar to or distinct from the underlying young ice and to identify organisms for future study that may be involved in the uptake or alteration of substrates at the ice-air interface. To place our results in the context of the broader polar marine environment we constructed a database of 1,865,271 16S rRNA gene sequences or sequence reads from 14 studies of Arctic and Antarctic environments to evaluate the distribution of clades of interest as defined by sequences generated in our clone libraries.



### 3.2 Results and Discussion

At the genus level the results of our microarray analysis, normalized to a fraction of total fluorescence for each sample (Fig. 3.2), depict significant differences ( $\alpha = 0.95$ ) between the young ice and frost flower (FF) samples by SIMPROF analysis of Euclidean distance (Clarke, Somerfield, & Gorley, 2008), with some genera notably more abundant in FF. In particular, the *Mesorhizobium*, *Rhizobium*, *Sphingomonas*, *Moorella*, and *Cellulophaga* had abundant fluorescent signals in FF samples but were reduced or undetected in young ice. Together the *Rhizobium* and *Mesorhizobium*, the most numerically abundant Rhizobiales in both the young ice and FF samples, comprised  $31.5 \pm 5.4$  % (mean, SD) of the fluorescent signal in FF compared to  $11.7 \pm 8.4$  % in young ice, a statistically significant difference by Student's T-test ( $p = 0.0076$ ,  $t = 3.94$ ). *Sulfitobacter*, *Levilinea*, *Anabaena*, *Bacillus*, and *Halanaerobium* were overall reduced in FF relative to young ice. Traces from the T-RFLP analysis also demonstrated a significant distinction between FF and young ice samples ( $p < 0.05$ ) using the SIMPROF test described above (Fig. S3.1). Most of the peaks could not be identified from the single restriction enzyme used in this study. The dominant peak from the FF samples, however, which averaged 74.7 % of total fluorescent area ( $n = 4$ ,  $SD = 5.5$  %), corresponded to a fragment length of 260 bp. This length is close to the 264 bp predicted for six clones classified as *Rhizobium* (barrow\_FF\_25-27, barrow\_FF\_29, barrow\_FF\_32, barrow\_FF\_37) from the FF library of 20 clones. A second abundant peak, at 224 bp, corresponded to a predicted fragment length of 228 bp for eight clones from the young ice clone library with 31 members (barrow\_YI\_33, barrow\_YI\_40, barrow\_YI\_44, barrow\_YI\_53, barrow\_YI\_54, barrow\_YI\_57, barrow\_YI\_59, and barrow\_YI\_62). These eight clones were classified as chloroplasts. Although the source of the 4 bp difference between the observed and predicted fragment lengths could not be identified, the

uniformity of the offset, lack of significant peaks at 224 and 264, and low probability of obtaining multiple clones from rare members of the community are consistent with the 264 peak belonging to the *Rhizobium* clades identified in the clone library.

Given the shallow nature of the clone library a further confirmation of rhizobial dominance was sought by extracting the 16S rRNA gene fragments from metagenomes derived from whole genome amplified DNA from one young ice sample (YI1) and one FF sample (FF4) (full data to be presented elsewhere). Of the 14,393 Bacterial 16S rRNA reads from young ice, 1,506 (10.5 %) classified as Rhizobiales. From FF 57,224 out of 73,915 (77.4 %) Bacterial 16S rRNA reads classified as Rhizobiales.

A phylogenetic placement approach was used to place sequences that classified as Rhizobiales from the clone libraries and metagenomes of this study, and from our 16S rRNA database for polar regions (Table 3.1), on a reference tree of Rhizobiales sequences from the Ribosomal Database Project (RDP) (Cole et al., 2007). This analysis demonstrated that three of our Barrow clones (barrow\_FF\_33, barrow\_FF\_40, and barrow\_YI\_46) are closely related to the RDP reference sequence S00021625, *Blastobacter aggregatus* (Fig. 3.1), along with 1,024 of our metagenomic reads (Table 3.1, Fig. 3.4). Also placed with this reference sequence were 66 reads from Galand et al. (2009), 60 reads from Kirchman et al. (2010), and 29 reads from SRP001216 (Table 3.1, Fig. 3.4). Four other Barrow clones (barrow\_FF\_27, barrow\_FF\_29, barrow\_FF\_32, and barrow\_FF\_37) were placed with the closely related reference sequence SR000015514, *B. capsulatus* (Fig. 3.1) along with 5,883 metagenomic reads, as were 405 reads from Kirchman et al. (2010), 251 reads from Galand et al. (2009), and 306 reads from SRP001216 (Fig. 3.4). None of the other polar studies considered (Table 3.1) yielded a sequence that placed with either *B. aggregatus* or *B. capsulatus*, but other members of the Rhizobiales were identified in the marine

(sea ice) studies of Collins et al. (2010) and Bowman et al. (2012), the marine studies of Gryzmanski et al. (2012) and De Corte et al. (2012), the terrestrial (soil) studies of Wallenstein et al. (2007), Männistö et al. (2009), and Chu et al. (2011), and the snow study of Larose et al. (2010). No Rhizobiales were reported in studies of seawater, soil, or snow by Kellogg and Deming (2009), Campbell et al. (2010), or Harding et al. (2011). Other Rhizobiales clones from our study clustered close to sequences on the reference tree but not to the Arctic Rhizobiales obtained in the other studies evaluated. Clones barrow\_YI\_45 and barrow\_FF\_26 are closely related to *Rhizobium giardinii*, while barrow\_FF\_28 and barrow\_FF\_31 are closely related to *Methylobacterium radiotolerans* (not shown). A significant number of metagenomic 16S rRNA reads clustered with *R. multihospitium* (Fig. 3.4).

The taxonomy of the genus *Rhizobium* is in a state of revision, with a proposal pending to include *Blastobacter aggregatus* in the *Rhizobium* as *R. aggregatum*, on the basis of GC content, fatty acid composition, and 16S rRNA gene sequence (Kaur, Verma, & Lal, 2011). The need for this revision is seen in our reference tree used for phylogenetic placement (Fig. 3.3), with the position of *B. capsulatus* in a monophyletic clade with *B. aggregatus*. To reflect the nomenclature of our reference tree, which relies on the nomenclature of RDP 10, we have used *B. aggregatus* here with the understanding that this 16S rRNA gene sequence is of the genus *Rhizobium*. Similarly, 16S rRNA gene sequences showing a close similarity to *B. capsulatus* are presumed to be *Rhizobium*, even though a proposal to rename *B. capsulatus* has not been published. Both microarray and T-RFLP results indicated a strong distinction between the microbial communities in FF samples and in underlying young ice. Coupled with the metagenomic reads and clone library results they point to enrichment at the young ice surface of members of the Rhizobiales. This clade is observed in the rare marine biosphere of both the

Arctic and Antarctic, but has not been described in either of these marine environments. The genomic (Flores et al., 1988; Young et al., 2006) and metabolic (Amachi, Kamagata, Kanagawa, & Muramatsu, 2001; Appanna & Preston, 1987; Coulter et al., 1999; Foster, Moy, & Rogers, 2000) flexibility of known Rhizobiales gives them biogeochemical relevance in many environments, but potential dominance at the surface of young ice suggests a previously unrecognized role for marine Rhizobiales in leads, polynyas and the expansive marginal sea ice zone.

Although further fieldwork is needed to conclude that the gene sequences attributed to *Blastobacter aggregatus* and *B. capsulatas* are abundant in frost flowers and on the surface of young sea ice in general, and not a unique case for our study site and time of sampling, the presence of similar 16S rRNA genes in a range of other polar marine studies implies broader significance. If *Rhizobium* species are exploiting an ecological niche in surface waters, and possibly at the surface of young sea ice where they appear to concentrate, then we would expect to find these clades in the background of the marine microbial community (as our literature survey showed) following ice melt or a shift in seasonal conditions. That no related sequences were found in the winter first-year ice study of Collins et al. (2010) or the summer multiyear ice study of Bowman et al. (2012) may reflect unique ecological opportunities present in springtime young ice (this study) and associated surface water.

Neither fine-scale temperature measurements nor evidence of bacterial activity on the surface of young ice are available, but we recorded temperatures at the ice surface, under a blanket of frost flowers, as warm as  $-6^{\circ}\text{C}$  at this and other study sites. Evidence for microbial activity in tank-grown sea ice at an average temperature of  $-7.5^{\circ}\text{C}$  was recently obtained by Aslam et al. (2012). Prior work has shown that activity in subzero water and bottom sea ice can

be robust (Alonso-Sáez, Sánchez, Gasol, Balagué, & Pedrós-Alio, 2008; Grossmann & Dieckmann, 1994; Kirchman, Elifantz, Dittel, Malmstrom, & Cottrell, 2007; Kirchman, Moran, & Ducklow, 2009; Rivkin, Anderson, & Lajzerowicz, 1996; Yager et al., 2001) and that metabolic activity can be expected at even colder temperatures. Cultured psychrophiles known from sea ice grow at temperatures as low as  $-12^{\circ}\text{C}$  (Breezee, Cady, & Staley, 2004; Wells & Deming, 2006) and respiratory activity was measured at  $-2$  to  $-20^{\circ}\text{C}$  in winter sea ice (Junge, Eicken, & Deming, 2004). Although to our knowledge no strictly psychrophilic member of the Rhizobiales has yet been described, numerous cold-adapted strains have been cultured (Bordeleau & Prévost, 1994; Prévost et al., 2003), suggesting the potential for metabolic activity of some marine species at the surface of young ice. Additional adaptations may make this clade well suited to persist at the young ice surface. *Rhizobium* spp. and closely related genera for example, have been noted for the ability to methylate halides (Amachi et al., 2001; Barloy-Hubler, Chéron, Hellégouarch, & Galibert, 2004), a possible mechanism for counteracting hydrogen peroxide (Bengtson, Bastviken, & Öberg, 2012) present at the young ice surface as a result of photochemistry or atmospheric deposition (Anastasio & Jordan, 2004).

Regardless of the potential for bacterial activity or preferential survival on the surface of young ice, bacterial growth in this study would have been limited by the short time (approximately one week) between initial ice formation and sampling. An enrichment of *Rhizobium* at the ice surface may instead be the result of a transport process that distributes bacteria unevenly within the ice during formation. As ice grows, bacteria are rejected with brine into the water column or upward onto the ice surface. The ability of the remaining bacteria to interact with the growing ice matrix, through ice-binding proteins (Raymond, Fritsen, & Shen, 2007) or the production of channel-clogging EPS (Christopher Krembs, Eicken, & Deming,

2011), or to interact with larger particles including diatoms, would contribute to the observed distribution. This hypothesis fits the model proposed by Grossman and Dieckmann (1994), that adherence to phytoplankton scavenged by frazil ice is the primary mechanism for incorporating marine bacteria into young ice (Grossmann & Dieckmann, 1994; Weissenberger & Grossmann, 1998). A key to understanding the distribution of bacteria in this system may be the epiphytic nature of some of the observed genera, including *Sulfitobacter* and *Rhizobium* (Hünken, Harder, & Kirst, 2008).

### **3.3 Conclusions**

The observation that members of the Rhizobiales dominated bacterial communities in frost flowers examined in this study brings attention to an understudied clade of marine bacteria and opens questions about possible metabolic function in an extreme environment of climatic relevance at the interface between ice and atmosphere. Taxonomic classification and phylogenetic placement of 16S rRNA gene sequences from other studies indicate that some of these genes belong to *Rhizobium* clades distributed across the Arctic and Antarctic suggesting significance beyond the study site in Barrow. The metabolic potential of the Rhizobiales, including symbiotic and asymbiotic nitrogen fixation (Elmerich, Dreyfus, Reyssset, & Aubert, 1982), the production of DMS (Todd et al., 2007), methylhalides (Amachi et al., 2001), and EPS (Appanna & Preston, 1987), implies that its polar clades could play a role in important biogeochemical processes. Work in other regions and seasons of new ice formation, and with strains brought into culture, is needed to further evaluate the enrichment effect in frost flowers and determine the functionality of observed clades in this expanding polar marine environment.

### **3.4 Supplementary Methods**

#### **3.4.1 Sample collection and DNA extraction**

Frost flowers (FF) and the underlying young ice were collected from an open lead located at 71.36816°N and 156.69784°W approximately 3.3 km offshore from Pt. Barrow. FF were collected from the young ice surface into sterile 16 L Whirlpak bags with an ethanol-rinsed aluminum shovel, an approach that allowed for sufficient volume collections (from each of four areas summing to approximately 300 m<sup>2</sup>) but also included surface brines (e.g., between FF) in the FF samples. The upper 2-4 cm of young ice were then cut and removed from each area that had been cleared of FF using an ethanol-rinsed saw and ice axe. Samples were stored at –20°C and transported to the University of Washington for further processing within two weeks.

Taking the measured ice surface temperature of –6°C as the upper bound for FF temperature, the FF internal brine salinity was at least 100.5 by the equations of Cox and Weeks (1983). To minimize osmotic shock during melting, each of the four FF and four young ice samples was melted at 2°C into 3 L of autoclaved, 0.2 µm filtered 228 ppt NaCl brine, contained in an ethanol-rinsed and UV-sterilized 20 L HDPE bucket, to achieve a final salinity of approximately 100 ppt. Melting to achieve this final salinity has been shown to minimize or prevent cell loss to osmotic shock in upper sea ice and in culture work (Ewert, Carpenter, Colangelo-Lillis, & Deming, 2012). Due to variability in sample volume and salinity, the final melt salinities achieved ranged between 76 and 98 ppt.

Melting ice samples into brine to conserve cellular DNA, as the target of this study, meant that bulk sample salinity was not measured conventionally (on samples allowed to melt directly). Instead, we estimated sample salinity by calculating it from the final melt salinity (measured by refractometer) according to Eqt. 1:

$$1. \quad S = \frac{(V_b + V_s)S_f - (V_b * S_b)}{V_s}$$

where  $S$  = sample salinity,  $V_b$  = volume of melt brine,  $V_s$  = sample volume,  $S_f$  = final salinity of sample melted into brine, and  $S_b$  = melt brine salinity. The accuracy of this approach is limited by variability in sample volume and salinity. Mean bulk salinities ( $\pm$  S.D., range) calculated for the young ice and FF samples were 8.2 ( $\pm$  0.76, 7.3–9.0) and 18 ( $\pm$  9.1, 2.9–26), respectively. These young ice values are consistent with values of 8–10 established for upper portions of young sea ice (Weeks and Lee, 1962), but we believe the FF values are underestimates due to inaccuracy in sample volume measurement. They are also lower than previously observed in the western Arctic in springtime (J. S. Bowman & Deming, 2010; Martin, Drucker, & Fort, 1995; Perovich & Richter-Menge, 1994) even accounting for brine drainage over the lifetime of the FF.

Cells were harvested from the melted samples using a 0.2  $\mu$ m Sterivex filter (Millipore). Cells were lysed in SET buffer (Kellogg & Deming, 2009) directly on the filter using 50 mg ml<sup>-1</sup> lysozyme and incubating for 30 min at 30°C. Following lysis, proteins were inactivated using 20 mg ml<sup>-1</sup> proteinase K and 90  $\mu$ l ml<sup>-1</sup> SDS. DNA was extracted using the phenol chloroform method as described in Kellogg and Deming (2009). Extracted DNA was quantified with Picogreen (Molecular Probes) following standard procedures. Aliquots (40 ml) of the melted samples were fixed with 0.2  $\mu$ m filtered 37 % formaldehyde to a final concentration of 2 %. Bacterial abundance was determined from these fixed samples by double staining with DAPI and AO and enumerating by epifluorescence microscopy, counting a minimum of 20 fields and 200 cells per sample. Mean ( $\pm$  S.D.) bacterial abundance was  $2.01 \times 10^5$  ( $\pm 7.42 \times 10^4$ ) for the FF samples and  $4.64 \times 10^4$  ( $\pm 9.27 \times 10^3$ ) for the young ice samples. This > 4-fold difference in bacterial abundance between FF and young ice (significant by a Student's T-test at  $p = 0.0061$ ) is consistent with prior observations (J. S. Bowman & Deming, 2010).

### **3.4.2 Clone library**



Two 16S rRNA gene clone libraries was developed from sample Y11 and FF4. DNA from the selected samples was PCR amplified using primers 27F (5'-AGAGTTTGATCMTGGCTCAG) and 1492R (5'-GGYTACCTTGTTACGACTT) under the following conditions: initial denaturation at 95°C for 150 s, then 32 cycles of 94°C for 30 s, 55°C for 1 min, and 72°C for 1 min, followed by a final extension at 72°C for 10 min. PCR products were cloned into the PCR4 vector and transformed into chemically competent cells using the TOPO TA cloning kit (Invitrogen) according to the manufacturer's instructions.

A total of 67 white and light blue colonies indicating successful transformants were picked and grown overnight on LB media with ampicillin. Cloned sequences were PCR amplified using primers M13F and M13R and submitted to the University of Washington High Throughput Sequencing Center for Sanger sequencing. Sequencing was carried out using primers T3, 519F, and T7. Contigs for 51 clones that successfully sequenced were assembled in DNA Dragon (Sequentix). The assembled sequences were checked for chimeras with the pintail algorithm in Mothur (Schloss et al., 2009) against a data base of 4,938 Bacterial and Archaeal sequences from greengenes (DeSantis et al., 2006) downloaded from the Mothur website ([www.mothur.org](http://www.mothur.org)). Chimera-checked sequences were classified with RDP Release 10 (Cole et al., 2007) using the online submission tool. Sequences for this study have been submitted to Genbank with accession numbers JX66803-JX668753.

### **3.4.3 Microarray analysis**

Splits of the extracted DNA from each of the four FF and young ice sample were shipped to the University of Lyon for microarray analysis of the 16S rRNA gene. DNA was PCR amplified using universal primers pA (5'-TAATACGACTCACTATAGAGAGTTTGATCCTGGCTCAG) and pH-T7 (5'-

AAGGAGGTGATCCAGCCGCA) and the Illustra Hot Start Mix RTG PCR kit (GE Healthcare). The PCR conditions were 3 min at 94°C, followed by 35 cycles of 45 s of denaturation at 94°C, 45 s of annealing at 55°C, and 90 s of elongation at 72°C. After a final 5-min extension at 72°C, PCR products were separated by 1%-agarose gel electrophoresis, purified using the NucleoSpin Extract II kit (Clontech), and transcribed. *In vitro* transcription was carried out at 37°C during 4 h with the following mix: T7 RNA buffer (5X), DDT (100 mM), 10 mM of each of the four NTPs, RNAsin (40 U  $\mu\text{L}^{-1}$ ), T7 RNA polymerase (1  $\mu\text{L}$ ) and UTP-Cy3 (5 mM). RNA was purified using the RNeasy mini kit (Qiagene) before undergoing chemical fragmentation by the addition of 5.7  $\mu\text{L}$  of a Tris Cl (1mM) and  $\text{ZnSO}_4$  (100mM) mix. Samples were incubated for 30 min at 60°C and fragmentation was stopped by placing the tubes on ice. EDTA (500 mM) was added to each tube (1.2  $\mu\text{L}$ ) followed by 1  $\mu\text{L}$  RNAsin (40 U  $\mu\text{L}^{-1}$ ) after a 1-min incubation period at 25°C. The RNA solution was then diluted to 5 ng  $\mu\text{L}^{-1}$  and a hybridization mix was prepared (v/v ratio) in a 50  $\mu\text{L}$  reaction with 2x GeX Hyb Buffer (Agilent). A total of 100 ng of RNA were then placed on the slide and incubated at 60°C for 4 h in the Agilent Hybridization Oven. Details on chip analysis can be found in Delmont et al. (2011).

#### **3.4.4 T-RFLP**

DNA from each of the four FF and four young ice samples was amplified in the same manner as for the clone library. PCR products were purified using a GeneJet purification kit (Fermentas) prior to digestion with 5 U of HAEIII at 37°C for 3 h. After digestion the DNA fragments were purified and concentrated using Sephadex 650 (Sigma-Aldrich), spun in a turbovap until dry, and redissolved in 2.0  $\mu\text{L}$  HPLC water (Sigma-Aldrich). The fragments were submitted to the Fred Hutchinson Cancer Research Center in Seattle for size fractionation and

quantification on an ABI Capillary Sequencer (Invitrogen). Fragment analysis was conducted in Gelquest (Sequentix) by aligning each sample to ladder 1200 LIZ (Invitrogen) and binning to a resolution of 1 bp.

To assign a putative classification to fragments of interest, 22,460 full length archaeal and bacterial sequences were downloaded from Silva (Pruesse et al., 2007). These sequences were trimmed to position 8 (corresponding to the start position for primer 27F) and digested *in silico* at GG|CC for restriction enzyme HAEIII. The length of the resulting fragments was compared to the fragment lengths obtained during T-RFLP. A second database was prepared in the same manner from the clone library reads produced by sequencing the forward portion of the cloned 16S rRNA gene.

### **3.4.5 Metagenomic analysis**

DNA was amplified using the SEQPLEX whole genome amplification kit (Sigma) following the manufacturer's protocol. Under conditions of stringent sterility 200-500 bp fragments were diluted in molecular grade water to approximately 100 pg  $\mu\text{l}^{-1}$  and amplified in triplicate reactions for 29 cycles. The reaction products and negative controls were visualized on a gel using standard methods. No product was observed in the two negative control lanes. The amplified material was purified using the GeneJet PCR purification kit (Fermentas) and the primers cleaved following the SEQPLEX protocol and provided reagents. Amplified material was purified an additional time with the GeneJet PCR purification kit. Primer free material was submitted to the Argonne National Lab for paired end sequencing on the Illumina platform with 100 bp read lengths utilizing one flow cell lane per sample.

Raw reads were paired by overlapping the reverse complement of the forward read with the reverse read using a custom script. Reads which did not overlap due to sequencing error

were eliminated from this preliminary analysis (24.3 % of young ice reads, 11.0 % of FF reads). Assembled reads were classified using Mothur (Schloss et al., 2009) in the manner as described for Fig. 3.4, with a cutoff of 60. Overlapped reads were also submitted to the MG-RAST server (Meyer et al., 2008). Taxonomic classification of reads by MG-RAST reports 56.2 % of totals reads as belonging to Rhizobiales for the FF sample and 7.2 % of total reads as belonging to Rhizobiales for the young ice sample.

### **3.4.6 Comparison with other data sets**

To compare the *Rhizobium* sequences observed in this study with sequences obtained in other Arctic and Antarctic studies, sequences were obtained from Genbank for the studies given in Table 3.1. Six polar marine studies and four studies of Arctic soil were used along with two studies of Arctic snow. These sequences were classified using Mothur against a data base of 84,414 Bacterial and Archaeal sequences from greengenes (DeSantis et al., 2006) obtained from the Mothur website ([www.mothur.org](http://www.mothur.org)). Sequences classified as Rhizobiales with a bootstrap score of 60 or higher were extracted, aligned to a 7,682 position alignment in Mothur, and trimmed to positions 299-6,046.

Sequence region varied between these studies and between the clones in our clone library. To resolve the issue of non-overlapping reads, a phylogenetic placement approach was selected for comparison. Near full-length, typed Rhizobiales sequences were obtained from the RDP. These sequences were aligned in Mothur, trimmed to positions 299-6,046, and redundant sequences were removed. FastTree (Price, Dehal, & Arkin, 2010) was used to create an approximate maximum-likelihood reference tree of these sequences. The phylogenetic placement program pplacer (Matsen, Kodner, & Armbrust, 2010) was then used to place the Arctic and Antarctic reads on the reference tree along with our Barrow reads. To aid in

visualization, reads from the larger Galand et al. (2009), Kirchman et al. (2010), and Antarctic data sets were viewed on the reference tree independently.

### **Acknowledgments**

This work was supported by NSF IGERT and EPA STAR fellowships to J.S.B and NSF OPP award 0908724 and Walters Endowed Professorship to J.W.D. The “Institut polaire français Paul Émile Victor “ (IPEV) helped support the phylogenetic microarray work. We are grateful to the Barrow Arctic Science Consortium and Umiaq for valuable assistance in the field, and in particular to Nok Acker and Lewis Bower for their guidance.

## References

- Alonso-Sáez, L., Sánchez, O., Gasol, J. M., Balagué, V., & Pedrós-Alio, C. (2008). Winter-to-summer changes in the composition and single-cell activity of near-surface Arctic prokaryotes. *Environ. Microbiol.*, *10*(9), 2444–2454. doi: 10.1111/j.1462-2920.2008.01674.x
- Amachi, S., Kamagata, Y., Kanagawa, T., & Muramatsu, Y. (2001). Bacteria mediate methylation of iodine in marine and terrestrial environments. *Appl. Environ. Microbiol.*, *67*(6), 2718–2722. doi: 10.1128/aem.67.6.2718-2722.2001
- Anastasio, C., & Jordan, A. L. (2004). Photoformation of hydroxyl radical and hydrogen peroxide in aerosol particles from Alert, Nunavut: implications for aerosol and snowpack chemistry in the Arctic. *Atmos. Environ.*, *38*(8), 1153–1166. doi: 10.1016/j.atmosenv.2003.11.016
- Appanna, V. D., & Preston, C. M. (1987). Manganese elicits the synthesis of a novel exopolysaccharide in an arctic *Rhizobium*. *FEBS Lett.*, *215*(1), 79–82. doi: 10.1016/0014-5793(87)80117-5
- Aslam, S., Underwood, G., Kaartokallio, H., Norman, L., Autio, R., Fischer, M., . . . Thomas, D. (2012). Dissolved extracellular polymeric substances (dEPS) dynamics and bacterial growth during sea ice formation in an ice tank study. *Pol. Biol.*, *35*(5), 661–676. doi: 10.1007/s00300-011-1112-0
- Barloy-Hubler, F., Chéron, A., Hellégouarch, A., & Galibert, F. (2004). Smc01944, a secreted peroxidase induced by oxidative stresses in *Sinorhizobium meliloti* 1021. *Microbiol.*, *150*(3), 657–664. doi: 10.1099/mic.0.26764-0
- Beine, H., Anastasio, C., Domine, F., Douglas, T., Barret, M., France, J., . . . Ullmann, K. (2012). Soluble chromophores in marine snow, seawater, sea ice and frost flowers near Barrow, Alaska. *J. Geophys. Res.*, *117*, D00R15. doi: 10.1029/2011jd016650
- Bengtson, P., Bastviken, D., & Öberg, G. (2012). Possible roles of reactive chlorine II: assessing biotic chlorination as a way for organisms to handle oxygen stress. *Environ. Microbiol.* doi: 10.1111/j.1462-2920.2012.02807.x
- Bordeleau, L., & Prévost, D. (1994). Nodulation and nitrogen fixation in extreme environments. *Plant and Soil*, *161*(1), 115–125. doi: 10.1007/bf02183092
- Bowman, J. S., & Deming, J. W. (2010). Elevated bacterial abundance and exopolymers in saline frost flowers with implications for atmospheric chemistry and microbial dispersal. *Geophys. Res. Lett.*, *37*(L13501).
- Bowman, J. S., Rasmussen, S., Blom, N., Deming, J. W., Rysgaard, S., & Sicheritz-Ponten, T. (2012). Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene. *ISME J*, *6*, 11–20.

- Breezee, J., Cady, N., & Staley, J. T. (2004). Subfreezing growth of the sea ice bacterium “*Psychromonas ingrahamii*”. *Microb Ecol*, 47(3), 300–304.
- Campbell, B. J., Polson, S. W., Hanson, T. E., Mack, M. C., & Schuur, E. A. G. (2010). The effect of nutrient deposition on bacterial communities in Arctic tundra soil. *Environ. Microbiol.*, 12(7), 1842–1854. doi: 10.1111/j.1462-2920.2010.02189.x
- Chu, H., Fierer, N., Lauber, C. L., Caporaso, J. G., Knight, R., & Grogan, P. (2011). Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environ. Microbiol.*, 12(11), 2998–3006. doi: 10.1111/j.1462-2920.2010.02277.x
- Clarke, K. R., Somerfield, P. J., & Gorley, R. N. (2008). Testing of null hypotheses in exploratory community analyses: similarity profiles and biota-environment linkage. *J. Exp. Mar. Biol. Ecol.*, 366(1–2), 56–69. doi: 10.1016/j.jembe.2008.07.009
- Cole, J. R., Chai, B., Farris, R. J., Wang, Q., A. S. Kulam-Syed-Mohideen, D. M. M., Bandela, A. M., . . . Tiedje, J. M. (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nuc. Acids Res.*, 35, D169–D172. doi: 10.1093/nar/gkl889
- Collins, R., & Deming, J. (2012a). Abundant dissolved genetic material in Arctic sea ice Part I: Extracellular DNA. *Pol. Biol.*, 34(12), 1819–1830. doi: 10.1007/s00300-011-1041-y
- Collins, R., & Deming, J. (2012b). Abundant dissolved genetic material in Arctic sea ice Part II: Viral dynamics during autumn freeze-up. *Pol. Biol.*, 34(12), 1831–1841. doi: 10.1007/s00300-011-1008-z
- Collins, R. E., Rocap, G., & Deming, J. W. (2010). Persistence of bacterial and archaeal communities in sea ice through an Arctic winter. *Environ. Microbiol.*, 12(7), 1828–1841.
- Coulter, C., Hamilton, J. T. G., McRoberts, W. C., Kulakov, L., Larkin, M. J., & Harper, D. B. (1999). Halomethane: bisulfide/halide ion methyltransferase, an unusual corrinoid enzyme of environmental significance isolated from an aerobic methylotroph using chloromethane as the sole carbon source. *Appl. Environ. Microbiol.*, 65(10), 4301–4312.
- Cox, G. F. N., & Weeks, W. F. (1983). Equations for determining the gas and brine volumes in sea-ice samples. *J. Glaciol.*, 29(102), 306–316.
- De Corte, D., Sintes, E., Yokokawa, T., & Herndl, G. J. (2012). Comparison between MICROCARD-FISH and 16S rRNA gene clone libraries to assess the active versus total bacterial community in the coastal Arctic. *Environ Microbiol Reports*, online in advance of print. doi: 10.1111/1758-2229.12013
- Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., . . . Vogel, T. M. (2011). Accessing the soil metagenome for studies of microbial diversity. *Appl. Environ. Microbiol.*, 77(4), 1315–1324. doi: 10.1128/aem.01526-10

- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., . . . Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, 72(7), 5069–5072. doi: 10.1128/aem.03006-05
- Douglas, T. A., Domine, F., Barret, M., Anastasio, C., Beine, H. J., Bottenheim, J., . . . Steffen, A. (2012). Frost flowers growing in the Arctic ocean-atmosphere-sea ice-snow interface: 1. Chemical composition. *J. Geophys. Res.*, 117, D00R09. doi: 10.1029/2011jd016460
- Douglas, T. A., Sturm, M., Simpson, W. R., Brooks, S., Lindberg, S. E., & Perovich, D. K. (2005). Elevated mercury measured in snow and frost flowers near Arctic sea ice leads. *Geophys. Res. Lett.*, 32. doi: 10.1029/2004gl022132
- Elmerich, C., Dreyfus, B., Reyssset, G., & Aubert, J. (1982). Genetic analysis of nitrogen fixation in a tropical fast-growing *Rhizobium*. *EMBO*, 1(4), 499–503.
- Ewert, M., Carpenter, S., Colangelo-Lillis, J., & Deming, J. (2012). Bacterial and extracellular polysaccharide content of brine-wetted snow over Arctic winter first-year sea ice. *J Geophys Res*, (accepted, pending revision).
- Flores, M., González, V., Pardo, M. A., Leija, A., Martínez, E., Romero, D., . . . Palacios, R. (1988). Genomic instability in *Rhizobium phaseoli*. *J. Bacteriol.*, 170(3), 1191–1196.
- Foster, L. J. R., Moy, Y. P., & Rogers, P. L. (2000). Metal binding capabilities of *Rhizobium etli* and its extracellular polymeric substances. *Biotechnol. Lett.*, 20, 857–863.
- Galand, P. E., Casamayor, E. O., Kirchman, D. L., & Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *PNAS*, 106(52), 22427–22432.
- Grossmann, S., & Dieckmann, G. S. (1994). Bacterial standing stock, activity, and carbon production during formation and growth of sea ice in the Weddell Sea, Antarctica. *Appl. Environ. Microbiol.*, 60(8), 2746–2753.
- Grzymski, J. J., Riesenfeld, C. S., Williams, T. J., Dussaq, A. M., Ducklow, H., Erickson, M., . . . Murray, A. E. (2012). A metagenomic assessment of winter and summer bacterioplankton from Antarctica Peninsula coastal surface waters. *ISME J*, 6(10), 1901–1915. doi: <http://www.nature.com/ismej/journal/v6/n10/suppinfo/ismej201231s1.html>
- Harding, T., Jungblut, A. D., Lovejoy, C., & Vincent, W. F. (2011). Microbes in High Arctic Snow and Implications for the Cold Biosphere. *Appl. Environ. Microbiol.*, 77(10), 3234–3243. doi: 10.1128/aem.02611-10
- Hünken, M., Harder, J., & Kirst, G. O. (2008). Epiphytic bacteria on the Antarctic ice diatom *Amphiprora kufferathii* Manguin cleave hydrogen peroxide produced during algal photosynthesis. *BMC Plant Biol.*, 10(4), 519–526. doi: 10.1111/j.1438-8677.2008.00040.x



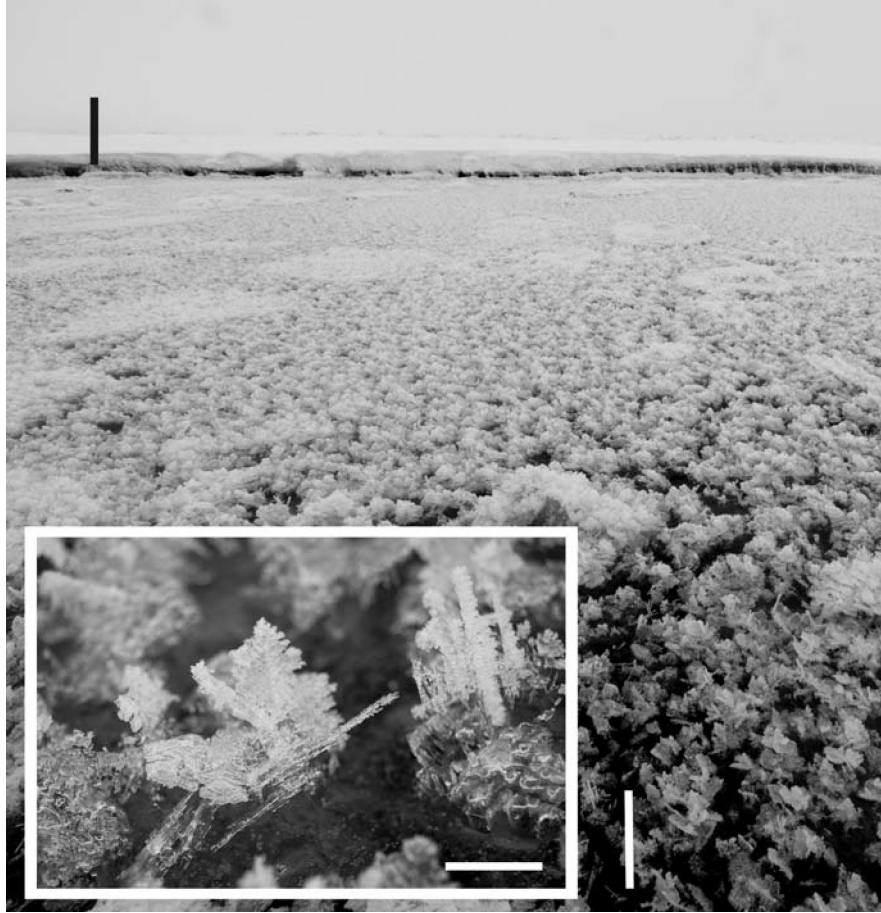
- Junge, K., Eicken, H., & Deming, J. W. (2004). Bacterial activity at  $-2$  to  $-20^{\circ}\text{C}$  in Arctic wintertime sea ice. *Appl. Environ. Microbiol.*, *70*(1), 550–557. doi: 10.1128/aem.70.1.550-557.2004
- Kaleschke, L., Richter, A., Burrows, J., Afe, O., Heygster, G., Notholt, J., . . . Jacobi, H. W. (2004). Frost flowers on sea ice as a source of sea salt and their influence on tropospheric halogen chemistry. *Geophys. Res. Lett.*, *31*. doi: 10.1029/2004gl020655
- Kaur, J., Verma, M., & Lal, R. (2011). *Rhizobium rosettiformans* sp. nov., isolated from a hexachlorocyclohexane dump site, and reclassification of *Blastobacter aggregatus* Hirsch and Muller 1986 as *Rhizobium aggregatum* comb. nov. *Int. J. Syst. Evol. Microbiol.*, *61*(5), 1218–1225. doi: 10.1099/ijs.0.017491-0
- Kellogg, C. T. E., & Deming, J. W. (2009). Comparison of free-living, suspended particle, and aggregate-associated bacterial and archaeal communities in the Laptev Sea. *Aquat. Microb. Ecol.*, *57*(1), 1–18. doi: 10.3354/ame01317
- Kirchman, D. L., Cottrell, M. T., & Lovejoy, C. (2010). The structure of bacterial communities in the western Arctic Ocean as revealed by pyrosequencing of 16S rRNA genes. *Environ. Microbiol.*, *12*(5), 1132–1143. doi: 10.1111/j.1462-2920.2010.02154.x
- Kirchman, D. L., Elifantz, H., Dittel, A. I., Malmstrom, R. R., & Cottrell, M. T. (2007). Standing stocks and activity of Archaea and Bacteria in the western Arctic Ocean. *Limnol Ocean*, *52*(2), 495–507.
- Kirchman, D. L., Moran, X. A. G., & Ducklow, H. (2009). Microbial growth in the polar oceans: role of temperature and potential impact of climate change. *Nat. Rev. Micro.*, *7*(6), 451–459.
- Krembs, C., Eicken, H., & Deming, J. W. (2011). Exopolymer alteration of physical properties of sea ice and implications for ice habitability and biogeochemistry in a warmer Arctic. *PNAS*, *108*(9), 3653–3658. doi: 10.1073/pnas.1100701108
- Krembs, C., Eicken, H., Junge, K., & Deming, J. W. (2002). High concentrations of exopolymeric substances in Arctic winter sea ice: implications for the polar ocean carbon cycle and cryoprotection of diatoms. *Deep Sea Res. Part I*, *49*(12), 2163–2181.
- Larose, C., Berger, S., Ferrari, C., Navarro, E., Dommergue, A., Schneider, D., & Vogel, T. (2010). Microbial sequences retrieved from environmental samples from seasonal Arctic snow and meltwater from Svalbard, Norway. *Extremophiles*, *14*(2), 205–212. doi: 10.1007/s00792-009-0299-2
- Männistö, M., Tirola, M., & Häggblom, M. (2009). Effect of freeze-thaw cycles on Bacterial communities of Arctic tundra soil. *Microb. Ecol.*, *58*(3), 621–631. doi: 10.1007/s00248-009-9516-x

- Martin, S., Drucker, R., & Fort, M. (1995). A laboratory study of frost flower growth on the surface of young sea ice. *J. Geophys. Res.*, *100*. doi: 10.1029/94jc03243
- Matsen, F., Kodner, R., & Armbrust, E. V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, *11*(1), 538.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., . . . Edwards, R. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, *9*(1), 386.
- Nghiem, S. V., Rigor, I. G., Richter, A., Burrows, J. P., Shepson, P. B., Bottenheim, J., . . . Asplin, M. G. (2012). Field and satellite observations of the formation and distribution of Arctic atmospheric bromine above a rejuvenated sea ice cover. *J. Geophys. Res.*, *117*(Online in advance of print).
- Perovich, D. K., & Richter-Menge, J. A. (1994). Surface characteristics of lead ice. *J. Geophys. Res.*, *99*.
- Perovich, D. K., & Richter-Menge, J. A. (2009). Loss of sea ice in the Arctic. *Ann. Rev. Mar. Sci.*, *1*, 417–441.
- Prévost, D., Drouin, P., Laberge, S., Bertrand, A., Cloutier, J., & Levesque, G. (2003). Cold-adapted rhizobia for nitrogen fixation in temperate regions. *Can. J. Botany*, *81*(12), 1153–1161.
- Price, M., Dehal, P., & Arkin, A. (2010). FastTree 2 - Approximate maximum likelihood trees for large alignments. *PLOS one*, *5*(3), e9490.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glockner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl. Acids Res.*, *35*(21), 7188–7196. doi: 10.1093/nar/gkm864
- Raymond, J. A., Fritsen, C., & Shen, K. (2007). An ice-binding protein from an Antarctic sea ice bacterium. *FEMS Microbiol. Ecol.*, *61*(2), 214–221. doi: 10.1111/j.1574-6941.2007.00345.x
- Riedel, A., Michel, C., Gosselin, M., & LeBlanc, B. (2007). Enrichment of nutrients, exopolymeric substances and microorganisms in newly formed sea ice on the Mackenzie shelf. *Mar. Ecol. Prog. Ser.*, *342*, 55–67. doi: 10.3354/meps342055
- Rivkin, R., Anderson, M., & Lajzerowicz, C. (1996). Microbial processes in cold oceans. I. Relationship between temperature and bacterial growth rate. *Aquat Microb Ecol*, *10*(3), 243–254.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-

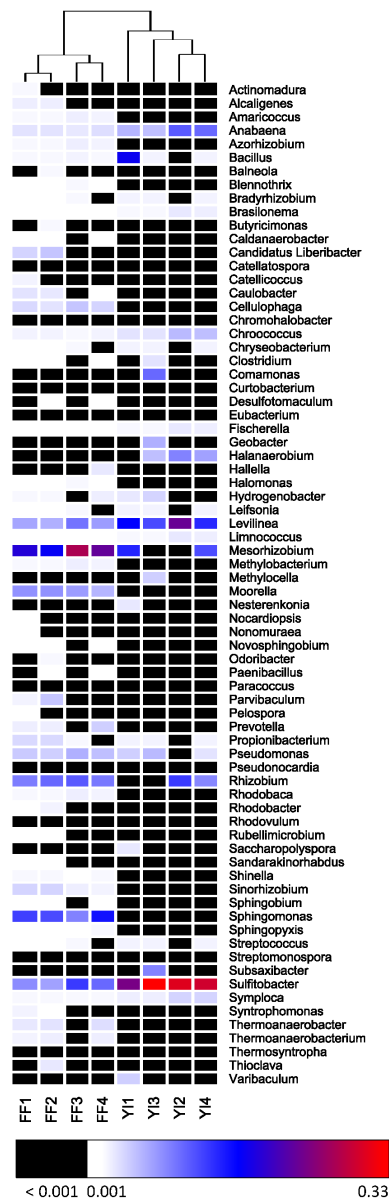
- supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, 75(23), 7537–7541. doi: 10.1128/aem.01541-09
- Team, R. C. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Todd, J. D., Rogers, R., Li, Y. G., Wexler, M., Bond, P. L., Sun, L., . . . Johnston, A. W. B. (2007). Structural and regulatory genes required to make the gas dimethylsulfide in Bacteria. *Science*, 315(5812), 666–669. doi: 10.1126/science.1135370
- Wallenstein, M. D., McMahon, S., & Schimel, J. (2007). Bacterial and fungal community structure in Arctic tundra tussock and shrub soils. *FEMS Microbiol. Ecol.*, 59(2), 428–435. doi: 10.1111/j.1574-6941.2006.00260.x
- Weissenberger, J., & Grossmann, S. (1998). Experimental formation of sea ice: importance of water circulation and wave action for incorporation of phytoplankton and bacteria. *Polar Biology*, 20(3), 178–188. doi: 10.1007/s003000050294
- Wells, L. E., & Deming, J. W. (2006). Characterization of a cold-active bacteriophage on two psychrophilic marine hosts. *Aquat Microb Ecol*(45), 15–29.
- Yager, P. L., Connelly, T. L., Mortazavi, B., Wommack, K. E., Bano, N., Bauer, J. E., . . . Hollibaugh, J. T. (2001). Dynamic bacterial and viral response to an algal bloom at subzero temperatures. *Limnol and Oceanogr*, 46(4), 790–801.
- Young, J. P., Crossman, L., Johnston, A., Thomson, N., Ghazoui, Z., Hull, K., . . . Parkhill, J. (2006). The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology*, 7(4), R34.

**Table 3.1.** Datasets used to evaluate the occurrence of marine Rhizobiales in polar regions.

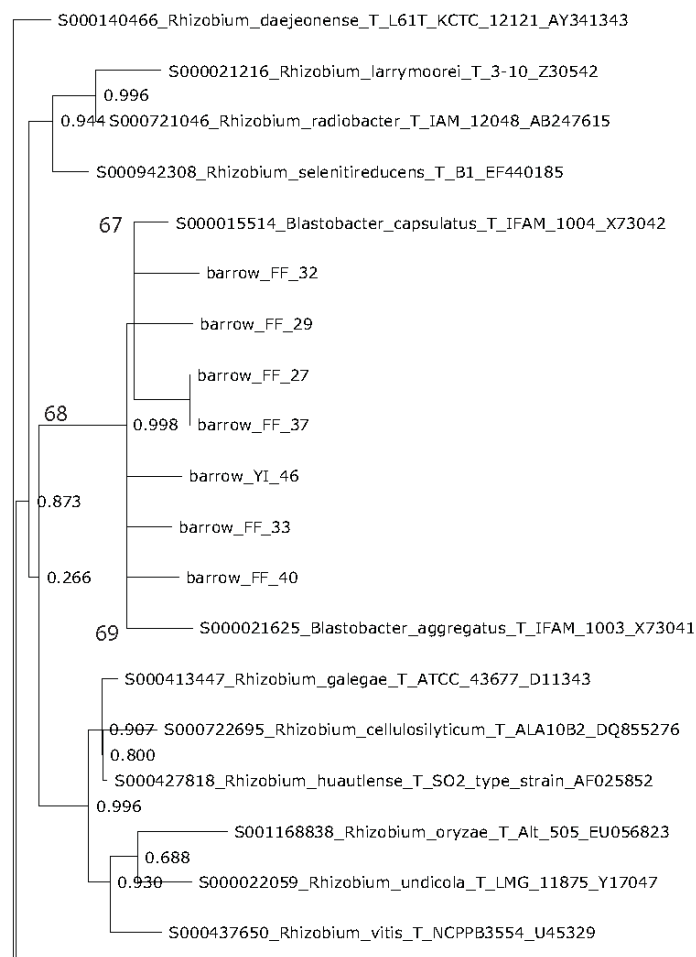
<b>Study</b>	<b>Environment</b>	<b>Number of sequences</b>	<b>Classified as Rhizobiales</b>	<b>Placed with <i>Blastobacter aggregatus</i></b>	<b>Placed with <i>Blastobacter capsulatus</i></b>
Wallenstein et al., 2007	Arctic soil	1,569	78	0	0
Galand et al., 2009	Arctic marine	885,187	2,592	66	405
Kellogg et al., 2009	Arctic marine	188	0	0	0
Männistö et al., 2009	Arctic soil	174	14	0	0
Campbell et al., 2010	Arctic soil	172	0	0	0
Collins et al., 2010	Arctic marine	166	4	0	0
Kirchman et al., 2010	Arctic marine	448,608	1,184	60	251
Larose et al., 2010	Arctic snow and meltwater	114	5	0	0
Bowman et al., 2011	Arctic marine	47,162	12	0	0
Chu et al., 2011	Arctic soil	17	1	0	0
Harding et al., 2011	Arctic snow	143	0	0	0
Genbank SRA SRP001216	Antarctic marine	478,831	577	29	306
Gryzmski et al., 2012	Antarctic marine	1,414	2	0	0
DeCorte et al., 2012	Arctic marine	222	3	0	0
This study, clone libraries	Arctic marine	51	14	4	3
This study, metagenomes	Arctic marine	88,308	58,730	1,024	5,883



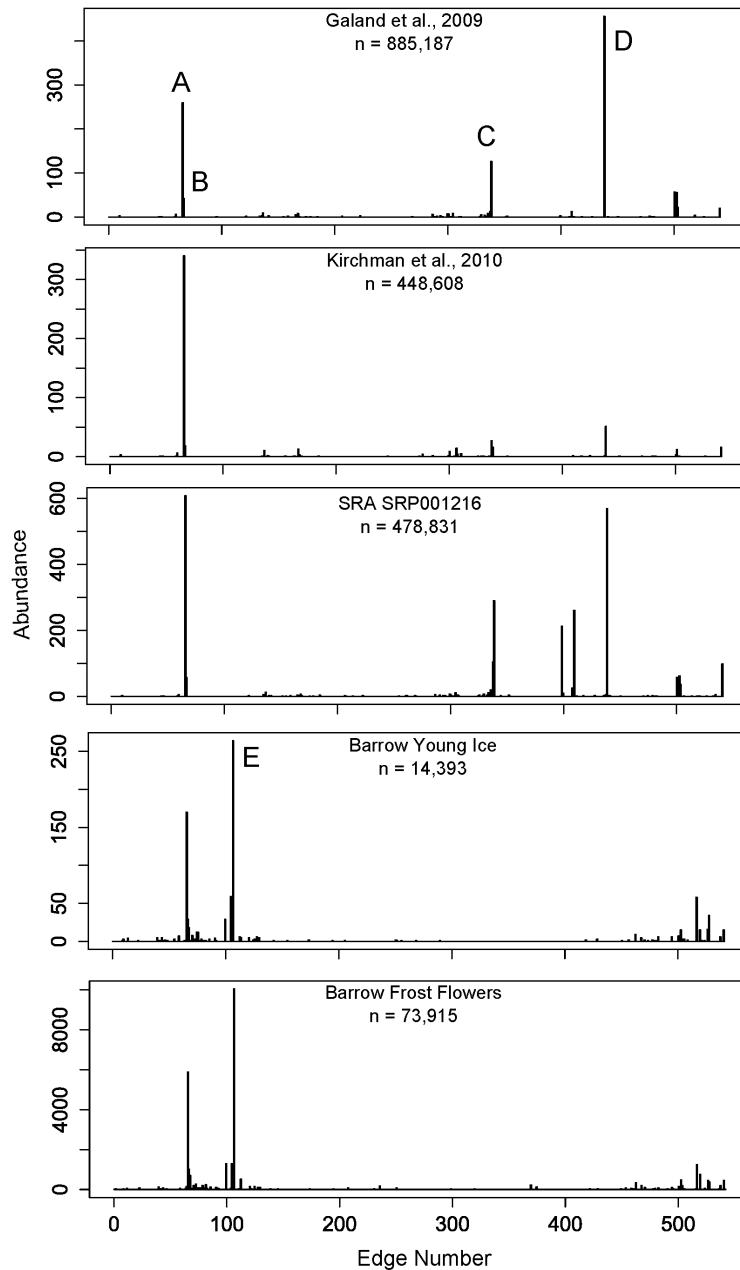
**Fig. 3.1. Frost flowers and young sea ice at the study site.** Samples were collected using an ethanol sterilized shovel (frost flowers) and hand saw (young ice) from four sites (approximately 4 m<sup>2</sup> each) in a 15 x 20 m area of a frozen lead 3.3 km offshore of Barrow, Alaska, on 21 April 2010. The lead, approximately 15 m across and several hundred meters long, ran perpendicular to the large perennial Barrow lead located just over the horizon in this image. Foreground (white) scale bar indicates a distance of 10 cm; background (black) scale bar, 1 m; inset scale bar, 1 cm. Air temperature at the time of sampling was  $-8^{\circ}\text{C}$ ; ice surface temperature was  $-6^{\circ}\text{C}$ .



**Fig. 3.2. Relative fluorescence of genera evaluated by microarray.** Diversity of the 16S rRNA gene from four frost flower (FF) and four young ice (YI) samples was evaluated following the methods of Delmont et al. (2011). Relative fluorescence is represented as the fraction of total fluorescence for each sample. Sample clustering was carried out using the complete agglomeration method on Euclidean distance between samples. Statistical analysis was conducted using the software package R (Team, 2012).



**Fig. 3.3. Placement of Barrow frost flower (FF) and young ice (YI) 16S rRNA gene clones on a Rhizobiales reference tree.** An approximate maximum-likelihood reference tree was created using FastTree (Price et al., 2010) from 352 type Rhizobiales sequences obtained from the RDP (Cole et al., 2007) and aligned to the greengenes core set (DeSantis et al., 2006). Chimera-checked sequences obtained in this study that classified as Rhizobiales were placed on this reference tree using pplacer (Matsen et al., 2010) and the wrapper script place\_it\_v1.01.py ([https://sites.google.com/site/bowmanjs/methods/python\\_scripts](https://sites.google.com/site/bowmanjs/methods/python_scripts)). Integers near select nodes correspond to edge numbers in Fig. 3.4.



**Fig. 3.4. Placement of 16S rRNA gene reads from available deep sequence libraries (Table 3.1).** Sequence reads for the western Arctic (Galand et al., 2009), across the circumpolar Arctic (Kirchman et al., 2010), and from Antarctic waters (SRP001216) were obtained from Genbank. Sequence reads for Barrow young ice and frost flowers were obtained from the metagenomes of this study. Reads that classified as Rhizobiales using Mothur (Schloss et al., 2009) with the RDP

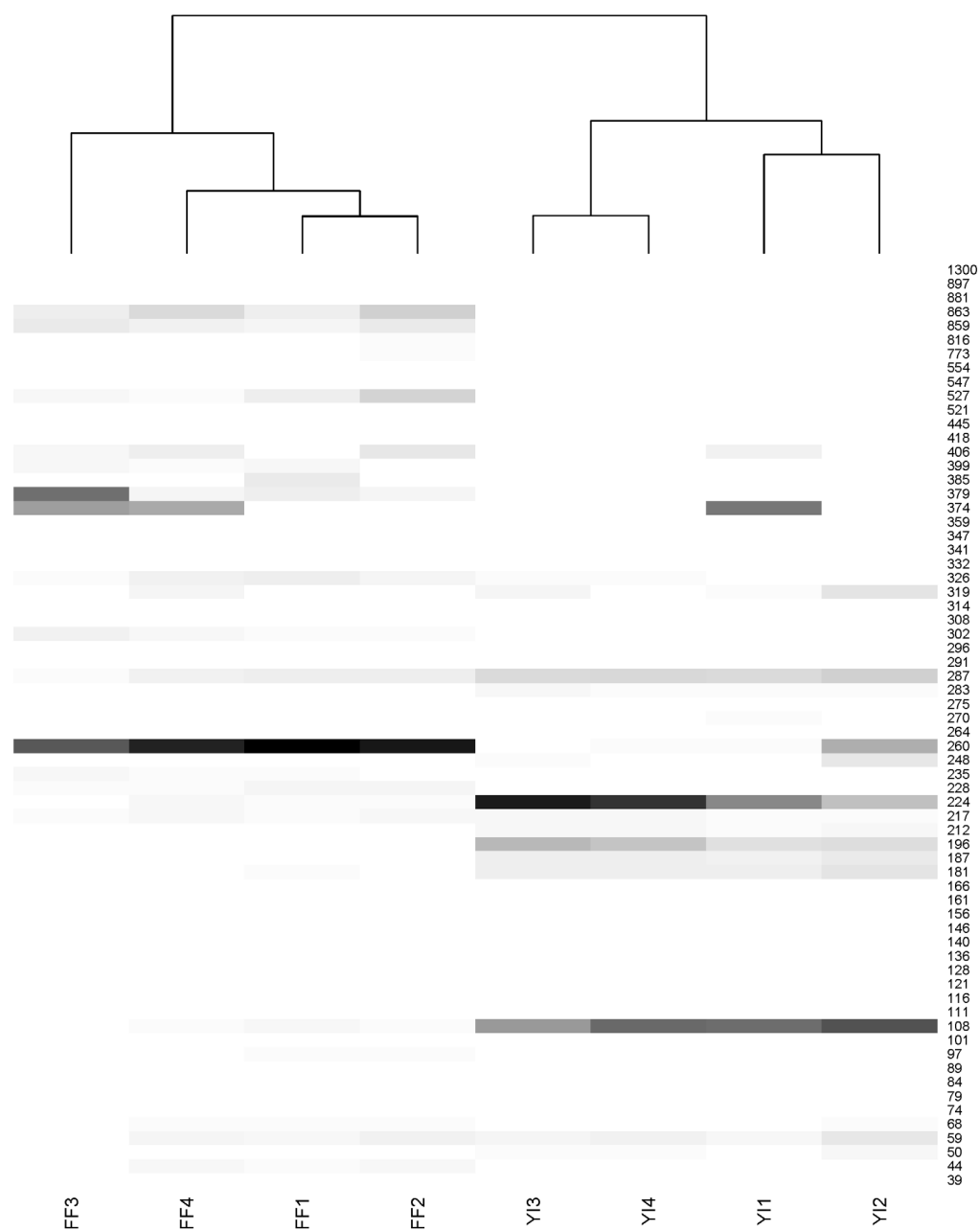


taxonomy and the silva.bacteria.fasta reference set

([http://www.mothur.org/wiki/Taxonomy\\_outline](http://www.mothur.org/wiki/Taxonomy_outline)) were placed on the reference tree using pplacer

(Matsen et al., 2010) and place\_it\_v1.01. Edge number refers to location of placements on the reference tree; abundance gives the number of placements at that location (note differences in scale). The total number of Bacterial 16S rRNA gene reads for each dataset is provided as n.

Notable edges are: A) *Blastobacter capsulatus* B) *B. aggregatus* C) *Labrys* spp. D) *Maritalea*, *Zhangella*, *Cucumibacter*, *Devosia*, and *Prosthecomicrobium* spp. E) *Rhizobium multihospitium*.



**Fig. S3.1.** Heatmap of T-RFLP fragment relative abundance. White indicates the minimal value (0), black indicates the maximum value (0.553). Column dendrogram represents clustering by Euclidean distance between samples (Fig. 3.1), the FF and young ice sample clusters are significant at  $p < 0.05$ . Row labels refer to fragment length in bp.

## Chapter 4

### **The genetic potential for key biogeochemical processes in Arctic frost flowers and young sea ice revealed by metagenomic analysis**

(Citation: Bowman, J. S., Berthiaume, C. T., Armbrust, E., & Deming, J. W. (2014). The genetic potential for key biogeochemical processes in Arctic frost flowers and young sea ice revealed by metagenomic analysis. *FEMS Microbiology Ecology*, 89: 376–387.)

#### ABSTRACT

Newly formed sea ice is a vast and biogeochemically active environment. Recently we reported an unusual microbial community dominated by members of the Rhizobiales in frost flowers at the surface of Arctic young sea ice based on the presence of 16S gene sequences related to these strains. Here we use metagenomic analysis of two samples, from a field of frost flowers and the underlying young sea ice, to explore the metabolic potential of this surface ice community. The analysis links genes for key biogeochemical processes to the Rhizobiales, including dimethylsulfide uptake, betaine glycine turnover, and halocarbon production. Nodulation and nitrogen fixation genes characteristic of terrestrial root-nodulating Rhizobiales were generally lacking from these metagenomes. Non-Rhizobiales clades at the ice surface had genes that would enable additional biogeochemical processes, including mercury reduction and dimethylsulfoniopropionate catabolism. Although the ultimate source of the observed microbial community is not known, considerations of the possible role of aeolian deposition or transport with particles entrained during ice formation favor a suspended particle source for this microbial community.

## 4.1 Introduction

Frost flowers are hypersaline ice structures common to the surface of newly formed (young) sea ice when atmospheric temperatures are sufficiently cold (Style & Worster, 2009). Frost flowers, having higher salinity, greater surface area (per unit mass), and lower temperature than the underlying young sea ice, have garnered significant attention for their potential influence on atmospheric chemistry (Alvarez-Aviles et al., 2008; Rankin, Wolff, & Martin, 2002). This influence is a product of their ability to concentrate salt and organic matter at the critical ice-atmosphere interface (Aslam et al., 2012; Beine et al., 2012; Bowman & Deming, 2010; Thomas A. Douglas et al., 2012; Perovich & Richter-Menge, 1994).

Frost flowers also concentrate bacteria, with the more saline frost flowers typically containing the largest number of bacteria, suggesting that they are populated from sea ice brines rejected to the surface of young sea ice during ice growth (Bowman & Deming, 2010). In an earlier study we used 16S rRNA gene sequences to investigate the microbial community composition and structure of frost flowers and underlying young sea ice growing on a small lead offshore of Barrow, Alaska (Bowman, Larose, Vogel, & Deming, 2013), part of the larger Barrow flaw-lead. The Barrow flaw-lead is itself an extension of the Arctic-wide circumpolar flaw-lead system. Surprisingly the microbial community was dominated by members of the order Rhizobiales. While the Rhizobiales are not unknown in the marine environment or in other extreme environments, they had not been previously reported in Arctic waters or in association with sea ice.

To further develop hypotheses regarding the source of the observed frost flower Rhizobiales, and to explore the potential biogeochemical impact of this community at the critical ice-ocean-atmosphere interface, we sequenced metagenomes from two of the samples from the

earlier study, one from a small field of frost flowers and the other from the underlying young sea ice. A quantitative approach was developed to assess the taxonomic composition of these communities and the presence of key functional genes. We also considered available data on wind magnitude, sea ice sediment concentration, and bacterial abundance to evaluate whether aerial deposition after ice formation or entrainment of suspended particles during ice formation, was the more likely delivery mechanism of the frost flower microbial community.

## **4.2 Materials and Methods**

### **4.2.1 Sample collection and sequencing**

Samples were collected in April, 2010, from a small lead offshore of Pt. Barrow, Alaska, USA, and the DNA extracted as described previously (Bowman et al., 2013). One young ice and one frost flower sample corresponding to samples YI4 and FF1 of Bowman et al. (2013) were selected for shotgun sequencing on the Illumina platform following whole genome amplification (WGA). For WGA, sheared DNA (200-500 bp) was amplified with the SEQPLEX kit (Sigma-Aldrich) following the manufacturer's recommended protocol, including the enzymatic primer removal step. WGA was carried out under conditions of high sterility to reduce the risk of contamination. Products from WGA were visualized on a gel and purified using the Gene-Jet PCR Purification Kit (Fermentas). Paired end sequencing at 2 x 125 bp with a 175 bp insert was conducted at the Argonne National Lab on the Illumina Hi-Seq platform using one lane for each metagenome.

### **4.2.2 Sequence processing and taxonomic analysis**

Raw sequence reads were screened for duplicated reads using the `fastq_nodup` command in SEASAR (Iverson et al., 2012). The command was called with the flags `-z -l 15 -d 1 -e 2 -v`. Remaining primer sequence was removed using an in-house script (`primer_purge.py`). De-

duplicated reads without primers were trimmed for quality using the SEAStAR command trimfastq with the flags -z --mates\_file -p 0.5 -l 34 -m 34 --add\_len. The resulting trimmed 16S rRNA gene reads were classified using the Bayesian classifier in Mothur (Schloss et al., 2009) with a kmer size of 8 against the greengenes reference taxonomy ([www.mothur.org/wiki/Greengenes-formatted\\_databases](http://www.mothur.org/wiki/Greengenes-formatted_databases)). Reads that classified at the level of phylum with a bootstrap score greater than 60 were considered to be 16S rRNA genes and the taxonomic classification was retained.

All quality controlled (QC'd) reads were assembled iteratively with the Velvet assembler (Zerbino, 2010) following the method of Iverson et al. (2012). Kmer lengths between 15 and 29 were evaluated, with the optimal assembly using a kmer length of 23. Following the initial round of assembly, QC'd reads were mapped to the assembled contigs using the BWA aligner (Li & Durbin, 2009) with the flags -n 0.001 -l 18. The next round of assembly used only the contigs assembled in the previous round and successfully mapped reads. After five iterations the N<sub>50</sub> value and maximum contig length stabilized at 1,905 and 18,102 respectively for the frost flower sample, and 329 and 4,992 for the young sea ice sample. Open reading frames (ORFs) longer than 300bp were identified in frost flower contigs longer than 10,000 bp using an in-house script (annotate\_contigs.py). ORFs were evaluated as potential coding sequences (CDS) by searching translations against the PFAM-A database (Punta et al., 2012) with the hmmscan command in HMMER v3.0 (Eddy, 1998) and delta blast (Boratyn et al., 2012) against the NCBI cdd database. Contigs were searched for 16S rRNA genes using blastn (Altschul, Gish, Miller, Myers, & Lipman, 1990) against the NCBI 16SMicrobial database.

#### **4.2.3 Metabolic Profiles**

We created five databases composed of high quality, full length or near full length protein sequences from UniProt (Bairoch et al., 2009) derived from 20 different genes or gene families diagnostic of 10 different metabolic processes (Table 4.1). All QC'd reads were aligned against these databases using blastx (Altschul et al., 1990). Reads that successfully mapped to a sequence in the database were extracted as amino acid translations of the alignment. For the frost flower metagenome, extracted read translations were further classified using the phylogenetic placement program pplacer (Matsen, Kodner, & Armbrust, 2010) and a reference tree of the database sequences. Alignment of the reference sequences, and the query sequences to the reference alignment, was carried out with three iterations in Clustalo v1.2 (Sievers et al., 2011). Reference trees were constructed using FastTree2 (Price, Dehal, & Arkin, 2010) or RAxML v7.2.8 (Stamatakis, 2006). The relative proportion of reference genes in the two metagenomes was estimated by dividing the combined length of the recruited read translations by the mean length of the reference sequences. Because the 16S rRNA gene often has multiple copies in bacterial genomes, we normalized coverage to proteins from the single copy *recA* gene. The significance of differences in normalized coverage between the two metagenomes was assessed using the two proportion z-test, with a sample size of 1,000. Thus a hypothetical RecA-normalized coverage value of 0.019 would represent 19 observations out of 1,000 in the z-test.

To further investigate both the metabolic potential of the frost flower and the young sea ice microbial communities, and the taxonomic affiliation of the detected strains, we used the BWA aligner to align all QC'd reads against the 4,710 prokaryotic genomes (chromosomes and plasmids) available from NCBI as complete assemblies as of October, 2012. The BWA aln command was called with the flags -n 0.001 -l 12 -k 2 following a call to the BWA index command using -a is. These alignments were conducted independently; that is, each index

created for the alignment contained only a single genome. Genome coverage and breadth, and the fraction of positions in the reference sequence with a mapped read, were calculated using in-house scripts (`tally_reads_in_sam.py`, `read_length_recruit_from_sam.py`, and `get_coverage_gaps.r`). To assess functional differences between the genomes present in the frost flower metagenome and sequenced Rhizobiales, we analyzed regions of the top five recruiting Rhizobiales reference genomes (according to coverage x breadth), and the genome of *Candidatus Pelagibacter ubique*, longer than 5,000 nt that did not recruit any reads using an in-house script (`get_coverage_gaps.py`). Translated ORFs from these gap regions were evaluated using blastp against the NCBI refseq\_protein database with an E-value cutoff of  $10^{-30}$ .

The metagenomic sequences were submitted to the MG-RAST database (Meyer et al., 2008) with MG-RAST ID numbers 4537103.3 – 4537105.3. Scripts used in this analysis are available at [https://github.com/bowmanjeffs/barrow\\_metagenome\\_scripts](https://github.com/bowmanjeffs/barrow_metagenome_scripts).

### 4.3 Results

Sequencing produced 21,012,180 read pairs for the frost flower sample and 9,882,191 read pairs for the young sea ice sample. Approximately 30% of the reads passed QC for both libraries, with an average length of 96 nt, producing 1.1 Gb for the frost flower library and 560 Mb from the young sea ice library. About 0.0012 % of the reads classified as 16S rRNA. A majority of the 16S sequences (52 %, 7,247) from the frost flower library classified as Rhizobiales whereas about 8 % (541) of those from young sea ice classified as Rhizobiales (Table 4.2). 16S rRNA reads classified as derived from chloroplasts were recovered from both the frost flower (464, 2 %) and young sea ice (1,815, 22 %) libraries. A large fraction of the young sea ice 16S rRNA reads could not be classified at the level of order.



Assembly of the frost flower reads produced 13 contigs longer than 10,000 bp, with the longest reaching 18,102 bp (Table 4.3). No 16S rRNA genes were recovered from these contigs. Two contigs displayed high similarity to sequences within the NCBI nt database. Contig\_1473 aligned to the genome of *Sinorhizobium meliloti* Rm41 (score = 4,947) and Contig\_171 aligned to plasmid pSymB of *S. meliloti* 1021 (score = 2,105).

Small proportions (frost flower: 1.68%, young sea ice: 1.9%) of the QC'd metagenomic reads mapped to available reference genomes. In both metagenomes, the chromosome of Candidatus *Pelagibacter ubique* was the deepest and most broadly covered. In the frost flower metagenome, *P. ubique* coverage was 3.62 and breadth was 0.60 (Fig. 4.1) whereas in the young sea ice, *P. ubique* coverage was 1.38 and breadth was 0.41, indicating the lower overall sequence depth of the young sea ice sample. When ranked by coverage x breadth, the other well covered genomes in the frost flower sample were either other members of the Rhizobiales, particularly *Ochrobactrum anthropi* ATCC 49188, *Sinorhizobium fredii* NGR234, and *Sinorhizobium meliloti* GR4, or the alphaproteobacterium Candidatus *Pelagibacter* sp. IMCC9063. Other than *Pelagibacter* spp. the first non-Rhizobiales in the mapping order was *Sphingobium japonicum* UT26S, which had sequence coverage in relatively few regions (coverage = 0.48, breadth = 0.01). The most highly covered genome in the young sea ice sample after *P. ubique* was Candidatus *Hodgkinia cicadicola* (coverage = 0.29, breadth = 0.01).

To further evaluate the taxonomic makeup of the community we considered only those reads that mapped uniquely to a single genome or to multiple genomes from a single species. These uniquely mapped reads were considered diagnostic of the presence of closely related strains within the metagenome samples. The species most covered by uniquely mapped reads from the frost flower metagenome was Candidatus *Pelagibacter ubique* at 2.61, followed by

Candidatus *Pelagibacter* sp. IMCC9063 (0.21), *Agrobacterium vitis* (0.08), *Rhodococcus erythropolis* (0.05), and Candidatus *Nitrosopumilus* sp. AR2 (0.05). The species most covered by uniquely mapped reads from the young sea ice metagenome was Candidatus *P. ubiquus* at 0.97, followed by *Pelagibacter* sp. IMCC9063 (0.10), and *Nitrosopumilus* sp. AR2 (0.03). Among the Rhizobiales many genes required for plant infection and symbiosis are contained on plasmids. By Student's T-test on both breadth ( $T = 2.53$ ,  $p = 0.012$ ,  $df = 254$ ) and coverage ( $T = 10.10$ ,  $p < 10^{-6}$ ,  $df = 235$ ), plasmids were poorly covered compared to chromosomes among the Rhizobiales (Fig. 4.1 B-D). The exceptions to this trend were plasmids from *Agrobacterium vitis* S4 (NC\_011986), *Ochrobactrum anthropi* ATCC49188 (NC\_009670 and NC\_009671), *Methylobacterium radiotolerans* JCM2831 (NC\_010510), and *Sinorhizobium fredii* HH103 (NT\_187146) (Fig. 4.2).

The distribution of reads across the reference genomes was analyzed to determine whether potential systematic differences existed between the organisms detected in our samples and their cultured relatives. As a negative control for this comparison we included Candidatus *Pelagibacter ubiquus*, with previously described variable genomic regions (Rusch et al., 2007; Wilhelm, Tripp, Givan, Smith, & Giovannoni, 2007). At a resolution of 5,000 bp, a bin larger than the average gene, we observed only 5 regions of low sequence coverage in the genome of *P. ubiquus* (Fig. 4.3), with the largest corresponding to the expected position of the HVR2 hypervariable region (Wilhelm et al., 2007). Genes within these regions primarily encoded proteins involved in substrate utilization, energy transfer, and biosynthesis. In contrast, the chromosome of *Sinorhizobium fredii* NGR234, the most covered root-nodulating Rhizobiales and fourth most covered genome overall, contained an abundance of coverage gaps distributed along its genome (Fig. 4.3). Similar to the *P. ubiquus* analysis, some genes within these coverage

gaps encoded proteins for biosynthesis and energy transfer. Importantly, when the gap regions from the five best covered Rhizobiales genomes were analyzed, 32 genes encoding nodulation factors were found to be lacking in the observed Rhizobiales, as well as 230 transposases, integrases, and other genes involved in DNA transfer (Table S4.1).

In Blast analysis of the nodulating proteins (NodA–J), RecA normalized coverage varied by protein (Table 4.1). NodIG were well covered at 2.242 and 0.43 respectively, while NodBCEJ were more moderately covered between 0.1 and 0.133. NodADFH had relatively low coverage. Several of the groups of predicted proteins were present in different proportions between the frost flower and young sea ice metagenomes. Considering RecA normalized coverage the NodDGJ proteins were more represented in the frost flower dataset (Table 4.1).

#### **4.3.1 Metabolic profiles**

Blastx analysis of the frost flower metagenome identified a large number of reads with significant homology to proteins of biogeochemical interest (Table 4.1). These alignments signify potential for the catabolism of glycine betaine (a compatible solute used under hypersaline conditions), dimethylsulfide (DMS), and dimethylsulfoniopropionate (DMSP), and for mercury resistance and methylhalide production. Relatively few reads aligned to the NifH nitrogenase, CmuA methyltransferase (which functions in methylhalide catabolism), or ice structuring databases (consisting of ice binding and ice nucleating proteins).

Phylogenetic placement of translated reads from the frost flower metagenome placed more than half of the haloperoxidase and DMS degradation reads with reference sequences from the Rhizobiales (Table 4.4). For DMS dehydrogenase and monooxygenase, which have relatively few representative sequences in UniProt, all Rhizobiales placements were to an edge corresponding to *Hyphomicrobium sulfonivorans*. For haloperoxidase the placements were more

varied, with edges corresponding to *Rhizobium etli* and *Agrobacterium* spp. accounting for most of the Rhizobiales placements. Proportionally fewer betaine methyltransferase reads were associated with the Rhizobiales, although the absolute number was still high. A large number of reads placed to a close homolog of the *Sinorhizobium meliloti* betaine methyltransferase in *Cupriavidus nectar* and to the bifurcating node for these sequences (Fig. 4.4). Fewer indoleacetamide hydrolase (IAH) reads placed with the Rhizobiales, and all of those placements were to *Brucella* spp. There are no reported Rhizobiales DMSP catabolism or degradation genes. Reads associated with DMSP lyase (which liberates the gas DMS) and DMSP demethylase were placed with non-Rhizobiales species, primarily *Ruegeria pomeroyi* (DddQ, DddW, DmdA), *Pelagibacter ubique* (DmdA), and *Aspergillus sydowii* (DddP). Considering RecA normalized coverage the proteins for DMS catabolism and haloperoxidases were more represented in the frost flower dataset. The ice structuring proteins were better covered in the young sea ice dataset (Table 4.1).

#### **4.4 Discussion**

Our analysis of the frost flower and young sea ice metagenomes suggests a partitioning of metabolic function between these two environments, with genes for DMS catabolism and haloperoxidase production abundant and elevated within frost flowers. The presence of these genes appears to be connected to the occurrence of bacteria belonging to the Rhizobiales, a bacterial order most commonly (but not exclusively) associated with soil. Despite the taxonomic overlap with soil the genetic composition of the frost flower microbial community is generally inconsistent with a terrestrial origin. Root-nodulating bacteria are found in tundra soil (Bordeleau & Prévost, 1994), but there is little evidence of nodulating bacteria in the frost flower metagenome. The complex process by which symbiotic rhizobia infect terrestrial plants and

carry out a mutualistic interaction is mediated by over 50 proteins (Downie, 1998). In this analysis we searched for genes encoding ten key nodulation proteins, NodA-J. Several of these proteins were almost entirely absent by our analysis, most notably NodABCDJ, which are present in all root-nodulating Rhizobiales (Downie, 1998). NodFH, respectively a lipo-chitin oligonucleotide production protein and a sulfation protein that are taxonomically restricted to the *Sinorhizobium* and *Rhizobium*, were also present only at low abundance. Of the remaining nodulation proteins NodGI were well represented in these metagenomes. NodG is a ribitol dehydrogenase and a close homolog to the housekeeping protein FadG (López-Lara & Geiger, 2001). Placement of translated frost flower NodG reads on a tree of NodG and FadG proteins suggests that most of the NodG blastx hits are in fact FadG read translations (Fig. 4.5). NodI, a membrane transporter, could represent a similar case of a protein required for, but not indicative of, symbiosis.

The observed Rhizobiales also appear deficient in genes encoding IAH and NifH. IAH catalyzes the final step in the conversion of tryptophan to indole-3-acetate (IAA), a known signaling molecule between bacteria and plants in the terrestrial environment (Patten & Glick, 1996). Although a number of translated reads classified with IAH (Table 4.1), none of them were placed with IAH proteins from the root-nodulating genera *Rhizobium*, *Bradyrhizobium*, or *Agrobacterium*. A majority of the IAH reads grouped with IAH homologs belonging to *Rhodococcus*, *Cupriavidus*, and *Ralstonia* spp. NifH enables nitrogen fixation, widely believed to be the primary benefit plants derive from root-nodulating bacteria. The few NifH read translations in the frost flower metagenome classified with NifH proteins from the Archaea and the genus *Paenibacillus*, suggesting that the frost flower Rhizobiales community was not equipped to fix nitrogen.

Many Rhizobiales genes required for plant infection and symbiosis are contained on plasmids. An overall absence of known symbiosis plasmids in our samples further supports an alternative ecology for these putatively marine Rhizobiales. Limited sequence homology was detected within our frost flower sample to predicted proteins encoded over a short segment of the symbiosis plasmid pSymB of *Sinorhizobium meliloti* 1021. The homologous proteins were amino acid transporters and a transcription regulator rather than proteins diagnostic of a root-nodulating lifestyle commonly found on the plasmids such as NodA-J. The assembled contig likely represents a conserved fragment of a much altered plasmid, or a chromosomal fragment in a bacterium that shares some ancestry (whether vertical or horizontal) with *S. meliloti* 1021.

Despite the general absence of key nodulation genes known from terrestrial plant symbioses, other evidence indicates that the microbial community present in frost flowers was equipped to live in association with an aquatic or marine photosynthetic community. This evidence includes detection of genes for the catabolism of the important algal exudates glycine betaine, DMSP, and DMS, all of which have important biogeochemical consequences (Welsh, 2000). Betaine homocysteine methyltransferase mediates the first step in the bacterial conversion of glycine betaine to pyruvate and ammonia. Frost flower betaine homocysteine methyltransferase reads classified primarily with *Sinorhizobium meliloti* and a close homolog in the  $\beta$ -proteobacterium *Cupriavidus necator* (a genus not otherwise present), suggesting that this function belonged predominantly to the Rhizobiales. DMSP catabolism is nearly ubiquitous among marine bacteria and our observed predicted DMSP sequences were associated with typical marine strains. DMS catabolism is less common and the DMS sequences were associated primarily with the aquatic Rhizobiales genus *Hyphomicrobium*.

There is some similarity between the frost flower microbial community and aquatic microbial communities. Previously we noted that 16S rRNA genes recovered from frost flowers are closely homologous to the free-living aquatic bacterium *Rhizobium aggregatum* and *Rhizobium capsulatus*, formerly of the genus *Blastobacter* (Bowman et al., 2013). Although genomes are not available for either of these *Rhizobium*, our recovery of DMS dehydrogenase reads with close homology to that of *Hyphomicrobium sulfinovorans*, a Rhizobiales known from Antarctic lake sediment (Moosvi, McDonald, Pearce, Kelly, & Wood, 2005), is suggestive, as is the observation of *Hyphomicrobium*-like morphologies on marine diatoms (Kaczmarek et al., 2005). Although we did not find 16S rRNA or *cmuA* gene sequence (Nadalig, Farhan, Haque, Roselli, & Schaller, 2011) evidence for *H. sulfinovorans* in our metagenomes, the taxonomic relationship with *Rhizobium*, and overlapping habitat with *R. aggregatum* and *R. capsulatus*, would have provided opportunities for the vertical or horizontal transfer of DMS dehydrogenase genes to the *Rhizobium* genus.

Because of the presence of genes for key processes, and the demonstrated association between Rhizobiales and phytoplankton in other environments (Bidle & Azam, 2001; Jasti, Sieracki, Poulton, Giewat, & Rooney-Varga, 2005; Kazamia et al., 2012; Makk, Beszteri, Ács, Márialigeti, & Szabó, 2003; Stevenson & Waterbury, 2006), we previously hypothesized that the observed Rhizobiales were originally associated with marine phytoplankton (Bowman et al., 2013). In this model, bacteria physically attached to phytoplankton cells or associated with the phycosphere (the region around a phytoplankton cell dominated by cell exudates) would have been entrained along with their hosts in newly formed sea ice. Once entrained the larger cells would become immobilized. A dissociation of hosts and symbionts, induced by stress or physical force, would leave these smaller cells free to transport upwards through the ice.

Physical attachment to large algal cells has long been considered a mechanism by which bacteria can enter sea ice (Grossmann & Dieckmann, 1994), but the adequacy of phycosphere association for bacterial entrainment remains to be tested. Analysis of the young sea ice 16S rRNA genes (Table 4.2) indicated the presence of a significant number of Stramenopile (diatom) chloroplasts that were much reduced by comparison in the frost flower metagenome.

An alternative to entrainment from the water column is aeolian transport. Although the available data on bacterial abundance of springtime Arctic air or dust is very limited, it is possible to make a rough estimate of these contributions. We found dust deposition in combination with airborne bacterial abundance to be inadequate to explain the observed enrichment of bacteria in frost flowers relative to the underlying young sea ice (Suppl. Materials). Supporting this conclusion are the low bacterial abundances in snow over sea ice adjacent to the frost flower field (Suppl. Materials) and in the vicinity a month earlier (Ewert, Carpenter, Colangelo-Lillis, & Deming, 2013). A similar calculation, however, suggests a role for lithogenic particles in the delivery of bacteria to young sea ice. Frazil ice can scavenge seafloor sediment along with biotic and abiotic suspended particles, delivering this material to young sea ice. Although frazil ice formation is typically considered to be a fall phenomenon (Kempema, Reimnitz, & Barnes, 1989), the water column was sufficiently cool to allow frazil ice formation during the frost flower growth period. A sediment concentration of  $100 \text{ g L}^{-1}$ , at the low end of the range of values given for visible sediment layers in Eicken et al. (2005), would be sufficient to explain the observed accumulation of bacteria. Considering that a sediment layer was not visible in the young sea ice, and that a number of chloroplasts were present in the young sea ice metagenome, a mix of less sediment and phytoplankton seems more plausible (see Suppl. Materials).



The surface of young sea ice can be a particularly extreme environment for bacteria regardless of their origin, especially in winter when temperatures are very low at the ice-air interface and salinities very high due to expelled brines (Bowman & Deming, 2010). The young ice surface is also a highly oxidizing environment (Thomas A. Douglas et al., 2012) and some Rhizobiales produce methylhalides as a byproduct of antioxidant activity (Amachi, Kamagata, Kanagawa, & Muramatsu, 2001; Bengtson, Bastviken, & Öberg, 2012). Although atmospheric halide chemistry above Arctic polynyas has received much attention, measurements of halocarbons were only recently made in young sea ice. In that study a biological source was invoked to explain the observed flux of halocarbons from newly formed sea ice (Granfors et al., 2013), though the responsible organisms were not identified. Our observation of bacterial haloperoxidases in both the young sea ice and frost flower metagenomes, and particularly associated with the Rhizobiales in frost flowers, is suggestive of a potential role in volatile halide production.

Heavy metal toxicity is another biological stressor noted by previous geochemical studies of the young sea ice environment, where mercury levels in particular were elevated in frost flowers (T. A. Douglas et al., 2005). Genes coding for MerA, the mercury reductase that confers mercury resistance, have been observed in sea ice (Møller et al., 2013; Poulain et al., 2007); the detoxification mechanism volatilizes the mercury, impacting atmospheric chemistry and the fate of the mercury. Consistent with the observations of MerA in sea ice, a large number of read translations aligned to MerA proteins in our blastx analysis, though none of these were homologous to known Rhizobiales MerA. Instead these read translations placed with the Archaeal genus *Nitrososphaera*.

## **4.5 Conclusions**

A comparative analysis of two metagenomes, one from a sampling of frost flowers and one from the underlying young sea ice, confirmed our previous 16S rRNA gene-based observation that the frost flower sample was enriched in members of the bacterial order Rhizobiales. Detailed analysis, including recruitment of reads to the available Rhizobiales genomes, suggests that the observed bacteria were genetically distinct from the well-studied strains responsible for root nodulation in terrestrial temperate environments. Specifically, the observed bacteria appeared to lack symbiosis plasmids and key genes for nitrogen fixation and root nodulation that characterize members of the terrestrial rhizosphere. Instead, the community appeared adapted to live in close association with marine or aquatic phytoplankton, carrying genes for the catabolism of common algal exudates.

Regardless of their origin the presence of the metabolically diverse Rhizobiales in the young sea ice environment could have implications for biogeochemical cycles there. The observed Rhizobiales have the potential for DMS and betaine glycine catabolism, and halocarbon production, with the larger community additionally capable of DMSP catabolism and mercury reduction. Our findings are limited by the fact that these metagenomes were sourced from samples that represent a single location in the flaw-lead system of the coastal Arctic. Observations of other microbial communities in the biologically under-studied young sea ice environment are needed to determine the broader implications of our results for the polar marine and atmospheric environments, the efficacy of the proposed transport mechanisms, and whether gene potential is translated to activity at the sea ice-atmosphere interface.

## **4.6 Supplemental Information**

### **4.6.1 Potential transport mechanism methods**

To examine aerial delivery as a potential source of bacteria to the frost flower field, hourly wind speed and direction for April 13–20, 2010 were obtained for the NOAA observatory at Barrow from [www.esrl.noaa.gov/gmd/obop/brw](http://www.esrl.noaa.gov/gmd/obop/brw). The observatory is located approximately 5.87 km from the sample site. Wind magnitude ( $M_{wind}$ ) was calculated as wind speed ( $\text{cm s}^{-1}$ ) times duration (s). Hourly wind magnitude was binned in 45 degree increments. Magnitude was summed for the easterly, southeasterly, and southerly bins, corresponding to directions with potential for overland transport. Given an airborne bacterial abundance ( $B_{air}$ ), the potential transport of bacteria ( $B_{FF}$ ) to a field of frost flowers with height ( $h_{FF}$ ), length ( $l_{FF}$ ), and width ( $w_{FF}$ ) is:

$$1. B_{FF} = \frac{B_{air} M_{wind} h_{FF} w_{FF}}{h_{FF} w_{FF} l_{FF}}$$

This relationship simplifies to:

$$2. B_{FF} = \frac{B_{air} M_{wind}}{l_{FF}}$$

To consider sea-ice entrainment of suspended sediments as a mechanism for delivering bacteria to frost flowers, we assumed that sea ice sediment concentrations from Eicken et al. (2005) provide a good approximation for young sea ice in springtime (no data on sediment concentration in the water column or sea ice was available for the study period). Further assuming that particles, whether lithogenic or biological, are as efficiently excluded from ice formation as salt, then the ratio of salt in frost flowers to salt in seawater can be used as a proxy for the ratio of particles expected. Given particle-associated bacterial abundance ( $B_p$ ), seawater salinity ( $S_{sw}$ ), and seawater particle concentration ( $P_{sw}$ ), the potential delivery of bacteria to a field of frost flowers of a given salinity ( $S_{FF}$ ) is:

$$3. B_{FF} = \frac{S_{FF}}{S_{sw}} P_{sw} B_p$$

Bacterial abundance was determined in snow as for the frost flower and young sea ice samples in Bowman et al. (2013). Snow on sea ice immediately adjacent to the frost flower field was sampled ( $n = 1$ ) on April 20, while fresh snow was sampled ( $n = 2$ ) overtop the frost flower field on April 21, following heavy snowfall the night of April 20.

#### **4.6.2 Potential transport mechanism results**

From April 13 to April 20, the week prior to sampling, wind magnitude at the nearby NOAA observatory was  $2.79 \times 10^6$  m. Northeasterly winds, which do not travel overland to reach the study site, accounted for 35 % of the calculated magnitude (Fig. S1). Easterly, southeasterly, and southerly winds, which travel overland before reaching the study site, accounted for 31 % of the calculated magnitude. Airborne bacterial abundance ( $B_{\text{air}}$ ) measurements have not been made for the wintertime coastal Arctic. Measurements at a high alpine site were reported to be  $0.96 \text{ bacteria ml}^{-1}$  (Bowers et al., 2009), a value at the high end of the range observed elsewhere (Deguillaume et al., 2008). Using the observed wind magnitude over land and observed frost flower height of 2 cm and field length of 20 m (Fig. 1 in Bowman et al., 2013) in Eqt. 2, the wind would have been sufficient to deliver  $4.2 \times 10^4 \text{ bacteria ml}^{-1}$  of frost flower, a factor of 4.8 lower than the observed frost flower bacterial abundance of  $2.01 \times 10^5 \pm 7.42 \times 10^4 \text{ ml}^{-1}$  (Bowman et al., 2013). The true delivery would be less, for this estimate is generous, not only due to use of a relatively high value for bacterial abundance in air but also because it assumes a transfer efficiency of 100 % between air and frost flowers (with no bacterial growth or mortality) and complete deposition of the bacterial load in the bottom 2 cm of the atmosphere to the first 20 m of the lead where sampling took place (i.e., that a strong gradient in frost flower bacterial abundance should be present parallel to the wind vector responsible for bacterial transport, which has not been tested). Rearranging Eqt. 2 to solve for  $B_{\text{air}}$  suggests that

3.6 bacteria  $\text{ml}^{-1}$  of air (4 times the higher reported value for air) would be necessary to provide the observed bacterial abundance in frost flowers, with all the same assumptions still in place. In contrast to frost flowers, bacterial abundances in snow immediately adjacent to the frost flower field contained  $2.83 \times 10^3$  bacteria  $\text{ml}^{-1}$ , while fresh snow deposited on the frost flower field just after sampling them contained  $1.66 \times 10^4 \pm 8.13 \times 10^3$  bacteria  $\text{ml}^{-1}$ .

To calculate the maximum possible transport due to sediment in the water column we used a sea ice sediment concentration of  $508 \text{ mg L}^{-1}$ , the largest mean value reported in Eicken et al. (2005). Large sediment transport events are associated with fall storms (Hume & Schalk, 1967), as is the production of frazil ice responsible for the entrainment of suspended sediment in sea ice (Kempema et al., 1989). In the spring, land-fast sea ice protects the shore from erosion, so the selected sediment concentration should represent a high estimate of the particle load at the time of sampling. Salinity values are not available for the sampled frost flowers. Assuming a salinity of 70 ppt, a reasonable value for frost flowers (Bowman & Deming, 2010), a salinity of 10 ppt for the underlying sea ice (Bowman et al., 2013), and  $10^9$  bacteria  $\text{g particle}^{-1}$  (Luna, Manini, & Danovaro, 2002), we estimate that  $8.2 \times 10^5$  bacteria could have been delivered from sediment particles for each milliliter of frost flower. The sediment concentration sufficient to explain the observed accumulation of bacteria in frost flowers is in fact much lower; i.e.,  $100 \text{ g L}^{-1}$ , at the low end of the range of values given for visible sediment layers in Eicken et al. (2005).

#### **4.6.3 Discussion of potential transport mechanisms**

Relative to the underlying sea ice, bacterial abundance was high within the sampled frost flowers, consistent with other measurements of bacterial abundance in natural and laboratory frost flowers (Bowman & Deming, 2010; Eronen-Rasimus et al., 2014). We considered two possible transport mechanisms that could lead to this accumulation of bacteria; aeolian transport

of tundra soil or exposed lake sediment, and the entrainment of suspended particles and subsequent transport of particle-associated bacteria with brine to the ice surface. In the case of aeolian transport, the dominant fraction of the wind magnitude during the frost flower growth period was attributable to northeasterly winds over the ocean. There was a significant magnitude, however, associated with winds from the east, southeast, and south over land. A major unknown is the abundance of bacteria in the lower atmosphere during late winter in the coastal Arctic. Even if the abundance of airborne bacteria was at the high end of the observed range in other regions and transfer efficiency was 100%, the wind was insufficient to deliver the observed frost flower bacterial abundance. The required airborne abundance of  $3.6 \text{ bacteria ml}^{-1}$  corresponds to  $3.6 \times 10^6 \text{ bacteria per m}^3$ , a value that exceeds any that we can find in the literature.

Limited measurements of dust deposition on sea ice are also available for Pt. Barrow. Darby et al. (1974) measured a dust fallout rate of  $8.9 \mu\text{g cm}^{-2} \text{ yr}^{-1}$  for an undisclosed site one mile from Pt. Barrow. Assuming that the dust contained  $10^9 \text{ bacteria g}^{-1}$ , a standard load for terrestrial soil (Buckeridge & Grogan, 2008), this rate of deposition is sufficient to deliver only  $10^6 \text{ bacteria cm}^{-2} \text{ week}^{-1}$  (the relevant timescale for this frost flower field). There is also no evidence of significant accumulation in nearby snow; fresh snow had more bacteria than snow sampled the day prior. Values from both days were consistent with other measurements at Barrow that fell between  $10^2$  and  $10^4 \text{ bacteria ml}^{-1}$  (Ewert et al., 2013), 1–3 orders of magnitude below that observed in frost flowers. If exposed (snow-free) tundra soil or lake or estuarine sediment, delivered by wind, is the source of the observed bacterial community, then frost flowers must be a unique spot for accumulation of these bacteria and the genes transported with them.

Suspended particles including sediment (Kempema et al., 1989) and phytoplankton (Gradinger & Ikävalko, 1998) accumulate in sea ice through scavenging by frazil ice and, for the algal cells, through growth in the basal ice. Although the formation of frazil ice is associated with the fall freeze-in, water temperatures recorded by the Barrow sea ice mass balance site (<http://seaice.alaska.edu/gi>) were below  $-1.9^{\circ}\text{C}$  during the week prior to sampling, suggesting springtime presence of the supercooled water required for frazil ice formation. If we allow a minimum value for a visible layer of sea ice sediment concentration, the inclusion of sediment by frazil ice formation alone is sufficient to explain the observed bacterial abundance. By late April, however, the water column likely contained less sediment and a significant number of phytoplankton cells, as evidenced by the number of Stramenopile chloroplasts that had accumulated in young sea ice. Applying a similar calculation for chloroplasts as we did for suspended sediment, only 2.8 bacteria per chloroplast would be needed to explain the observed accumulation. Although the number of chloroplasts contained in a diatom cell is variable, 2.8 bacteria chloroplast<sup>-1</sup> is a reasonable number of bacteria to be within the phycosphere of a diatom. Based on the genetic and taxonomic composition of the frost flower microbial community we consider aeolian deposition to be highly unlikely, and particle inclusion to be very likely. Better constraints on the nature of suspended particles during spring, and the number of bacteria associated with those particles, are required to refine our interpretation, but it seems reasonable that the observed frost flower microbial community was sourced from both phytoplankton and inorganic sediment particles.

### **Acknowledgements**

This project was funded by NSF OPP award 0908724 to JWD and by the Walters Endowed Professorship. JSB was supported by an NSF IGERT fellowship through the UW

Astrobiology Program and an EPA STAR fellowship. EVA was supported by the Moore Foundation Marine Microbial Investigator program. Field support was provided by Umiaq and the Barrow Arctic Science Consortium; we are particularly grateful to Nok Acker and Lewis Brower for their assistance. The manuscript was substantially improved by input from two anonymous reviewers.



## References

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.*, 215(3), 403–410. doi: [http://dx.doi.org/10.1016/S0022-2836\(05\)80360-2](http://dx.doi.org/10.1016/S0022-2836(05)80360-2)
- Alvarez-Aviles, L., Simpson, W. R., Douglas, T. A., Sturm, M., Perovich, D., & Domine, F. (2008). Frost flower chemical composition during growth and its implications for aerosol production and bromine activation. *J. Geophys. Res.*, 113, D21304. doi: 10.1029/2008jd010277
- Amachi, S., Kamagata, Y., Kanagawa, T., & Muramatsu, Y. (2001). Bacteria mediate methylation of iodine in marine and terrestrial environments. *Appl. Environ. Microbiol.*, 67(6), 2718–2722. doi: 10.1128/aem.67.6.2718-2722.2001
- Aslam, S., Underwood, G., Kaartokallio, H., Norman, L., Autio, R., Fischer, M., . . . Thomas, D. (2012). Dissolved extracellular polymeric substances (dEPS) dynamics and bacterial growth during sea ice formation in an ice tank study. *Pol. Biol.*, 35(5), 661–676. doi: 10.1007/s00300-011-1112-0
- Bairoch, A., Bougueleret, L., Altairac, S., Amendolia, V., Auchincloss, A., Argoud-Puy, G., . . . Boeckmann, B. (2009). The Universal Protein Resource (UniProt) 2009. *Nuc. Acids Res.*, 37(Database issue), D169–174.
- Beine, H., Anastasio, C., Domine, F., Douglas, T., Barret, M., France, J., . . . Ullmann, K. (2012). Soluble chromophores in marine snow, seawater, sea ice and frost flowers near Barrow, Alaska. *J. Geophys. Res.*, 117, D00R15. doi: 10.1029/2011jd016650
- Bengtson, P., Bastviken, D., & Öberg, G. (2012). Possible roles of reactive chlorine II: assessing biotic chlorination as a way for organisms to handle oxygen stress. *Environ. Microbiol.*, 15, 991–1000. doi: 10.1111/j.1462-2920.2012.02807.x
- Bidle, K. D., & Azam, F. (2001). Bacterial control of silicon regeneration from diatom detritus: significance of bacterial ectohydrolases and species identity. *Limnol Oceanogr.*, 46(7), 1606–1623.
- Boratyn, G. M., Schaffer, A., Agarwala, R., Altschul, S. F., Lipman, D. J., & Madden, T. L. (2012). Domain enhanced lookup time accelerated BLAST. *Biol. Direct*, 7(1), 12.
- Bordeleau, L., & Prévost, D. (1994). Nodulation and nitrogen fixation in extreme environments. *Plant and Soil*, 161(1), 115–125. doi: 10.1007/bf02183092
- Bowers, R. M., Lauber, C. L., Wiedinmyer, C., Hamady, M., Hallar, A. G., Fall, R., . . . Fierer, N. (2009). Characterization of Airborne Microbial Communities at a High-Elevation Site and Their Potential To Act as Atmospheric Ice Nuclei. *Appl. Environ. Microbiol.*, 75(15), 5121–5130. doi: 10.1128/aem.00447-09

- Bowman, J. S., & Deming, J. W. (2010). Elevated bacterial abundance and exopolymers in saline frost flowers with implications for atmospheric chemistry and microbial dispersal. *Geophys. Res. Lett.*, 37(L13501).
- Bowman, J. S., Larose, C., Vogel, T. M., & Deming, J. W. (2013). Selective occurrence of Rhizobiales in frost flowers on the surface of young sea ice near Barrow, Alaska and distribution in the polar marine rare biosphere. *Environ. Microbiol. Rep.*, 5, 575–582. doi: 10.1111/1758-2229.12047
- Buckeridge, K. M., & Grogan, P. (2008). Deepened snow alters soil microbial nutrient limitations in arctic birch hummock tundra. *Appl. Soil Ecol.*, 39(2), 210–222.
- Darby, D., Burckle, L., & Clark, D. (1974). Airborne dust on the Arctic pack ice, its composition and fallout rate. *Earth and Planetary Science Letters*, 24, 166–172.
- Deguillaume, L., Leriche, M., Amato, P., Ariya, P. A., Chaumerliac, N., Bauer, H., . . . Morris, C. E. (2008). Microbiology and atmospheric processes: chemical interactions of primary biological aerosols. *Biogeosciences Discuss.*, 5, 841–870.
- Douglas, T. A., Domine, F., Barret, M., Anastasio, C., Beine, H. J., Bottenheim, J., . . . Steffen, A. (2012). Frost flowers growing in the Arctic ocean-atmosphere-sea ice-snow interface: 1. Chemical composition. *J. Geophys. Res.*, 117, D00R09. doi: 10.1029/2011jd016460
- Douglas, T. A., Sturm, M., Simpson, W. R., Brooks, S., Lindberg, S. E., & Perovich, D. K. (2005). Elevated mercury measured in snow and frost flowers near Arctic sea ice leads. *Geophys. Res. Lett.*, 32, L04502. doi: 10.1029/2004gl022132
- Downie, J. A. (1998). Functions of rhizobial nodulation genes. In H. P. Spaink, A. Kondorosi & P. J. J. Hooykaas (Eds.), *The Rhizobiaceae* (pp. 387–402). Netherlands: Springer.
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, 14(9), 755–763.
- Eicken, H., Gradinger, R., Gaylord, A., Mahoney, A., Rigor, I., & Melling, H. (2005). Sediment transport by sea ice in the Chukchi and Beaufort Seas: Increasing importance due to changing ice conditions? *Deep Sea Research Part II: Topical Studies in Oceanography*, 52(24), 3281–3302.
- Eronen-Rasimus, E., Kaartokallio, H., Lyra, C., Autio, R., Kuosa, H., Dieckmann, G. S., & Thomas, D. N. (2014). Bacterial community dynamics and activity in relation to dissolved organic matter availability during sea ice formation in a mesocosm experiment. *MicrobiologyOpen*.
- Ewert, M., Carpenter, S., Colangelo-Lillis, J., & Deming, J. (2013). Bacterial and extracellular polysaccharide content of brine-wetted snow over Arctic winter first-year sea ice. *J. Geophys. Res-Oceans*, 118, 726–735.
- Gradinger, R., & Ikävalko, J. (1998). Organism incorporation into newly forming Arctic sea ice in the Greenland Sea. *J. Plankton Res.*, 20(5), 871–886. doi: 10.1093/plankt/20.5.871

- Granfors, A., Andersson, M., Chierici, M., Fransson, A., Gårdfeldt, K., Torstensson, A., . . . Abrahamsson, K. (2013). Biogenic Halocarbons in Young Arctic Sea Ice and Frost Flowers. *Mar. Chem.*, 155, 124–134.
- Grossmann, S., & Dieckmann, G. S. (1994). Bacterial standing stock, activity, and carbon production during formation and growth of sea ice in the Weddell Sea, Antarctica. *Appl. Environ. Microbiol.*, 60(8), 2746–2753.
- Hume, J. D., & Schalk, M. (1967). Shoreline processes near Barrow, Alaska: a comparison of the normal and the catastrophic. *Arctic*, 20(2), 86–103.
- Iverson, V., Morris, R. M., Frazar, C. D., Berthiaume, C. T., Morales, R. L., & Armbrust, E. V. (2012). Untangling genomes from metagenomes: revealing an uncultured class of marine Euryarchaeota. *Science*, 335(6068), 587–590.
- Jasti, S., Sieracki, M. E., Poulton, N. J., Giewat, M. W., & Rooney-Varga, J. N. (2005). Phylogenetic diversity and specificity of bacteria closely associated with *Alexandrium* spp. and other phytoplankton. *App. Environ. Microbiol.*, 71(7), 3483–3494.
- Kaczmarska, I., Ehrman, J. M., Bates, S. S., Green, D. H., Léger, C., & Harris, J. (2005). Diversity and distribution of epibiotic bacteria on *Pseudonitzschia multiseries* (Bacillariophyceae) in culture, and comparison with those on diatoms in native seawater. *Harmful Algae*, 4, 725–741.
- Kazamia, E., Czesnick, H., Nguyen, T. T. V., Croft, M. T., Sherwood, E., Sasso, S., . . . Smith, A. G. (2012). Mutualistic interactions between vitamin B12 - dependent algae and heterotrophic bacteria exhibit regulation. *Environ. Microbiol.*, 14(6), 1466–1476.
- Kempema, E., Reimnitz, E., & Barnes, P. W. (1989). Sea ice sediment entrainment and rafting in the Arctic. *Journal of Sedimentary Research*, 59(2), 308–317.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, 25(14), 1754–1760.
- López-Lara, I. M., & Geiger, O. (2001). The nodulation protein NodG shows the enzymatic activity of an 3-oxoacyl-acyl carrier protein reductase. *Mol. Plant Microbe In.*, 14(3), 349–357.
- Luna, G., Manini, E., & Danovaro, R. (2002). Large fraction of dead and inactive bacteria in coastal marine sediments: comparison of protocols for determination and ecological significance. *App. Environ. Microbiol.*, 68(7), 3509–3513.
- Makk, J., Beszteri, B., Ács, É., Márialigeti, K., & Szabó, K. (2003). Investigations on diatom-associated bacterial communities colonizing an artificial substratum in the River Danube. *Arch. Hydrobiol.*, 14(3–4), 249–265.

- Matsen, F., Kodner, R., & Armbrust, E. V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11(1), 538.
- Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E., Kubal, M., . . . Edwards, R. (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*, 9(1), 386.
- Møller, A., Barkay, T., Hansen, M., Norman, A., Hansen, L., Sørensen, S., . . . Kroer, N. (2013). Mercuric reductase genes (*merA*) and mercury resistance plasmids in high Arctic snow, freshwater and sea-ice brine. *FEMS Microb. Ecol.*, 73, 2230–2238.
- Moosvi, A. S., McDonald, I. R., Pearce, D. A., Kelly, D. P., & Wood, A. P. (2005). Molecular detection and isolation from Antarctica of methylotrophic bacteria able to grow with methylated sulfur compounds. *Systematic and Applied Microbiology*, 28, 541-554.
- Nadalig, T., Farhan, M., Haque, U., Roselli, S., & Schaller, H. (2011). Detection and isolation of chloromethane-degrading bacteria from the Arabidopsis thaliana phyllosphere, and characterization of chloromethane utilization genes. *FEMS Microb Ecol*, 77, 438-448.
- Patten, C. L., & Glick, B. R. (1996). Bacterial biosynthesis of indole-3-acetic acid. *Can. J. Microbiol.*, 42(3), 207-220.
- Perovich, D. K., & Richter-Menge, J. A. (1994). Surface characteristics of lead ice. *J. Geophys. Res.*, 99, 16341–16350.
- Poulain, A. J., Ni Chadhain, S. M., Ariya, P. A., Amyot, M., Garcia, E., Campbell, P. G. C., . . . Barkay, T. (2007). Potential for mercury reduction by microbes in the high Arctic. *Appl. Environ. Microbiol.*, 73(7), 2230-2238. doi: 10.1128/aem.02701-06
- Price, M., Dehal, P., & Arkin, A. (2010). FastTree 2 - Approximate maximum likelihood trees for large alignments. *PLOS one*, 5(3), e9490.
- Punta, M., Coghill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R. D. (2012). The Pfam protein families database. *Nuc. Acids Res.*, 40(D1), D290–D301. doi: 10.1093/nar/gkr1065
- Rankin, A. M., Wolff, E. W., & Martin, S. (2002). Frost flowers: Implications for tropospheric chemistry and ice core interpretation. *J. Geophys. Res.*, 107, 4683–4695. doi: 10.1029/2002jd002492
- Rusch, D. B., Halpern, A. L., Sutton, G., Heidelberg, K. B., Williamson, S., Yooseph, S., . . . Venter, J. C. (2007). The *Sorcerer II* global ocean sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol*, 5(3), e77. doi: 10.1371/journal.pbio.0050077
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-

- supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, 75(23), 7537–7541. doi: 10.1128/aem.01541-09
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, 7(1).
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, 22(21), 2688–2690. doi: 10.1093/bioinformatics/btl446
- Stevenson, B. S., & Waterbury, J. B. (2006). Isolation and identification of an epibiotic bacterium associated with heterocystous *Anabaena* cells. *Biol. Bull.*, 210(2), 73–77.
- Style, R. W., & Worster, M. G. (2009). Frost flower formation on sea ice and lake ice. *Geophys. Res. Lett.*, 36(11), L11501.
- Welsh, D. T. (2000). Ecological significance of compatible solute accumulation by micro-organisms: from single cells to global climate. *FEMS Microb. Rev.*, 24(3), 263–290.
- Wilhelm, L., Tripp, H. J., Givan, S., Smith, D., & Giovannoni, S. (2007). Natural variation in SAR11 marine bacterioplankton genomes inferred from metagenomic data. *Biology Direct*, 2(1), 27.
- Zerbino, D. R. (2010). Using the Velvet de novo Assembler for Short-Read Sequencing Technologies *Current Protocols in Bioinformatics*: John Wiley & Sons, Inc.

**Table 4.1.** Number of reads, coverage, and normalized coverage by protein groups (left column). Protein groups with statistically significant differences in abundance are shown in bold. Negative differences indicate protein groups with higher normalized abundance in the young sea ice dataset. Protein group function shown below.

	Mean length	Frost flower			Young sea ice			Comparison		
		Translated reads	Coverage	Normalized coverage	Translated reads	Coverage	Normalized coverage	$\Delta$	z	p
Betaine methyltransferase <sup>a</sup>	337	1450	137	0.187	436	41	0.198	-0.011	-0.624	0.5326
CmuA <sup>b</sup>	356	69	6	0.008	25	2	0.01	-0.002	-0.474	0.6355
DMSP Lyase/demethylase <sup>c</sup>	260	1425	175	0.239	451	55	0.266	-0.027	-1.39	0.1645
<b>DMS catabolism<sup>d</sup></b>	<b>465</b>	<b>2187</b>	<b>150</b>	<b>0.205</b>	<b>354</b>	<b>24</b>	<b>0.116</b>	<b>0.089</b>	<b>5.422</b>	<b>&lt; 0.0002</b>
<b>Haloperoxidase<sup>e</sup></b>	<b>240</b>	<b>889</b>	<b>118</b>	<b>0.161</b>	<b>105</b>	<b>14</b>	<b>0.068</b>	<b>0.093</b>	<b>6.531</b>	<b>&lt; 0.0002</b>
<b>Ice structuring<sup>f</sup></b>	<b>337</b>	<b>46</b>	<b>4</b>	<b>0.005</b>	<b>105</b>	<b>9</b>	<b>0.043</b>	<b>-0.038</b>	<b>-5.552</b>	<b>&lt; 0.0002</b>
Indolacetamide hydrolase <sup>g</sup>	457	626	43	0.059	198	13	0.063	-0.004	-0.374	0.7084
MerA <sup>h</sup>	454	1417	99	0.135	369	26	0.126	0.009	0.597	0.5505
NifH <sup>i</sup>	279	51	5	0.007	30	3	0.014	-0.007	-1.536	0.1245
NodA <sup>j</sup>	160	62	12	0.016	2	0	0	0.016	na	na
<b>NodB<sup>k</sup></b>	<b>136</b>	<b>638</b>	<b>150</b>	<b>0.205</b>	<b>57</b>	<b>13</b>	<b>0.063</b>	<b>0.142</b>	<b>9.321</b>	<b>&lt; 0.0002</b>
NodC <sup>l</sup>	245	743	97	0.133	211	27	0.13	0.003	0.133	0.8942
<b>NodD<sup>m</sup></b>	<b>249</b>	<b>118</b>	<b>15</b>	<b>0.021</b>	<b>21</b>	<b>2</b>	<b>0.01</b>	<b>0.011</b>	<b>1.991</b>	<b>0.0465</b>
<b>NodE<sup>n</sup></b>	<b>312</b>	<b>719</b>	<b>73</b>	<b>0.1</b>	<b>277</b>	<b>28</b>	<b>0.135</b>	<b>-0.035</b>	<b>-2.43</b>	<b>0.0151</b>
NodF <sup>o</sup>	90	12	4	0.005	5	1	0.005	0	0	1
<b>NodG<sup>p</sup></b>	<b>194</b>	<b>1905</b>	<b>314</b>	<b>0.43</b>	<b>422</b>	<b>69</b>	<b>0.333</b>	<b>0.097</b>	<b>4.465</b>	<b>&lt; 0.0002</b>
NodH <sup>q</sup>	242	13	1	0.001	5	0	0	0.001	na	na
NodI <sup>r</sup>	309	15836	1639	2.242	2609	270	1.304	0.938	na	na
<b>NodJ<sup>s</sup></b>	<b>268</b>	<b>702</b>	<b>83</b>	<b>0.114</b>	<b>150</b>	<b>17</b>	<b>0.082</b>	<b>0.032</b>	<b>2.407</b>	<b>0.0161</b>
RecA <sup>t</sup>	276	6306	731	1	1789	207	1	0	na	na

<sup>a</sup>glycine betaine catabolism <sup>b</sup>methylhalide uptake <sup>c</sup>dimethylsulfoniopropionate catabolism

<sup>d</sup>dimethylsulfide uptake <sup>e</sup>methylhalide production <sup>f</sup>ice binding/nucleating <sup>g</sup>indole-3-acetate production

<sup>h</sup>mercury reduction <sup>i</sup>nitrogen fixation <sup>j-s</sup>nodulation factor production <sup>t</sup>DNA repair

**Table 4.2.** Most abundant orders according to classification of 16S rRNA gene reads.

<b>Order</b>	<b>Number of reads</b>	<b>Percent of total 16S reads</b>
<i><b>Frost flower</b></i>		
Rhizobiales	7247	51.96
Actinomycetales	332	2.38
Flavobacteriales	181	1.30
Rickettsiales	177	1.27
Stramenopiles	134	0.96
Euglenozoa	127	0.91
Oceanospirillales	111	0.80
Rhodobacterales	80	0.57
Sphingomonadales	41	0.29
Mycoplasmatales	41	0.29
unclassified	4676	33.53
<i><b>Young sea ice</b></i>		
Stramenopiles	1441	21.56
Actinomycetales	585	8.75
Rhizobiales	541	8.10
Euglenozoa	128	1.92
Rickettsiales	64	0.96
Flavobacteriales	61	0.91
Oceanospirillales	35	0.52
Chlorophyta	31	0.46
Clostridiales	19	0.28
Mycoplasmatales	15	0.22
unclassified	3763	56.30

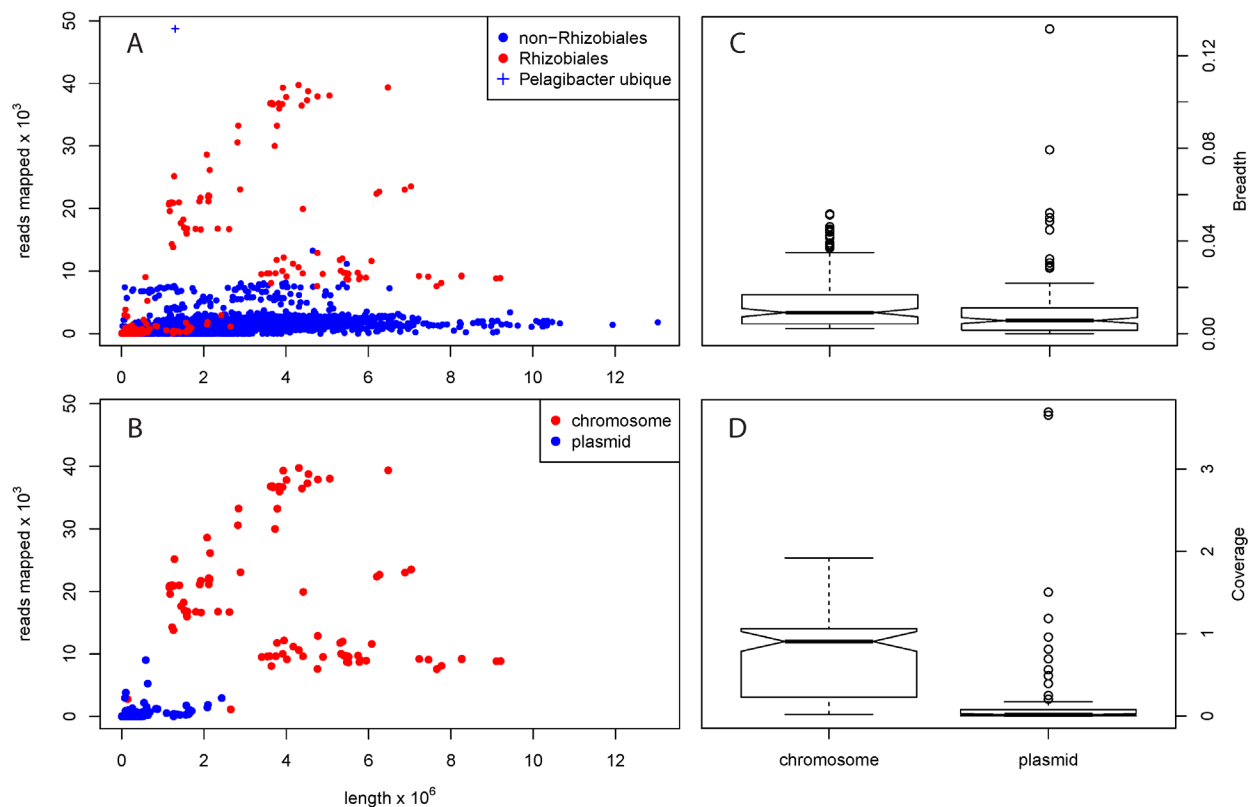
**Table 4.3.** Blastn results for assembled contigs longer than 10,000 bp, searched against nt as single queries.

<b>Contig</b>	<b>Number of putative CDS</b>	<b>Top hit in nt</b>	<b>Max score</b>
NODE_1473	13	<i>Sinorhizobium meliloti</i> Rm41 complete genome	4947
NODE_171	11	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymB, complete sequence	2105
NODE_229	6	<i>Rhodopseudomonas palustris</i> DX-1, complete genome	1274
NODE_448	8	<i>Agrobacterium tumefaciens</i> str. C58 circular chromosome, complete sequence	874
NODE_870	17	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymA, complete sequence	807
NODE_2	10	<i>Rhizobium tropici</i> CIAT 899, complete genome	740
NODE_100	11	<i>Sinorhizobium meliloti</i> 1021 plasmid pSymB, complete sequence	621
NODE_213	8	<i>Rhizobium leguminosarum</i> insertion sequence ISR1F7-2 putative transposase (Tnp) gene, complete cds	522
NODE_525	8	<i>Rhizobium leguminosarum</i> bv. <i>viciae</i> plasmid pRL7 complete genome, strain 3841	515
NODE_14	12	<i>Sinorhizobium fredii</i> NGR234, complete genome	497
NODE_389	6	<i>Caulobacter</i> sp. K31 plasmid pCAUL01, complete sequence	345
NODE_411	8	<i>Sinorhizobium meliloti</i> SM11, complete genome	345
NODE_388	2	<i>Burkholderia vietnamiensis</i> G4 plasmid pBVIE03, complete sequence	69.8

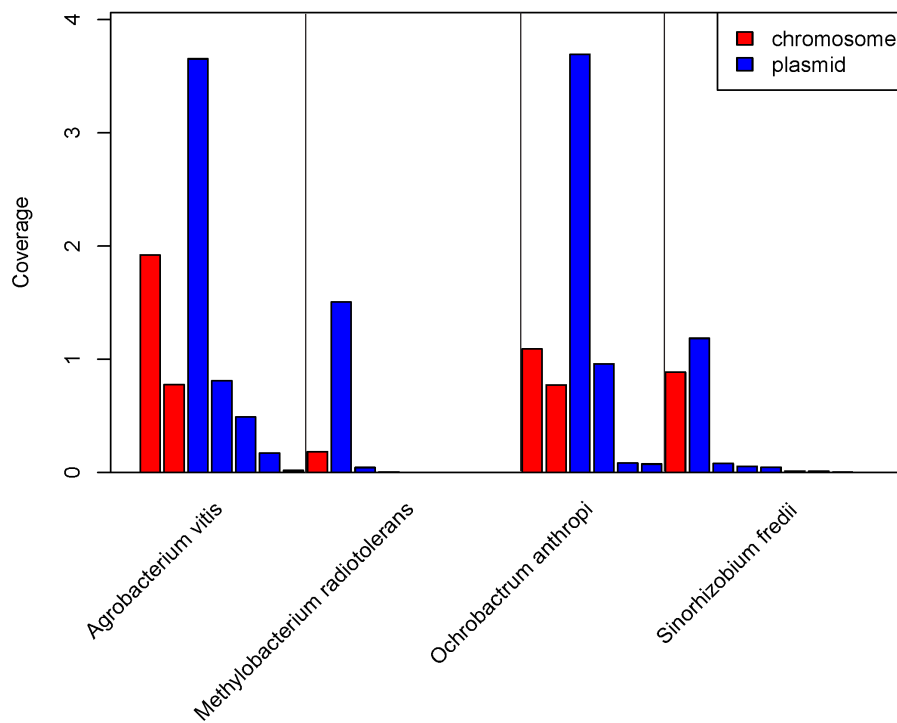


**Table 4.4.** Putative peptides placed on edges corresponding to proteins from the Rhizobiales.

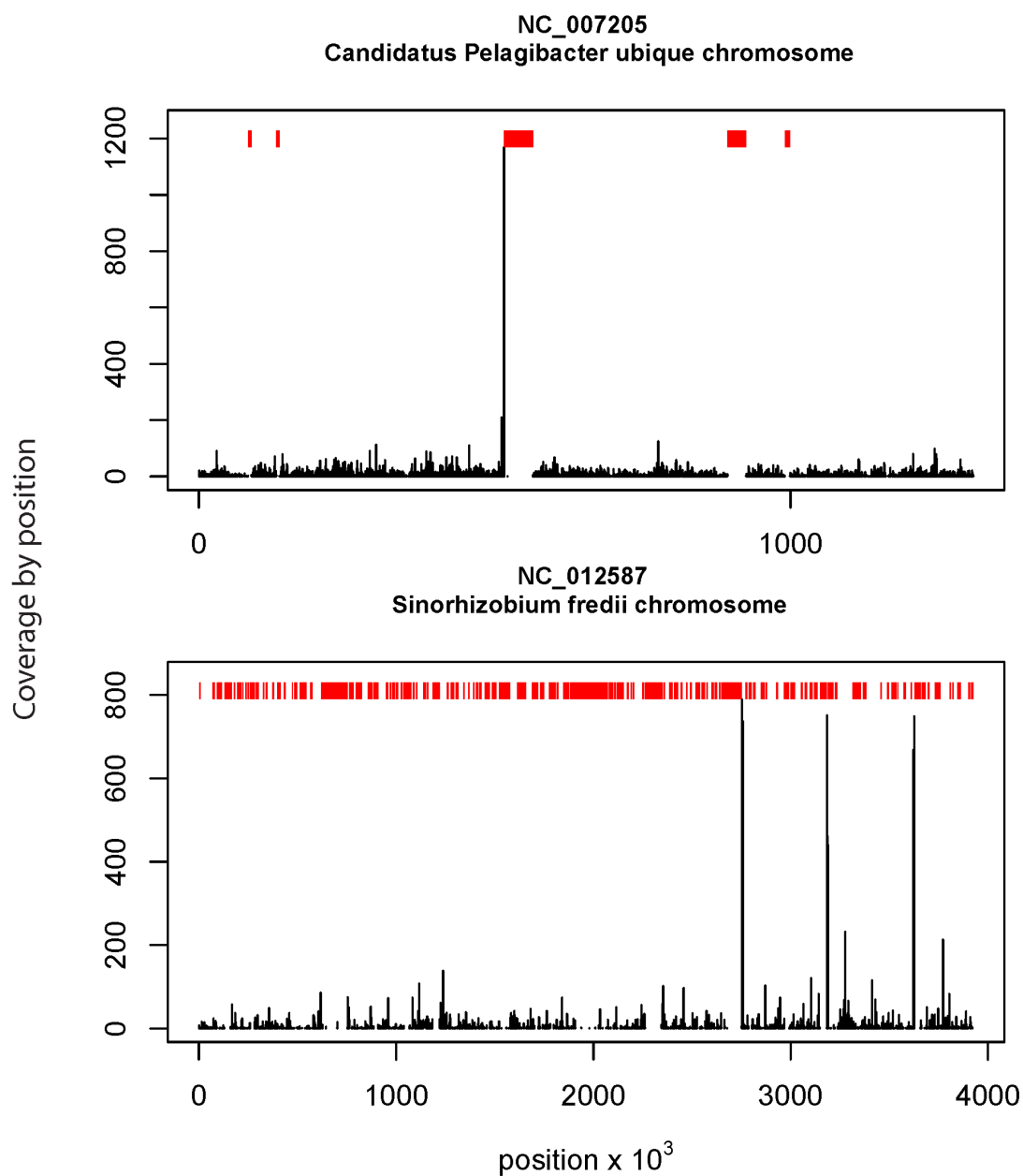
<b>Protein group</b>	<b>Number of reads</b>	<b>Fraction of total</b>
indoleacetomide	73	0.116
hydrolase		
betaine methyltransferase	406	0.280
dms catabolism	1,137	0.520
haloperoxidase	467	0.525
MerA	0	0



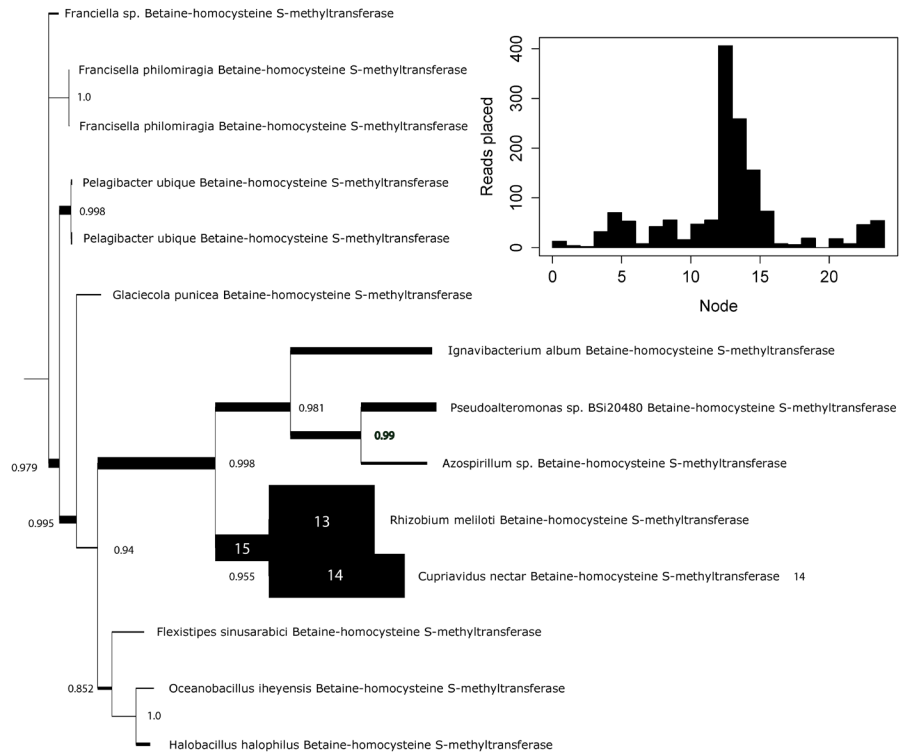
**Fig. 4.1. Coverage of chromosomes and plasmids in the frost flower metagenome.** (A) all available chromosomes and plasmids; (B) all Rhizobiales chromosomes and plasmids; (C) distribution of breadth for chromosomes and plasmids of the Rhizobiales; (D) distribution of coverage for chromosomes and plasmids of the Rhizobiales.



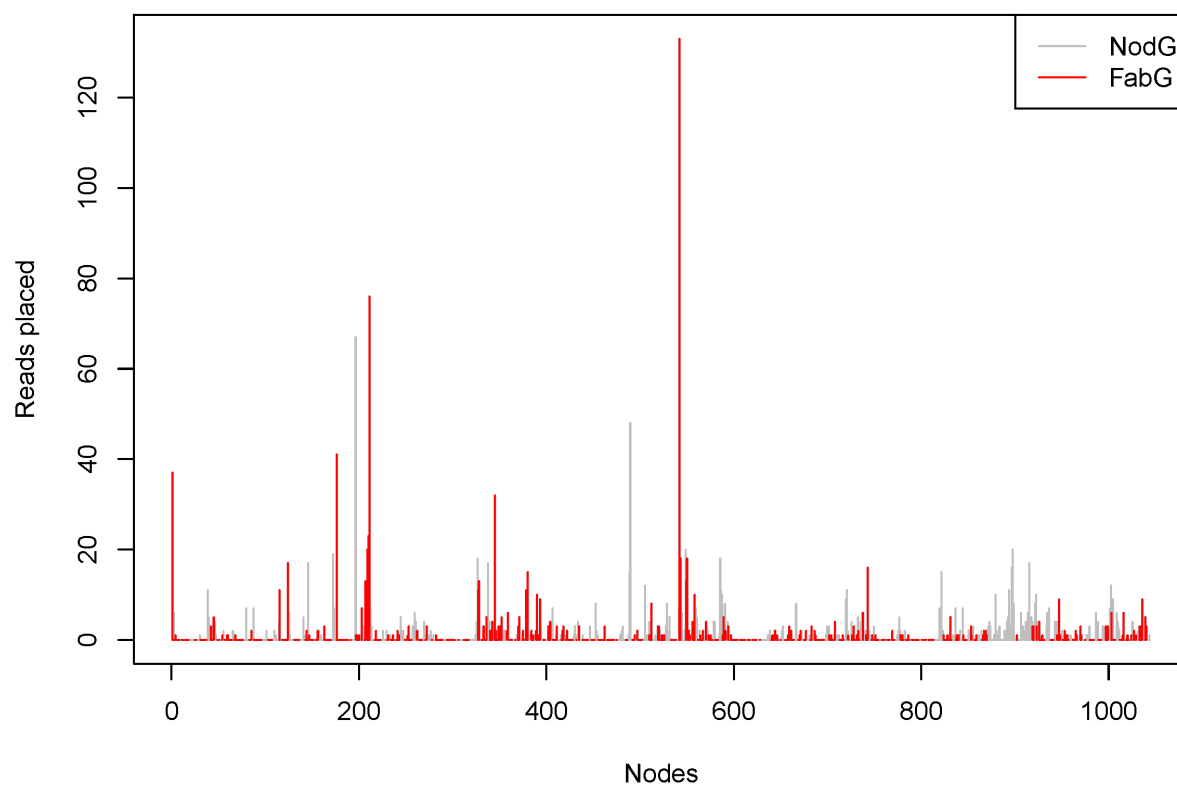
**Fig. 4.2. Coverage for chromosomes and plasmids for strains with well covered plasmids.**



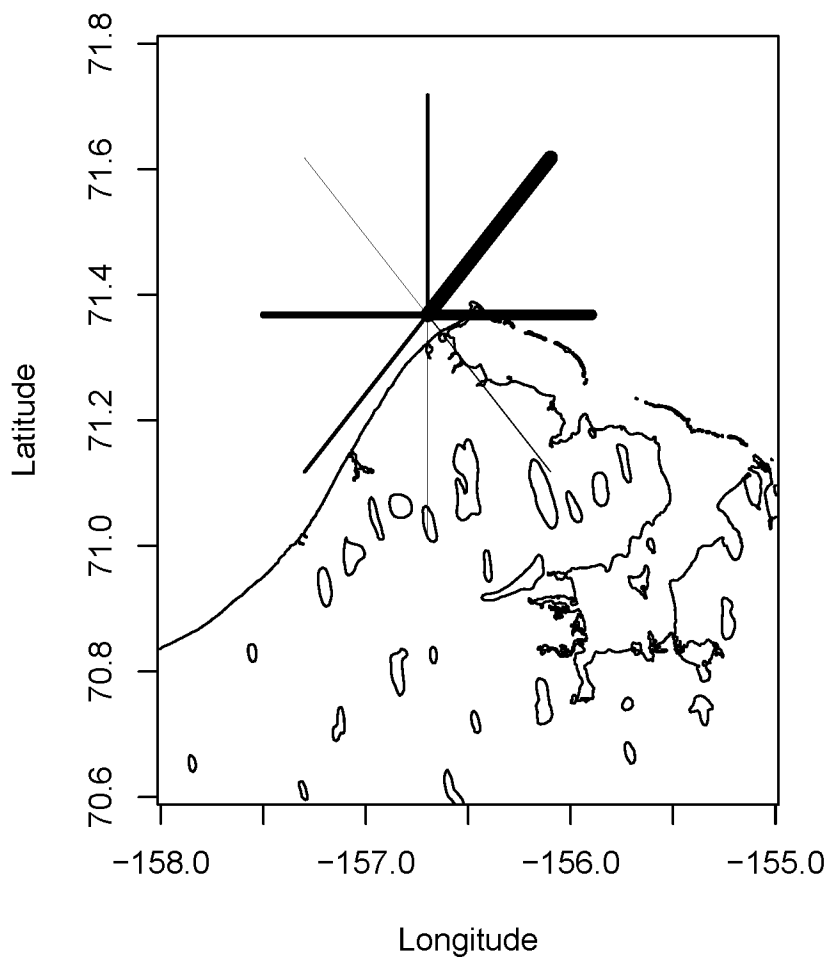
**Fig. 4.3.** Coverage across the chromosomes of *Candidatus Pelagibacter ubique* (top) and *Sinorhizobium fredii* HH103 (bottom). Gap regions, highlighted by the red boxes near to the top of the plot, are defined as any region greater than 5,000 nt with zero coverage.



**Fig. 4.4. Phylogenetic placement of betaine methyltransferase read translations from the frost flower metagenome on a reference tree of UniProt sequences.** The width of each edge is proportional to the number of reads placed on the edge. Exact values are given in the histogram, by node number on the reference tree. Node numbers for select locations are given in white on the reference tree. Confidence values for the reference tree are given in black at each bifurcation.



**Fig. 4.5. Phylogenetic placement of NodG and FabG read translations from the frost flower metagenome on a reference tree of UniProt sequences.** Reference sequences annotated as FabG are shown in red, reference sequences annotated as NodG are shown in grey, along with ambiguous nodes.



**Fig. S4.1. Wind magnitude at the Barrow NOAA observatory for April 14 – 21, 2010.**

Wind magnitude as a fraction of the total is given by the width of the directional lines. The directional lines converge at the sample site.

## Chapter 5

### **Evidence for enhanced horizontal gene transfer in psychrophile genomes and possible links to Phanerozoic climate**

#### ABSTRACT

Prokaryotic genomes can acquire genes horizontally through a number of mechanisms, including conjugation with plasmids or transposable elements, incorporation with temperate phage as prophage, and infection by gene transfer agents. The opportunities for horizontal gene transfer (HGT) may be elevated in cold environments given the adoption of lysogeny as a phage survival mechanism, increased contact rates between bacteria and HGT vectors in sympagic media such as sea ice or permafrost, and strong selective pressure. To test this hypothesis, we assessed the degree of genomic plasticity in 20 psychrophile, and in 20 taxonomically related mesophile genomes, as the divergence of their predicted proteomes relative to their 16S rRNA genes. As predicted, the psychrophile genomes appeared to be more divergent than their mesophile counterparts (Mann-Whitney test;  $p = 3.5 \times 10^{-6}$ ). To put these results in the context of bacterial evolution, we estimated the date of acquisition for coding sequences with anomalous GC content and compared them to the Phanerozoic climate record. We found that the acquisition of horizontally transferred genes retained in the genomes of contemporary psychrophiles was elevated during periods when the global mean temperature was below the current mean. In contrast, we found no connection between gene acquisition and retention and cold periods for mesophile genomes. These results suggest a strong role for large-scale climate events in the evolution of microorganisms adapted to the cold.



## 5.1 Introduction

Horizontal gene transfer (HGT) plays an important role in the evolution of prokaryotic genomes and can play a key role in the transfer of ecosystem functions between clades. Recent genomic studies have suggested that HGT in cold environments may be responsible for the acquisition of diverse functions, including pollutant degradation (Ma, Wang, & Shao, 2006), toxin resistance (Feng, Powell, Wilson, & Bowman, 2014), the production of protective extracellular polymers (EPS) (J. P. Bowman, 2008; Allen et al., 2009; Feng et al., 2014), ice affinity (Kiko, 2010; Raymond & Kim, 2012), and acquisition and catabolism of compatible solutes (R. E. Collins & Deming, 2013). These genomic studies are supported by reports of high numbers of HGT vectors in a variety of cold environments. Studies of extracellular DNA (R. Collins & Deming, 2012a) and virus-like particles in young sea ice (R. Collins & Deming, 2012b), an environment characterized by sharp temporal and spatial gradients in temperature and salinity (J. S. Bowman & Deming, 2010), have indicated high concentrations of both of these HGT vectors. Many other cold environments, including mature sea ice (Gowing et al., 2004), cryoconite holes and freshwater polar lakes (Anesio, Mindl, Laybourn-Parry, Hodson, & Sattler, 2007; S  wstr  m, Anesio, Gran  li, & Laybourn-Parry, 2007; S  wstr  m, Gran  li, Laybourn-Parry, & Anesio, 2007), hypersaline polar lakes (Lisle & Priscu, 2004; Laybourn-Parry, Marshall, & Madan, 2007), and polar marine waters (Guixa-Boixereu, Vaqu  , Gasol, S  nchez-C  mara, & Pedr  s-Ali  , 2002; Wells & Deming, 2006c), also contain high numbers of virus-like particles or infected bacterial cells. Analyses of genomes of psychrophilic (maximum growth temperature < 20  C) Bacteria and Archaea have revealed the presence of prophage and transposable elements (M  the et al., 2005; Wang et al., 2008; Zhao, 2011; Freese et al., 2013).

The ease with which horizontal gene transfer proceeds in a given microbial environment is influenced by both biological and physical factors. In very cold microbial habitats a combination of unique factors may combine to support high levels of HGT, with these factors reaching a maximum in cold sympagic environments such as sea ice, glacial ice, and permafrost. There, cold temperatures stabilize both phage and extracellular DNA, while contact rates between bacteria and phage, and between bacteria and bacteria, increase as temperature and, correspondingly, the volume of the liquid phase of the ice decrease (Wells & Deming, 2006b). Because of this reduction in brine volume, decreasing temperatures in ice are also coupled to elevated osmotic stress, increasing membrane permeability and providing strong selective pressure for the acquisition of survival genes (Feng et al., 2014). In sea ice and other sunlit polar environments, UV radiation and oxygen stress add additional selective pressure (Ewert & Deming, 2013).

The adoption of lysogeny and pseudolysogeny as a phage survival strategy may also play an important role in HGT. The temperate phage lifestyle has been proposed as a viral survival mechanism under oligotrophic conditions (Wells & Deming, 2006a; Paul, 2008), a strategy that may also apply to low temperature environments where microbial growth is often dependent on transient or seasonally permissive conditions (Wells & Deming, 2006a). Although rates of lysogeny and pseudolysogeny in polar environments are not well known, initial estimates suggest that they are high (S  wstr  m, Gran  li, et al., 2007; S  wstr  m, Lisle, Anesio, Priscu, & Laybourn-Parry, 2008), which would explain the high prophage and transposable element content of some psychrophile genomes.

Despite the potential importance of HGT among cold adapted microbial communities, and the favorable physical and biological environment, an investigation of HGT among the

available cold adapted genomes has not been undertaken. Here we estimate the degree of genomic plasticity between 20 psychrophile genomes and a control group of 20 taxonomically related mesophile genomes. We used a method based on the divergence of the putative proteomes in each group with respect to the divergence of a conserved gene in each group, in this case the 16S rRNA gene (Fig. 5.1). This approach provides a framework for future analysis as the collection of psychrophile genomes continues to grow. More broadly, it can be used to identify hotspots of genomic plasticity within any population of genomes.

We then considered possible connections between temperature and HGT from the perspective of global climate. Earth's climate has been highly variable over the last billion years, with significant portions of this time period being cooler than present (Shaviv & Veizer, 2003). These cold periods may have served as key opportunities for adaptation and the environmental expansion of psychrophiles as the range of cold sympagic environments increased along with the prevalence of temperate phage. To test the hypothesis that the rate of HGT among psychrophiles is enhanced during geologic cold periods, we applied the GC amelioration dating technique (Lawrence & Ochman, 1997) to coding sequences (CDS) with anomalous GC content, representing likely HGT events, in psychrophile and mesophile genomes, analyzing the distribution of these events over the last 650 My.

## **5.2 Results**

### **5.2.1 Genome divergence**

The ratio of genome divergence to 16S rRNA gene divergence ( $\text{div}_G$ ) was higher for the psychrophile group than the mesophile group (Fig. 5.2) in an analysis sensitive enough to identify a difference of five coding sequences (CDS) between two genomes (Fig. S5.1).

Although this parameter did not meet the Anderson-Darling test for normality (Thode, 2002), the

data fit a normal distribution model better than a gamma distribution model (log-likelihood ratio = 75.26,  $p \ll 0.001$ ); thus the Mann-Whitney test was used to compare the two groups and means were calculated as for normally distributed data (psychrophile mean = 0.118, mesophile mean = 0.146,  $w = 267$ ,  $df = 798$ ,  $p = 3.5 \times 10^{-6}$ ; Fig. 5.2). The mechanism underlying the observed genome divergence is evident in the difference between the two groups for the normalized compositional vector and the 16S rRNA gene distances (Figs. S5.2, S5.3). The former differed significantly between psychrophile and mesophile genomes ( $w = 89,068$ ,  $p = 2.49 \times 10^{-8}$ ), while the latter did not ( $w = 73,312$ ,  $p = 0.7134$ ). Mean  $\text{div}_G$  values for each strain ( $\bar{x}\text{div}_G$ ) in the psychrophile group differed significantly from the mesophile group (psychrophile mean = 0.188, mesophile mean = 0.146,  $w = 293$ ,  $p = 0.01$ ), as did the median  $\bar{x}\text{div}_G$  values of the two groups by an alternate test (Chambers, Cleveland, Kleiner, & Turkey, 1983). All but four of the strains in the psychrophile group had a higher  $\bar{x}\text{div}_G$  value than their counterpart in the mesophile group (Table 5.1).

The highest  $\text{div}_G$  value was obtained between the genomes of *Octadecabacter arcticus* 238 and *O. antarcticus* 307 (raw|normalized distance: 16S rRNA gene = 0.007|0.016, compositional vector = 6,121|0.585,  $\text{div}_G = 0.569$ ). A high value was also obtained for the comparison between *Psychroflexus torquis* and *P. gondwandensis* (raw|normalized distance: 16S rRNA = 0.011|0.024, compositional vector = 5,672|0.449,  $\text{div}_G = 0.425$ ), though compositional vector distance and thus  $\text{div}_G$  are likely influenced by the draft state of the latter genome. The most plastic genome in our analysis, according to  $\bar{x}\text{div}_G$ , was the mesophilic Euryarchaeota *Methanosarcina mazei* Tuc01 ( $\bar{x}\text{div}_G = 8.05$ ), followed by the related psychrophile *Methanococcoides burtonii* DSM 6242 ( $\bar{x}\text{div}_G = 7.62$ ) and the psychrophilic

Gammaproteobacteria *Aliivibrio salmonicida* LFI1238 ( $\bar{x}\text{div}_G = 5.54$ ) and *O. arcticus* 238 ( $\bar{x}\text{div}_G = 5.09$ ).

To examine the mechanism underlying the largest  $\text{div}_G$  value more closely, we conducted a tblastx (Altschul et al., 1997) analysis of the CDS of *Octadecabacter antarcticus* 307 against the CDS of *O. arcticus* 238. This analysis indicated that 1,800 CDS in *O. arcticus* 238 are absent from *O. antarcticus* 307, with 831 CDS found in *O. antarcticus* but not *O. arcticus*. Evaluation of the Genbank annotations for both strains identified 942 HGT specific genes (coding for transposases, integrases, recombinases, or phage-specific proteins). Of these 942 genes, 11 were not found in *O. arcticus* 238, while 626 were absent from *O. antarcticus* 307 (Fig. S5.4).

### 5.2.2 CDS with anomalous GC content

As determined by the GC anomaly method, the psychrophile genomes had a total of 3,308 anomalous CDS (4.94 % of all CDS) (Fig. S5.5). The mesophile group had a total of 3,180 anomalous CDS (4.78 % of all CDS) (Fig. S5.6). In the psychrophile group the anomalous CDS represent 1,469 unique protein families (pfams), compared to the 1,369 pfams for the mesophile group. Several families of transposases (Y1\_Tnp, DDE\_Tnp\_ISL3, HTH\_Tnp\_ISL3), representing likely HGT events by GC anomaly, were present more often in the psychrophile group than in the mesophile group, while other families of transposases and some phage integrases, also suggestive of HGT events, were present more often in the mesophiles (DDE\_Tnp\_IS66, Phage\_int\_SAM\_4, Phage\_integrase) (Table S5.2). Some of the pfams (listed in Table S5.2) found more frequently among the psychrophiles were broadly distributed among those strains. Of particular interest is DUF1602 (a domain of unknown function), detected as a likely HGT event across all psychrophile strains and all mesophile strains

except for *Acinetobacter baumannii* AYE and *Methanosarcina mazei* Go1. In these two strains a non-anomalous CDS codes for this pfam. Other widely distributed pfams among the likely psychrophile HGT genes are Response\_reg (a sensory domain), PP-binding (which functions in fatty-acid biosynthesis), and HTH\_1 (found in regulatory proteins).

Of the 6,488 CDS with anomalous GC content evaluated, the 2,805 CDS that yielded positive GC1 and GC2 acquisition date estimates differing by more than 100 My were discarded. For the remaining 3,683 CDS, 37 acquisition date estimates were high confidence (both dates passed a similarity test, as described in Methods), 945 were medium confidence (both dates were within 100 My), and 2,787 were low confidence (only GC1 or GC2 provided a positive estimated date of acquisition). Dates of acquisition for the CDS with anomalous GC content suggest that an excess of HGT events have been retained in the psychrophile genomes, but not the mesophile genomes, from periods when the global mean temperature was predicted to be below the current mean (Fig. 5.4). Approximately 57 % of the last 650 My has been colder than the present (Veizer, Godderis, & François, 2000), with a sum total of 562.74 anomalies for the psychrophile group and -757.61 for the mesophile group during cold periods. Monte Carlo simulations of random distributions of normalized HGT events indicated that these distributions are not random; in  $10^6$  iterations the anomaly was not exceeded for the psychrophiles yet was always exceeded for the mesophiles.

To explore the relationship between increases and decreases in bacterial diversity (manifested as the rate of HGT) and the diversity of higher taxa, we compared the occurrence of HGT events in the psychrophile and mesophile groups to the rates of species origination and extinction in the fossil compendium of Sepkoski et al. (Sepkoski, Jablonski, & Foote, 2002) as calculated by Foote (Foote, 2000) and obtained from <http://strata.geology.wisc.edu/jack/start.php>

(Fig. 5.4). Because the HGT occurrence bins differ from the Sepkoski curve dates, a spline function was used on the Sepkoski curve data to estimate values at the HGT occurrence time points. No correlation was found between the number of putative HGT transfers in the psychrophile or mesophile groups and the rate of extinction or origination of marine metazoan genera. The mass extinction events plotted in Fig. 5.4 rarely co-occurred with large positive or negative HGT anomalies. A modest positive anomaly occurred with the Cretaceous-Paleogene mass extinction (65 Ma) and a larger positive anomaly preceded the Triassic-Jurassic extinction (200 Ma) in both groups, while a positive anomaly co-occurred with the Late Devonian event (375 Ma) in the psychrophile group.

### 5.3 Discussion

In a comparison of predicted proteome and 16S rRNA gene distance between taxonomically similar psychrophile and mesophile groups, we found evidence that the psychrophile proteomes are more divergent, relative to their 16S rRNA genes, than their mesophile counterparts. There is a clear increase in the  $\text{div}_G$  values of the psychrophile group over the mesophile group driven by the increased compositional vector distance for the psychrophiles. Although we found a significant difference between the mean  $\bar{x}\text{div}_G$  for the two groups, the test was limited by the relatively small number of available psychrophile genomes and narrow taxonomic distribution. Some mesophiles, for example, exhibited a high degree of genomic plasticity, particularly *Methanosarcina mazei* Tuc01 (Table 5.1). As more psychrophile genomes are published, applying this test to a larger, more diverse group of genomes will be possible.

The significant differences in mean and median  $\bar{x}\text{div}_G$  values and mean  $\text{div}_G$  values provide strong evidence of increased genomic plasticity among psychrophiles, yet alone are not

definitive of increased HGT. Genome streamlining, or the loss of functional genes from the genome, would have a similar impact on proteome divergence, as would genome duplication events. Genome rearrangement with no loss or gain of genes, however, would have no effect on compositional vectors. We did not assess the degree of genome duplication directly, but the numbers of CDS in the two groups were very similar (mean psychrophile = 3,302, mean mesophile = 3,329,  $t = -0.1181$ ,  $p = 0.91$ ), suggesting that gene loss without replacement (by HGT or duplication) was not the cause of the observed difference in divergence.

The idea behind the  $\bar{x}div_G$  value is to describe the relative plasticity of all genomes in an experimental group (in our case psychrophiles) relative to those in a taxonomically related mesophile group. For the top scoring strains in our analysis, the  $\bar{x}div_G$  value is well supported by previous studies. The genomes of *Methanosarcina mazei* Tuc01 and *Methanococcoides burtonii* DSM were reported to contain in excess of 50 % and 40 % horizontally transferred genes, respectively (Allen et al., 2009). Although the same study found no connection between temperature and the degree of HGT among various Archaea, the authors noted that the parametric HGT detection methods used are prone to error. That *M. mazei* Tuc01 had a higher  $\bar{x}div_G$  value than *M. burtonii* DSM in our analysis supports the idea that many factors can trump temperature as a driver of HGT, but more genomes of psychophilic Archaea are needed to determine if temperature is less significant for that domain than it appears to be for Bacteria. The Bacterium *Aliivibrio salmonicida* LFI12388, with the highest  $\bar{x}div_G$  value of either domain, has also been noted as having a highly plastic genome containing a large number of insertion sequences (Hjerde et al., 2008). Genomic plasticity in this strain is thought to be related to its pathogenic nature, though other pathogens in our analysis had less extreme  $\bar{x}div_G$  values.



The  $\bar{x}\text{div}_G$  score between *Octadecabacter arcticus* 238 (third highest  $\bar{x}\text{div}_G$  value for the psychrophile group) and closely related *Octadecabacter antarcticus* 307 revealed an extraordinary degree of genome divergence over a short taxonomic distance (Vollmers et al., 2013). Although the genome of *O. antarcticus* 307 hosts a number of transposable elements, the high  $\text{div}_G$  is driven by gene acquisition in *O. arcticus* 238, probably related to the large number of transposable elements found in that genome (Fig. S5.4). Why two bacteria with shared ancestry and temperature class and ostensibly the same habitat (sea ice) have genomes shaped by such differing levels of HGT is unclear. Not enough psychrophile genomes are currently available to consider endemism expressed as an Arctic vs. Antarctic difference in HGT, but the different climatic histories and thus selective pressures inherent to these regions present intriguing possibilities.

The genome of the Flavobacteriaceae *Psychroflexus torquis* ATCC 00755, isolated from Antarctic sea ice, has also been reported as strongly influenced by HGT (Feng et al., 2014). Although this strain has a high  $\text{div}_G$  value compared with *P. gondwanensis* ACAM44, its genome is not exceptionally plastic among the analyzed psychrophiles, ranking 13 out of 20 for  $\bar{x}\text{div}_G$ . It is one of only four psychrophile strains, however, to have a lower  $\bar{x}\text{div}_G$  value than its mesophilic counterpart (*Flavobacteriales bacterium* HTCC2170,  $\bar{x}\text{div}_G = 3.61$ ). Because *P. gondwanensis* ACAM44 is not psychrophilic (though it was isolated from a cold environment) and its genome is in draft state, we could not include it in either group to calculate a  $\bar{x}\text{div}_G$  value, precluding a direct comparison of genome plasticity between the two *Psychroflexus* strains.

Prior speculation on the potential role of cold glacial periods in enhancing HGT considered the Neoproterozoic period, over 580 million years ago, when Earth was thought to be nearly or completely covered with ice (Anesio & Bellas, 2011). By examining the series of more

recent and well-established cold geologic periods and applying a quantitative genomic approach, we found a significant correlation between the normalized abundance of HGT events for the psychrophile genomes and these cold periods. We remain cautious about drawing broad conclusions from this analysis, because the limited resolution of GC amelioration times makes it difficult to correlate positive HGT anomalies with specific events during the Phanerozoic. Specific geological events such as glaciations can happen rapidly (Royer, 2006) relative to the resolution of our analysis or to estimates of global mean temperature, as can gene acquisition events and subsequent sweeps through a population (Nowell, Green, Laue, & Sharp, 2014). Further compounding the analysis are the assumptions inherent to use of a molecular clock based on inferred rates of divergence between *Escherichia coli* and *Salmonella enterica* (Lawrence & Ochman, 1997), which may differ from rates between other strains. As molecular clocks become more sophisticated, and more sequenced genomes become available, the correlation between HGT and geologic cold periods can be evaluated more precisely. Within the constraints of current knowledge, our analysis compensates for existing limitations by basing the probable distribution of acquisition times on a large number of anomalous CDS.

Despite the lack of a positive correlation between mesophile gene acquisition and geologic temperature, the overall pattern of anomalies appears generally similar between the two groups. The major difference occurs at the transition from the neo-Proterozoic to the Phanerozoic (500–600 Ma), where the mesophiles have a positive anomaly and the psychrophiles have a negative anomaly. One explanation may be correspondence with the Cambrian Period (541–485 Ma), notable for extremely warm conditions and the rapid diversification of complex life known as the Cambrian Explosion. Because many of the mesophile genomes come from strains associated, either directly or taxonomically, with a

symbiotic lifestyle, this rapid metazoan diversification may have been accompanied by a rapid diversification of prokaryotic symbionts. In contrast to the main conclusions of this paper, we found an accumulation of negative anomalies for both groups of genomes in the period 270 to 370 Ma (early Permian to late Devonian Era), a time marked by glaciation.

The early Jurassic, from approximately 160 Ma to 220 Ma, was the period with the strongest positive HGT rate anomaly for the psychrophiles and a significant positive anomaly for the mesophiles. Although the Jurassic is often considered a warm period, ice-rafted debris suggests the presence of at least some glaciers or seasonal sea ice; positive calcite  $\delta^{18}\text{O}$  values indicate that the global mean temperature was cooler than present (Veizer et al., 2000; Shaviv & Veizer, 2003). The Jurassic follows an extreme warm period from 220 to 270 Ma, the breakup of the supercontinent Pangaea, and the Triassic-Jurassic extinction event. A short-lived “super greenhouse” induced by magma extrusions has been proposed as a mechanism for this extinction (Whiteside, Olsen, Eglinton, Brookfield, & Sambrotto, 2010). This event was followed by two successive spikes in atmospheric  $\text{CO}_2$  concentrations (Royer, 2006) that correspond with large HGT anomalies in both groups of genomes. Thus a large and rapid shift in global temperature with additional environmental stressors may have provided strong selective pressure for horizontal gene transfer throughout the microbial realm.

Despite limitations on HGT date estimates by the GC amelioration technique, the relationship detected between gene acquisition and cold periods during the last 650 My supports the influence of environmental temperature on gene acquisition and retention by psychrophiles. This link is logical: during cold periods psychrophiles have a greatly expanded range that would expose them to new environmental conditions and new pools of genetic material. It is also reasonable to expect an increase in temperate phage during cold periods, as the extent of

geographical regions with strong seasonal shifts in temperature and metabolic activity grows. Our survey of twenty psychrophile genomes suggests an elevated level of genomic plasticity that is consistent with this hypothesis. Despite the generally understood decrease in growth rates at low temperatures, a high rate of evolution has been sustained in this group of cold-adapted microorganisms.

## **5.4 Materials and Methods**

### **5.4.1 Genome selection and proteome prediction**

Twenty completed psychrophile genomes were identified from the hima database (<http://eric.org/work/hima/>), temperature annotations in GOLD (<http://www.genomesonline.org/cgi-bin/GOLD/index.cgi>), and reviews by Casanueva et al. (2010) and Siddiqui et al. (2013) (Fig. S5.7). Only true psychrophiles or stenopsychrophiles, defined as having a temperature growth maximum below 20°C, were used. This approach precluded genomes from several interesting cold-tolerant strains from our analysis, such as the sea ice bacterium *Shewanella frigidum*, but was necessary to constrain our analysis along the temperature tolerance continuum. The available psychrophile genomes have a strong taxonomic bias toward the Gammaproteobacteria (Fig. S5.7). To reduce the effects of this bias, the mesophile set was selected from the GOLD database in a taxonomically aware fashion using the nearest taxonomic neighbor annotated as a mesophile for each psychrophile strain. Ties between nearest mesophile neighbors were broken by random selection. For both the mesophile and psychrophile groups care was taken to include all plasmids and additional chromosomes in the analyzed genome.

For each genome a putative proteome was determined by translating all open reading frames (ORFs), defined as 300 or more bases without a stop codon. The hmmscan command in

HMMER 3.0 (Eddy, 1998) was used to search the product of each ORF against the PFAM-A database (Punta et al., 2012). ORF products with at least one recognized protein domain at  $E < e^{-5}$  were considered to be coding sequences (CDS) and retained in the putative proteome. Only CDS that met this criterion were used in downstream analyses, except for the extraction of 16S rRNA genes from each complete genome. Because our final dataset is limited to proteins with known domains in the PFAM-A database, some true proteins were not considered. We chose to use this protein subset in place of the existing annotations to provide a consistent annotation across all of the genomes used in the analysis.

#### **5.4.2 Genome divergence**

From the putative proteome a normalized compositional vector was created following the method of Qi et al. (2004) using 5 amino acid kmers. A compositional vector was created independently for each putative protein in the proteome; each proteome is represented as a vector that is the sum of its individual protein vectors, with normalization applied to this summed vector to account for random background mutation (Qi et al., 2004). A distance matrix was constructed from all proteome compositional vectors using the Euclidean measure of distance.

The 16S rRNA gene was identified in each genome by blastn megablast search against the NCBI 16S microbial database. Only one match was allowed for each genome. Although most of the genomes contained multiple copies of the 16S rRNA gene, few sequence differences are expected between these different copies (Coenye & Vandamme, 2003). The identified 16S genes were extracted and aligned using the align.seqs command in Mothur (Schloss et al., 2009) against the greengenes reference alignment available on the Mothur website ([http://www.mothur.org/wiki/Greengenes-formatted\\_databases](http://www.mothur.org/wiki/Greengenes-formatted_databases)). The alignments were filtered to

the earliest start position (5') and latest end position (3') prior to distance matrix construction using the `dist.seqs` command.

Because the two measures of distance have a different range, the matrices were transformed to have a mean of 0 and a variance of 1 and then scaled to the range of 0 to 1. The  $\text{div}_G$  score was calculated as the difference between the normalized matrices (Fig. 5.3). The mean of each row in the  $\text{div}_G$  matrix was taken to generate a relative measure of the degree of plasticity for each genome ( $\bar{x}\text{div}_G$ ). The distances used in the  $\text{div}_G$  calculation were derived collectively on the two groups, while  $\bar{x}\text{div}_G$  was calculated only from other members in each strain's group. To evaluate the sensitivity of this method to genomic plasticity, random sets of five genes were successively added to the genome of *Colwellia psychrerythraea* 34H and  $\text{div}_G$  recalculated for 50 "generations" assuming no change to the 16S rRNA gene (Fig. S5.1).

Because the genomes of *Octadecabacter antarcticus* 307 and *O. arcticus* 238 were observed to have an unusually high  $\text{div}_G$  value, genes uniquely present in the larger genome of *O. arcticus* 238 were identified by searching the Genbank `ffn` file (all CDS identified in the Genbank annotation) for *O. arcticus* 238 using `tblastx` against a database generated from the `ffn` file for *O. antarcticus* 307. Genes likely to be involved in HGT were identified in *O. antarcticus* 307 by keyword search against the Genbank annotation for the *O. antarcticus* 307 genome using "transposase", "recombinase", "Is110", "integrase", "capsid", "tail", "phage", and "viral".

### **CDS with anomalous GC content**

For each genome, CDS with anomalous GC content were identified by comparing the GC content of each CDS with the mean GC content of all CDS. CDS with a GC content exceeding 2 standard deviations from the mean were classified as anomalous and flagged as likely HGT events (Figs. 5.4, 5.5). A time of acquisition was calculated for each HGT event using the

method of Lawrence and Ochman (1997) as implemented by Collins et al. (2013) using the R script *gcamel* with the following modifications: jackknifed (random sampling of 50 % of codons) age estimates in 1 My timesteps were calculated independently for both the first and second codon positions (GC1 and GC2). Because the GC content of the third codon position (GC3) estimated by reverse amelioration often fails to converge with the estimate from the GC1 and GC2 reverse amelioration, we allowed convergence from forward amelioration, resulting in a negative (future) acquisition date. Although these dates clearly do not reflect the actual date of acquisition, including them made it possible to generate approximately normal distributions of age estimates from the jackknife iterations and thus calculate a mean estimated date of acquisition. We also re-calculated the model parameters for estimating GC1 and GC2 from GC3, using 2,302 genomes from Genbank. The set of genomes was selected from the set of all available completed genomes and draft genomes with scaffolds and disallowing re-use of the first and second word in the genome name. This approach eliminated most repeat strains for each species present in the database.

The jackknife dates for the two independent time estimates were smoothed using a 10 My moving average and tested for similarity using the Kolmogorov-Smirnov test (Conover, 1971), selected for its insensitivity to non-normal distributions. Confidence for the date of acquisition was assigned as follows. If the distribution of GC1 and GC2 bootstrap estimates passed the similarity test, then confidence was high and the acquisition date was reported as the mean of both groups. If the mean of the GC1 and GC2 distributions differed by less than 100 My and both were positive, then the true mean was assumed to fall between the distribution means; confidence was medium and the acquisition date was reported as the mean of both means. If only one distribution had a positive mean, then confidence was low and the date of acquisition

was reported as the positive mean. CDS with GC1 and GC2 distribution means differing by more than 100 My were not used further. To develop a probability distribution of HGT events, each event was represented in downstream analysis as a 10 My window centered around the mean for high and low confidence dates, and as a window between the two means for medium confidence dates. This method of representation assumes no correlation between window size and time, which was confirmed through linear regression (Fig. S5.8).

The GC amelioration of horizontally transferred genes makes it more difficult to detect very old HGT events by GC anomaly. Date estimate windows were detrended by subtracting a spline function generated from all age estimates for each temperature class with the R package *dplr* (Bunn, 2008) (Fig. 5.4A). Positive values after detrending suggest positive anomalies in the rate of HGT. The time value for each bin was compared to the estimated timing of Phanerozoic cold periods, defined as periods with a global mean temperature below the current mean as calculated by Shaviv and Veizer (2003), and the approximate timing of the final Neoproterozoic snowball Earth glaciation (Pierrehumbert, Abbot, Voigt, & Koll, 2011). The first two snowball Earth glaciations, occurring between 660 and 720 and 730 and 775 Ma, are beyond the horizon for dating gene acquisition by GC amelioration. The probability that the high positive net anomaly for cold periods was random was assessed in a Monte Carlo simulation with  $10^6$  permutations. For each permutation, the dates were randomized and the net anomalies for cold and warm periods were re-calculated. Jackknife age estimates for the anomalous CDS and scripts for this project can be found at [https://github.com/bowmanjeffs/cold\\_HGT](https://github.com/bowmanjeffs/cold_HGT).

## **Acknowledgments**



JSB was supported by an EPA STAR fellowship and the University of Washington Astrobiology Program. JWD was supported by the Walters Endowed Professorship and NSF PLR 1203267. ERC was supported under NSF PLR 1203267.

## References

- Allen, M. A., Lauro, F. M., Williams, T. J., Burg, D., Siddiqui, K. S., De Francisci, D., . . . De Maere, M. Z. (2009). The genome sequence of the psychrophilic archaeon, *Methanococcoides burtonii*: the role of genome evolution in cold adaptation. *ISME J.*, 3(9), 1012–1035.
- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nuc. Acids Res.*, 25(17), 3389-3402.
- Anesio, A. M., & Bellas, C. M. (2011). Are low temperature habitats hot spots of microbial evolution driven by viruses? *Trends in Microbiol.*, 19(2), 52–57.
- Anesio, A. M., Mindl, B., Laybourn-Parry, J., Hodson, A. J., & Sattler, B. (2007). Viral dynamics in cryoconite holes on a high Arctic glacier (Svalbard). *J. Geophys Res.*, 112(G4).
- Bowman, J. P. (2008). Genomic analysis of psychrophilic prokaryotes. In R. Margesin, F. Schinner, J.-C. Marx & C. Gerday (Eds.), *Psychrophiles from Biodiversity to Biotechnology*. Berlin: Springer.
- Bowman, J. S., & Deming, J. W. (2010). Elevated bacterial abundance and exopolymers in saline frost flowers with implications for atmospheric chemistry and microbial dispersal. *Geophys. Res. Lett.*, 37(L13501).
- Bunn, A. G. (2008). A dendrochronology program library in R (dplR). *Dendrochronologia*, 26(2), 115–124.
- Casanueva, A., Tuffin, M., Cary, C., & Cowan, D. A. (2010). Molecular adaptations to psychrophily: the impact of ‘omic’ technologies. *Trends in Microbiol.*, 18(8), 374–381. doi: 10.1016/j.tim.2010.05.002
- Chambers, J., Cleveland, W., Kleiner, B., & Turkey, P. (1983). *Graphical methods for data analysis*: Wadsworth & Brooks/Cole.
- Coenye, T., & Vandamme, P. (2003). Intragenomic heterogeneity between multiple 16S ribosomal RNA operons in sequenced bacterial genomes. *FEMS Microb. Lett.*, 228(1), 45–49.
- Collins, R., & Deming, J. (2012a). Abundant dissolved genetic material in Arctic sea ice Part I: Extracellular DNA. *Pol. Biol.*, 34(12), 1819–1830. doi: 10.1007/s00300-011-1041-y

- Collins, R., & Deming, J. (2012b). Abundant dissolved genetic material in Arctic sea ice Part II: Viral dynamics during autumn freeze-up. *Pol. Biol.*, 34(12), 1831–1841. doi: 10.1007/s00300-011-1008-z
- Collins, R. E., & Deming, J. W. (2013). An inter-order horizontal gene transfer event enables the catabolism of compatible solutes by *Colwellia psychrerythraea* 34H. *Extremophiles*, 1-10.
- Conover, W. J. (1971). *Practical Nonparametric Statistics*. New York: John Wiley and Sons.
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, 14(9), 755–763.
- Ewert, M., & Deming, J. W. (2013). Sea Ice Microorganisms: Environmental Constraints and Extracellular Responses. *Biology*, 2(2), 603–628.
- Feng, S., Powell, S. M., Wilson, R., & Bowman, J. P. (2014). Extensive gene acquisition in the extremely psychrophilic bacterial species *Psychroflexus torquis* and the link to sea-ice ecosystem specialism. *Gen. Bio. Evol.*, 6(1), 133–148.
- Foote, M. (2000). Origination and extinction components of taxonomic diversity: general problems. *Paleobiology*, 26(sp4), 74–102.
- Freese, H. M., Dalingault, H., Petersen, J., Pradella, S., Davenport, K., Teshima, H., . . . Goodwin, L. A. (2013). Genome sequence of the phage-gene rich marine *Phaeobacter arcticus* type strain DSM 23566 T. *Standards in Genomic Sciences*, 8(3).
- Gowing, M. M., Garrison, D. L., Gibson, A. H., Krupp, J. M., Jeffries, M. O., & Fritsen, C. H. (2004). Bacterial and viral abundance in Ross Sea summer pack ice communities. *Mar. Ecol. Prog. Ser.*, 279, 3–12.
- Guixa-Boixereu, N., Vaqué, D., Gasol, J. M., Sánchez-Cámara, J., & Pedrós-Alió, C. (2002). Viral distribution and activity in Antarctic waters. *Deep-Sea Res. Pt. II*, 49(4), 827–845.
- Hjerde, E., Lorentzen, M., Holden, M., Seeger, K., Paulsen, S., Bason, N., . . . Thomson, N. (2008). The genome sequence of the fish pathogen *Aliivibrio salmonicida* strain LFI1238 shows extensive evidence of gene decay. *BMC Genomics*, 9(1), 616.
- Kiko, R. (2010). Acquisition of freeze protection in a sea-ice crustacean through horizontal gene transfer? *Pol. Biol.*, 33(4), 543-556.
- Lawrence, J. G., & Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.*, 44(4), 383–397.

- Laybourn-Parry, J., Marshall, W. A., & Madan, N. J. (2007). Viral dynamics and patterns of lysogeny in saline Antarctic lakes. *Pol. Biol.*, 30(3), 351–358.
- Lisle, J., & Priscu, J. (2004). The occurrence of lysogenic bacteria and microbial aggregates in the lakes of the McMurdo Dry Valleys, Antarctica. *Microb. Ecol.*, 47(4), 427–439.
- Ma, Y., Wang, L., & Shao, Z. (2006). Pseudomonas, the dominant polycyclic aromatic hydrocarbon-degrading bacteria isolated from Antarctic soils and the role of large plasmids in horizontal gene transfer. *Environ. Microbiol.*, 8(3), 455–465.
- Méthe, B. A., Nelson, K. E., Deming, J. W., Momen, B., Melamud, E., Zhang, X., . . . Fraser, C. M. (2005). The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *P. Natl. Acad. Sci.*, 102(31), 10913–10918. doi: 10.1073/pnas.0504766102
- Nowell, R. W., Green, S., Laue, B. E., & Sharp, P. M. (2014). The extent of genome flux and its role in the differentiation of bacterial lineages. *Gen. biol. evol., Online in advance of print*. doi: 10.1093/gbe/evu123
- Paul, J. H. (2008). Prophages in marine bacteria: dangerous molecular time bombs or the key to survival in the seas? *ISME J.*, 2(6), 579–589.
- Pierrehumbert, R., Abbot, D., Voigt, A., & Koll, D. (2011). Climate of the Neoproterozoic. *Annu. Rev. Earth Pl. Sc.*, 39, 417–460.
- Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R. D. (2012). The Pfam protein families database. *Nuc. Acids Res.*, 40(D1), D290–D301. doi: 10.1093/nar/gkr1065
- Qi, J., Wang, B., & Hao, B.-I. (2004). Whole proteome prokaryote phylogeny without sequence alignment: a K-string composition approach. *J. Mol. Evol.*, 58(1), 1–11.
- Raymond, J., & Kim, H. (2012). Possible role of horizontal gene transfer in the colonization of sea ice algae. *PLOS one*, 7(5), e35968.
- Royer, D. L. (2006). CO<sub>2</sub>-forced climate thresholds during the Phanerozoic. *Geochimica et Cosmochimica Acta*, 70(23), 5665–5675.
- Säwström, C., Anesio, M. A., Granéli, W., & Laybourn-Parry, J. (2007). Seasonal viral loop dynamics in two large ultraoligotrophic Antarctic freshwater lakes. *Microb. Ecol.*, 53(1), 1–11.

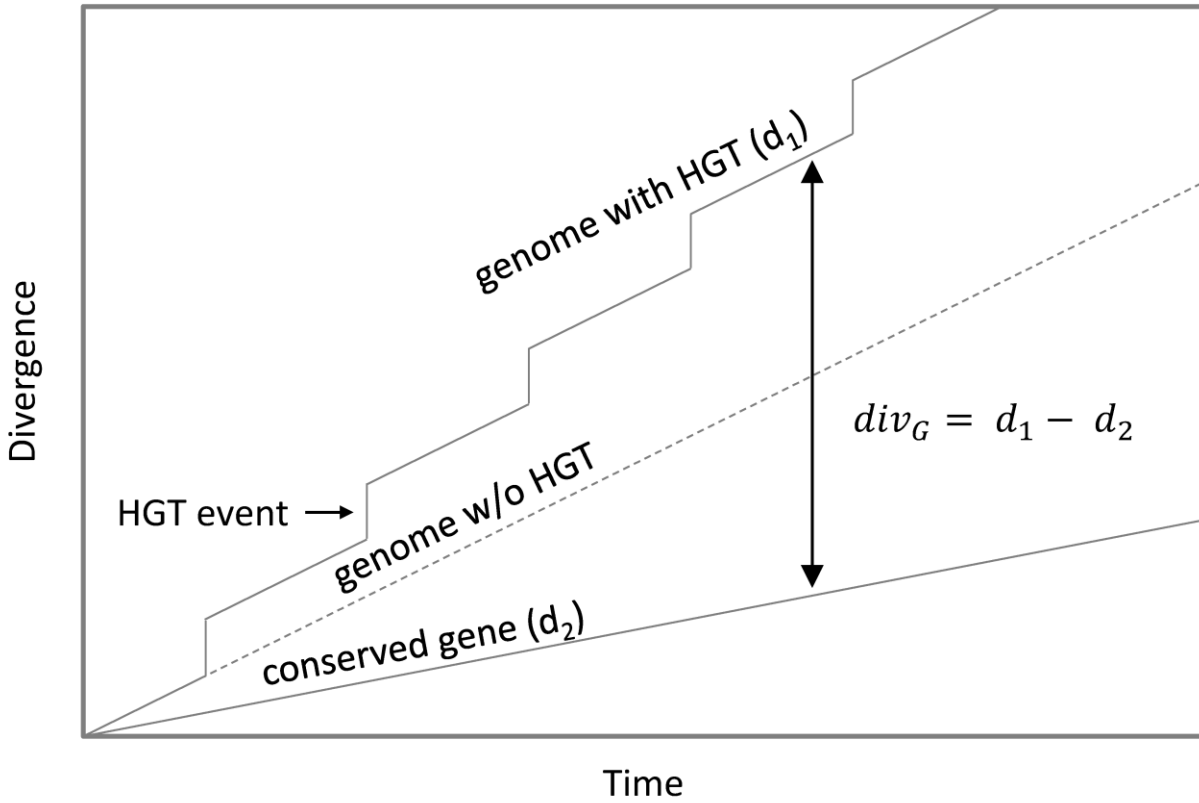
- Säwström, C., Granéli, W., Laybourn-Parry, J., & Anesio, A. M. (2007). High viral infection rates in Antarctic and Arctic bacterioplankton. *Environ. Microbiol.*, 9(1), 250–255.
- Säwström, C., Lisle, J., Anesio, A. M., Priscu, J. C., & Laybourn-Parry, J. (2008). Bacteriophage in polar inland waters. *Extremophiles*, 12(2), 167–175.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, 75(23), 7537–7541. doi: 10.1128/aem.01541-09
- Sepkoski, J. J., Jablonski, D., & Foote, M. (2002). A compendium of fossil marine animal genera. *Bull. Am. Paleontol.*, 363(1).
- Shaviv, N. J., & Veizer, J. (2003). Celestial driver of Phanerozoic climate? *GSA today*, 13(7), 4–10.
- Siddiqui, K. S., Williams, T. J., Wilkins, D., Yau, S., Allen, M. a., Brown, M. V., . . . Cavicchioli, R. (2013). Psychrophiles. *Ann. Rev. Earth Pl. Sc.*, 41(1), 87–115.
- Thode, H. C. (2002). *Testing for normality* (Vol. 164): CRC Press.
- Veizer, J., Godderis, Y., & François, L. M. (2000). Evidence for decoupling of atmospheric CO<sub>2</sub> and global climate during the Phanerozoic eon. *Nature*, 408(6813), 698–701.
- Vollmers, J., Voget, S., Dietrich, S., Gollnow, K., Smits, M., Meyer, K., . . . Daniel, R. (2013). Poles apart: arctic and antarctic *Octadecabacter* strains share high genome plasticity and a new type of xanthorhodopsin. *PLoS ONE*, 8(5), e63422.
- Wang, F., Wang, J., Jian, H., Zhang, B., Li, S., Wang, F., . . . Yu, J. (2008). Environmental adaptation: genomic analysis of the piezotolerant and psychrotolerant deep-sea iron reducing bacterium *Shewanella piezotolerans* WP3. *PLoS ONE*, 3(4), e1937.
- Wells, L. E., & Deming, J. W. (2006a). Characterization of a cold-active bacteriophage on two psychrophilic marine hosts. *Aquat. Microb. Ecol.*(45), 15–29.
- Wells, L. E., & Deming, J. W. (2006b). Modelled and measured dynamics of viruses in Arctic winter sea-ice brines. *Environ. Microbiol.*, 8(6), 1115–1121.
- Wells, L. E., & Deming, J. W. (2006c). Significance of bacterivory and viral lysis in bottom waters of Franklin Bay, Canadian Arctic, during winter. *Aquat. Microb. Ecol.*, 43(3), 209–221. doi: 10.3354/ame043209

- Whiteside, J. H., Olsen, P. E., Eglinton, T., Brookfield, M. E., & Sambrotto, R. N. (2010). Compound-specific carbon isotopes from Earth's largest flood basalt eruptions directly linked to the end-Triassic mass extinction. *Proc. Natl. Acad. Sci.*, 107(15), 6721–6725.
- Zhao, T. (2011). *Genome Sequencing and Analysis of the Psychrophilic Anoxygenic Phototrophic Bacterium Rhodoferax antarcticus sp. ANT. BR.* Arizona State University.

## Tables and Figures

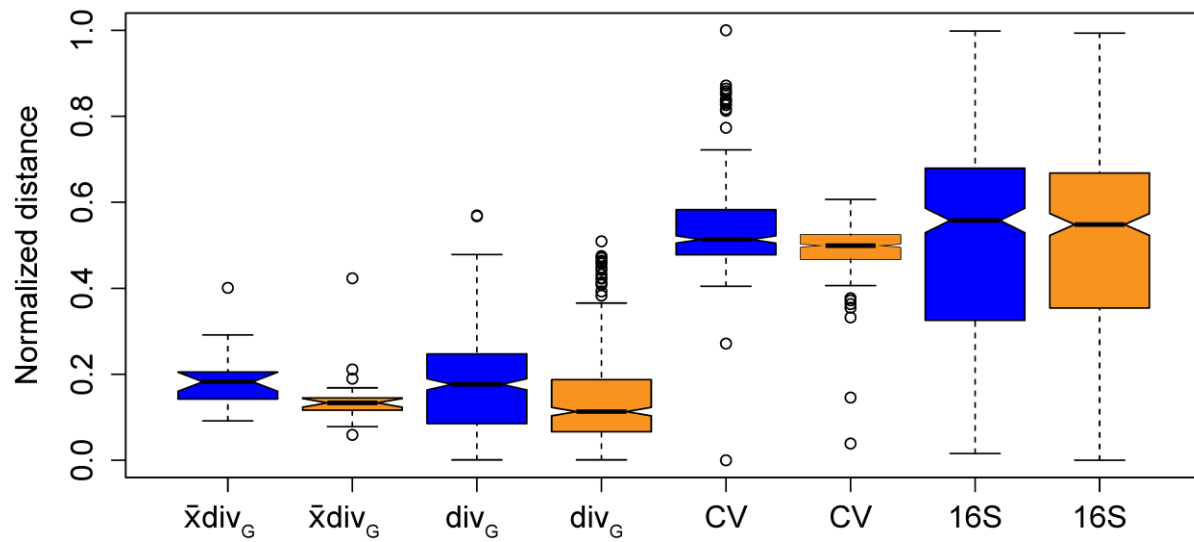
**Table 5.1.** Strains used in this study and key values. Each row represents a taxonomically similar psychrophile and mesophile pair. Bold (positive) values in the  $\Delta \bar{X}div_G$  column indicate pairs where the  $\bar{X}div_G$  value is larger for the psychrophilic strain.

Psychrophile strain	$\bar{X}div_G$	Mesophile strain	$\bar{X}div_G$	$\Delta \bar{X}div_G$
<i>Aeromonas salmonicida</i> A449	0.18	<i>Aeromonas veronii</i> B565	0.13	<b>0.05</b>
<i>Aliivibrio salmonicida</i> LFI1238	0.28	<i>Vibrio fischeri</i> ES114	0.12	<b>0.16</b>
<i>Colwellia psychrerythraea</i> 34H	0.16	<i>Alteromonas macleodii</i> Deep ecotype	0.16	0.00
<i>Desulfotalea psychrophila</i> LSv54	0.11	<i>Desulfocapsa sulfexigens</i> DSM 10523	0.09	<b>0.02</b>
<i>Flavobacterium psychrophilum</i> JIP02 86	0.19	<i>Flavobacterium indicum</i> GPTSA100 9	0.20	-0.01
<i>Glaciecola psychrophila</i> 170	0.13	<i>Glaciecola agarilytica</i> 4H 3 7 YE 5	0.12	<b>0.01</b>
<i>Methanococcoides burtonii</i> DSM 6242	0.38	<i>Methanosarcina mazei</i> Tuc01	0.40	-0.02
<i>Octadecabacter antarcticus</i> 307	0.09	<i>Rhodobacter sphaeroides</i> ATCC 17025	0.06	<b>0.03</b>
<i>Octadecabacter arcticus</i> 238	0.25	<i>Ketogulonicigenium vulgare</i> Y25	0.07	<b>0.18</b>
<i>Photobacterium profundum</i> SS9	0.20	<i>Vibrio vulnificus</i> YJ016	0.13	<b>0.07</b>
<i>Pseudoalteromonas haloplanktis</i> TAC125	0.17	<i>Pseudoalteromonas atlantica</i> T6c	0.13	<b>0.04</b>
<i>Psychrobacter arcticum</i> 273 4	0.11	<i>Acinetobacter baumannii</i> ACICU	0.12	-0.01
<i>Psychrobacter cryohalolentis</i> K5	0.11	<i>Acinetobacter baumannii</i> AYE	0.11	0.00
<i>Psychroflexus torquis</i> ATCC 700755	0.17	<i>Flavobacteriales bacterium</i> HTCC2170	0.18	-0.01
<i>Psychromonas</i> CNPT3	0.18	<i>Marinobacter aquaeolei</i> VT8	0.12	<b>0.06</b>
<i>Psychromonas ingrahamii</i> 37	0.17	<i>Alteromonas macleodii</i> English Channel 673	0.13	<b>0.04</b>
<i>Shewanella halifaxensis</i> HAW EB4	0.18	<i>Shewanella</i> MR 7	0.13	<b>0.05</b>
<i>Shewanella sediminis</i> HAW EB3	0.17	<i>Shewanella denitrificans</i> OS217	0.14	<b>0.03</b>
<i>Shewanella violacea</i> DSS12	0.21	<i>Shewanella putrefaciens</i> 200	0.14	<b>0.07</b>
<i>Terriglobus saanensis</i> SP1PR4	0.14	<i>Terriglobus roseus</i> DSM 18391	0.10	<b>0.04</b>

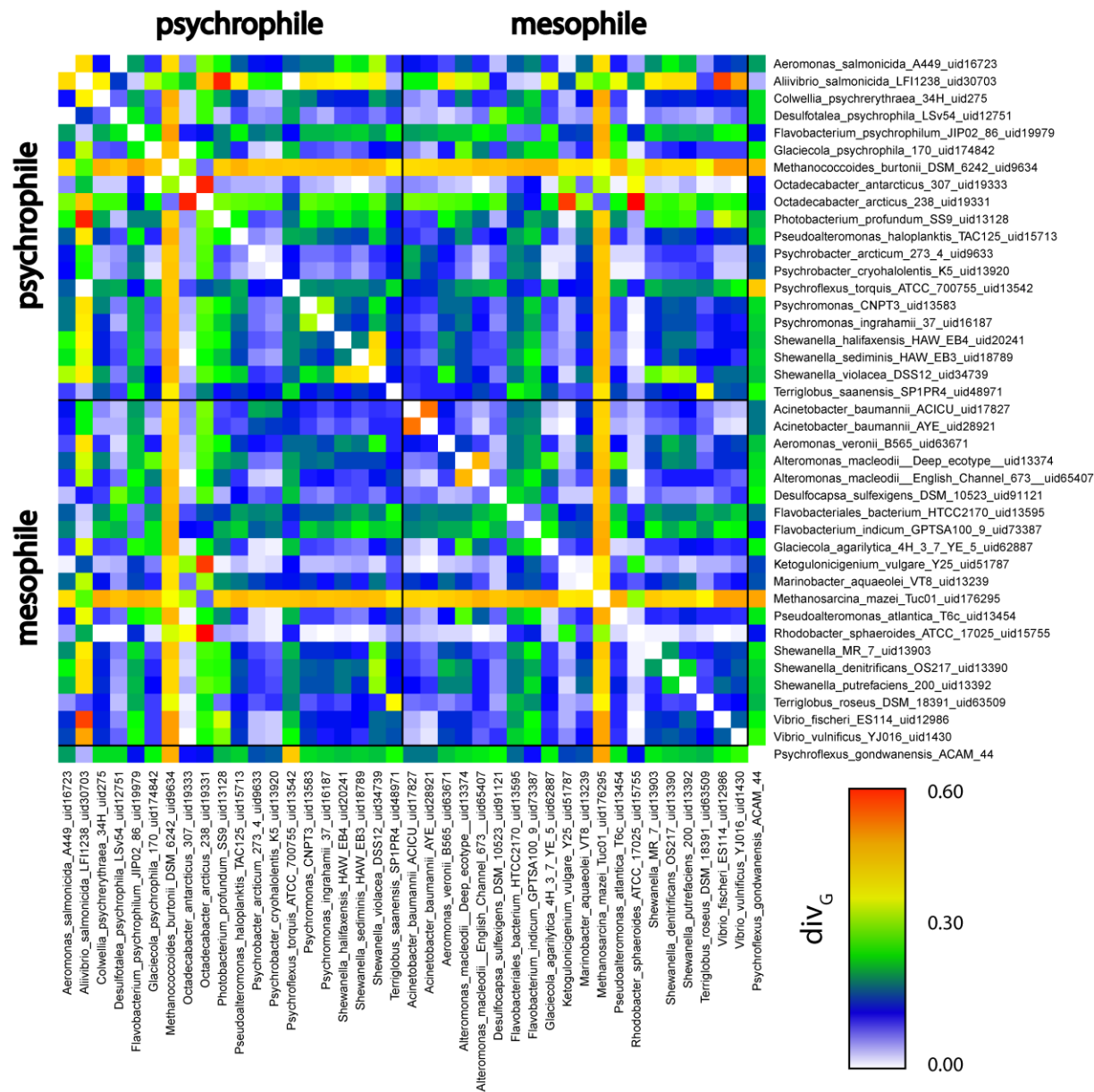


**Fig. 5.1. Proteome vs. 16S rRNA gene divergence.** Hypothetical relationship between proteome divergence with HGT, proteome divergence without HGT, and conserved gene (16S rRNA) divergence with respect to time. Hypothetical slopes are for illustrative purposes only.

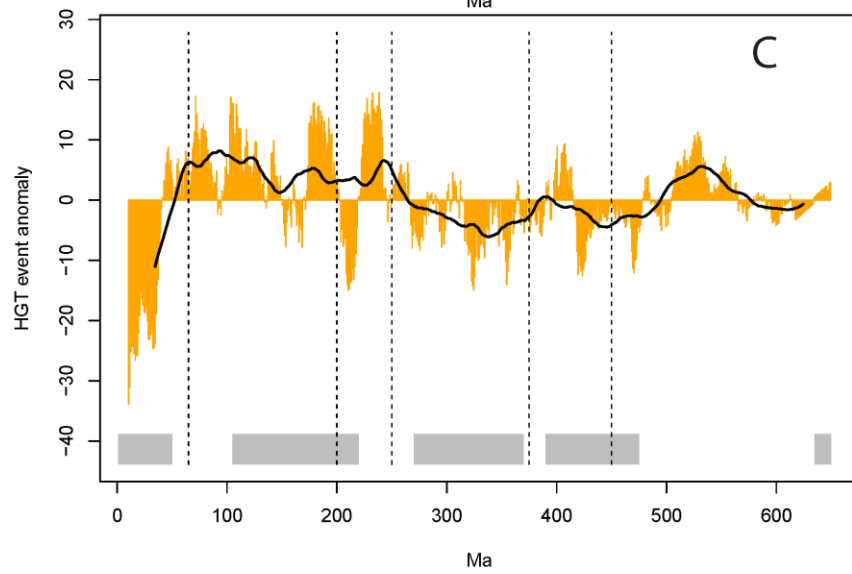
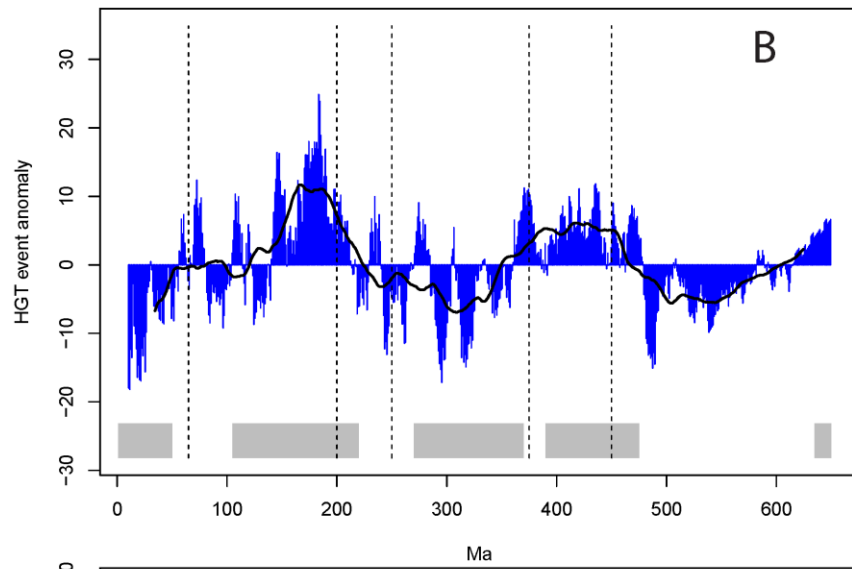
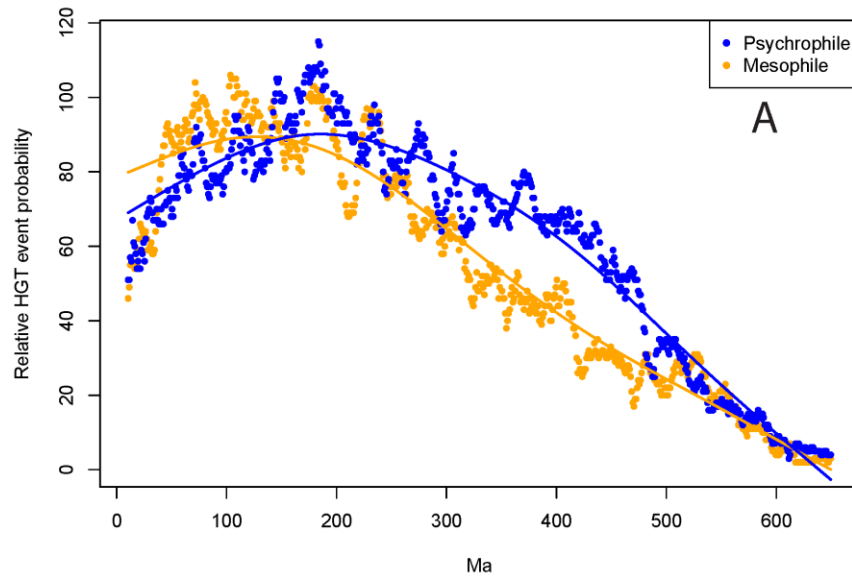




**Fig. 5.2. Measures of divergence between groups.** Values are shown for  $\bar{x}div_G$ ,  $div_G$ , normalized compositional vector distance (CV) and normalized 16S rRNA gene distance (16S). Where box notches of the two groups do not overlap, support for a difference in the median value is strong. Psychrophiles are shown in blue; mesophiles, in orange.



**Fig. 5.3.** Div<sub>G</sub> values between strains in the psychrophile and mesophile groups. Also included is the genome of the psychrotolerant strain *Psychroflexus gondwanensis* ACAM44.



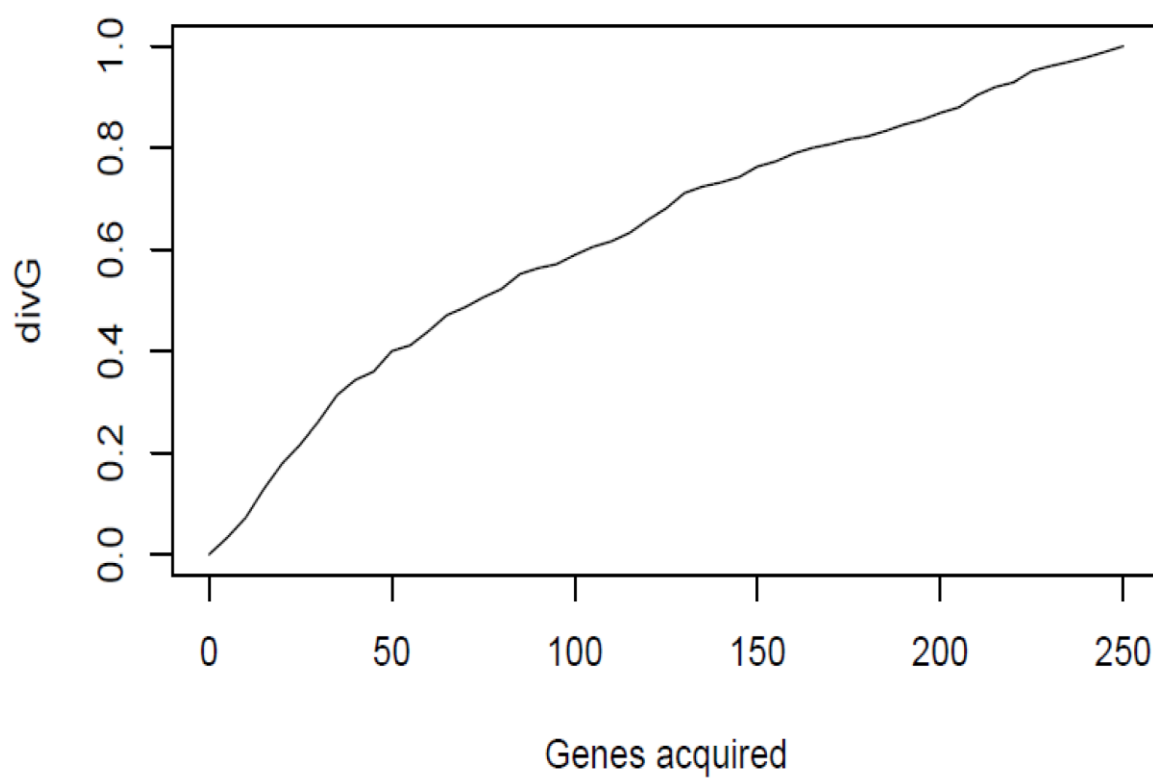
**Fig. 5.4. Frequency of HGT events across geologic time for the psychrophile and mesophile groups.** A) The timing (in Ma) of HGT events detected within the psychrophile genomes and the spline fit to the data shown in blue; same for mesophiles shown in orange. B) Detrended HGT frequency for the psychrophiles (blue) with a 10 My moving average (black line). C) Detrended HGT frequency for the mesophiles (orange) with a 10 My moving average (black line). Time periods with a negative global mean temperature anomaly, according to the 10/50 resolution of Veizer et al. (Veizer et al., 2000) relative to the present, are indicated by grey bands, along with the last major Neoproterozoic glaciation (at ~ 650 Ma). Dotted vertical lines indicate the timing of the five major mass extinctions.

**Table S5.1.** Isolation environment,  $\bar{x}div_G$ , and the number of anomalous CDS for each evaluated strain.

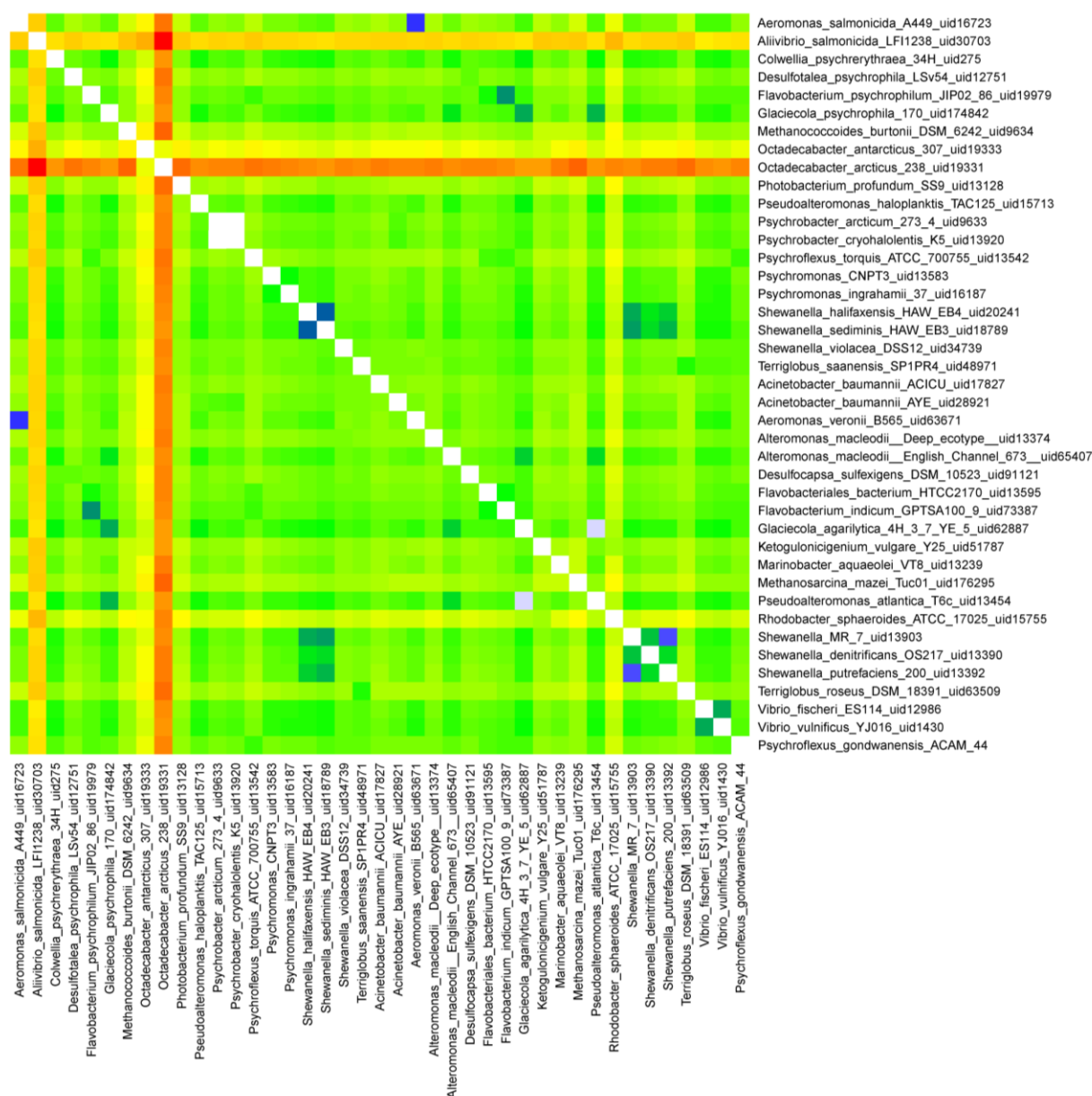
Psychrophilic Strain	Isolation		Anomalous		Mesophile Strain	Isolation		Anomalous		$\Delta \bar{x}div_G$
	Environment	$\bar{x}div_G$	CDS			Environment	$\bar{x}div_G$	CDS		
<i>Aeromonas salmonicida</i> A449	pathogenic	3.65	187		<i>Aeromonas veronii</i> B565	pond sediment	2.67	160		<b>0.98</b>
<i>Allivibrio salmonicida</i> LF11238	pathogenic	5.54	171		<i>Vibrio fischeri</i> ES114	squid symbiont	2.49	160		<b>3.05</b>
<i>Colwellia psychrerythraea</i> 34H	sediment	3.17	187		<i>Alteromonas macleodii</i> Deep ecotype	deep sea	3.20	161		-0.03
<i>Desulfotalea psychrophila</i> L5v54	permafrost	2.12	146		<i>Desulfocapsa sulfexigens</i> DSM 10523	tidal flat	1.79	158		<b>0.33</b>
<i>Flavobacterium psychrophilum</i> JIP02.86	pathogenic	3.77	88		<i>Flavobacterium indicum</i> GPTSA100.9	spring water	4.02	98		-0.24
<i>Glaciecola psychrophila</i> 170	sea ice	2.62	203		<i>Glaciecola agarilytica</i> 4H 3.7 YE 5	sediment	2.49	192		<b>0.13</b>
<i>Methanococcoides burtonii</i> DSM 6242	saline lake	7.62	106		<i>Methanosarcina mazei</i> Tuc01	sediment	8.05	126		-0.42
<i>Octadecabacter antarcticus</i> 307	sea ice	1.75	204		<i>Rhodobacter sphaeroides</i> ATCC 17025	unknown	1.12	172		<b>0.62</b>
<i>Octadecabacter arcticus</i> 238	sea ice	5.09	267		<i>Ketogulonicigenium vulgare</i> Y25	unknown	1.49	147		<b>3.6</b>
<i>Photobacterium profundum</i> SS9	deep sea	4.03	263		<i>Vibrio vulnificus</i> YJ016	pathogenic	2.58	195		<b>1.45</b>
<i>Pseudoalteromonas haloplanktis</i> TAC125	seawater	3.48	143		<i>Pseudoalteromonas atlantica</i> T6c	marine algae	2.51	191		<b>0.97</b>
<i>Psychrobacter arcticum</i> 273.4	permafrost	2.20	101		<i>Acinetobacter baumannii</i> ACICU	pathogenic	2.31	132		-0.11
<i>Psychrobacter cryohalolentis</i> K5	permafrost	2.10	113		<i>Acinetobacter baumannii</i> AYE	pathogenic	2.12	125		-0.03
<i>Psychroflexus torquis</i> ATCC 700755	sea ice	3.34	156		<i>Flavobacteriales bacterium</i> HTCC2170	seawater	3.61	138		-0.27
<i>Psychromonas CNPT3</i>	sea ice	3.61	114		<i>Marinobacter aquaeolei</i> VT8	oil well	2.43	202		<b>1.18</b>
<i>Psychromonas ingrahamii</i> 37	deep sea	3.43	173		<i>Alteromonas macleodii</i> English Channel 673	seawater	2.63	170		<b>0.81</b>
<i>Shewanella halifaxensis</i> HAW EB4	seawater	3.60	190		<i>Shewanella</i> MR 7	seawater	2.57	170		<b>1.03</b>
<i>Shewanella sediminis</i> HAW EB3	sediment	3.46	202		<i>Shewanella denitrificans</i> OS217	seawater	2.73	161		<b>0.73</b>
<i>Shewanella violacea</i> DSS12	deep sea	4.21	155		<i>Shewanella putrefaciens</i> 200	crude oil	2.79	163		<b>1.42</b>
<i>Terrioglobus saanensis</i> SP1PR4	tundra soil	2.79	139		<i>Terrioglobus roseus</i> DSM 18391	soil	2.05	159		<b>0.74</b>

**Table S5.2.** Occurrence of Pfams as likely HGT events in the psychrophile and mesophile groups.

<b>Pfam</b>	<b>Difference between populations (psychrophile-mesophile)</b>	<b>Number of psychrophile strains in which pfam occurs as likely HGT event</b>
Y1_Tnp	39	8
DDE_Tnp_ISL3	19	3
DUF1602	18	20
HTH_Tnp_ISL3	14	1
CSD	-13	0
DDE_Tnp_IS66	-13	0
DUF4096	-12	0
DUF4323	12	2
Phage_int_SAM_4	-12	0
DUF2971	11	10
Phage_integrase	-11	0
DUF3265	-10	0
GGDEF	10	9
HTH_31	-9	0
NTP_transf_3	-9	0
PP-binding	9	12
TetR_N	-9	0
BLUF	8	3
HTH_1	8	10
Mrr_cat	8	6
Acyl_transf_3	-7	0
Beta-lactamase	-7	0
CarboxypepD_reg	-7	0
Glyco_trans_1_4	-7	0
GntR	7	0
Methylase_S	-7	0
Polysacc_synt_3	-7	0
PP-binding_2	-7	0
Response_reg	7	15
VPEP	-7	0
Wzy_C	-7	0
Acetyltransf_3	-6	0
Aldedh	6	5
AMP-binding	6	5
CcdA	-6	0
DUF3738	-6	0
HNH	6	6
MerR	-6	0
NA37	-6	0
Thiolase_C	6	4
TPR_11	6	6
Usp	-6	0
YbjQ_1	6	3
Acetyltransf_4	-5	0
Biotin_lipoyl	-5	0
Cation_efflux	-5	0
DUF3296	-5	0
DUF3732	5	4
DUF465	5	8
Glyco_trans_4_4	5	6

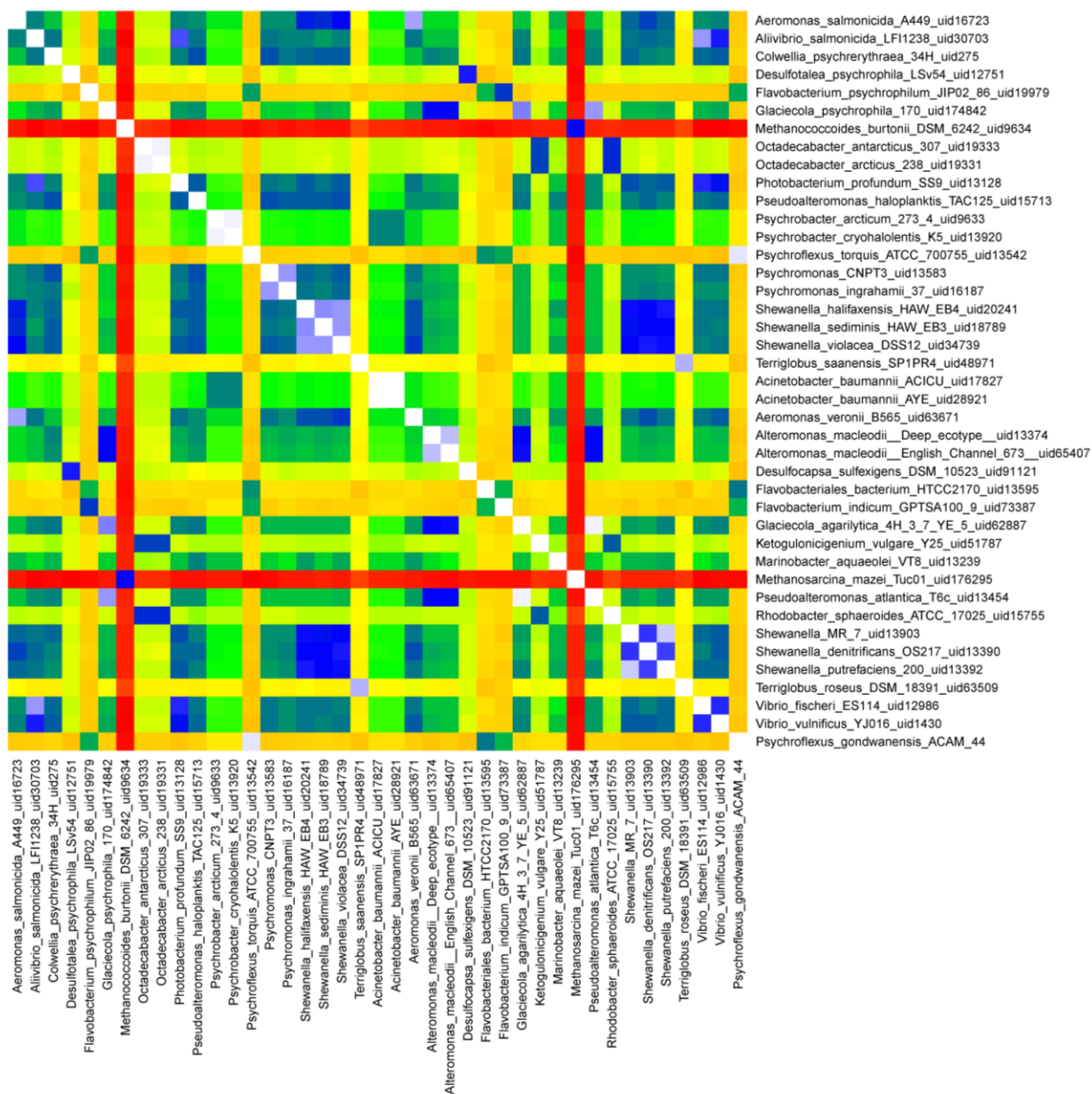


**Fig. S5.1. Sensitivity of  $\text{div}_G$  to gene acquisition.** For hypothetical acquisitions of genes by the genome of *Colwellia psychrerythraea* 34H. See text for description of sensitivity analysis.

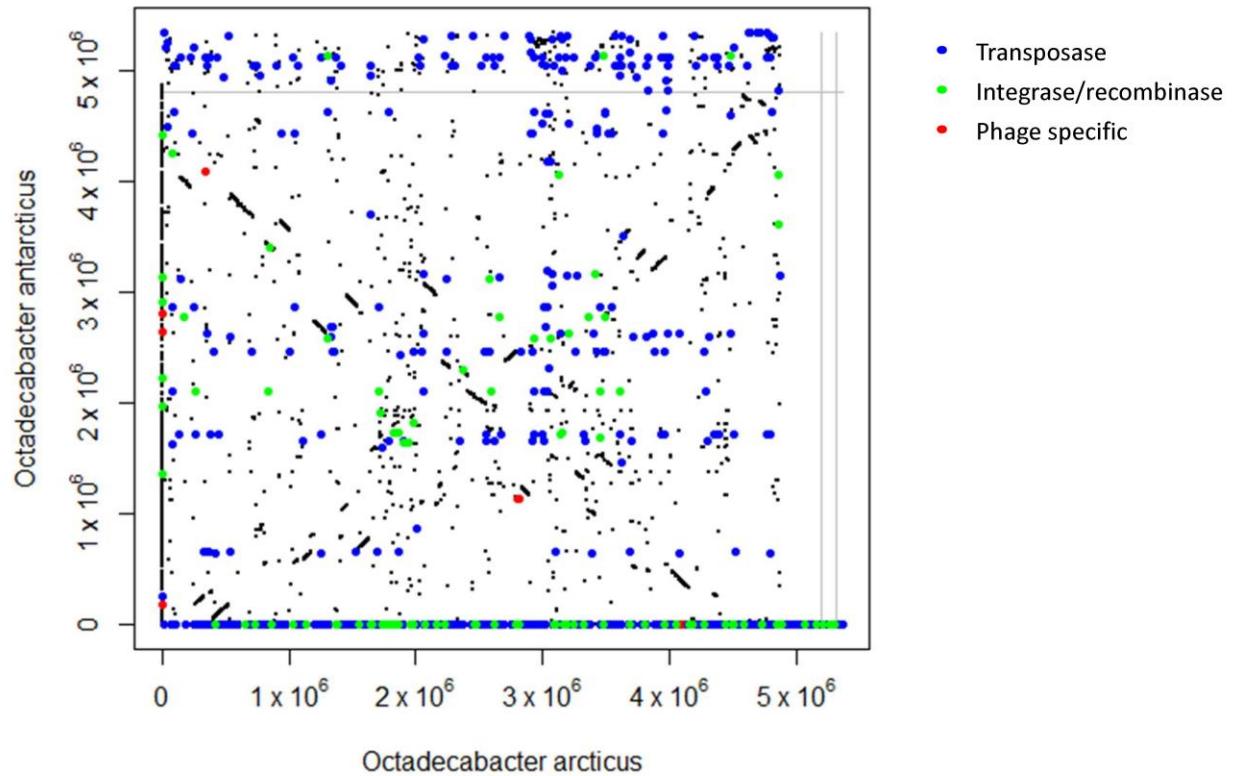


**Fig. S5.2. Normalized compositional vector distance between strains in the psychrophile and mesophile groups.** Also included is the genome of the psychrotolerant strain *Psychroflexus gondwanensis* ACAM44.



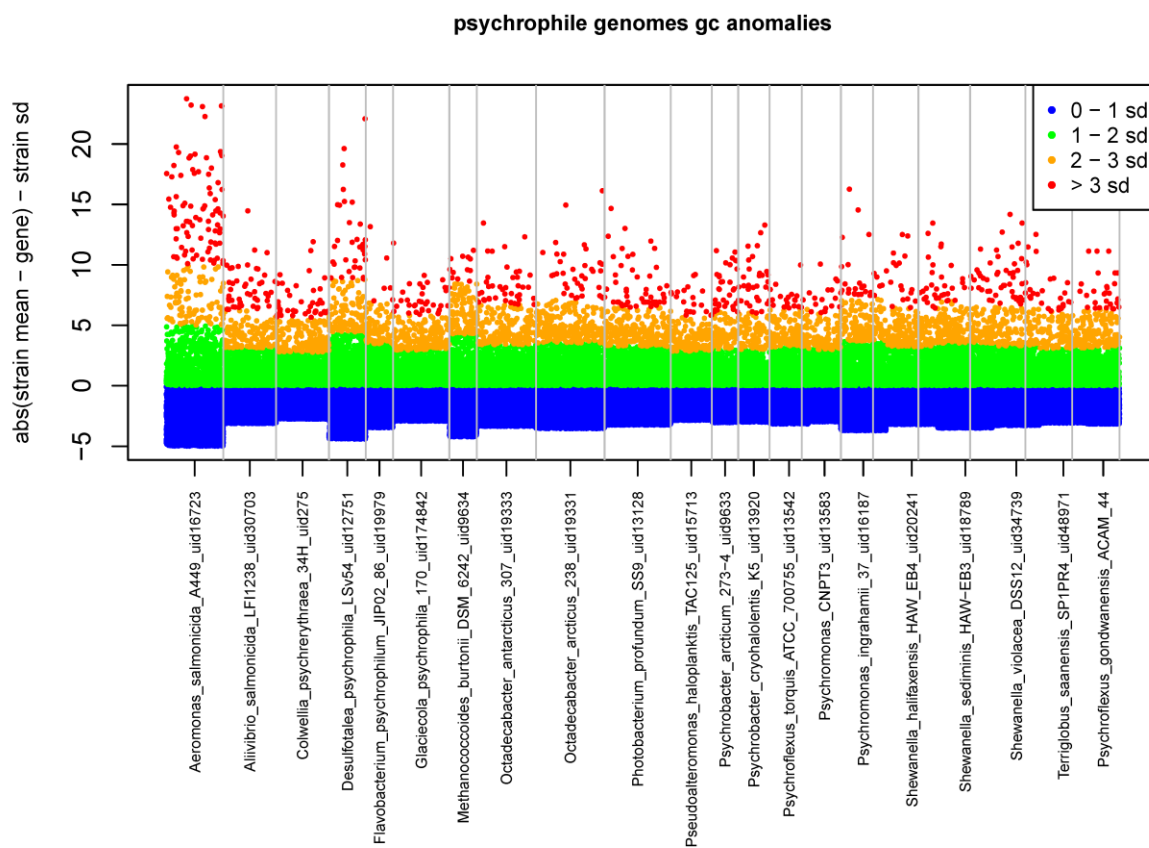


**Fig. S5.3. 16S rRNA gene distance between strains in the psychrophile and mesophile groups.** Also included is the genome of the psychrotolerant strain *Psychroflexus gondwanensis* ACAM44.

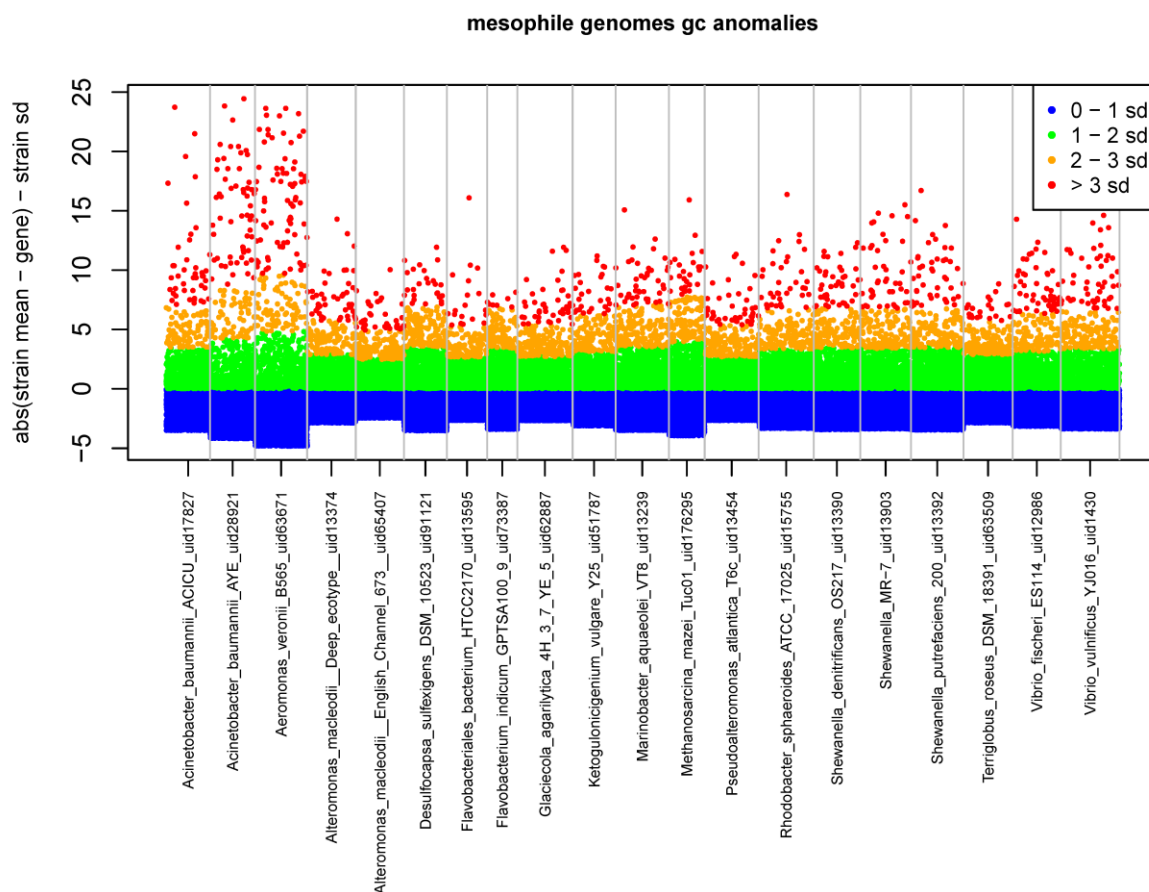


**Fig. S5.4. Synteny plot of *Octadecabacter antarcticus* 307 and *Octadecabacter arcticus* 238.**

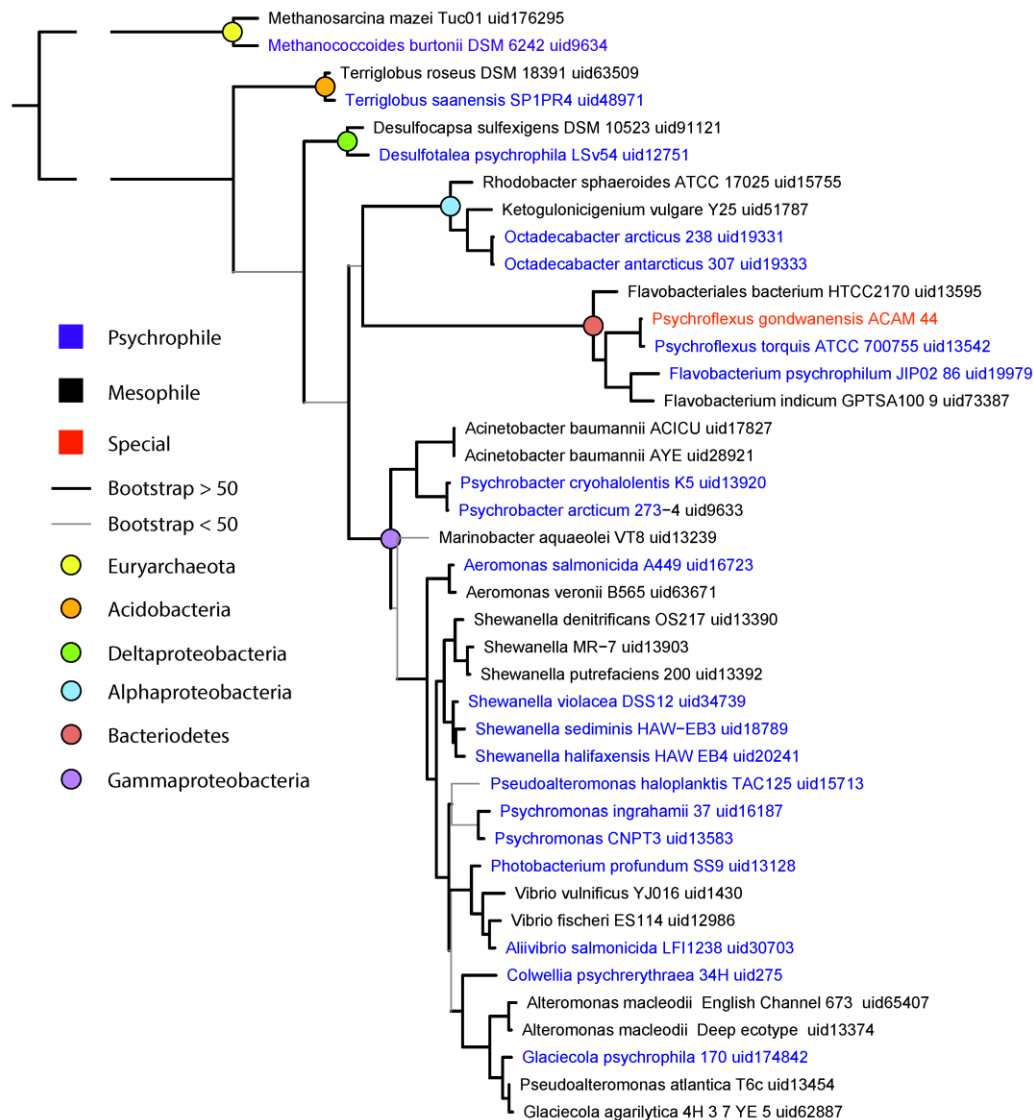
The genomes were analyzed as chromosomes plus plasmids; boundaries between these different genetic elements are shown by gray lines. Coding sequences located at  $x = 0$  do not appear in the genome of *O. arcticus* 238, while those located at  $y = 0$  do not appear in *O. antarcticus* 307. Blue points indicate transposases, red points indicate phage specific genes, green points indicate integrases and recombinases.



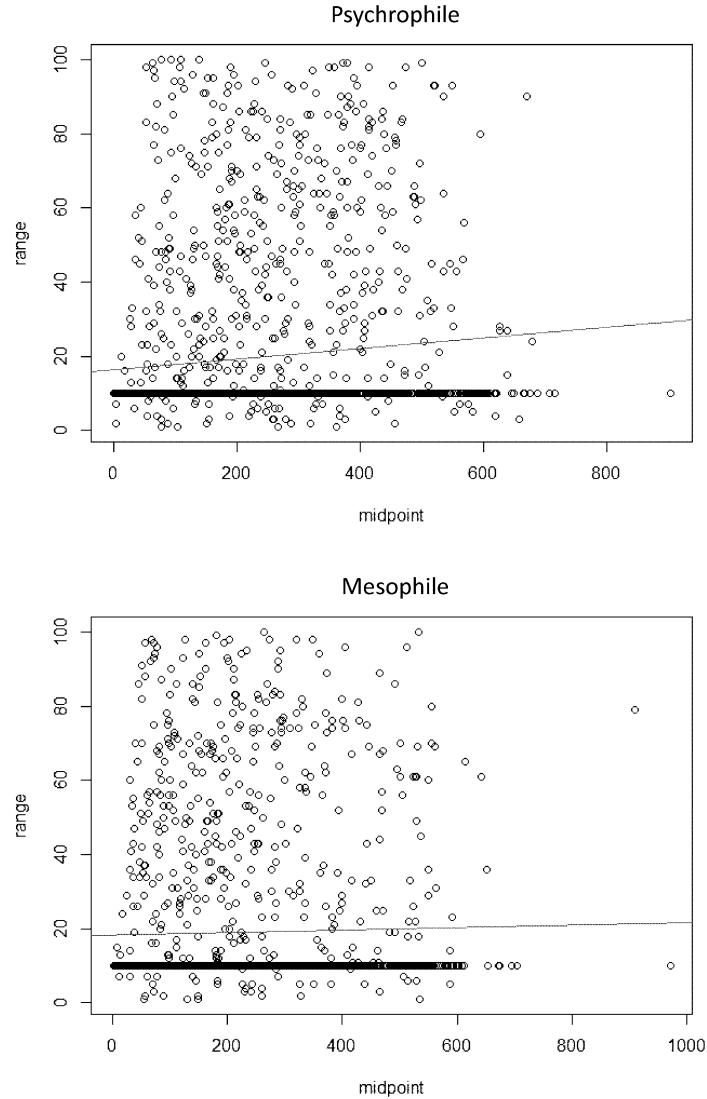
**Fig. S5.5. GC anomalies of coding sequences in the psychrophile genomes.** Vertical gray lines indicate the boundaries between the analyzed genomes.



**Fig. S5.6. GC anomalies of coding sequences in the mesophile genomes.** Vertical gray lines indicate the boundaries between the analyzed genomes.



**Fig. S5.7. Phylogenetic tree of strains used in this study.** The tree was constructed with RAxML v7.7.9 from the 16S rRNA alignment used in the distance matrix calculation. RAxML was called as `raxmlHPC-PTHREADS -f a -T 8 -p 12345 -x 12345 -# 100 -m GTRGAMMA -n [tree name] -s [alignment name]`. Note the misclassification of *Pseudoalteromonas atlantica* Tc6, which falls within the *Glaciecola* on our tree. Although *P. atlantica* Tc6 was added as the control for the psychrophile *Pseudoalteromonas haloplanktis* TAC125, taxonomic balance is retained at the order level.



**Fig. S5.8. Change of HGT time since acquisition probability window size with time.**

Regression lines are shown in red. No correlation was observed between window size and date for the mesophile genomes ( $R^2 = 4.52 \times 10^{-5}$ ,  $p = 0.30$ ). For the psychrophile genomes, a weak correlation was observed ( $R^2 = 0.0097$ ,  $p = 8.83 \times 10^{-6}$ ) with a slope of 0.014. This weak correlation should have resulted in a slight bias in anomaly size in later time periods for the psychrophiles. Because of the preponderance of cold periods in the more recent half of the analyzed time period, however, this bias would work against the reported correlation of HGT anomalies and cold periods.

## Chapter 6

### **Amino acid preferences in the proteomes of psychrophilic bacteria:**

#### **Is serine the answer to enzyme adaptation to low temperature and high salinity?**

#### ABSTRACT

Psychrophilic microorganisms, adapted to life at low temperature ( $< 20^{\circ}\text{C}$ ), are a diverse and abundant group encompassing members from all three domains of life. The enzymes that enable life at low temperature exhibit specific adaptations, often identifiable in comparative analyses as substitutions of specific amino acid residues. Because low water activity exerts an additional strong selective pressure in many cold environments, we considered that co-optimization of proteins for low water activity and low temperature might select for different amino acid substitutions than low temperature alone. We evaluated protein parameters and amino acid preferences in the proteomes of 20 psychrophilic bacteria relative to those of taxonomically similar mesophiles, finding hydropathy, aromaticity, and aliphatic index to be elevated in the psychrophilic bacteria relative to mesophiles. Increased flexibility, widely believed to be a common adaptation to low temperature, was observed only in the helix region of psychrophilic proteins. Elevated levels of isoleucine, leucine, methionine, asparagine, serine, and tyrosine were observed across all secondary structure elements within this set of psychrophile genomes. To further explore how an enzyme might evolve under selective pressure from multiple environmental parameters we developed and applied a simple model, the Protein Evolution Parameter Calculator (PEPC). Using PEPC, mutations of arginine to lysine led to small improvements in flexibility and a slightly reduced isoelectric point, a common feature of halophilic proteins. Mutations to serine, far more beneficial to optimizing a protein for both

flexibility and isoelectric point, were less common except when the protein was optimized for flexibility, isoelectric point, and hydropathy.

Expanding the genome basis for examining psychrophilic parameters in proteins can lead to a refined view of the structural flexibility understood to define cold adaptation in enzymes. Although some conventional traits were observed, indications of flexibility may be more limited than previously considered. The abundance of lysine relative to serine in proteins from psychrophilic bacteria suggests that the protein fitness landscape has only partially been explored in nature since the last inter-glacial period.



## 6.1 Introduction

A large fraction of the space available for microbial colonization consists of cold habitats including the deep sea, polar oceans, sea and glacial ice, permafrost, high altitudes, and seasonal snow. Despite the thermodynamic constraints that low temperatures impose on metabolism in these environments, single and multicellular organisms from all three domains of life thrive in them, with members of the Bacteria observed active to temperatures of  $-33\text{ }^{\circ}\text{C}$  (Bakermans & Skidmore, 2011) and capable of growth to  $-12\text{ }^{\circ}\text{C}$  (Breezee, Cady, & Staley, 2004; Wells & Deming, 2006). Unlike (hyper)thermophiles, adapted to life at high temperature, psychrophiles do not form distinct taxonomic groups according to 16S rRNA gene phylogeny, with many genera hosting both mesophilic and psychrophilic members (Fig. 6.1). A number of studies have identified alterations at the genome and gene level that enable life at low temperatures. At the genome level the presence of genes involved in compatible solute acquisition (R. E. Collins & Deming, 2013), energy conversion (Methe et al., 2005), increased membrane fluidity (Chattopadhyay & Jagannadham, 2001), and the production of cryoprotectant compounds (Marx, Carpenter, & Deming, 2009) reflect known adaptations to low temperature. At the gene level mutations that give rise to a number of enzyme modifications play an important role in low temperature adaptation. The number of known enzyme modifications is extensive but has rarely been studied systematically. As reviewed by D'Amico et al. (2006) known modifications are sometimes enzyme-specific but generally favor increased structural flexibility, as acquired through a lower content of proline and arginine, a reduction in hydrophobic interactions, an increase in solvent (water) interactions, and a reduced number of ion pairs.

Metpally and Reddy (2009) undertook an evaluation of the compositional differences between the sequenced proteomes of six psychrophilic bacteria and closely related mesophiles

that were available at the time (2009) to shed further light on parameters that characterize cold-activity. Consistent with earlier observations that temperature optimization and enzyme flexibility are closely linked (Bae & Phillips, 2004; Coquelle, Fioravanti, Weik, Vellieux, & Madern, 2007; Johns & Somero, 2004; Kim et al., 1999; Leiros et al., 2007; Methe et al., 2005; Russell, Gerike, Danson, Hough, & Taylor, 1998; Zecchinon et al., 2005), they reported an enrichment of residues in proteins from the psychrophiles that might result in enhanced flexibility. These enrichments included neutral residues and those classified as tiny or small (see Table 6.1 for classification). They also noted a reduction in aromatic, charged, basic, and hydrophilic residues. Comparing by protein secondary structure element, the authors reported additional amino acid preferences between the two temperature classes.

During an evaluation of cold active potential in putative alkane hydroxylase proteins coded in the genomes of twenty psychrophilic bacteria (Bowman & Deming, 2014) we were motivated to revisit the proteome analysis of Metpally and Reddy (Metpally & Reddy, 2009) and undertake an expanded version of it, adding newly available psychrophile genomes. In the process, we identified ways to make the statistical support for differences in protein composition between psychrophiles and mesophiles more rigorous. We then reanalyzed the same (or closely related) genomes used in Metpally and Reddy [9], applying an alternate statistical test that does not make the assumption of normality. We also evaluated differences in amino acid composition by temperature class rather than by strain, an approach that results in many more degrees of freedom and a normal distribution of amino acid composition in most cases. Finally we expanded the mesophile and psychrophile datasets to include twenty genomes each, with a similar taxonomic profile in the two temperature classes. Analysis of the expanded dataset represents the most comprehensive analysis of compositional difference between psychrophilic

and mesophilic proteins to date. The results of this comparative effort are expected to aid investigations into the differences in molecular dynamics between enzymes from the two temperature classes and the rational design of enzymes with enhanced function at low temperature.

Most previous investigations into enzyme optimization for low temperature have focused on temperature as the exclusive selective parameter, though in nature low temperature is often accompanied by additional co-selecting stressors. These include high pressure, as found throughout the deep ocean and subglacial environments, and high salinity, a ubiquitous feature of the liquid phase of ice formations. Because solutes are excluded from the ice matrix during freezing, they concentrate in interstitial brines, with the size of the interstitial space dependent on the quantity and specific properties of the solute. These brines are perceived to host life in glacial ice (Price, 2009) and are known to host a rich and active microbial community in sea ice at temperatures to  $-20^{\circ}\text{C}$  (Junge, Eicken, & Deming, 2004), equivalent to a brine salinity of 209 ppt (Cox & Weeks, 1983). To conceptualize the challenge of evolving an enzyme optimized for multiple environmental parameters we developed a model of protein evolution, the Protein Evolution Parameter Calculator (PEPC). We used PEPC to explore likely residue mutations leading to high flexibility, low isoelectric point, and high hydrophathy (becoming more hydrophobic) within the helix regions of a homologous pair of psychrophile and mesophile proteins.

## **6.2 Methods**

### **6.2.1 Genome selection and annotation**

Genomes corresponding to the 12 proteomes reported in Metpally and Reddy (2009) were downloaded from Genbank. Because five (of six) mesophilic strains used by Metpally and

Reddy could not be found in Genbank, alternate mesophilic strains of the same species were selected. Original and substituted strains, respectively, are: *Haemophilus influenzae* Hinf-M1 and *Haemophilus influenzae* Rd KW20; *Synechocystis* PCC680 (*Spec-M2*) and *Synechocystis* sp. PCC 6803; *Escherichia coli* K12 Ecol-M4 and *Escherichia coli* str. K-12 substr. MDS42; *Vibrio cholera* Vcho-M5 and *Vibrio cholerae* O1 biovar eltor str. N16961; and *Lactobacillus salivarius* UCCI 1 Lsal-M6 and *Lactobacillus salivarius* UCC118. The expanded dataset contained twenty psychrophilic strains (including the six used by Metpally and Reddy (Metpally & Reddy, 2009)) and twenty closely related mesophilic strains selected from Genbank (Fig. 6.1). Only true psychrophiles ( $T_{\max}$  for growth below 20 °C) were chosen according to temperature annotations in the GOLD (Bernal, Ear, & Kyrpidesa, 2001) and HIMA (E. Collins, 2013) databases and to reviews by Casanueva et al. (Casanueva, Tuffin, Cary, & Cowan, 2010) and Siddiqui et al. (Siddiqui et al., 2013). Care was taken to include all plasmids and chromosomes with the genome of each strain. To insure a consistent annotation across all analyzed genomes, a *de novo* annotation of the genomes was conducted, with open reading frames (ORFs) identified as genomic regions longer than 150 bp without a stop codon. ORFs were translated and searched for conserved protein domains against the Pfam-A database (Punta et al., 2012) using hmmscan in HMMER v3.0 (Eddy, 1998). Coding sequences (CDS) were defined as an open reading frame with at least one conserved domain with an E-value below  $10^{-5}$ . All downstream analysis proceeded on the translated coding sequences identified for each strain.

### 6.2.2 Secondary structure prediction

Secondary structure was predicted for all CDS products with PSIPRED v3 (McGuffin, Bryson, & Jones, 2000). The PSI-Blast portion of PSIPRED was run against a protein blast database generated from Pfam-A.fasta. In a limited comparison of PSIPRED predictions with

3D-protein structures predicted using Phyre2 (Kelley & Sternberg, 2009) we determined that this method was at least as accurate as PSIPRED v3 predictions based on PSI-Blast against the Uniref-90 protein blast database (data not shown).

### **6.2.3 Analysis of protein parameters**

The protein parameters of flexibility, grand average of hydropathy (GRAVY), isoelectric point, aliphatic index, aromaticity and amino acid composition were estimated for whole proteins and for secondary structure elements using the Bio.SeqUtils ProtParam module of Biopython (Cock et al., 2009). The parameters were first analyzed for the whole protein, then for a nine-residue sliding window starting at residue position 5. Window size was selected for consistency with the window used in flexibility calculations, based on the work of Vihinen, 1994 (Vihinen, Torkkila, & Riikonen, 1994). Window means were then binned to the secondary structure element (coil,  $\alpha$ -helix, or  $\beta$ -sheet) of their central residue. This method was selected to include neighbor effects on the parameters of a residue at a given position in the protein.

We evaluated differences in protein parameters between psychrophiles and mesophiles in two ways. First, we employed the by-strain approach of Metpally and Reddy (Metpally & Reddy, 2009), comparing the mean of each strain for each parameter between the two groups. Next, we disregarded strain means and compared the parameter means across each temperature class. Residue property groups were defined as in Metpally and Reddy (Metpally & Reddy, 2009) as indicated in Table 6.1.

Two statistical tests were applied to determine whether the parameters were significantly different between the two populations. The Student's t-test was used where the distribution of parameters for the two populations was normal by visual inspection of a histogram. If the distribution was not observed to be normal, then the Wilcoxon rank-sum test was used.

Significance for all tests was determined by comparing p-values to a Holm-Bonferroni corrected significance level (Holm, 1979).

#### 6.2.4 PEPC

To explore the co-evolution of physical parameters of psychrophilic and mesophilic proteins we developed a Python based tool, the Protein Evolution Parameter Calculator (PEPC). PEPC models the evolution of a protein in a guided fashion. First, secondary structure is determined using PSIPRED v3 (McGuffin et al., 2000). Second, rate variation across sites is estimated from blastp (Altschul et al., 1997) alignments to Pfam-A (Punta et al., 2012). The fraction of alignments with a mutation at a given position is determined and the fraction is used in the later mutation step. For every generation the protein is allowed a single mutation based on the probabilities determined in Eqt. 1, where  $P_{ij}$  = the probability of residue i mutation to j,  $p_i$  = the fraction of blast alignments with a mutation at the position corresponding to residue i,  $f_i$  = the natural frequency of i,  $f_j$  = the natural frequency of j, and  $B_{ij}$  = the Blosum80 (Henikoff & Henikoff, 1992) scoring matrix value for i mutating to j.

$$Eq. 1 \quad P_{ij} = p_i f_i f_j 10^{B_{ij}}$$

After each mutation the protein parameters are re-evaluated. Improvements in a parameter space selected by the user are perceived as increases in fitness and the new protein advances to the next generation. Worsening in the user-defined parameter space is perceived as a decrease in fitness and the new protein does not advance to the next generation (to introduce variability a fraction of proteins with reduced fitness may be allowed to survive). Protein parameters are determined as described previously. PEPC produces a table of protein parameters at each generation, a position-wise list of mutations, and the surviving sequences, and can be obtained from github at <https://github.com/jeffsbowman/PEPC>.

We used PEPC to evaluate the potential evolutionary trajectory of two aminopeptidases, one shown to be cold-active (Huston, Methe, & Deming, 2004) from the psychrophilic Gammaproteobacterium *Colwellia psychrerythraea* 34H (Genbank accession YP270144.1) and one from the mesophilic Gammaproteobacterium *Shewanella oneidensis* MR-1 (Genbank accession AAN55048.2). In the first experiment fitness was defined as increased flexibility in the helix region. In a second experiment fitness was defined as increased flexibility and decreased isoelectric point in the helix region. In the third experiment fitness was defined as increased flexibility, decreased isoelectric point, and increased hydrophathy in the helix region. In all three experiments PEPC was run for 30,000 generations with no survival by proteins exhibiting reduced fitness. In a fourth experiment PEPC was run as for the third experiment, but for  $10^6$  generations.

## **6.3 Results**

### **6.3.1 Metpally and Reddy dataset**

In revisiting the Metpally and Reddy [9] dataset, we analyzed 18,556 proteins in six psychrophile genomes and 14,420 proteins in six mesophile genomes. When amino acid abundances were tallied by strain, no significant difference was observed between temperature classes following the significance correction we applied, although some differences met the significance threshold before correction (Table S6.1). When proteins were grouped by temperature class of their source strain, 104 significant differences were observed, with the corrected significance threshold falling at  $p = 0.0001$  (Table S6.2). The most significant differences between the two temperature classes were in the increased abundance in the psychrophile proteins of tiny residues (psychrophile mean = 29.3 %, mesophile mean = 27.7 %), small residues (psychrophile mean = 49.6 %, mesophile mean = 47.8 %), and serine

(psychrophile mean = 6.8 %, mesophile mean = 6.1 %) (Table S6.2). Because our annotation method included pseudogenes not under selective pressure (and some mesophile genomes that were not used by Metpally and Reddy (Metpally & Reddy, 2009)), we used linear regression to compare the parameter means with the means from Metpally and Reddy (Metpally & Reddy, 2009). Despite the difference in annotation method and grouping (by temperature class and by strain) there was very little difference between means (Fig. 6.2).

### **6.3.2 Expanded dataset**

The expanded dataset consisted of 66,880 proteins from psychrophiles and 66,430 proteins from mesophiles. For the by-strain analysis of this dataset no parameter fell below  $p = 0.0003$ , the minimum threshold for significance. The parameter with the most significant difference was valine ( $p = 0.0095$ ,  $W = 105$ , psychrophile mean = 14.9 %, mesophile mean = 15.7 %) (Table S6.3). Grouped by temperature class, 100 parameters met the significance threshold (Tables 6.1, 6.2). The most significant p-values were found across the whole protein for the amino acids isoleucine (psychrophile mean = 6.8 %, mesophile mean = 6.3 %), proline (psychrophile mean = 4.0 %, mesophile mean = 4.3 %), and alanine (psychrophile mean = 8.6 %, mesophile mean = 9.2 %) (Table S6.2).

The large sample size for this analysis made it extremely sensitive to differences between the populations. While this sensitivity was helpful for identifying significant but small differences in parameters between the proteins of psychrophiles and mesophiles, some of the largest differences were masked by more significant but smaller differences. We ranked the parameters by the magnitude of the difference between temperature classes by dividing the difference in population means by the range across both populations. Scored in this way the most significant parameter, isoleucine for the whole protein, was also the parameter with the



largest difference between the two populations, followed by lysine for the whole protein (psychrophile mean = 6.0 %, mesophile mean = 5.6 %) and alanine for the whole protein (psychrophile mean = 8.6 %, mesophile mean = 9.1 %) (Table S6.4).

Our analysis necessarily aggregated trends across the diverse array of proteins coded by these genomes. Although this aggregation allowed us to generate robust statistical support, it blurs adaptations to low temperature that may be unique to only a subset of proteins. Despite this caveat we found that 792 Protein families out of 5,151 (comprising 50,854 proteins out of 135,671) fit the general pattern of increased hydrophathy, increased aromaticity, and increased aliphatic index across the whole protein.

To evaluate the specific challenges of adaptation to low temperature and low water activity we explored the relationship between flexibility, presumed to be an important adaptation to low temperature, and isoelectric point, presumed to be an important adaptation to low water activity (Kiraga et al., 2007). As expected the analyzed proteins exhibited a bimodal distribution for both parameters, resulting in four distinct clusters in a plot of flexibility vs. isoelectric point (Fig. 6.3). The local minimum for isoelectric point was determined to be 7.55 and the local minimum for flexibility to be 0.988. Interestingly, the relationship between flexibility and isoelectric point differed between clusters. For the region defined by high flexibility and low isoelectric point (Fig. 6.4), linear models fit to each temperature class indicated a strong negative correlation between the two parameters (psychrophile proteins:  $R^2 = 0.03284$ ,  $df = 38,976$ ,  $p \approx 0$ , mesophile proteins:  $R^2 = 0.05113$ ,  $df = 40,485$ ,  $p \approx 0$ ). Model slopes were determined to be significantly different with a z-test (psychrophile =  $-0.001044$ , mesophile =  $-0.001342$ ,  $z = -7.44$ ,  $p < 0.0001$ ) (Paternoster, Brame, Mazerolle, & Piquero, 1998). Slopes for the remaining

clusters were not significantly different between temperature classes and were positive for high isoelectric point and high flexibility and negative for both low flexibility clusters.

### 6.3.3 PEPC

After 30,000 iterations the aminopeptidase from *C. psychrerythraea* had acquired 954 mutations compared to 961 mutations acquired by the aminopeptidase from *S. oneidensis* when flexibility was the sole selective parameter (Fig. 6.5). When increased flexibility was combined with decreased isoelectric point, however, *C. psychrerythraea* acquired only 195 mutations compared to 211 by *S. oneidensis*. Adding hydrophathy as a selective pressure reduced the mutations in *C. psychrerythraea* to 27 and the mutations in *S. oneidensis* to 25. Increasing the number of iterations to  $10^6$  under the same selective pressure yielded an additional 28 mutations for *S. oneidensis* and 15 for *C. psychrerythraea*. The final flexibility and isoelectric point reached varied widely between the three experiments run for 30,000 iterations, but not between the two proteins (Fig. 6.6). In the flexibility experiment isoelectric point increased in near-sync with flexibility as a result of mutations to lysine, though there were significant numbers of mutations to glutamic acid, methionine, and serine, all with relatively low isoelectric points (Fig. 6.7). Without constraint on the isoelectric point both proteins rapidly evolved outside of the functional space defined by the high flexibility/low isoelectric point quadrant of Fig. 6.4. When decreased isoelectric point was added as a selective parameter, the proteins evolved within the observed functional space, showing considerable improvements in flexibility and reduced isoelectric point through mutations to glutamic acid, aspartic acid, methionine, serine, tyrosine and, to a much lesser extent than in the previous experiment, lysine (Fig. 6.7). In the experiment utilizing increased flexibility, decreased isoelectric point, and increased hydrophathy as selective pressure, only modest improvements were observed in flexibility and isoelectric point. These

improvements resulted primarily from mutations of arginine to lysine, with a few mutations of glutamine to serine and arginine to glutamine (Fig. 6.7). Extending the iterations to  $10^6$  did not change this pattern, but did increase the total number of mutations of arginine to glutamine and glutamine to serine. Less common beneficial mutations included histidine to serine and glutamine to valine.

## 6.4 Discussion

Our revisitation of the Metpally and Reddy (Metpally & Reddy, 2009) analysis identified far fewer statistically significant differences than originally reported. Metpally and Reddy (2009) had relied primarily on the Student's T-test with few degrees of freedom ( $n = 12$ ,  $df = 10$ ); the Student's T-test is appropriate for determining the difference in means between two normally distributed populations, but amino acid composition is not normally distributed across the small dataset involved. A large number (131) of comparative tests were performed in their study, but the p-values were not adjusted to control for multiple comparisons; in the absence of adjustments, the selected cutoff for reporting significance ( $\alpha = 0.90$ ) results in a higher likelihood of false positives. By re-interpreting the reported t-statistics and applying the Holm-Bonferroni correction (Holm, 1979) to account for the many comparisons in their study, we found that the significance threshold for  $\alpha = 0.90$  was  $p = 0.00025$ . Using this more rigorous definition, only serine emerged as a statistically significant difference between the two populations examined, and only for the coil region of the protein. The elevation of serine in the coil region is consistent with our re-analysis of the Metpally and Reddy dataset and of the extended dataset, where we found serine preferred in psychrophile proteins for all structures (Table 6.1).

Enhanced enzyme function at low temperature is generally considered to be a function of increased flexibility, which comes at the cost of reduced protein stability (D'Amico et al., 2006).

Protein flexibility can be considered on local or global scales, depending on whether residues are interacting with near neighbors or distant neighbors across folds. We did not incorporate 3-D structures in our analyses; as a result they are blind to interactions at the global scale.

Considering only local interactions, and based on amino acid mutation frequencies and natural abundances, the easiest way to improve local flexibility is to replace glutamic acid with lysine (D'Amico et al., 2006). In contrast with previous work (Metpally & Reddy, 2009), we found an elevated abundance of lysine and reduced abundance of glutamic acid in the genomes of psychrophiles relative to mesophiles. Although highly “probable” by the BLOSUM scoring matrices, this substitution has the disadvantage of minimizing hydropathy, which we observed to be elevated in psychrophile proteins (only arginine, disfavored in psychrophile proteins, has a lower hydropathy value than lysine). In our analysis, 1,646 Pfams containing the majority of analyzed proteins (90,116 out of 135,671) had elevated hydropathy in the psychrophile proteins. We could not find a published link between organism ecology and protein hydropathy; it is not yet clear what environmental parameter might be selecting for increased hydropathy.

The situation is made more complicated by reduced water activity in environments below the freezing point of water, where hygroscopic solutes are concentrated in the same interstitial spaces as microorganisms. Although not all psychrophiles are exposed to low water activity, those associated with permafrost, sea ice, glacial ice, snow, and more exotic environments such as polar saline lakes and springs, are. Although we did not find a significant difference in isoelectric point between all proteins from the psychrophiles and mesophiles examined, the difference was highly significant between acidic, high flexibility proteins ( $t = -10.2374$ ,  $df = 79,201$ ,  $p \approx 0$ , psychrophile mean = 5.59, mesophile mean = 5.64) (Fig. 6.4). A more acidic proteome is an observed feature of high-salt adapted microbes (Kiraga et al., 2007), thus a

lowered isoelectric point might help psychrophiles to achieve co-adaptation to low temperature and high salinity. Our model runs indicate that, while it is possible to evolve a protein with improvements in this parameter space through glutamic acid and methionine substitutions, it is more difficult than responding to selective pressure for increased flexibility alone. Further evidence of this trade-off can be seen in the observed relationship between these parameters for acidic, high flexibility proteins. Surprisingly, the slope of the relationship between these parameters is reduced in psychrophiles, suggesting an evolved preference for reduced isoelectric point at the cost of reduced flexibility. This trend may indicate that for some psychrophiles the selective pressure of low water activity trumps the selective pressure of low temperature. If this priority is the case, it may help to explain the presumed rarity of extreme halophiles that are also psychrophilic and their absence from culture collections.

As in the case of hydrophathy, lysine substitutions generally shift the isoelectric point away from the ideal (only arginine is more basic). Optimizing for both hydrophathy and isoelectric point strongly biases against lysine; in this scenario only arginine residues may be replaced by lysine. As demonstrated by the BLOSUM scoring matrices and our PEPC model runs, arginine to lysine substitutions are not difficult to achieve in nature. Lysine however, is an unlikely candidate for low temperature protein optimization, being hydrophilic, charged, and polar, all properties identified in previous studies as unfavorable for cold activity in proteins. In our analysis of the expanded genome dataset, we observed an increase in lysine and reduction in arginine in proteins from psychrophiles, suggesting that this substitution may have a role in protein optimization. In our model runs we observed a much smaller number of mutations of arginine to serine or glutamine. Both of these mutations impart a greater fitness improvement (much greater for serine), but are nonetheless harder to achieve. Serine is clearly preferred in the

analysis by Metpally and Reddy (Metpally & Reddy, 2009), in our re-analysis of the same sized dataset, and in the new analysis of the expanded dataset (Tables S6.1, S6.2). Thus lysine may represent the “first stop” on the road to full enzyme optimization that ends with serine.

Despite the “untapped” evolutionary potential of lysine residues, variation in mutation rate across sites may greatly reduce further optimization. Our collector’s curves (Fig. 6.5) suggest that for all six model runs the mutational potential of the protein was nearly exhausted. The beneficial and probable mutations were made early on. The remaining arginine and lysine residues may be contained within conserved regions of the protein, with further optimization proceeding only over considerable time. As the collection of psychrophile genomes grows, it may become possible to examine natural enzyme optimization from the perspective of both time and ecology. Considering the extreme difficulty of achieving anywhere near full optimization in the time afforded between geological thermal maxima (e.g. the Paleocene-Eocene Thermal Maximum 55 Mya (Zachos et al., 2003)), when many psychrophilic niches are destroyed, it is likely that most contemporary natural cold active proteins are far from optimal. Using a predicted substitution rate of 0.0245 % per nonsynonymous base position per My (Lawrence & Ochman, 1997), a protein coded by a 600 nucleotide gene might experience only 5 amino acid substitutions in 55 My. This substitution rate is an overestimate if population size is small or growth rate is slow (likely for psychrophiles). The cold deep sea, the most continuous low temperature environment but with limited psychrophilic representatives in culture, may represent the best environment for finding naturally occurring enzymes more completely optimized for cold activity. Because the deep sea is an ice-free environment of relatively stable salinity however, we do not expect these enzymes to be dually optimized to low temperature and low water activity.

## References

- Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nuc. Acids Res.*, 25(17), 3389-3402.
- Bae, E., & Phillips, G. N. (2004). Structures and analysis of highly homologous psychrophilic, mesophilic, and thermophilic adenylate kinases. *J. Biol. Chem.*, 279(27), 28202–28208.
- Bakermans, C., & Skidmore, M. (2011). Microbial respiration in ice at subzero temperatures (−4°C to −33°C). *Environ. Microbiol. Reports*, 3(6), 774-782. doi: 10.1111/j.1758-2229.2011.00298.x
- Bernal, A., Ear, U., & Kyrpides, N. (2001). Genomes OnLine Database (GOLD): a monitor of genome projects world-wide. *Nuc. Acids Res.*, 29, 126–127.
- Bowman, J. S., & Deming, J. W. (2014). Alkane hydroxylase genes in psychrophile genomes and their potential for cold activity. *BMC Genomics*, submitted.
- Breeze, J., Cady, N., & Staley, J. T. (2004). Subfreezing growth of the sea ice bacterium “*Psychromonas ingrahamii*”. *Microb Ecol*, 47(3), 300–304.
- Casanueva, A., Tuffin, M., Cary, C., & Cowan, D. A. (2010). Molecular adaptations to psychrophily: the impact of ‘omic’ technologies. *Trends in Microbiol.*, 18(8), 374–381. doi: 10.1016/j.tim.2010.05.002
- Chattopadhyay, M., & Jagannadham, M. (2001). Maintenance of membrane fluidity in Antarctic bacteria. *Pol. Biol.*, 24(5), 386–388.
- Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., . . . Wilczynski, B. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422–1423.
- Collins, E. (2013). hima: A Meta-database of Low Temperature Genomes and Metagenomes, from <http://relic.org/work/hima/>
- Collins, R. E., & Deming, J. W. (2013). An inter-order horizontal gene transfer event enables the catabolism of compatible solutes by *Colwellia psychrerythraea* 34H. *Extremophiles*, 1-10.
- Coquelle, N., Fioravanti, E., Weik, M., Vellieux, F., & Madern, D. (2007). Activity, stability and structural studies of lactate dehydrogenases adapted to extreme thermal environments. *J. Mol. Biol.*, 374(2), 547–562.
- Cox, G. F. N., & Weeks, W. F. (1983). Equations for determining the gas and brine volumes in sea-ice samples. *J. Glaciol.*, 29(102), 306–316.
- D'Amico, S., Collins, T., Marx, J. C., Feller, G., & Gerday, C. (2006). Psychrophilic microorganisms: challenges for life. *EMBO reports*, 7(4), 385–389.

- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, 14(9), 755–763.
- Henikoff, S., & Henikoff, J. G. (1992). Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.*, 89(22), 10915–10919.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, 6, 65–70.
- Huston, A. L., Methe, B., & Deming, J. W. (2004). Purification, characterization, and sequencing of an extracellular cold-active aminopeptidase produced by marine psychrophile *Colwellia psychrerythraea* strain 34H. *Appl. Environ. Microbiol.*, 70(6), 3321–3328.
- Johns, G. C., & Somero, G. N. (2004). Evolutionary convergence in adaptation of proteins to temperature: A4-lactate dehydrogenases of Pacific damselfishes (*Chromis* spp.). *Mol. Biol. Evol.*, 21(2), 314–320.
- Junge, K., Eicken, H., & Deming, J. W. (2004). Bacterial activity at –2 to –20°C in Arctic wintertime sea ice. *Appl. Environ. Microbiol.*, 70(1), 550–557. doi: 10.1128/aem.70.1.550-557.2004
- Kelley, L. A., & Sternberg, M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protocols*, 4(3), 363–371.
- Kim, S.-Y., Hwang, K. Y., Kim, S.-H., Sung, H.-C., Han, Y. S., & Cho, Y. (1999). Structural basis for cold adaptation sequence, biochemical properties, and crystal structure of malate dehydrogenase from a psychrophile *Aquaspirillum arcticum*. *J. Biol. Chem.*, 274(17), 11761–11767.
- Kiraga, J., Mackiewicz, P., Mackiewicz, D., Kowalczyk, M., Biecek, P., Polak, N., . . . Cebrat, S. (2007). The relationships between the isoelectric point and: length of proteins, taxonomy and ecology of organisms. *BMC Genomics*, 8(1), 163.
- Lawrence, J. G., & Ochman, H. (1997). Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.*, 44(4), 383–397.
- Leiros, H.-K. S., Pey, A. L., Innselset, M., Moe, E., Leiros, I., Steen, I. H., & Martinez, A. (2007). Structure of phenylalanine hydroxylase from *Colwellia psychrerythraea* 34H, a monomeric cold active enzyme with local flexibility around the active site and high overall stability. *J. Biol. Chem.*, 282(30), 21973–21986.
- Letunic, I., & Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nuc. Acids Res.*, 39(suppl 2), W475–W478.
- Marx, J. G., Carpenter, S. D., & Deming, J. W. (2009). Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium *Colwellia psychrerythraea* strain 34H under extreme conditions. *Can. J. Microbiol.*, 55, 63–72.



- McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, *16*(4), 404–405.
- Methe, B. A., Nelson, K. E., Deming, J. W., Momen, B., Melamud, E., Zhang, X., . . . Fraser, C. M. (2005). The psychrophilic lifestyle as revealed by the genome sequence of *Colwellia psychrerythraea* 34H through genomic and proteomic analyses. *P. Natl. Acad. Sci.*, *102*(31), 10913–10918. doi: 10.1073/pnas.0504766102
- Metpally, R., & Reddy, B. (2009). Comparative proteome analysis of psychrophilic versus mesophilic bacterial species: Insights into the molecular basis of cold adaptation of proteins. *BMC Genomics*, *10*(1), 11.
- Paternoster, R., Brame, R., Mazerolle, P., & Piquero, A. (1998). Using the correct statistical test for the equality of regression coefficients. *Criminology*, *36*(4), 859–866.
- Price, P. B. (2009). Microbial genesis, life and death in glacial ice. *Can. J. Microbiol.*, *55*, 1–11.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glockner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl. Acids Res.*, *35*(21), 7188–7196. doi: 10.1093/nar/gkm864
- Punta, M., Coghill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R. D. (2012). The Pfam protein families database. *Nuc. Acids Res.*, *40*(D1), D290–D301. doi: 10.1093/nar/gkr1065
- Russell, R. J., Gerike, U., Danson, M. J., Hough, D. W., & Taylor, G. L. (1998). Structural adaptations of the cold-active citrate synthase from an Antarctic bacterium. *Structure*, *6*(3), 351–361.
- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, *75*(23), 7537–7541. doi: 10.1128/aem.01541-09
- Siddiqui, K. S., Williams, T. J., Wilkins, D., Yau, S., Allen, M. a., Brown, M. V., . . . Cavicchioli, R. (2013). Psychrophiles. *Ann. Rev. Earth Pl. Sc.*, *41*(1), 87–115.
- Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics*, *22*(21), 2688–2690. doi: 10.1093/bioinformatics/btl446
- Vihinen, M., Torkkila, E., & Riikonen, P. (1994). Accuracy of protein flexibility predictions. *Proteins*, *19*(2), 141–149.
- Wells, L. E., & Deming, J. W. (2006). Characterization of a cold-active bacteriophage on two psychrophilic marine hosts. *Aquat. Microb. Ecol.*(45), 15–29.

- Zachos, J. C., Wara, M. W., Bohaty, S., Delaney, M. L., Petrizzo, M. R., Brill, A., . . . Premoli-Silva, I. (2003). A transient rise in tropical sea surface temperature during the Paleocene-Eocene thermal maximum. *Science*, 302(5650), 1551–1554.
- Zecchinon, L., Oriol, A., Netzel, U., Svennberg, J., Gerardin-Otthiers, N., & Feller, G. (2005). Stability Domains, Substrate-induced Conformational Changes, and Hinge-bending Motions in a Psychrophilic Phosphoglycerate Kinase: A microcalorimetric study. *J. Biol. Syst.*, 280(50), 41307–41314.

## Tables and Figures

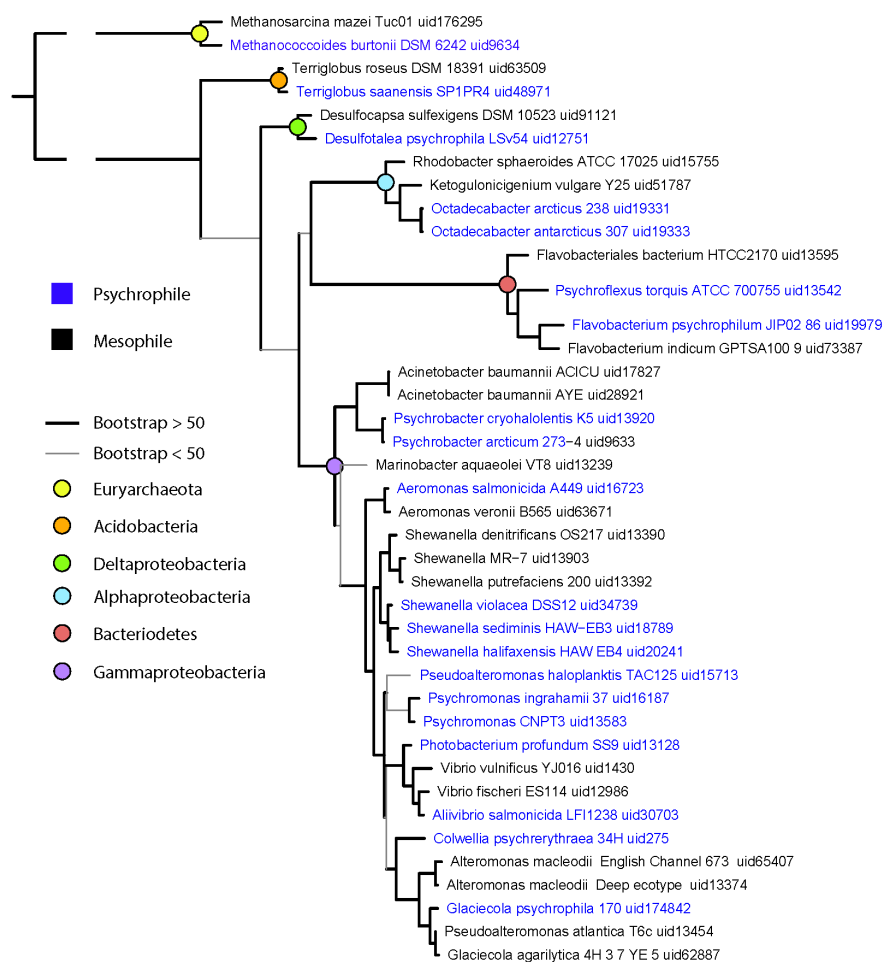
**Table 6.1.** Psychrophile amino acid preferences as reported in Metpally and Reddy (MR)

(Metpally & Reddy, 2009), the reanalysis (This analysis, MR), and the expanded dataset (This analysis, EX). X indicates residue property groups; +, psychrophile preferences in whole proteins and secondary structure elements (coil,  $\beta$ -sheet,  $\alpha$ -helix); and –, amino acids disfavored by psychrophiles. Gray shaded cells indicate a sign change between MR and the extended dataset. Red shaded cell indicates the preference was significant after the re-interpretation of the t-values in Metpally and Reddy (Metpally & Reddy, 2009).

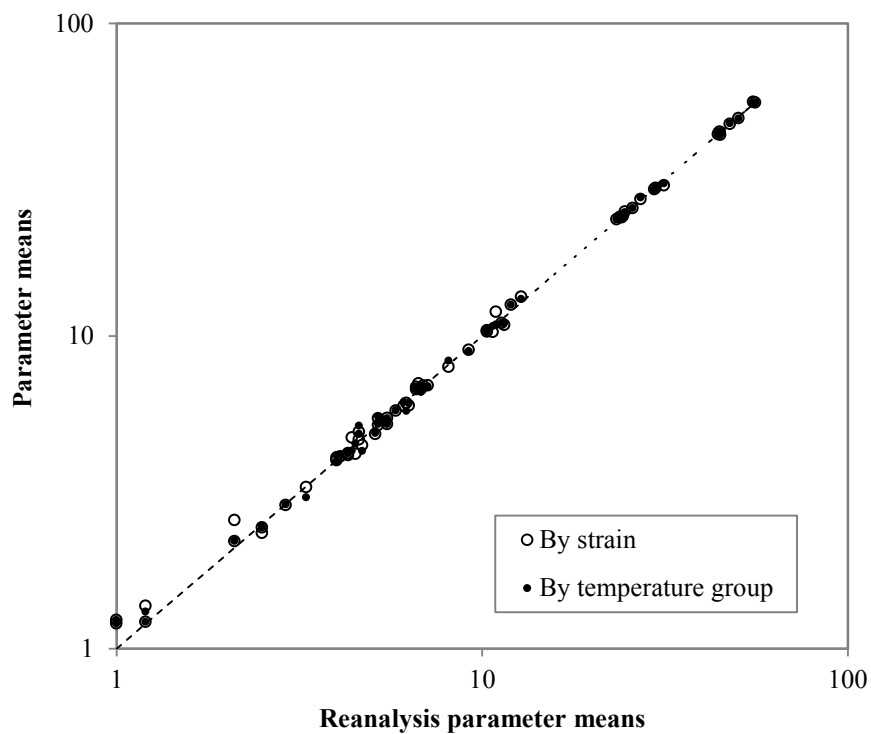
		Amino acid properties												MR				This analysis, MR				This analysis, EX				
Alanine	A	X		X	X	X	X							+	+	+	+	+	+	+	+	+	+	+	+	+
Cysteine	C	X		X	X	X	X															+		+	+	
Aspartic acid	D	X		X	X			X	X					+	+		+	+	+	+	+	+	+	+	+	
Glutamic acid	E	X		X	X			X						–	–		–	–	–	–	–	–	–	–	–	
Phenylalanine	F	X			X		X										–	–	–			+			+	
Glycine	G					X	X		X	X				+		+	+				–	–			–	
Histidine	H	X		X	X		X		X								+	–	–			–				
Isoleucine	I	X				X		X									+	+		+	+	+	+	+	+	
Lysine	K			X	X	X		X									–	–	–	–		+	+	+	+	
Leucine	L	X				X		X						–	–		–	–	–	–						
Methionine	M				X		X														+	+	+	+	+	
Asparagine	N				X			X	X					–			–	–			+	+	+	+	+	
Proline	P						X		X								–	–			–	–	–	–	–	
Glutamine	Q				X		X	X									–	–	–		–	–				
Arginine	R			X	X	X		X									–	–	–	–	–	–	–	–	–	
Serine	S					X		X	X	X				+	+		+	+	+	+	+	+	+	+	+	
Threonine	T					X		X	X	X				+	+		+	+	+	+	+	+	+		+	
Valine	V	X				X		X	X					+		+	+	+		+	–	–	–	–	–	
Tryptophan	W	X				X		X									–			–	–	–		–	–	
Tyrosine	Y	X				X		X									–	–	–	–	+	+	+	+	+	
		acidic	aliphatic	aromatic	basic	charged	hydrophilic	hydrophobic	neutral	nonpolar	polar	small	tiny	whole	coil	beta	helix	whole	coil	beta	helix	whole	coil	beta	helix	

**Table 6.2.** Psychrophile protein parameter preferences in Metpally and Reddy (MR) (Metpally & Reddy, 2009), the reanalysis (This analysis, MR), and the expanded dataset (This analysis, EX). + indicates psychrophile preferences; –, parameters disfavored by psychrophiles. Shaded cells indicate a sign change between MR and the extended dataset.

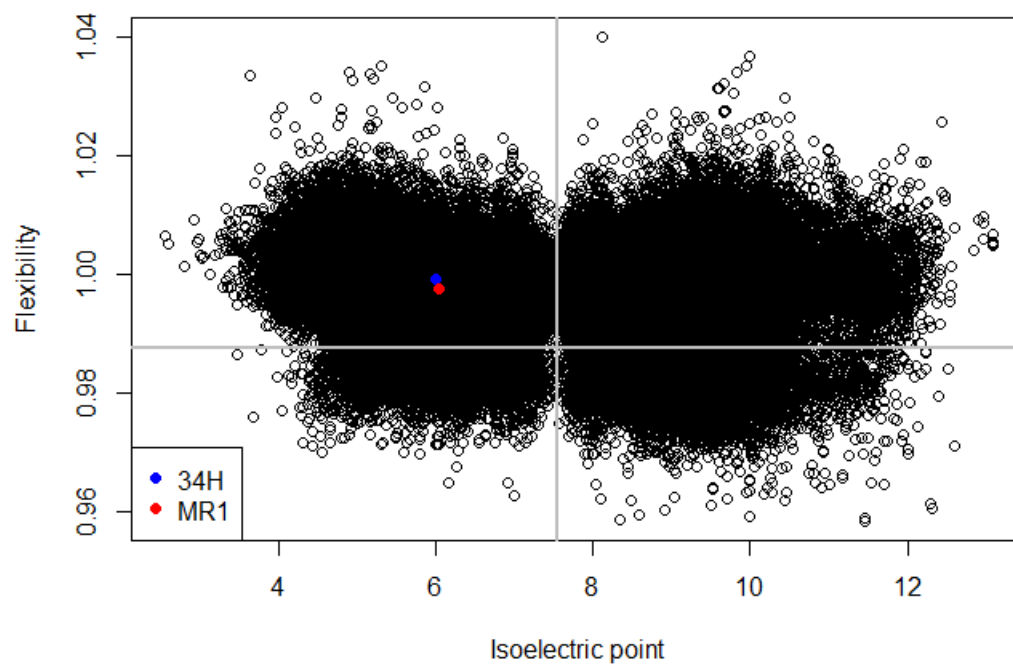
	MR				This analysis, MR				This analysis, EX			
tiny	+		+	+	+	+	+	+	–	–	–	–
small	+			+	+	+	+	+	–	–	–	–
aliphatic		–		+					+	+		+
aromatic	–			+	–	–	–	–	+			+
nonpolar		–				–	–	+	–	–	–	–
polar		–				+	+	–	+	+	+	+
charged	–	–			–	–	–	–				
basic	–	–			–	–	–	–				
acidic		–		+		+		–	–			–
neutral	+			+	+	+	+	+	+			+
hydrophilic	–	–			–	–	–	–		+		
hydrophobic		–		+		+	–	+				
flexibility						+						+
gravy					+	+	+	+	+	+		
isoelectric												
point					–	–	–	–				
aromaticity					–	–	–	–	+	+		+
aliphatic												
index									+	+	+	
	whole	coil	beta	helix	whole	coil	beta	helix	whole	coil	beta	helix



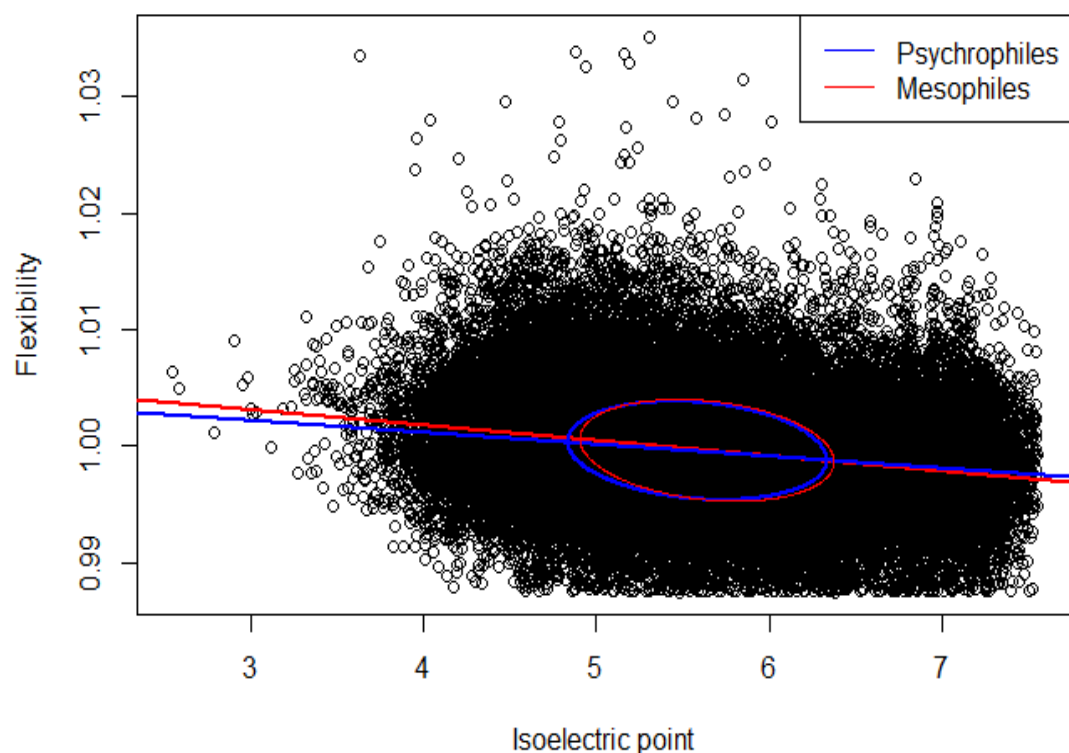
**Fig. 6.1. Maximum likelihood tree of 16S rRNA genes from Bacteria and Archaea in the expanded dataset.** The 16S rRNA genes were aligned in Mothur (Schloss et al., 2009) against the Silva reference set (Pruesse et al., 2007). The tree was constructed using RAxML v2.8 (Stamatakis, 2006) and visualized with iTOL (Letunic & Bork, 2011).



**Fig. 6.2. Correlation between parameter means (amino acid composition) in Metpally and Reddy (Metpally & Reddy, 2009) and in our reanalysis.** Dotted line indicates unity. Open circles indicate analysis by strain; solid circles, by temperature group.

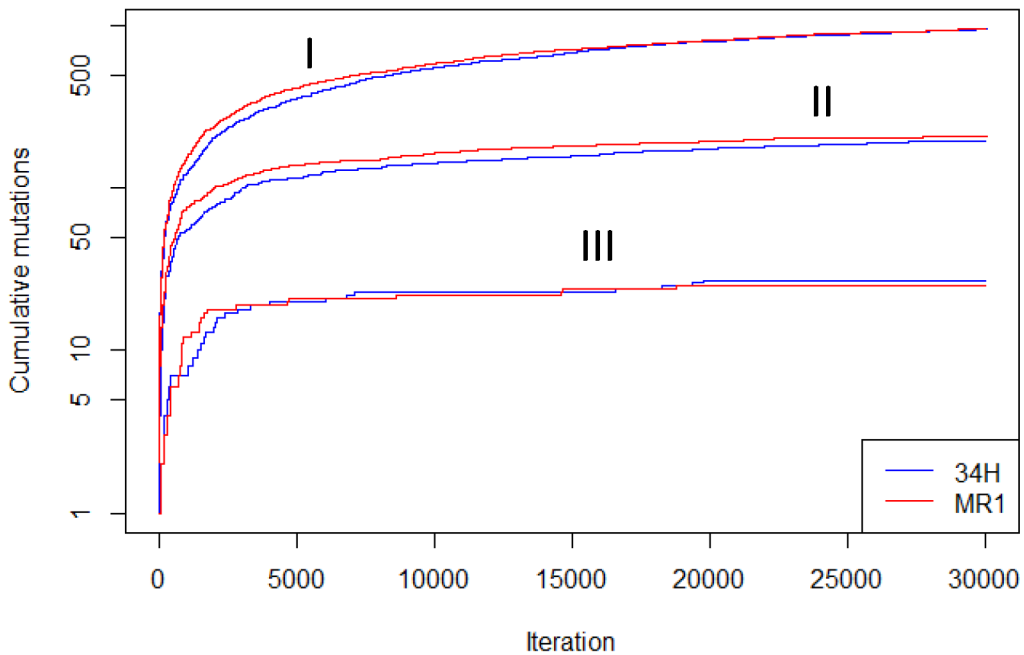


**Fig. 6.3. Flexibility vs. isoelectric point for all proteins in the expanded dataset.** 34H (blue) indicates *Colwellia psychrerythraea* 34H; MR1 (red) indicates *Shewanella oneidensis* MR-1. Gray lines indicate the local minima for isoelectric point and flexibility.

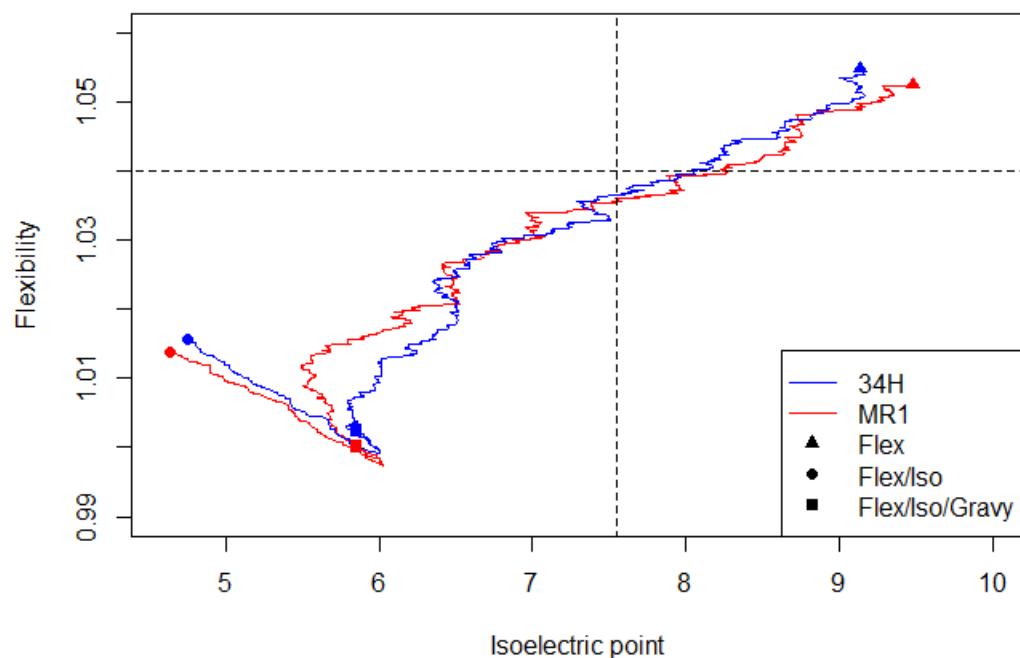


**Fig. 6.4. Flexibility vs. isoelectric point for proteins in the high flexibility/low isoelectric point quadrant (Fig. 6.3).** All proteins from the extended dataset that fall in the parameter space are plotted. Red line is a linear model fit to the mesophile proteins ( $R^2 = 0.05113$ ,  $p \approx 0$ ), with the red circle showing the mean and standard deviation for the two parameters. Blue line is a linear model fit to the psychrophile proteins ( $R^2 = 0.03284$ ,  $p \approx 0$ ), with the blue circle showing the mean and standard deviation for the two proteins.

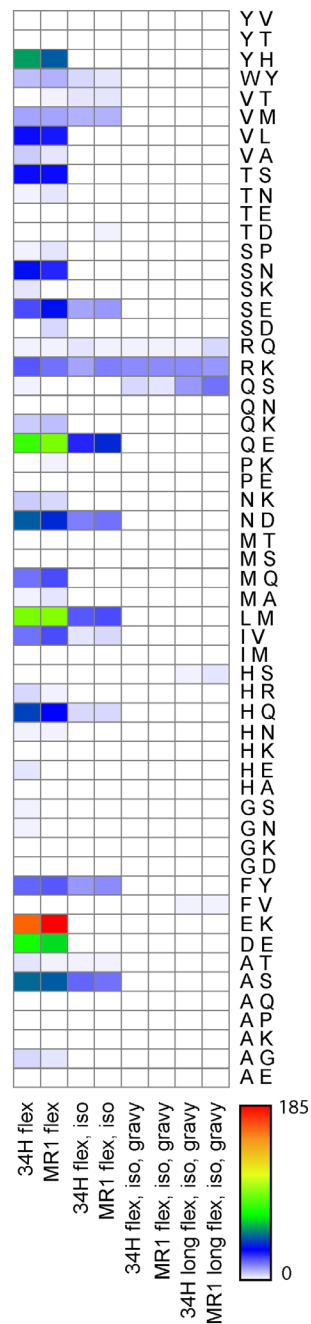




**Fig. 6.5. Mutation collector's curves generated with the PEPC model.** I) Increased flexibility is the only selective pressure. II) Increased flexibility and decreased isoelectric point are the selective pressures. III) Increased flexibility, decreased isoelectric point, and increased hydropathy are the selective pressures. 34H: an aminopeptidase from *Colwellia psychrerythraea* 34H, MR1: a homologous aminopeptidase from *Shewanella oneidensis* MR1.



**Fig. 6.6. Location in parameter space defined by flexibility and isoelectric point during PEPC model runs.** End points of these 30,000 iteration runs for a psychrophilic (blue) and mesophilic (red) aminopeptidase are given by shapes; start point is the point of convergence for the model runs. Dashed lines indicate the maximum flexibility and isoelectric point for high flexibility/low isoelectric point proteins 34H (blue) and MR1 (red) indicate source strains, as given in Fig. 6.3.



**Fig. 6.7. Mutations preferred in each PEPC model run.** Original and replacement residue are given as row names, where letters are defined as in Table 6.1. Column names follow secondary structure element parameters in Table 6.2, for strains 34H and MRI, as given in Fig. 6.3. Long indicates model run for  $10^6$  iterations, other model runs were for 30,000 iterations.

**Table S6.1.** Psychrophile protein parameter preferences for the Metpally and Reddy (Metpally & Reddy, 2009) dataset determined using the by-strain method as reported by Metpally and Reddy (bulk protein only) and as determined in this analysis (by secondary structure element and by bulk protein). See supplementary file S6.1.tsv.

**Table S6.2.** Psychrophile protein parameter preferences for the Metpally and Reddy (Metpally & Reddy, 2009) dataset determined using the by-class method as determined in this analysis (by secondary structure element and by bulk protein). See supplementary file S6.2.tsv.

**Table S6.3.** Psychrophile protein parameter preferences for the extended dataset using the by-strain method. See supplementary file S6.3.tsv.

**Table S6.4.** Psychrophile and mesophile protein parameter preferences for the extended dataset using the by-class method. See supplementary file S6.4.tsv.

## Chapter 7

### **Alkane hydroxylase genes in psychrophile genomes and the potential for cold active catalysis**

#### ABSTRACT

Psychrophiles are presumed to play a large role in the catabolism of alkanes and other components of crude oil in natural low temperature environments. In this study we analyzed the functional diversity of genes for alkane hydroxylases, the enzymes responsible for converting alkanes to more labile alcohols, as found in psychrophile genomes. In seven of the analyzed psychrophile genomes we found 27 candidate alkane hydroxylase genes, of which only two are currently annotated as alkane hydroxylase. These candidates were mostly related to the AlkB and cytochrome p450 alkane hydroxylases, but several homologues of the LadA and AlmaA enzymes, significant for their ability to degrade long-chain alkanes, were also detected. To identify possible mechanisms of low temperature optimization we compared putative alkane hydroxylases from psychrophiles with homologues from taxonomically related mesophiles, finding significant differences in primary structure, with preferences for specific amino acids and increased flexibility on loops, bends, and  $\alpha$ -helices for psychrophile proteins. These findings are consistent with observations of microbial degradation of crude oil in cold environments and identify proteins that can be targeted in rate studies and in the design of molecular tools for low temperature bioremediation.

## 7.1 Introduction

In the natural environment crude oil, a complex mixture of light and heavy hydrocarbons and inorganic compounds, is degraded by members of the Bacteria and Archaea, as well as by certain plants and fungi. Significant work has been done to identify the taxonomic groups and pathways involved in the bioremediation of crude oil, motivated in part by the need to improve predictions of *in situ* degradation rates of the oil and of targeted hydrocarbon compounds. The 2010 Macondo Well blowout in the Gulf of Mexico highlighted some of the gaps in our understanding of how crude oil is degraded *in situ* in the pelagic marine environment, as both the hydrocarbons emerging from the damaged well and the *in situ* marine microbial community reacted in ways that were unpredicted (Hazen et al., 2010; Redmond & Valentine, 2012; Valentine et al., 2010). In that case, components of the crude oil advecting in a deep plume below 1000 m in the Gulf were consumed *in situ* by the marine microbial community, reducing ecological disturbance at the sea surface.

The extent of the microbial crude oil catabolism at the relatively cold temperature (approximately 5°C) of the deep plume was considered surprising (Redmond & Valentine, 2012). It has long been recognized however, that bacteria can respond quickly to crude oil even in near-freezing seawater (Colwell, Walker, & Cooney, 1977). Even sea ice microbial communities, living at temperatures below the freezing point of seawater, can respond to inputs of diesel fuel and crude oil (Brakstad, Nonstad, Faksness, & Brandvik, 2008; D. Delille, Basseres, & Dessommes, 1996; Gerdes, Brinkmeyer, Dieckmann, & Helmke, 2005). Low temperature crude oil degradation has also been observed in polar and alpine soil (Bell et al., 2013; D Delille, 2000; Margesin, Labbe, Schinner, Greer, & Whyte, 2003), and by several Bacterial strains in culture (Powell, Bowman, & Snape, 2004; Whyte, Bourbonniere, & Greer,

1997; Whyte et al., 1998). Despite these and other advances in understanding the potential for low temperature bioremediation, the presence of crude oil degradation genes specifically among the available psychrophile genomes has not been investigated, though recent work has suggested that these genes might be broadly distributed across the domains of Bacteria and Archaea (Nie et al., 2014). By identifying such genes and evaluating differences between gene products and homologues from mesophilic Bacteria and Archaea, we hoped to identify structural differences that may enable crude oil catabolism at low temperatures. In addition to improving our ability to predict *in situ* bioremediation in cold environments, this knowledge paves the way for the rational design or modification of enzymes for improved function at *in situ* temperature in polar and sub-polar environments. These considerations are important for small scale, reduced energy environmental clean-up strategies involving bioreactors and other technologies. Rational protein manipulation has already resulted in enzymes of potential value for environmental cleanup and industrial processes (Glieder, Farinas, & Arnold, 2002; Harford-Cross et al., 2000), however, this work has been limited to possible terrestrial, not marine, applications at standard conditions of room temperature and pressure.

By mass a considerable fraction of crude oil is n-alkanes (alkanes); straight chain, saturated hydrocarbons with no cyclic functional groups. The shortest and most volatile alkanes are the natural gas components methane, ethane, butane, and propane, all of which are important substrates for a variety of Bacteria and Archaea. Even approaching the freezing point of water these small alkanes remain in the gas phase and are therefore highly bioavailable.

Bioavailability decreases with an increasing number of carbons in an alkane molecule, reaching a minimum with large, extremely hydrophobic waxes (Colwell et al., 1977). At mesophilic growth temperatures alkanes larger than C<sub>16</sub> are solid, necessitating the use of emulsifiers to improve

bioavailability (Wentzel, Ellingsen, Kotlar, Zotchev, & Throne-Holst, 2007). To degrade alkanes of different lengths Bacteria and Archaea have evolved a diverse array of enzymatic tools, collectively termed alkane hydroxylases. All alkane hydroxylases function by oxidizing the terminal or subterminal carbon, converting the alkane into an alcohol (Beilen & Funhoff, 2007). This conversion “activates” the alkane for processing by downstream enzymes, starting with alcohol dehydrogenase.

The diversity of alkane hydroxylases, described in recent reviews (Beilen & Funhoff, 2007; Ji, Mao, Wang, & Bartlam, 2013; Van Beilen, Li, Duetz, Smits, & Witholt, 2003; Wentzel et al., 2007), are briefly summarized here. Operating on the lowest molecular weight alkanes (approximately C<sub>1</sub>-C<sub>4</sub>) are the soluble methane monooxygenase (SMMO), particulate methane monooxygenase (PMMO), and propane/butane monooxygenase (P/BMO) enzymes. Acting on mid-weight alkanes (roughly C<sub>5</sub>-C<sub>16</sub>) are a group of alkane hydroxylases belonging to the cytochrome p450 family of enzymes and the membrane-bound non-heme AlkB enzymes. Less is known about the degradation of long chain alkanes, but two enzymes, AlmA and LadA, have been identified that utilize alkanes large than C<sub>20</sub> (Feng et al., 2007; Throne-Holst, Wentzel, Ellingsen, Kotlar, & Zotchev, 2007).

To explore the diversity of alkane hydroxylases in the genomes of psychrophilic Bacteria and Archaea we conducted a *de novo* annotation of twenty psychrophile genomes, searching for homologues of known alkane hydroxylase genes. To evaluate what properties of these proteins might enable catalytic function at low temperature we compared protein parameters between putative alkane hydroxylases from psychrophiles and mesophiles averaged across the whole protein, within secondary structure elements, and, for protein flexibility, within specific residues along the length of the protein. We were guided in part by results from a related analysis seeking



generic indicators of cold adaptation in proteins of the same psychrophile genomes (Bowman & Deming, 2014).

## **7.2 Methods**

### **7.2.1 Identifying alkane hydroxylases**

Proteins representative of alkane hydroxylases were identified in the UniProt database (Bairoch et al., 2009) by protein name search for ‘alkane hydroxylase’, ‘methane monooxygenase’, ‘propane monooxygenase’, ‘butane monooxygenase’, ‘LadA’, and ‘AlmA’. Proteins belonging to uncultured organisms or identified as fragments were excluded from further analysis; duplicated names or sequences were reduced to a single copy. An exception was made to allow fragments for AlmA, as all AlmA proteins in the database were described as fragments yet were of similar length. Conserved domains were identified in the representative alkane hydroxylases by hmmscan in HMMER v3.0 (Eddy, 1998) against Pfam-A (Punta et al., 2012) with an e-value cutoff of  $10^{-5}$ .

Twenty psychrophilic strains (maximum growth temperature  $< 20^{\circ}\text{C}$ ) and twenty closely related mesophilic strains were selected from Genbank and translated as described separately (Bowman & Deming, 2014). ORFs were translated and searched for conserved protein domains against the Pfam-A database (Punta et al., 2012) using hmmscan in HMMER v3.0 (Eddy, 1998) and an E-value cutoff of  $10^{-5}$ . Coding sequences (CDS, ORFs containing a pfam domain) with a hit to a pfam present in alkane hydroxylases were extracted for further analysis.

Complete records for diagnostic pfams were downloaded as fasta files from the Protein Family website (Pfam; <http://pfam.xfam.org/>). For Pfam datasets larger than 5,000 sequences, 5,000 sequences were randomly selected for analysis. For each pfam, the Pfam dataset was combined with the proteins of that family from the UniProt, psychrophile, and mesophile

datasets. These combined protein sets were aligned using three iterative alignments in Clustalo v1.2 (Sievers et al., 2011). The alignments were then filtered using an in-house script (`filter_seqs_selective.py`) which trims the alignment to the last start and earliest end position of the proteins from the UniProt dataset. Proteins from the psychrophile, mesophile, or Pfam datasets that did not meet a minimum length guideline after filtering were eliminated from further analysis. After filtering, the sequences were aligned one more time, and a distance matrix of each pfam was created using the `--full` and `--use-kimura` flags in Clustalo v1.2.

To describe sequence similarity within pfams we used nonmetric multidimensional scaling (NMDS) of Kimura corrected genetic distance (Kimura, 1980) in the R package Vegan (Oksanen et al., 2012). This method was selected over phylogenetic trees based on the ease with which points in a region of interest on a 2D NMDS plot can be selected programmatically, compared to selecting branches on a phylogenetic tree. Although NMDS plots have been used to describe protein homology previously (Pelé, Abdi, Moreau, Thybert, & Chabbert, 2011), this method is not in wide use. To validate the NDMS approach to describing sequence similarity we compared the Euclidean distance between NMDS points in the first and second dimension, maximum likelihood distance from a phylogenetic tree based on the same alignment, and bit scores from a reciprocal blastp search. We used the combined protein dataset for the FA\_desaturase pfam for this analysis and generated a tree of the filtered alignment using FastTree OpenMP v2 (Price, Dehal, & Arkin, 2010) with the JTT+CAT model. Summed branch lengths between all branch tips were extracted from the tree with an in-house script (`dist_from_tree.py`) using the Phylo package in Biopython (Cock et al., 2009). To describe the relationship between phylogenetic tree distance, bit score, and NMDS distance, linear models were fit to a randomly selected subset of the data ( $n = 10,000$ ) in log-linear space for NMDS and

phylogenetic distance and log-log space for NMDS and bit score distance. Goodness of fit was further evaluated by exploring the distribution of the residuals.

For NMDS analysis we determined the ideal number of dimensions to be three for fewer than 3,000 sequences, four for between 3,000 and 6,000 sequences, and five for more than 6,000 sequences. Sequences that placed very far from the majority of points in a 2D plot of the NMDS analysis, and thus prevented the identification of distinct clusters for the majority of points, were culled from the original alignment and a new distance matrix was constructed before re-running the NMDS analysis. Clusters on the final 2D NMDS plots that contained proteins from the UniProt, psychrophile, and mesophile datasets were selected for further analysis.

### **7.2.2 Analysis of Protein Parameters**

The flexibility, grand average of hydropathy (GRAVY), isoelectric point, and aromaticity parameters were calculated as described separately (Bowman & Deming, 2014). In brief, parameters were calculated with the ProtParam module in BioPython (Cock et al., 2009) and the aliphatic index calculation method of Ikai (Ikai, 1980). To determine the parameters by secondary structure, the  $\alpha$ -helix,  $\beta$ -strand, and coil region for each protein was determined by the standalone version of psipred (McGuffin, Bryson, & Jones, 2000) and the runpsipredplus script. The best database for secondary structure prediction was evaluated by comparing predictions using the NCBI nr database, uniref90, and Pfam-A for one candidate alkane hydroxylase against predictions obtained from an intensive 3-D structural prediction with Phyre2 (Kelley & Sternberg, 2009). Both databases achieved a prediction accuracy of 71.8 %, just below the prediction of psipred as implemented by the Phyre2 server (72.7 %). We used Pfam-A for further predictions due to the smaller size of that database. Protein parameters were recalculated using a 9 residue window (selected for consistency with the flexibility calculation), and the per-

position parameter was taken as the mean of the window centered on that position. Per-position values for each parameter were then extracted for comparison according to secondary structure.

To evaluate differences in flexibility, widely considered important for cold activity, between putative alkane hydroxylases from psychrophiles and mesophiles on a by-position basis we aligned the flexibility parameters for all proteins in each cluster according to a multiple sequence alignment generated in Clustalo v2 (Sievers et al., 2011) using in-house scripts (align\_params.py, align\_params.r). For each position in the alignment the mean flexibility and standard deviation were calculated for psychrophile and mesophile proteins. Positions in the alignment where the difference in means (psychrophile proteins – mesophile proteins) between the two groups exceeded the sum of the standard deviations were flagged as sites of significant deviation. To place these findings in the context of protein tertiary structure, 3D models were constructed of a representative psychrophile protein in each cluster using the intensive modeling option in Phyre2 (Kelley & Sternberg, 2009). Residues with significant differences in flexibility were highlighted in the models using Discovery Studio Visualizer (Accelrys).

All in-house scripts can be obtained from [https://github.com/bowmanjeffs/cold\\_ah](https://github.com/bowmanjeffs/cold_ah).

### **7.3 Results**

The UniProt searches collectively returned 939 alkane hydroxylase proteins after culling. These proteins belonged to 16 pfams of which seven were determined to have a regulatory or electron carrier binding function, or to be the result of an erroneous classification (MmoB\_DmpM, ADH\_zinc, LXG, Nol1\_Nop2\_Fmu, DUF900, NAD\_binding\_1, FAD\_binding\_6). Four of the remaining nine pfams were represented in the psychrophile genomes (Table 7.1). Among these four was FA\_desaturase, used to show the correlations between Euclidean distance in 2D NMDS space and phylogenetic tree distance ( $R^2 = 0.4232$ )

and bit score ( $R^2 = 0.5029$ ) (Fig. 7.1). NMDS plots of all four pfams contained clusters with both psychrophilic and UniProt proteins, indicating close sequence similarity (Fig. 7.2, Table 7.2). The Pyr\_redox\_3 and Bac\_luciferase pfams each had only one cluster, corresponding to AlmA and LadA respectively. FA\_desaturase had two clusters; cluster 0 corresponds to the AlkB group of membrane bound alkane hydroxylases, while cluster 1 is defined by only a single UniProt protein annotated as alkane-1 monooxygenase. The p450 family also had two clusters; cluster 0 corresponds to the Bacterial p450 alkane hydroxylase, while cluster 1 corresponds to the Eukaryotic p450 alkane hydroxylase. A total of 26 putative alkane hydroxylases were identified in the psychrophile genomes and 41 in the mesophile genomes (Table S7.1).

The Pyr\_redox\_3, Bacterial\_luciferase, FA\_desaturase, and p450 families contained sufficient psychrophile and mesophile proteins to allow a comparison of protein parameters within these two families. Considering parameters averaged across the protein, no pfam had a statistically significant difference in any parameter between the psychrophile and mesophile populations. We thus used “trending” to describe differences in the parameter space when  $p \ll 0.05$  by the Wilcox Test, but when differences did not meet that significance threshold after applying the Holm-Bonferroni method to correct for multiple comparisons (Holm, 1979). Trending differences were observed in three of the four pfams. Flexibility and tryptophan content trended lower in psychrophile FA\_desaturase, while GRAVY and lysine content trended higher. For p450, threonine content trended higher in psychrophiles. For Bac\_luciferase, alanine, isoleucine, and lysine trended lower in psychrophiles, while cytosine, methionine, arginine, and tyrosine trended higher.

Comparing between taxon pairs (Table 7.2) revealed more differences for parameters averaged across the whole protein (Table 7.3). For p450, four psychrophile-mesophile taxon

pairs were available for analysis: *Octadecabacter antarcticus* 307 and *Rhodobacter sphaeroides* ATCC 17025, *Glaciecola psychrophila* 170 and *Glaciecola agarylita* 4H37YE5, *Octadecabacter arcticus* 238 and *Ketogulonicigenium vulgare* Y25, and *Terriglobus roseus* DSM18391 and *Terriglobus saanensis* SP1PR4. Because multiple genes were present in some of these genomes, a total of nine pairwise comparisons were possible. In all nine comparisons the psychrophile protein had a lower isoelectric point and arginine content than the mesophile protein, while valine was elevated in psychrophiles in all comparisons. Asparagine and tyrosine were elevated in psychrophile proteins in all but one comparison.

There were only two psychrophile-mesophile pairs available for the analysis of putative alkane hydroxylases from the Pyr\_redox\_3 family: *G. psychrophila* 170 and *G. agarylita* 4H37YE5 and *Psychrobacter cryohalolentis* K5 and *Acinetobacter baumonii* AYE. Due to the large number of putative hydroxylases belonging to this pfam in *G. psychrophila* 170 and *A. baumonii* AYE, however, 11 comparisons were possible. Cysteine and valine were elevated in the psychrophile proteins for all but one comparison; glutamic acid was reduced in the psychrophile proteins for all comparisons. For FA\_desaturase only one taxon pair was available for analysis: *O. antarcticus* 307 and *R. sphaeroides* ATCC 17025, with four possible comparisons. Given the limited number of comparisons pairwise FA\_desaturase parameters were not explored further.

Considering protein parameters by the secondary structure elements  $\alpha$ -helix,  $\beta$ -sheet, or coil also revealed no statistically significant differences in protein physical parameters. The strongest trends were observed for psychrophile FA\_desaturases: lowered flexibility in the coil and  $\alpha$ -helix regions and reduced acidic residues and lysine in the  $\alpha$ -helices. Considering taxon pairs, for p450 isoleucine was generally reduced in psychrophile  $\alpha$ -helices and  $\beta$ -sheets but

elevated in coils. Asparagine and valine were always higher in the coil region for psychrophiles. Flexibility, isoelectric point, alanine, glycine, and proline were all generally elevated in  $\beta$ -sheets. Flexibility and aspartic acid were elevated in  $\alpha$ -helices, while arginine was reduced. For psychrophile Pyr\_redox\_3, aspartic acid was reduced in the coil while glycine was elevated. In psychrophile  $\beta$ -sheets, GRAVY and cysteine were both elevated. Glutamic acid and asparagine were elevated in psychrophile  $\alpha$ -helices while alanine was reduced.

The Phyre2 protein fold prediction server produced high confidence (90 % or more residues modeled at 90 % or greater confidence) for the Pyr\_redox\_3, Bac\_luciferase, and p450 pfams. FA\_desaturase could not be modeled with high confidence.

#### **7.4 Discussion**

Alkanes are ubiquitous in marine and soil environments, occurring as by-products of cell metabolism and death; they also enter the marine environment from natural hydrocarbon seeps and anthropogenic sources. As a result alkane degradation is likely widespread among heterotrophic bacteria (Nie et al., 2014). Several studies have demonstrated that some psychrotolerant Bacteria (growth at 0°C but maximum growth temperature > 20°C), which overlap ecologically and geographically with psychrophilic bacteria, can catabolize alkanes (Cao et al., 2011; Trotsenko & Khmelenina, 2002; Whyte et al., 1997; Whyte et al., 1998), while crude oil degradation has been observed in a variety of cold environments (Bell et al., 2013; Brakstad et al., 2008; D Delille, 2000; Gerdes et al., 2005; Margesin et al., 2003; Redmond & Valentine, 2012). Despite the ubiquity of low molecular weight alkanes, the AMO, AmoC, MeMO\_Hyd\_G, Monooxygenase\_B, and Phenol\_Hydrox pfams were not detected in the psychrophile genomes examined. Because these pfams are known to be restricted to relatively few taxonomic groups, their absence in the analyzed psychrophiles, though indicative of an

inability to catabolize low molecular weight alkanes in this group, may not be surprising. The twenty psychrophile genomes available to explore in this analysis, however, represent only a very small sampling of psychrophile functional diversity. Psychrophiles are known, for example, to undergo C1 metabolism (Trotsenko & Khmelenina, 2002), yet none of these strains has been targeted for genome sequencing. As more psychrophile genomes are sequenced and published, they can be explored for additional alkane catabolism pathways predicted from environmental evidence but missing in the current set of analyzed genomes. Given that the diversity of enzymes involved in alkane degradation is also not fully explored, the current genomes may well contain alkane hydroxylases lacking sufficient sequence homology to known alkane hydroxylases. This possibility was highlighted by a recent genomic study of a cold-adapted *Colwellia* strain obtained from the deep hydrocarbon plume in the Gulf of Mexico (Mason, Han, Woyke, & Jansson, 2014b). No genes for known alkane hydroxylase were identified in this strain despite abundant ancillary data linking *Colwellia* to short chain alkane degradation following the Macondo Well blowout (Mason, Han, Woyke, & Jansson, 2014a).

Because alkane bioavailability is positively correlated with temperature and negatively correlated with chain length, the preferential degradation of short chain alkanes is expected in cold environments. Surprisingly, we found several candidate alkane hydroxylases homologous to LadA and AlmA, enzymes associated with the degradation of long-chain alkanes. These putative long-chain alkane hydroxylases are ecologically diverse, occurring in the genomes of sea ice Bacteria (*Octadecabacter arcticus*, *O. antarcticus*, and *Glaciecola psychrophila*) and tundra soil Bacteria (*Terroglus saanensis*). Confirming the ability of these strains to degrade long-chain alkanes or similar substrates will be a priority in future work. Because *O. arcticus*, *O. antarcticus*, and *G. psychrophila* are all associated with sea ice, which in springtime hosts a high



density of ice algae, their hypothesized ability to degrade long-chain alkanes may result from a preference for ice-algal lipids. The bioavailability of lipids and long-chain alkanes can be enhanced at low temperature by naturally occurring surfactants, most notably microbially produced exopolymers (EPS) (Gutierrez et al., 2013). Sea ice is rich in EPS (Krembs, Eicken, Junge, & Deming, 2002) which may enable the catabolism of these compounds even at low temperature.

A considerable body of literature is dedicated to determining what protein modifications enable enzymatic function at low temperature. At low temperatures water molecules interact more tightly with the protein surface, reducing the overall flexibility of the protein. To counter the effect of low temperature on enzyme function cold active proteins make use of a variety of amino acid substitutions. The sum impact of these different substitutions, including their interactions and feedbacks, is difficult to predict. Compounding this difficulty is the co-occurrence of low temperature and low water activity, as found in virtually all ice-matrices (e.g., permafrost, glacial ice, and sea ice). Optimization to low water activity and low temperature may be more difficult than optimization to low temperature alone.

Recently we reported a broad trend toward serine enrichment within psychrophile proteins, as replacements of glutamine and histidine with serine have the combined advantage of increasing flexibility and hydrophathy (for cold activity) and lowering the protein isoelectric point (for low water activity) (Bowman & Deming, 2014). Despite the advantages of serine enrichment, the putative cold active alkane hydroxylases explored here depart from this trend; no significant enrichment for serine was observed for psychrophiles in any of the analyzed pfams or in the pairwise comparisons. Although all pfams showed some differences between psychrophiles and mesophiles for other parameters, no coherent overall optimization strategy

was evident. The clearest trends appeared in our pairwise comparisons, which were limited Pyr\_redox\_3 and p450 and Pyr\_redox\_3. High flexibility and low isoelectric point appear to be important for cold adaptation in p450, where asparagine, tyrosine, and valine were all enriched and arginine reduced. For Pry\_redox\_3 cysteine and valine were enriched and glutamic acid was reduced. With the exception of valine, which is usually reduced in psychrophiles, these amino acid preferences are typical for cold active proteins (Bowman & Deming, 2014).

One challenge to evaluating protein temperature optimization is the localization of some parameters. Although changes to isoelectric point and hydropathy are likely to be globally distributed, at least among secondary structure elements or sites of a given solvent accessibility, optimized flexibility may come about through the modification of only specific residues (Harvilla, Wolcott, & Magyar, 2014; Leiros et al., 2007). Regions of consistently increased flexibility were present in alignments from all six putative alkane hydroxylase clusters, with the exception of Pyr\_redox\_3 cluster 0 (regions of consistently decreased flexibility were present in all alignments) (Fig. 7.3). The generalized location of increased flexibility varied between cluster representatives. P450 cluster 0 had several regions of increased flexibility at probable hinge points on bends, loops, and in the coil region (Fig. 7.4). Interestingly, three of these were centered on methionine residues (Met8, Met269, Met295 in the cluster 0 p450 representative from *Glacielecola psychrophila*). Methionine is known to play a role in low temperature optimization of other heme-binding proteins by providing alternate heme binding sites in the event of partial denaturation (Harvilla et al., 2014). In the *G. psychrophila* p450, however, most of these sites were located toward the exterior of the protein and are unlikely to interact with heme. P450 cluster 1 had no evidence of increased flexibility on loops or coils, but did have regions of increased flexibility in the core of the protein. Bac\_luciferase cluster 0 had large

differences in local flexibility between the psychrophile and mesophile proteins. Regions of increased flexibility included bends likely to function as hinge points and residues near the protein active site.

Although the total number of putative alkane hydroxylases in the analyzed psychrophiles was smaller than in a taxonomically related group of mesophiles, the metabolic potential for alkane degradation in the psychrophiles is clear. These findings are consistent with observations of crude oil degradation in sea ice, permafrost, and most recently the cold deep ocean. As in other cold active enzymes the putative alkane hydroxylases show clear and, within clusters, consistent differences in amino acid composition and protein parameters from mesophilic homologues. These proteins are good candidates for rate studies and rational manipulations.

### **Acknowledgements**

JSB is supported by an EPA STAR Fellowship. JWD is supported by the Walters Endowed Professorship and NSF PLR 1203267.

## References

- Bairoch, A., Bougueleret, L., Altairac, S., Amendolia, V., Auchincloss, A., Argoud-Puy, G., . . . Boeckmann, B. (2009). The Universal Protein Resource (UniProt) 2009. *Nuc. Acids Res.*, 37(Database issue), D169–174.
- Beilen, J. B. v., & Funhoff, E. G. (2007). Alkane hydroxylases involved in microbial alkane degradation. *Appl. Microbiol. Biotechnol.*, 74(1), 13–21.
- Bell, T. H., Yergeau, E., Maynard, C., Juck, D., Whyte, L. G., & Greer, C. W. (2013). Predictable bacterial composition and hydrocarbon degradation in Arctic soils following diesel and nutrient disturbance. *ISME J*, 7(6), 1200–1210.
- Bowman, J. S., & Deming, J. W. (2014). Amino acid preferences in psychrophilic genomes: Is serine the answer to low temperature and high salinity? *BMC Genomics*, submitted.
- Brakstad, O., Nonstad, I., Faksness, L.-G., & Brandvik, P. (2008). Responses of microbial communities in Arctic Sea Ice after contamination by crude petroleum oil. *Microb. Ecol.*, 55(3), 540–552. doi: 10.1007/s00248-007-9299-x
- Cao, B., Ma, T., Ren, Y., Ren, Y., Li, G., Li, P., . . . Feng, L. (2011). Complete genome sequence of *Pusillimonas* sp. T7-7, a cold-tolerant diesel oil-degrading bacterium isolated from the Bohai Sea in China. *J. Bacteriol.*, 193(15), 4021–4022.
- Cock, P. J., Antao, T., Chang, J. T., Chapman, B. A., Cox, C. J., Dalke, A., . . . Wilczynski, B. (2009). Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics*, 25(11), 1422–1423.
- Colwell, R. R., Walker, J. D., & Cooney, J. J. (1977). Ecological aspects of microbial degradation of petroleum in the marine environment. *CRC Cr. Rev. Microbiol.*, 5(4), 423–445.
- Delille, D. (2000). Response of Antarctic soil bacterial assemblages to contamination by diesel fuel and crude oil. *Microb. Ecol.*, 40(2), 159–168.
- Delille, D., Basseres, A., & Dessommes, A. (1996). Seasonal variation of bacterial in sea ice contaminated by diesel fuel and dispersed crude oil. *Microb. Ecol.*, 33, 97–105.
- Eddy, S. R. (1998). Profile hidden Markov models. *Bioinformatics*, 14(9), 755–763.
- Feng, L., Wang, W., Cheng, J., Ren, Y., Zhao, G., Gao, C., . . . Peng, X. (2007). Genome and proteome of long-chain alkane degrading *Geobacillus thermodenitrificans* NG80-2 isolated from a deep-subsurface oil reservoir. *Proc. Natl. Acad. Sci.*, 104(13), 5602–5607.

- Gerdes, B., Brinkmeyer, R., Dieckmann, G., & Helmke, E. (2005). Influence of crude oil on changes of bacterial communities in Arctic sea-ice. *FEMS Microb. Ecol.*, 53(1), 129–139. doi: 10.1016/j.femsec.2004.11.010
- Glieder, A., Farinas, E. T., & Arnold, F. H. (2002). Laboratory evolution of a soluble, self-sufficient, highly active alkane hydroxylase. *Nat. Biotech.*, 20(11), 1135–1139.
- Gutierrez, T., Berry, D., Yang, T., Mishamandani, S., McKay, L., Teske, A., & Aitken, M. D. (2013). Role of bacterial exopolysaccharides (EPS) in the fate of the oil released during the deepwater horizon oil spill. *PLoS ONE*, 8(6), e67717.
- Harford-Cross, C. F., Carmichael, A. B., Allan, F. K., England, P. A., Rouch, D. A., & Wong, L.-L. (2000). Protein engineering of cytochrome P450cam (CYP101) for the oxidation of polycyclic aromatic hydrocarbons. *Protein Eng.*, 13(2), 121–128.
- Harvilla, P. B., Wolcott, H. N., & Magyar, J. S. (2014). The structure of ferricytochrome c 552 from the psychrophilic marine bacterium *Colwellia psychrerythraea* 34H. *Metallomics*, 6, 1126–1130.
- Hazen, T. C., Dubinsky, E. A., DeSantis, T. Z., Andersen, G. L., Piceno, Y. M., Singh, N., . . . Mason, O. U. (2010). Deep-sea oil plume enriches indigenous oil-degrading bacteria. *Science*, 330(6001), 204–208. doi: 10.1126/science.1195979
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scand. J. Stat.*, 6, 65–70.
- Ikai, A. (1980). Thermostability and aliphatic index of globular proteins. *J. Biochem.*, 88(6), 1895–1898.
- Ji, Y., Mao, G., Wang, Y., & Bartlam, M. (2013). Structural insights into diversity and n-alkane biodegradation mechanisms of alkane hydroxylases. *Frontiers in Microbiology*, 4.
- Kelley, L. A., & Sternberg, M. J. (2009). Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protocols*, 4(3), 363–371.
- Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.*, 16(2), 111–120. doi: 10.1007/bf01731581
- Krembs, C., Eicken, H., Junge, K., & Deming, J. W. (2002). High concentrations of exopolymeric substances in Arctic winter sea ice: implications for the polar ocean carbon cycle and cryoprotection of diatoms. *Deep Sea Res. Part I*, 49(12), 2163–2181.
- Leiros, H.-K. S., Pey, A. L., Innselset, M., Moe, E., Leiros, I., Steen, I. H., & Martinez, A. (2007). Structure of phenylalanine hydroxylase from *Colwellia psychrerythraea* 34H, a monomeric cold active enzyme with local flexibility around the active site and high overall stability. *J. Biol. Chem.*, 282(30), 21973–21986.

- Margesin, R., Labbe, D., Schinner, F., Greer, C., & Whyte, L. (2003). Characterization of hydrocarbon-degrading microbial populations in contaminated and pristine alpine soils. *Appl. Environ. Microbiol.*, 69(6), 3085–3092.
- Mason, O., Han, J., Woyke, T., & Jansson, J. (2014a). Single-cell genomics reveals features of a *Colwellia* species that was dominant during the Deepwater Horizon oil spill. *Frontiers in Microbiology*, 5, 332.
- Mason, O., Han, J., Woyke, T., & Jansson, J. (2014b). Single-cell genomics reveals features of *Colwellia* species dominant during the Deepwater Horizon oil spill. *Frontiers in Microbiology*, 5(332).
- McGuffin, L. J., Bryson, K., & Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), 404–405.
- Nie, Y., Chi, C.-Q., Fang, H., Liang, J.-L., Lu, S.-L., Lai, G.-L., . . . Wu, X.-L. (2014). Diverse alkane hydroxylase genes in microorganisms and environments. *Sci. Reports*, 4.
- Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'Hara, R. B., . . . Wagner, H. (2012). *vegan: Community Ecology Package*. Retrieved from <http://CRAN.R-project.org/package=vegan>
- Pelé, J., Abdi, H., Moreau, M., Thybert, D., & Chabbert, M. (2011). Multidimensional scaling reveals the main evolutionary pathways of class A G-protein-coupled receptors. *PLoS ONE*, 6(4), e19094. doi: 10.1371/journal.pone.0019094
- Powell, S., Bowman, J., & Snape, I. (2004). Degradation of nonane by bacteria from Antarctic marine sediment. *Pol. Biol.*, 27(10), 573–578.
- Price, M., Dehal, P., & Arkin, A. (2010). FastTree 2 - Approximate maximum likelihood trees for large alignments. *PLOS one*, 5(3), e9490.
- Punta, M., Coggill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., . . . Finn, R. D. (2012). The Pfam protein families database. *Nuc. Acids Res.*, 40(D1), D290–D301. doi: 10.1093/nar/gkr1065
- Redmond, M. C., & Valentine, D. L. (2012). Natural gas and temperature structured a microbial community response to the Deepwater Horizon oil spill. *P. Natl. Acad. Sci.*, 109(50), 20292–20297.
- Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., . . . Söding, J. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.*, 7(1).
- Throne-Holst, M., Wentzel, A., Ellingsen, T. E., Kotlar, H.-K., & Zotchev, S. B. (2007). Identification of novel genes involved in long-chain n-alkane degradation by *Acinetobacter* sp. strain DSM 17874. *Appl. Environ. Microbiol.*, 73(10), 3327–3332.

- Trotsenko, Y. A., & Khmelenina, V. N. (2002). Biology of extremophilic and extremotolerant methanotrophs. *Arch. Microbiol.*, 177(2), 123–131.
- Valentine, D. L., Kessler, J. D., Redmond, M. C., Mendes, S. D., Heintz, M. B., Farwell, C., . . . Villanueva, C. J. (2010). Propane respiration jump-starts microbial response to a deep oil spill. *Science*, 330(6001), 208–211. doi: 10.1126/science.1196830
- Van Beilen, J. B., Li, Z., Duetz, W. A., Smits, T. H., & Witholt, B. (2003). Diversity of alkane hydroxylase systems in the environment. *Oil Gas Sci. Technol.*, 58(4), 427–440.
- Wentzel, A., Ellingsen, T., Kotlar, H.-K., Zotchev, S., & Throne-Holst, M. (2007). Bacterial metabolism of long-chain n-alkanes. *Appl. Microbiol. Biotechnol.*, 76(6), 1209–1221. doi: 10.1007/s00253-007-1119-1
- Whyte, L. G., Bourbonniere, L., & Greer, C. W. (1997). Biodegradation of petroleum hydrocarbons by psychrotrophic *Pseudomonas* strains possessing both alkane (alk) and naphthalene (nah) catabolic pathways. *Appl. Environ. Microbiol.*, 63(9), 3719–3723.
- Whyte, L. G., Hawari, J., Zhou, E., Bourbonnière, L., Inniss, W. E., & Greer, C. W. (1998). Biodegradation of variable-chain-length alkanes at low temperatures by a psychrotrophic *Rhodococcus* sp. *Appl. Environ. Microbiol.*, 64(7), 2578–2584.

## Tables and Figures

**Table 7.1.** Occurrence in each dataset of conserved protein family (pfam) domains linked to alkane hydroxylases (AH).

<b>Pfam</b>	<b>AH example</b>	<b>Psychrophile</b>	<b>Mesophile</b>	<b>UniProt</b>	<b>AH candidate, psychrophile</b>	<b>AH candidate, mesophile</b>
AMO	PMMO subunit A	0	0	48	0	0
AmoC	PMMO subunit C	0	0	63	0	0
Bac_luciferase	LadA	18	51	7	3	4
FA_desaturase	AlkB	32	40	320	8	10
MeMO_Hyd_G	SMMO subunit G	0	0	17	0	0
Monooxygenase_B	PMMO subunit B	0	0	46	0	0
p450	p450	11	10	145	9	9
Phenol_Hydrox	PMO small subunit	0	0	164	0	0
Pyr_redox_3	AlmA	176	196	35	6	18

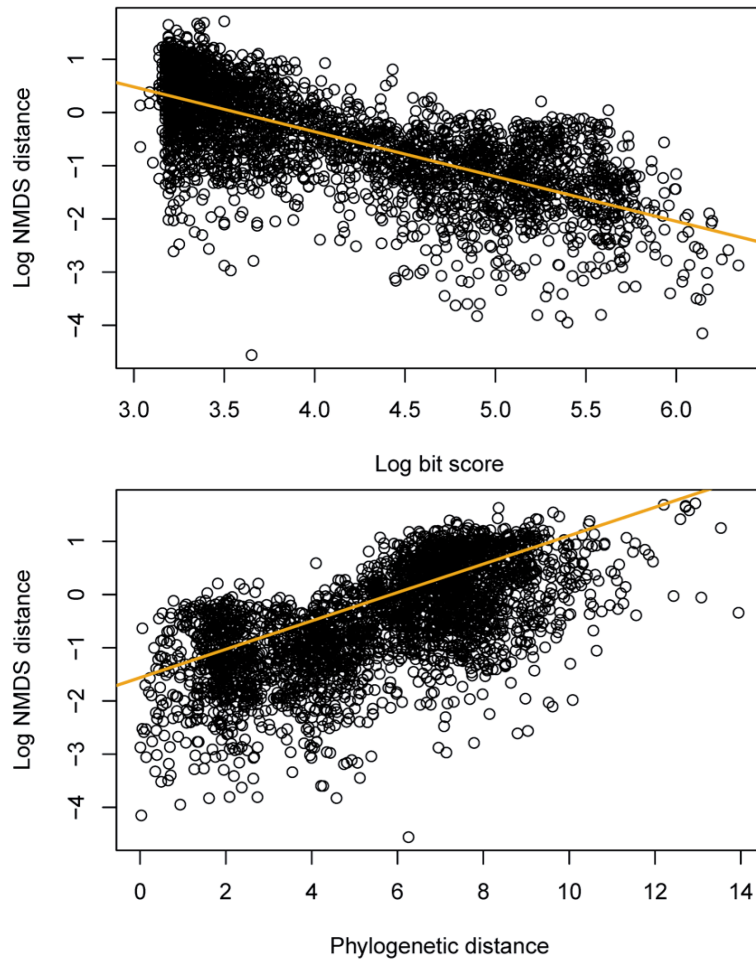


**Table 7.2.** Number of candidate alkane hydroxylases observed in each of the psychrophile and mesophile genomes examined. Pair indicates the taxonomic pairing used for the pairwise analysis. Strain with at least one candidate alkane hydroxylase are shown in bold.

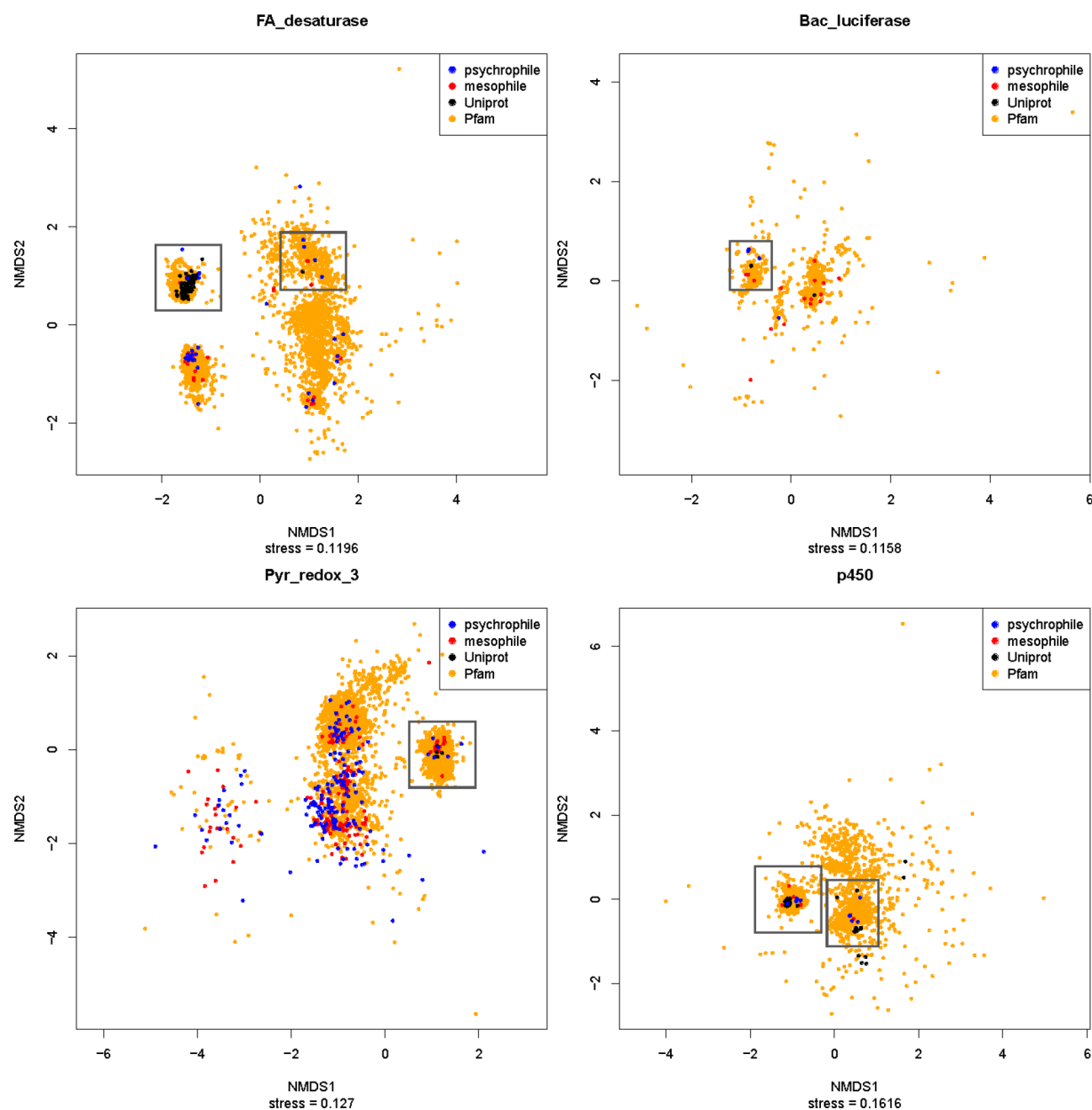
Strain	Pair	Accession	Bac luciferase	Pyr_redox 3	FA_desaturase cluster 0	FA_desaturase cluster 0	p450 cluster 0	p450 cluster 1
<b>Psychrophiles</b>								
<i>Aeromonas salmonicida</i> A449	A	CP000644	0	0	0	0	0	0
<i>Aliivibrio salmonicida</i> LFH1238	B	FM178379	0	0	0	0	0	0
<i>Colwellia psychrerythraea</i> 34H	C	CP000083	0	0	0	0	0	0
<i>Desulfotalea psychrophila</i> LSv54	D	CR522870	0	0	0	0	0	0
<i>Flavobacterium psychrophilum</i> JIP02 86	E	AM398681	0	0	0	0	0	0
<b><i>Glaciecola psychrophila</i> 170</b>	<b>F</b>	<b>CP003837</b>	<b>0</b>	<b>5</b>	<b>0</b>	<b>1</b>	<b>1</b>	<b>0</b>
<i>Methanococcoides burtonii</i> DSM 6242	G	CP000300	0	0	0	0	0	0
<b><i>Octadecabacter antarcticus</i> 307</b>	<b>H</b>	<b>CP003740</b>	<b>1</b>	<b>0</b>	<b>2</b>	<b>2</b>	<b>1</b>	<b>1</b>
<b><i>Octadecabacter arcticus</i> 238</b>	<b>I</b>	<b>CP003742</b>	<b>1</b>	<b>0</b>	<b>2</b>	<b>1</b>	<b>1</b>	<b>1</b>
<b><i>Photobacter profundus</i> SS9</b>	<b>J</b>	<b>CR354532</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
<i>Pseudoalteromonas haloplanktis</i> TAC125	K	CR954246	0	0	0	0	0	0
<i>Psychrobacter arcticum</i> 273-4	L	CP000082	0	0	0	0	0	0
<b><i>Psychrobacter cryohalolentis</i> K5</b>	<b>M</b>	<b>CP000323</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b><i>Psychroflexus torquis</i> ATCC700755</b>	<b>N</b>	<b>CP003879</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
<i>Psychromonas</i> CNPT3	O	CP004404	0	0	0	0	0	0
<i>Psychromonas ingrahamii</i> 37	P	CP000510	0	0	0	0	0	0
<i>Shewanella halifaxensis</i> HAW EB4	Q	CP000931	0	0	0	0	0	0
<i>Shewanella sediminis</i> HAW-EB3	R	CP000821	0	0	0	0	0	0
<i>Shewanella violacea</i> DSS12	S	AP011177	0	0	0	0	0	0
<b><i>Terroglolus saanensis</i> SP1PR4</b>	<b>T</b>	<b>CP002467</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>2</b>	<b>1</b>
<b>Mesophiles</b>								
<i>Aeromonas veronii</i> B565	A	CP002607	0	0	0	0	0	0
<i>Vibrio fischeri</i> ES114	B	CP000020	0	0	0	0	0	0
<b><i>Alteromonas macleodii</i> Deep ecotype</b>	<b>C</b>	<b>CP001103</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<i>Desulfocapsa sulfexigens</i> DSM 10523	D	CP003985	0	0	0	0	0	0
<i>Flavobacterium indicum</i> GPTSA100 9	E	HE774682	0	0	0	0	0	0
<b><i>Glaciecola agaritica</i> 4H37Ye5</b>	<b>F</b>	<b>CP002526</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>3</b>	<b>0</b>
<i>Methanosarcina mazei</i> Tuc01	G	CP004144	0	0	0	0	0	0
<b><i>Rhodobacter sphaeroides</i> ATCC 17025</b>	<b>H</b>	<b>CP000661</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>1</b>	<b>0</b>
<b><i>Ketogulonicigenium vulgare</i> Y25</b>	<b>I</b>	<b>CP002224</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>
<i>Vibrio vulnificus</i> YJ016	J	AP005352	0	0	0	0	0	0
<b><i>Pseudoalteromonas atlantica</i> T6c</b>	<b>K</b>	<b>CP000388</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>
<i>Acinetobacter baumannii</i> ACICU	L	CP000863	2	4	1	1	0	0
<i>Acinetobacter baumannii</i> AYE	M	CU459141	2	6	1	1	0	0
<b><i>Flavobacteriales bacterium</i> HTCC2170</b>	<b>N</b>	<b>CP002157</b>	<b>0</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b><i>Marinobacter aquaeolei</i> VT8</b>	<b>O</b>	<b>CP000514</b>	<b>0</b>	<b>3</b>	<b>3</b>	<b>1</b>	<b>2</b>	<b>0</b>
<b><i>Alteromonas macleodii</i> English Channel ecotype</b>	<b>P</b>	<b>CP003844</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<i>Shewanella</i> MR-7	Q	CP000444	0	0	0	0	0	0
<b><i>Shewanella denitrificans</i> OS217</b>	<b>R</b>	<b>CP000302</b>	<b>0</b>	<b>1</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<i>Shewanella putrefaciens</i> 200	S	CP002457	0	0	0	0	0	0
<b><i>Terroglolus roseus</i> DSM18391</b>	<b>T</b>	<b>CP003379</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>1</b>

**Table 7.3.** Pairwise parameters for candidate alkane hydroxylase with the p450 and Pyr\_redox\_3 conserved domains. Region refers to: whole; entire protein, C; coil, E;  $\beta$ -sheet, H;  $\alpha$ -helix. Letters in the parameter column are standard abbreviations for amino acids.

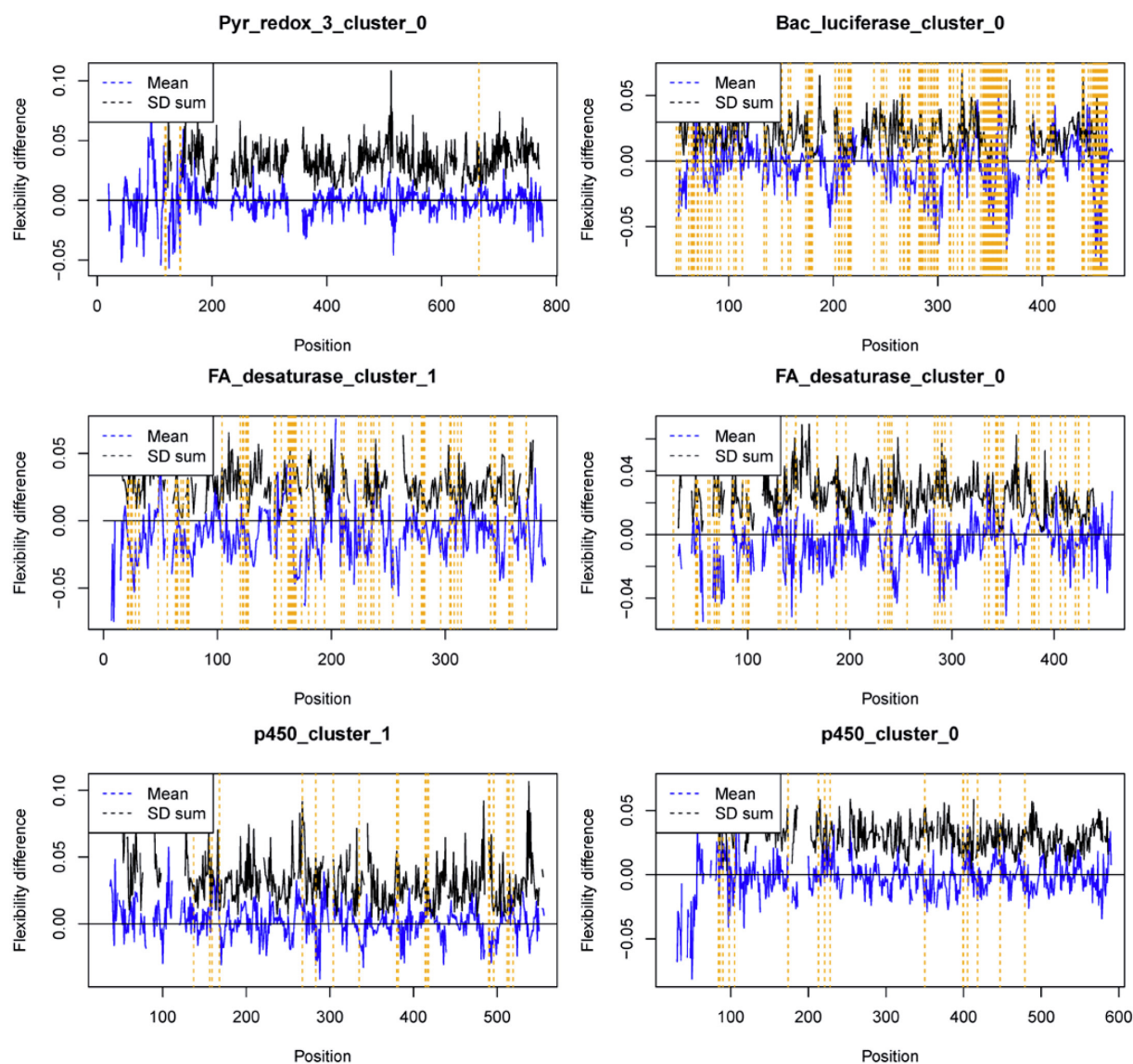
<b>pfam</b>	<b>Region</b>	<b>Parameter</b>	<b>Comparisons where psychrophile is higher</b>	<b>Total comparisons</b>
p450	whole	isoelectric point	9	9
p450	whole	N	1	9
p450	whole	R	9	9
p450	whole	T	1	9
p450	whole	V	0	9
p450	C	I	1	9
p450	C	N	0	9
p450	C	V	0	9
p450	E	flexibility	1	9
p450	E	isoelectric point	1	9
p450	E	A	1	9
p450	E	G	0	9
p450	E	I	8	9
p450	E	P	0	9
p450	H	flexibility	1	9
p450	H	D	0	9
p450	H	I	9	9
p450	H	R	9	9
Pyr_redox_3	whole	C	1	11
Pyr_redox_3	whole	E	11	11
Pyr_redox_3	whole	V	1	11
Pyr_redox_3	C	D	1	11
Pyr_redox_3	C	G	10	11
Pyr_redox_3	E	GRAVY	1	11
Pyr_redox_3	E	C	0	11
Pyr_redox_3	H	A	1	11
Pyr_redox_3	H	E	11	11
Pyr_redox_3	H	N	10	11



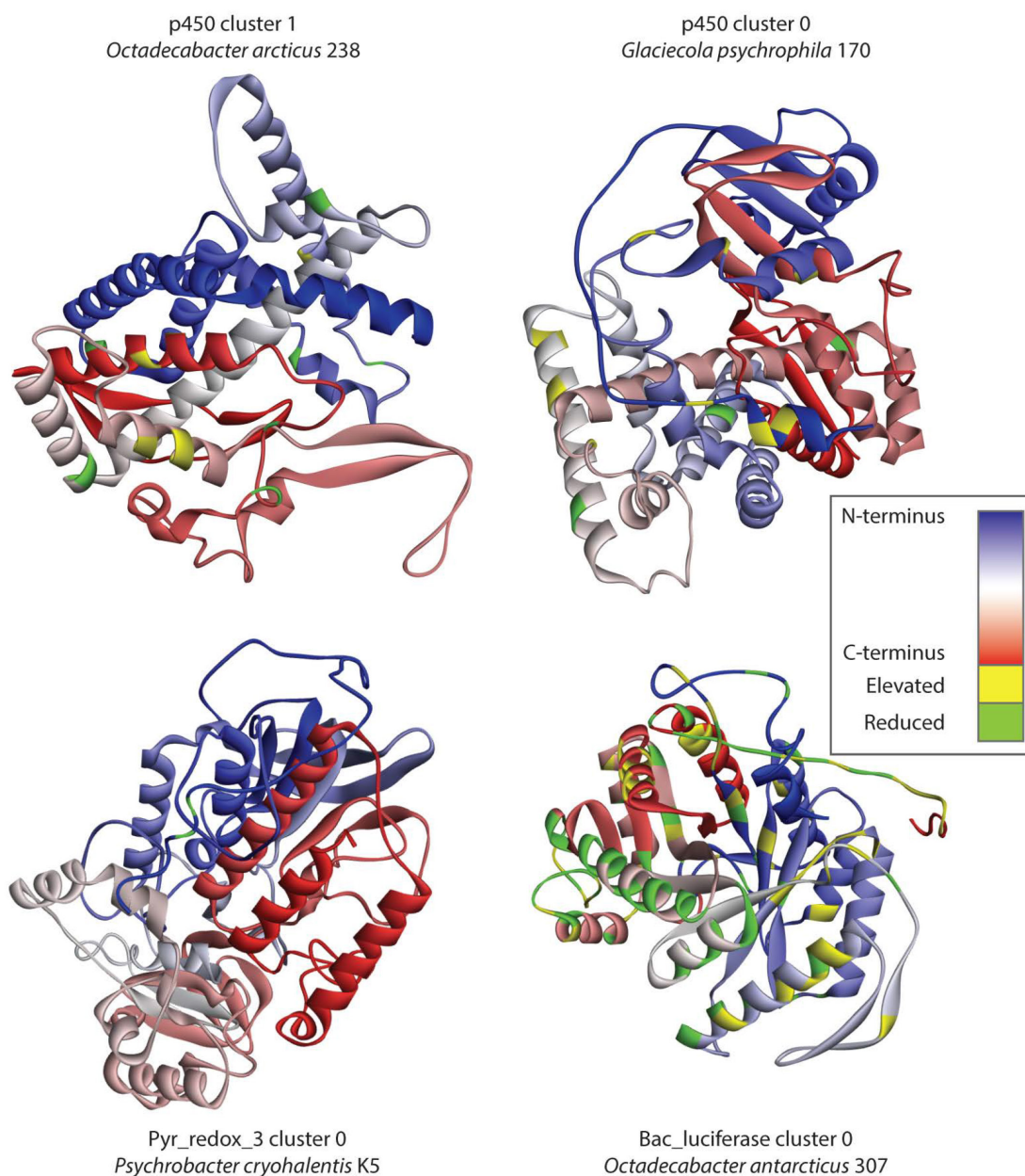
**Fig. 7.1. Euclidean distance in 2D NMDS space as a function of bit score and phylogenetic distance.** Euclidean distance in the FA\_desaturase pfam is strongly correlated with bit score ( $R^2 = 0.4232$ ,  $n = 2,439,512$ ), obtained from reciprocal blast, and with phylogenetic distance ( $R^2 = 0.5029$ ,  $n = 2,439,512$ ), as summed branch lengths from a maximum-likelihood tree. Orange lines are linear models fit to the complete data sets; only 10,000 randomly selected data points are plotted.



**Fig. 7.2. NMDS plots of genetic distance within four protein families (pfams).** The distance between two points on the plot is proportional to their sequence similarity. Clusters of points identified as candidate alkane hydroxylases due to the presence of UniProt alkane hydroxylases are outlined with gray boxes.



**Fig. 7.3. Alignment of the flexibility parameter between putative alkane hydroxylases in psychrophiles and mesophiles.** Center residues for windows with a significant difference in flexibility between psychrophiles and mesophiles are indicated by vertical orange lines. Positive values for mean (blue line) indicate positions in the alignment where the flexibility was greater for the psychrophile proteins. Gaps in the data reflect gaps in the alignment that prevent the calculation of the mean or standard deviation. SD sum in the legend refers to the sum of the standard deviation for each population.



**Fig. 7.4. Predicted 3D structure for psychrophile representatives of the four clusters with high confidence predictions.** Positions indicated in orange in Fig. 7.3 are highlighted as yellow (increased flexibility in psychrophiles) or green (reduced flexibility in psychrophiles).

**Table S7.1.** Genbank accession and annotation for the protein records corresponding to the gene products evaluated in this study. See supplementary file S7.1.pdf.

## Appendix 1

### **Elevated bacterial abundance and exopolymers in saline frost flowers and implications for atmospheric chemistry and microbial dispersal**

(Citation: Bowman, J. S., & Deming, J. W. (2010). Elevated bacterial abundance and exopolymers in saline frost flowers and implications for atmospheric chemistry and microbial dispersal. *Geophysical Research Letters*, 37(13), L13501.)

#### ABSTRACT

Frost flowers (FF) have been studied for their potential influence on ice-surface reflectivity and roles in atmospheric chemistry, but not as microbial habitats. We examined FF grown in a freezer laboratory from a bacteria-containing saline solution and FF formed naturally in the coastal (April) and central Arctic Ocean (September). All FF contained bacteria (up to  $3.46 \times 10^6 \text{ ml}^{-1}$  in natural FF) with densities 3–6-fold higher than in underlying ice. Bacterial abundance correlated strongly with salinity in FF ( $p$  values  $\leq 0.001$ ), a correlation that held for all components of the surface-ice environment ( $p < 0.0001$ , coastal samples). Concentrations of extracellular polysaccharides were also elevated in FF and brine skim relative to underlying ice (up to 74-fold higher). Here we consider implications of finding microbes and exopolymers within the chemically reactive surface-ice environment to the photolytic production of oxidants and long-range transport of potential ice-nucleating particles in the atmosphere.

## A1.1 Introduction

Frost flowers (FF), centimeter-scale structures composed of atmospheric hoar and liquid derived from sea ice brines, grow abundantly on newly formed sea ice (Alvarez-Aviles et al., 2008; Perovich & Richter-Menge, 1994). They have been the focus of numerous recent investigations due to their ability to transport salts (and possibly other materials) concentrated in sea ice brines to the atmosphere. This transport mechanism, combined with the high specific surface area of FF (but see Domine et al. (2005)), may play a significant role in tropospheric ozone depletion events common during spring polar sunrise (Alvarez-Aviles et al., 2008; Rankin, Wolff, & Martin, 2002). Elevated mercury concentrations have also been measured in FF, exceeding that of coastal snow 9-fold (Douglas et al., 2005).

FF are ephemeral, existing exposed on the ice surface for hours to days. Their fate is burial by snow or erosion by wind (Perovich & Richter-Menge, 1994). Aerosol material can be sourced to FF by a higher ratio of Na to  $\text{SO}_4$  than found in (unfrozen) seawater. This signature is the result of mirabilite ( $\text{Na}_2\text{SO}_4 \cdot 10 \text{H}_2\text{O}$ ) precipitation at  $-8^\circ\text{C}$ , near the upper temperature limit of FF growth. Using this tracer, FF have been identified as a significant source of salt to glacial ice (Rankin et al., 2002).

Despite their location at the frozen interface between ocean and atmosphere and their demonstrated role in concentrating and transporting sea salts, the biology of FF has not been investigated. We reasoned that FF, in wicking brine from the underlying ice or surface brine skim (a thin film of brine expelled upwards from the ice matrix), must also bring sea ice bacteria (Bacteria and Archaea (R. E. Collins, Rocap, & Deming, 2010)) into their structures, given that bacteria concentrate in the liquid brine phase (Junge, Krembs, Deming, Stierle, & Eicken, 2001). A growing body of work on the microbiology of winter sea ice suggests that bacteria can remain



active in carbon cycling (Junge, Eicken, & Deming, 2004; Wells & Deming, 2006) to temperatures as low as  $-26^{\circ}\text{C}$  (R. Eric Collins, Carpenter, & Deming, 2008) in part due to the presence of cryoprotectant exopolymers (EPS), organic exudates produced by bacteria and phytoplankton (R. Eric Collins et al., 2008; Krembs, Eicken, Junge, & Deming, 2002; Marx, Carpenter, & Deming, 2009). Experiments with microorganisms from other frozen environments suggest the persistence of metabolism to  $-35^{\circ}\text{C}$  (Panikov & Sizova, 2007), leaving open the possibility for microbial activity in very cold FF. In the atmosphere, when temperatures are low enough to suppress metabolic activity, bacteria from a variety of environments can still influence physical processes by functioning as ice nucleators (Christner, Morris, Foreman, Cai, & Sands, 2008; Jayaweera & Flanagan, 1982).

Motivated by preliminary observations of FF during the overwintering CFL-IPY 2007–2008 expedition (Deming, 2010), we established a system for growing FF in the laboratory and collected natural FF from coastal and central regions of the Arctic Ocean, analyzing for bacterial content and salinity. By collecting other features of the ice-surface environment, we sought to broaden the emerging relationship between bacterial and salt content in FF. At our coastal site we also analyzed for EPS, anticipating that biogenic materials other than intact microbial cells are wicked into FF. Finding elevated concentrations of bacteria and exopolymers in FF and brine skim has important implications for the transport of biogenic material from sea ice to atmosphere and for possible chemical reactivity.

## **A1.2 Methods**

FF were grown in a laboratory freezer room at  $-21^{\circ}\text{C}$  in a manner similar to that described by Style (2007). A NaCl solution of 35 ppt, amended with bacteria (to  $2 \times 10^6 \text{ ml}^{-1}$ ), was used instead of seawater. To ensure passive transport of the bacteria, a mesophilic bacterium

(*Halomonas pacifica*) incapable of growth or motility under the test conditions was used.

Individual FF and brine skim were harvested from the ice surface using an ethanol-sterilized spatula. Ice was sampled by removing a small section (the top 1–2 cm) with an ethanol-sterilized chisel. The samples were monitored while melting at room temperature (1–2 min); as the last ice crystals melted, subsamples were immediately fixed to 2 % v/v with paraformaldehyde.

Salinity was determined on unfixed aliquots of the melted samples using a handheld refractometer. Although an apparent departure from the convention for seawater, reporting in these units is made necessary by use of a refractometer (calibrated against ppt) and by sample salinities above the linear range for which the practical salinity scale is defined (Williams & Sherwood, 1994). Bacteria were enumerated by epifluorescence microscopy using the DNA-specific stain, 4',6-diamidino-2-phenylindole, counterstained with acridine orange. A minimum of 20 fields and 200 cells were counted for each sample. FF age was determined by time-lapse photography using a Nikon D70s camera controlled by a laptop running the Nikon Capture software outside the cold room. Photographs were taken at 1-min intervals throughout a typical experiment of 24–48 h. A bacterial enrichment index ( $I_B$ ) (Gradinger & Ikävalko, 1998) was calculated to determine the transport efficiency of bacteria into ice and FF (or back into solution) relative to salt. Significance of the mean deviation from  $I_B = 1$  (indicating equal incorporation of salt and bacteria) was determined using a Student's T-test.

Individual FF, brine skim, sea ice, and other components of the surface ice environment were collected from shore-fast sea ice near Barrow, Alaska, on April 6 and 7, 2009. Respective daily air temperatures during sampling were  $-23^{\circ}\text{C}$  and  $-32^{\circ}\text{C}$ . Tidal action replenished surface ponds over this ice surface, allowing growth of new FF each night on a saline surface ice layer.

The FF and all other surface ice samples were collected using an ethanol-sterilized spatula. Brine skim was sampled beneath each FF, followed by the upper 1–2 cm of underlying sea ice. Other samples included older FF and brine skim, buried by 5 cm snow (estimated age of 1 week), FF from atop a pressure ridge, and brine (collected in a sterile bottle) from a naturally occurring fissure in sea ice. Samples for bacterial counts were treated as for the laboratory grown FF, except that formaldehyde was used to fix the cells. Salinity was determined in the same manner as for laboratory grown FF. Samples for particulate EPS (pEPS) were transported frozen to the University of Washington in Seattle, where they were melted as before and filtered through a 0.4  $\mu\text{m}$  polycarbonate filter. The concentration of pEPS was determined using the phenol-sulfuric acid (PSA) assay previously described (Marx et al., 2009).

Individual FF were also collected from lead ice in the central Arctic Ocean during the LOMROG II expedition of the Swedish icebreaker *Oden* on September 2, 2009, at 84.84°N latitude and 14.67°E longitude. The air temperature was  $-8^{\circ}\text{C}$ . Samples were melted (1:1 v/v) in 0.2  $\mu\text{m}$ -filtered artificial seawater comprised of the major sea salts (35 ppt) and containing 4 % formaldehyde (post-melting concentration of 2 %) to minimize cell loss to osmotic shock during melting. A portion of each FF was melted separately for determining salinity by refractometer. Bacterial counts were conducted as for laboratory-grown FF and corrected for the artificial seawater dilution.

### **A1.3 Results**

In laboratory experiments, the abundance of *H. pacifica* cells in the saline ice that formed ranged from  $7.64 \times 10^5$  to  $1.13 \times 10^6 \text{ ml}^{-1}$ , while in FF growing on top of the ice it was higher, ranging from  $1.15 \times 10^6$  to  $5.56 \times 10^6 \text{ ml}^{-1}$  (Fig. A1.1A). In field-collected FF (Fig. A1.1B, C), the natural bacterial abundance was also elevated relative to underlying sea ice, ranging from

$1.28 \times 10^5 \text{ ml}^{-1}$  (Barrow) to  $3.46 \times 10^6 \text{ ml}^{-1}$  (central Arctic Ocean) in FF compared to  $0.28\text{--}3.83 \times 10^5 \text{ ml}^{-1}$  in underlying sea ice (Barrow, Fig. A1.1B). In all cases, bacterial abundance in FF correlated strongly with salinity (Fig. A1.1A-C). The correlation remained strong when other components of the surface sea ice environment were included (Fig. A1.1B). For cases where comparable data were obtained, pEPS concentration though variable was always higher in FF and brine skim, reaching maxima of 725 and 1,420  $\mu\text{g}$  glucose equivalents (gluceq)  $\text{ml}^{-1}$ , respectively, than in the underlying sea ice with its maximum of 36.5  $\mu\text{g}$  gluceq  $\text{ml}^{-1}$  (Fig. A1.2). The concentration of pEPS did not correlate significantly with either salinity or bacterial abundance.

Although prior work has shown that FF become increasingly saline as they age (Martin, Drucker, & Fort, 1995; Perovich & Richter-Menge, 1994), we did not find significant relationships between FF age and salinity ( $df = 12$ ,  $r = 0.10$ ,  $p = 0.73$ ) or bacterial abundance ( $df = 12$ ,  $r = 0.37$ ,  $p = 0.19$ ) in our relatively short laboratory experiments. Correlations between FF age, salinity and bacterial abundance appeared likely in our field data (Fig A1.2B), but age-based sampling was insufficient to test for these relationships. Calculated  $I_B$  for laboratory-grown FF showed variability in the efficiency of bacterial transport into FF relative to salt transport, but departure of the mean of these values from  $I_B = 1.0$  was not statistically significant (mean = 0.99,  $df = 22$ ,  $t = 0.306$ ,  $p = 0.76$ ), suggesting that bacteria were not preferentially enriched over salt in FF or brine skim.

#### **A1.4 Discussion**

Finding elevated bacterial abundance and high concentrations of pEPS in saline FF identifies a new icy habitat to explore biologically, with potentially important ramifications for atmospheric chemistry. For example, bacterial cells and pEPS may serve as substrate for the

photolytic production of oxidants and simple organic compounds, including formaldehyde and hydrogen peroxide. The production of these compounds in snow has been well studied due to the role they play in oxidation reactions within the troposphere (Jacobi et al., 2002).

Concentrations of both formaldehyde (Barret et al., 2009) and hydrogen peroxide (Beine, Pattern, & Anastasio, 2009) are elevated within FF. The mean value for pEPS in FF at Barrow ( $133 \mu\text{g glucose ml}^{-1}$ ) represents  $4.44 \times 10^{-6} \text{ mol C ml}^{-1}$ , sufficient carbon to support the production (assuming complete conversion) of  $133 \mu\text{g ml}^{-1}$  of HCHO (typically measured in the  $\text{pg ml}^{-1}$  range in snow and marine waters). Other types of organic compounds beside pEPS can also be expected within saline FF, given high levels of DOC (Thomas, Lara, Eicken, Kattner, & Skoog, 1995) and the potential for viral lysis of bacterial cells (Wells & Deming, 2006) in winter sea ice brines. Though not yet pursued, our preliminary observations of winter FF have revealed the presence of viruses (Deming, 2010).

Microbial activity was not measured in this study, but recent work has demonstrated aerobic metabolism within the top centimeter of early spring sea ice (similar to our Barrow ice) correlated to the presence of pEPS (Meiners, Krembs, & Gradinger, 2008). Whether this relationship holds for FF remains to be determined, but the presence of both pEPS and bacteria in elevated concentrations in FF is promising. Depending on type of activity, measurements of microbial activity in FF under severe winter conditions could define new (lower) temperature limits for the activity.

Considering the ultimate fate of bacteria within FF raises other important issues. If buried and insulated by snow, the microbial community originating from FF may be more dynamic than when exposed to the atmosphere, undergoing microbial succession (species-specific growth and mortality) within a saline snow layer on the surface of the ice. This

succession may resemble that understood to occur within the ice (R. E. Collins et al., 2010; Deming, 2010) and even influence underlying sea ice biology and chemistry via infiltration during warming (Brinkmeyer, Glöckner, Helmke, & Amann, 2004).

A different fate awaits microbes contained in FF that erode from the ice surface by wind. Biological materials in the atmosphere, collectively known as biogenic aerosols, have been investigated increasingly for their roles in atmospheric chemistry and cloud properties. Biogenic aerosols over the central Arctic Ocean contain bacteria and EPS (and associated viruses), considered to be cloud condensation nuclei (Leck & Bigg, 2005). Ice-nucleating bacteria have been ascribed an important role in some snowfall events (Christner et al., 2008). Although the impact of biological ice nucleation increases with temperature (decreasing the importance of ice nucleators derived from FF during Arctic and Antarctic winters), the upper temperature for FF growth (about  $-8^{\circ}\text{C}$ ) is warm enough to allow a role during early fall and late spring. Aerosol collections by aircraft over the Arctic Ocean suggest that biogenic aerosols can be transported substantial distances (Jayaweera & Flanagan, 1982). Wind erosion of FF filled with microbes and EPS may represent a unique dispersal mechanism for the organisms and their byproducts. If FF bacteria can function as ice nuclei, their long range dispersal becomes a means of delivering ice nucleators to lower latitudes, where warmer temperatures increase the significance of these particles to precipitation events.

A more detailed analysis of the possible fate of bacteria in wind-eroded FF can be made where ion concentrations have been analyzed within glacial ice. In Antarctic glacial ice as in Arctic snow, saline FF have been recognized as a significant source of sea salt through their characteristic signature of  $\text{SO}_4$  depletion relative to Na (Beaudon & Moore, 2009; Wolff, Rankin, & Rathlisberger, 2003). Using an average value for FF salinity, the abundance of Na within the

Dome C ice core gives an approximation of the flux of FF material to the interior of the Antarctic continent. With our measured bacterial abundances in FF and assuming bacterial transport coincident with salt, the flux of FF bacteria into glacial ice can be estimated. We take 100 ppt as a typical salinity for saline FF (mean for Barrow FF = 99.2,  $n = 12$ ,  $SD = 25.1$ ), which agrees well with previous studies (Martin et al., 1995; Martin, Yu, & Drucker, 1996; Perovich & Richter-Menge, 1994), and  $5 \times 10^5$  bacteria  $\text{ml}^{-1}$  as a typical value for bacterial abundance (mean for Barrow FF =  $4.8 \times 10^5$   $\text{ml}^{-1}$ ,  $n = 12$ ,  $SD = 2.8 \times 10^5$ ). Na flux varies with FF production, approaching  $500 \mu\text{g m}^{-2} \text{yr}^{-1}$  in recent times (Wolff et al., 2003). In a FF entirely depleted in  $\text{SO}_4$ , Na accounts for 27 % of the mass of all ions. Thus a Na flux of  $500 \mu\text{g m}^{-2} \text{yr}^{-1}$  is equivalent to  $1,850 \mu\text{g m}^{-2} \text{yr}^{-1}$  total ions. For FF of 100 this represents a mass of  $2.0 \times 10^{-5}$  kg, or 0.018 ml. This volume of FF would contain 8,900 bacteria for a flux on the order of 8,900 FF bacteria  $\text{m}^{-2} \text{yr}^{-1}$  to Dome C.

Modern molecular techniques may be able to detect this flux of bacteria from the marine environment to the continental interior. Although the bulk salinity of glacial ice is substantially lower than that of sea ice (and the interior brine volume substantially smaller), eutectic freezing does allow the presence of liquid films in glacial ice. Furthermore, the water activity ( $A_w$ ) of these veins will reflect the  $A_w$  of brine inclusions in sea ice of the same temperature (though the concentration of solutes and pH will differ). This similarity suggests that saline FF bacteria may be able to resist lysis due to osmotic stress within glacial ice. Total bacterial abundance in bulk glacial ice is on the order of  $10^2$ – $10^3$  bacteria  $\text{ml}^{-1}$  (Mader, Pettitt, Wadham, Wolff, & Parkes, 2006), the majority of which are assumed to be delivered with aeolian dust in snow. Given an average snow accumulation rate of  $3.6 \text{ g cm}^{-2} \text{yr}^{-1}$  at Dome C (Petit, Jouzel, Pourchet, & Merlivat, 1982) and our estimate of FF bacterial flux to the site, bacteria from FF would account

for one cell in every 5 ml of bulk ice. If, as in sea ice, bacteria partition into the liquid phase of glacial ice, a concentration factor between  $10^4$  and  $10^5$  would pertain (Mader et al., 2006) resulting in between  $2 \times 10^3$  and  $2 \times 10^4$  FF bacteria  $\text{ml}^{-1}$  of glacial brine.

### **A1.5 Conclusions**

FF represent a critical link between materials contained within sea ice brines and the atmosphere. Although previous studies have investigated the importance of interactions between inorganic and presumably abiogenic materials concentrated within FF and the atmosphere, the biology of FF has been overlooked until now. We have demonstrated that bacteria and exopolymers are concentrated in saline FF relative to other components of the surface sea ice environment. This concentration within FF may allow bacteria to be involved directly in a number of atmospheric processes, including photochemistry and long-range transport by wind. A biogeochemical role for FF bacteria will depend in part on the ability of these organisms to metabolize at very low temperatures and low  $A_w$ , providing incentive to characterize the microbial community and assess metabolic activity within FF.

### **Acknowledgements**

We thank Hans-Werner Jacobi for stimulating input and the invitation to sample FF at Barrow, Alaska, Soeren Rysgaard and Christian Marcussen for the invitation to join LOMROGII, Manuel Barret and Matthias Wietz for assistance with field sampling, Shelly Carpenter for laboratory support, and Seelye Martin, Eric Collins, and Marcela Ewert for helpful discussion. This research was supported by NSF-OPP award 0908724 to JWD, NSF-IGERT support to JSB, and the UW Astrobiology Program.



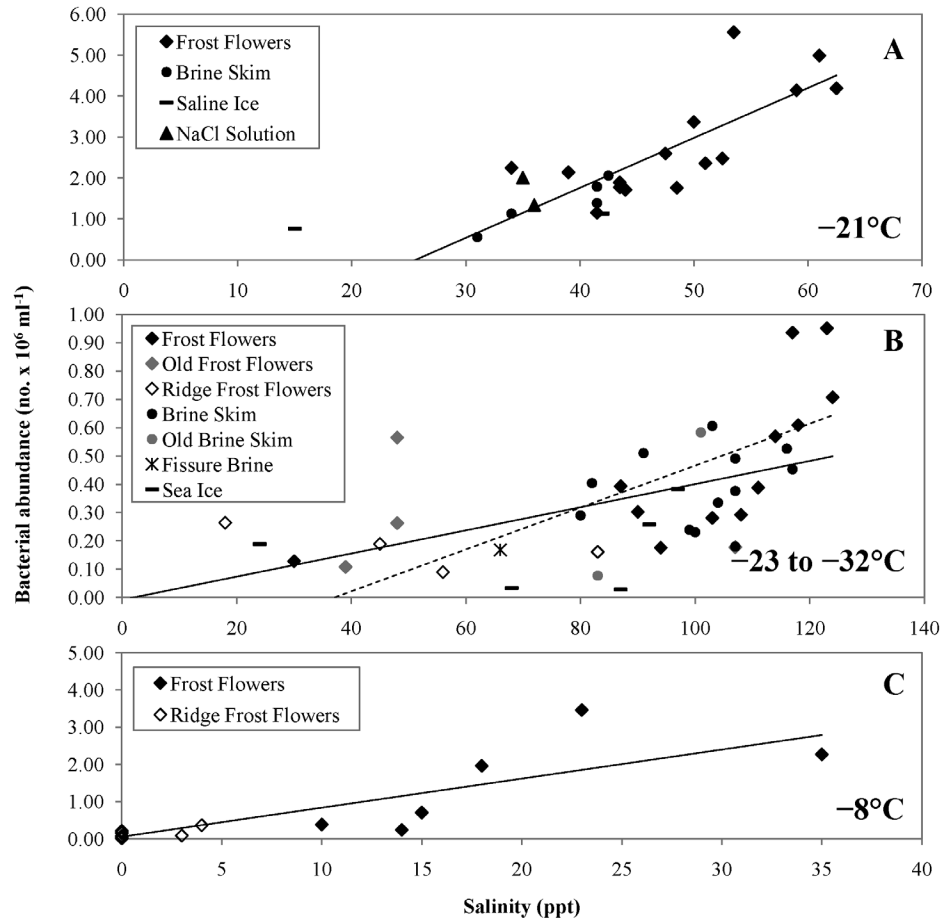
## References

- Alvarez-Aviles, L., Simpson, W. R., Douglas, T. A., Sturm, M., Perovich, D., & Domine, F. (2008). Frost flower chemical composition during growth and its implications for aerosol production and bromine activation. *J. Geophys. Res.*, *113*. doi: 10.1029/2008jd010277
- Barret, M., Houdier, S., Gallet, J.-C., Domine, F., Beine, H., Jacobi, H.-W., . . . Richter, D. (2009). Aldehydes in Arctic snow at Barrow (AK) during Barrow 2009 field campaign. *Eos*, *90*(52), Fall Meet. Suppl.
- Beaudon, E., & Moore, J. (2009). Frost flower chemical signature in winter snow on Vestfonna ice cap (Nordaustlandet, Svalbard). *The Cryos. Discuss.*, *3*(1), 159–180.
- Beine, H., Pattern, K., & Anastasio, C. (2009). Absorption spectra of surface snow, sea ice, and frost flowers. *Eos*, *90*(52), Fall Meet. Suppl.
- Brinkmeyer, R., Glöckner, F.-O., Helmke, E., & Amann, R. (2004). Predominance of  $\beta$ -Proteobacteria in summer melt pools on Arctic pack ice. *Limnol. and Oceanogr.*, *49*(4), 1013–1021.
- Christner, B. C., Morris, C. E., Foreman, C. M., Cai, R., & Sands, D. C. (2008). Ubiquity of biological ice nucleators in snowfall. *Science*, *319*(5867), 1214. doi: 10.1126/science.1149757
- Collins, R. E., Carpenter, S. D., & Deming, J. W. (2008). Spatial heterogeneity and temporal dynamics of particles, bacteria, and pEPS in Arctic winter sea ice. *J. Mar. Systems*, *74*(3–4), 902–917.
- Collins, R. E., Rocap, G., & Deming, J. W. (2010). Persistence of bacterial and archaeal communities in sea ice through an Arctic winter. *Environ. Microbiol.*, *12*(7), 1828–1841.
- Deming, J. W. (2010). Sea ice bacteria and viruses. In D. N. Thomas & G. S. Deickmann (Eds.), *Sea Ice - An introduction to its Physics, Chemistry, Biology, and Geology* (pp. 247–282). Oxford: Blackwell Science Ltd.
- Domine, F., Taillandier, A. S., Simpson, W. R., & Severin, K. (2005). Specific surface area, density and microstructure of frost flowers. *Geophys. Res. Lett.*, *32*. doi: 10.1029/2005gl023245
- Douglas, T. A., Sturm, M., Simpson, W. R., Brooks, S., Lindberg, S. E., & Perovich, D. K. (2005). Elevated mercury measured in snow and frost flowers near Arctic sea ice leads. *Geophys. Res. Lett.*, *32*. doi: 10.1029/2004gl022132

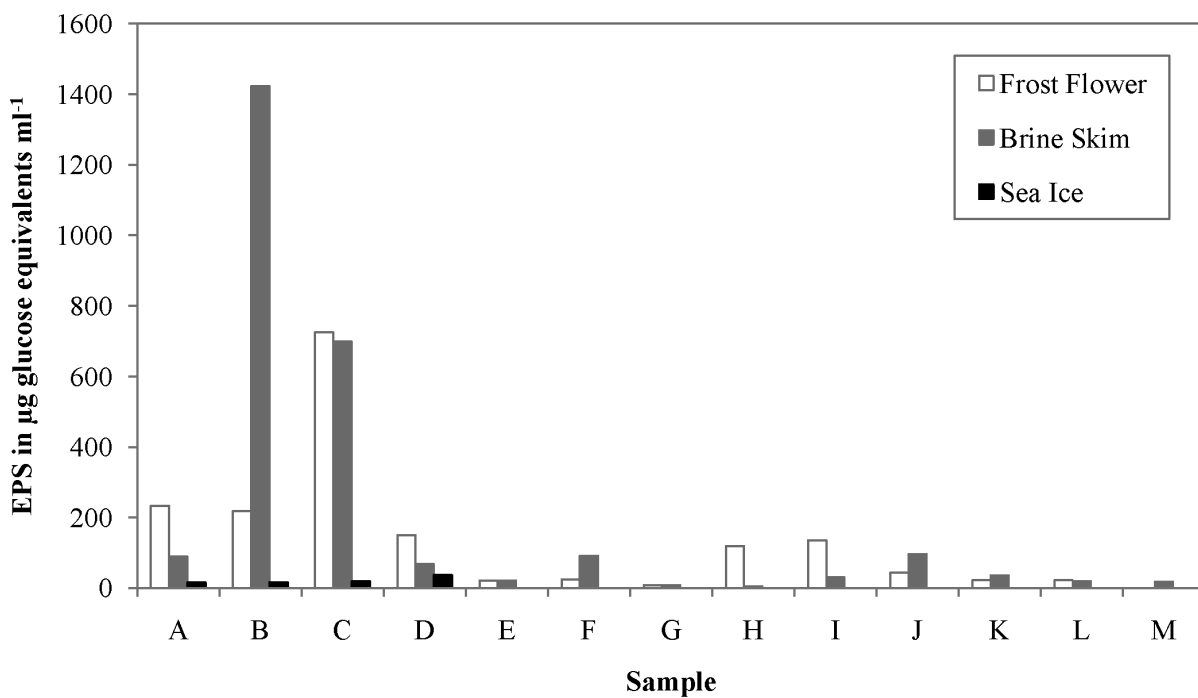
- Gradinger, R., & Ikävalko, J. (1998). Organism incorporation into newly forming Arctic sea ice in the Greenland Sea. *J. Plankton Res.*, 20(5), 871–886. doi: 10.1093/plankt/20.5.871
- Jacobi, H.-W., Frey, M. M., Hutterli, M. A., Bales, R. C., Schrems, O., Cullen, N. J., . . . Koehler, C. (2002). Measurements of hydrogen peroxide and formaldehyde exchange between the atmosphere and surface snow at Summit, Greenland. *Atmos. Environ.*, 36(15-16), 2619–2628.
- Jayaweera, K., & Flanagan, P. (1982). Investigations on biogenic ice nuclei in the Arctic atmosphere. *Geophys. Res. Lett.*, 9.
- Junge, K., Eicken, H., & Deming, J. W. (2004). Bacterial activity at –2 to –20°C in Arctic wintertime sea ice. *Appl. Environ. Microbiol.*, 70(1), 550–557. doi: 10.1128/aem.70.1.550-557.2004
- Junge, K., Krembs, C., Deming, J., Stierle, A., & Eicken, H. (2001). A microscopic approach to investigate bacteria under in situ conditions in sea-ice samples. *Ann. of Glaciol.*, 33, 304–310.
- Krembs, C., Eicken, H., Junge, K., & Deming, J. W. (2002). High concentrations of exopolymeric substances in Arctic winter sea ice: implications for the polar ocean carbon cycle and cryoprotection of diatoms. *Deep Sea Res. Part I*, 49(12), 2163–2181.
- Leck, C., & Bigg, E. K. (2005). Biogenic particles in the surface microlayer and overlaying atmosphere in the central Arctic Ocean during summer. *Tellus B*, 57, 305–316.
- Mader, H. M., Pettitt, M. E., Wadham, J. L., Wolff, E. W., & Parkes, R. J. (2006). Subsurface ice as a microbial habitat. *Geology*, 34(3), 169–172. doi: 10.1130/g22096.1
- Martin, S., Drucker, R., & Fort, M. (1995). A laboratory study of frost flower growth on the surface of young sea ice. *J. Geophys. Res.*, 100. doi: 10.1029/94jc03243
- Martin, S., Yu, Y., & Drucker, R. (1996). The temperature dependence of frost flower growth on laboratory sea ice and the effect of the flowers on infrared observations of the surface. *J. Geophys. Res.*, 101. doi: 10.1029/96jc00208
- Marx, J. G., Carpenter, S. D., & Deming, J. W. (2009). Production of cryoprotectant extracellular polysaccharide substances (EPS) by the marine psychrophilic bacterium *Colwellia psychrerythraea* strain 34H under extreme conditions. *Can. J. Microbiol.*, 55, 63–72.
- Meiners, K., Krembs, C., & Gradinger, R. (2008). Exopolymer particles: microbial hotspots of enhanced bacterial activity in Arctic fast ice (Chukchi Sea). *Aquat. Microb. Ecol.*, 52(2), 195–207. doi: 10.3354/ame01214

- Panikov, N. S., & Sizova, M. V. (2007). Growth kinetics of microorganisms isolated from Alaskan soil and permafrost in solid media frozen down to  $-35^{\circ}\text{C}$ . *FEMS Microbiol. Ecol.*, 59(2), 500–512.
- Perovich, D. K., & Richter-Menge, J. A. (1994). Surface characteristics of lead ice. *J. Geophys. Res.*, 99.
- Petit, J. R., Jouzel, J., Pourchet, M., & Merlivat, L. (1982). A detailed study of snow accumulation and stable isotope content in Dome C (Antarctica). *J. Geophys. Res.*, 87, 4301–4308. doi: 10.1029/JC087iC06p04301
- Rankin, A. M., Wolff, E. W., & Martin, S. (2002). Frost flowers: Implications for tropospheric chemistry and ice core interpretation. *J. Geophys. Res.*, 107. doi: 10.1029/2002jd002492
- Style, R. W. (2007). *The formation and evolution of frost flowers and related phenomena*. (PhD), St. Catharine's College, Cambridge.
- Thomas, D. N., Lara, R. J., Eicken, H., Kattner, G., & Skoog, A. (1995). Dissolved organic matter in Arctic multi-year sea ice during winter: major components and relationship to ice characteristics. *Pol. Biol.*, 15(7), 477–483.
- Wells, L. E., & Deming, J. W. (2006). Modelled and measured dynamics of viruses in Arctic winter sea-ice brines. *Environ. Microbiol.*, 8(6), 1115–1121.
- Williams, W. D., & Sherwood, J. E. (1994). Definition and measurement of salinity in salt lakes. *Int. J. Salt Lake Res.*, 3, 53 – 63.
- Wolff, E. W., Rankin, A. M., & Rathlisberger, R. (2003). An ice core indicator of Antarctic sea ice production? *Geophys. Res. Lett.*, 30. doi: 10.1029/2003gl018454

## Tables and Figures



**Fig. A1.1. Correlations between bacterial abundance and salinity (of melted samples) for laboratory-grown (A) and naturally occurring FF from Barrow (B) and the central Arctic Ocean (C).** In (A), data from laboratory brine skim, saline ice, and the ending NaCl solution containing *Halomonas pacifica* are also shown. The regression line is fit to FF data (df = 13,  $r = 0.75$ ,  $P = 0.0013$ ), but the brine skim data fall closely along this line. In (B) and (C), data from other components of the natural surface ice environment (see text) are also plotted, with regressions fit to FF (solid line, df = 12,  $r = 0.69$ ,  $P = 0.0065$ ) and to all data points (dashed line, df = 42,  $r = 0.60$ ,  $P < 0.0001$ ). Temperature at time of sampling is shown for A, B, and C; in B, the two values reflect sampling on successive days.



**Fig. A1.2. Concentration of pEPS in FF, brine skim, and sea ice at Barrow, AK.**

Comparative values are shown for frost flowers and the underlying brine skim and sea ice. Sea ice data are only available for sample sets A–D.

## Appendix 2

### **Bacterial phylotypes shared between the young sea ice and supraglacial environments**

#### ABSTRACT

Frost flowers growing on young sea ice are thought to be a major source of the sulfate-depleted sea salts observed in coastal glaciers. To test the hypothesis that biological material is transported aurally along with sea salt, we evaluated the microbial community composition and structure of frost flowers, young sea ice, and seawater from two sites north of Ross Island, Antarctica, and the Wilson Piedmont Glacier, Taylor Glacier, and snow by deep sequencing of the V4 region of the 16S rRNA gene. Although sequence reads most abundant in frost flowers, presumed to be the most mobile element of the sea ice environment, were rarely found in the terrestrial samples, Cyanobacterial reads associated with the freshwater genus *Pseudanabaena* were abundant in the glacial ice samples and comprised up to 14.1 % of the frost flower community. This finding suggests that biological material from glaciers and lake ice is readily transported to the young sea ice surface. Two operational taxonomic units (OTUs) associated with a sulfur-oxidizing, halophilic clade of the Gammaproteobacteria were also more abundant in frost flowers than in any other sample type. The source of these OTUs remains unknown; one possible explanation is transport from hydrothermally influenced habitats at nearby Mt. Erebus.

## A2.1 Introduction

In polar coastal environments glaciers and sea ice, two ecologically distinct ice environments, exist side by side. Although the bulk salinity of glacial ice (i.e. the salinity of glacial meltwater) is very low, glaciers do contain trace amounts of sea salt. The ionic composition of this sea salt is often fractionated; i.e., deficient in sulfate relative to chloride (Wagenbach et al., 1998; A. M. Rankin, Auld, & Wolff, 2000; Andrew M. Rankin, Wolff, & Martin, 2002; Beaudon & Moore, 2009; Pasteris et al., 2014). This fractionation has been attributed to the precipitation of mirabilite ( $\text{Na}_2\text{SO}_4 \cdot 10\text{H}_2\text{O}$ ) prior to the aerosolization of sea salt, which suggests that the dominant source of the salt reaches a temperature of  $-6.3^\circ\text{C}$  (the eutectic point for mirabilite) (Marion, Farren, & Komrowski, 1999) before reaching the atmosphere (Wagenbach et al., 1998). The low temperature necessary to precipitate mirabilite excludes bubble bursting at the sea surface as the source of these aerosols. Frost flowers, highly saline structures common to the surface of young sea ice, have been proposed as a source of fractionated sea salt (A. M. Rankin et al., 2000; Beaudon & Moore, 2009). Although limited laboratory observations have been unable to form aerosols from these structures (Roscoe et al., 2011), their delicate shape and eventual evolution to saline snow, another potential transport mechanism (Domine, Sarapani, Lanniello, & Beine, 2004; Ewert, Carpenter, Colangelo-Lillis, & Deming, 2013), are suggestive of a transport capability.

Several studies have shown that bacteria are enriched in frost flowers along with salt (Bowman & Deming, 2010; Eronen-Rasimus et al., 2014). This frost flower bacterial community can be quite different from the community observed in the underlying sea ice, though the mechanisms structuring these communities are not well understood (Bowman, Larose, Vogel, & Deming, 2013; Bowman, Berthiaume, Armbrust, & Deming, 2014). Bacterial enrichment in

frost flowers, and potential frost flower transport to the glacial environment, led us to consider frost flowers as a mechanism for transporting genetic material, dissolved and particular organic carbon, and bacteria themselves to the glacier surface. To test this hypothesis we sampled frost flowers, young sea ice, seawater, freshly fallen snow over land, and the surface of the Wilson Piedmont and Taylor Glaciers near McMurdo Sound, Antarctica, during Austral spring 2011 for analysis of community composition and structure using Illumina sequencing of the 16S rRNA gene.

## **A2.2 Methods**

### **A2.2.1 Sample collection and processing**

Samples of frost flowers, young sea ice, and seawater were collected from near Cape Crozier, Antarctica, at 77.38048 °S and 169.02777 °E on October 22, 2011, and from Lewis Bay, Antarctica at 77.14084 °S and 167.82689 °E on October 24, 2011 (Fig. A2.1). The Wilson Piedmont Glacier was sampled on October 17, 2011, at 77.38744 °S and 163.58801 °E. Snow from the terminus of the Wilson Piedmont Glacier was sampled approximately 200 m east of this location on October 25, 2011. Taylor Glacier was sampled on October 28, 2011, at 77.7398 °S and 162.1397 °E. Snow was sampled with an ethanol-rinsed shovel into UV and ethanol-rinsed plastic bins. Glacier ice was sampled using an ethanol-rinsed Kovacs corer and UV and ethanol-rinsed plastic bins. Frost flowers were sampled using an ethanol-rinsed shovel into 16 L whirlpack bags. Young sea ice was sampled using an ethanol-rinsed Kovacs corer into 16 L whirlpack bags. Seawater was sampled through Kovacs core holes into sterile 1 L bottles.

Frost flowers and young sea ice were melted to an approximate final melt salinity of 100 ppt by the addition of 0.2 µm and tangential-flow-filtered NaCl solution (approximately 280 ppt). Glacier ice and snow were melted without the addition of sterile brine. Melts took place at



room temperature and were carefully monitored to prevent warming following ice melt. After melting, meltwater was moved to a 1 °C cold room for filtration through 3.0, 1.0, and 0.2 µm filters. To evaluate contamination in the brine solution two melt brine blanks were prepared by filtering 5 L of brine through 0.2 µm Sterivex filters. All filters were stored in RNAlater (Ambion) at –80 °C until further processing.

DNA was extracted from the filters using the phenol-chloroform extraction method and standard protocols (Kellogg & Deming, 2009). Two procedural blanks were prepared by extracting from SET buffer without the addition of any filter. Extracted DNA was quantified using the picogreen assay and shipped to the Argonne National Lab sequencing center for amplification of the V4 region of the 16S rRNA gene, paired-end barcoded library preparation, and sequencing using a dedicated lane on the Illumina MiSeq platform following the methods of Caporaso et al. (Caporaso et al., 2011).

#### **A2.2.2 Data analysis**

All sequence processing was conducted with Mothur v1.33.2 (Schloss et al., 2009) except where indicated. First raw reads were trimmed to a mean Phred score of 25 using a custom Python script. Contigs were constructed from trimmed reads using the command `make.contigs` with 1 difference allowed. Contigs not 253 bp in length, with ambiguous sequences, or with more than 6 homopolymers were discarded. Contig quality was assessed outside of Mothur using SolexaQA (Cox, Peterson, & Biggs, 2010). Based on this assessment the first and last 10 positions were trimmed from each contig to improve contig quality.

Redundant contigs were tallied and combined with `unique.seqs`, nonredundant contigs were aligned against the combined Silva Bacteria, Archaea, and Eukarya reference alignments (Pruesse et al., 2007). After alignment `pcr.seqs` was used to generate a custom reference

alignment matching the region of best alignment for the query sequences, which were positions 13,000 to 24,000 in the Silva alignment. A new alignment of the quality controlled contigs was made to this reference alignment. Aligned reads were screened using the options `optimize=start-end-minlength` and `criteria=90`, and positions in the alignment containing no information, or occurring before the last start or after the first end, were discarded using the `filter.seqs` command. Redundant sequences were again tallied and combined with `align.seqs`. Chimeras were then identified and removed using `chimera.uchime` and `remove.seqs`.

Quality controlled contigs were classified against a no-gap version of the reference alignment using the Greengenes taxonomy (DeSantis et al., 2006), available from the Mothur website ([http://www.mothur.org/wiki/Taxonomy\\_outline](http://www.mothur.org/wiki/Taxonomy_outline)), and a bootstrap cutoff of 60. Reads that classified as Mitochondrial, unknown, Archaea, or Eukaryota were removed from the dataset. At this point in the analysis the 1  $\mu$ m fraction from the Wilson Piedmont Glacier had very few contigs and was removed from the dataset. The remaining samples had at least 3,000 contigs, sufficient for a comparative analysis of community composition (Caporaso et al., 2011). To ensure an even comparison between communities each sample was subsampled to 3,000 contigs. The commands `dist.seqs` and `cluster` were used to generate a distance matrix of the subsampled alignment and cluster contigs into OTUs. A consensus classification for OTUs at a distance of 0.03 was generated from the contig classifications. OTUs present in any blanks at > 30 contigs (greater than 1 % of the total population) were removed from all samples prior to further analysis.

The removal of two additional samples (a 1  $\mu$ m fraction from young sea ice sampled on 2011 October 22 and a 3.0  $\mu$ m fraction from young sea ice sampled on 2011 October 24) allowed subsampling to a depth of 23,000 contigs. All post-subsampling steps applied earlier

were repeated for this depth, except that for this deeper analysis all OTUs that appeared in any abundance in the blank samples were eliminated. Further statistical analyses and graphical description for both depths were carried out in R (Team, 2012).

All contigs from OTUs of particular interest (much more abundant in frost flowers than other sample types, or shared between the supraglacial and young sea ice environment) were evaluated further by phylogenetic placement using pplacer and guppy (Matsen, Kodner, & Armbrust, 2010). Gammaproteobacterial OTUs of interest were placed on a reference tree constructed from all type Gammaproteobacterial strains in the Ribosomal Database Project (RDP) (Cole et al., 2007). Cyanobacterial OTUs of interest were placed on a reference tree constructed from all Cyanobacterial isolates in RDP. To conduct the placement reference alignments were generated by aligning all reference sequences in Mothur (Schloss et al., 2009) as already described. From each reference alignment a reference tree was constructed using the GTR model in the OpenMP version of FastTree v2 (Price, Dehal, & Arkin, 2010). Guppy and pplacer were both run in posterior probability mode. Phyloxml format trees generated with guppy were visualized using Archaeopteryx v0.9891 (Han & Zmasek, 2009).

## **A2.3 Results**

### **A2.3.1 Environmental**

Young sea ice sampled on 2011 October 22 from near Cape Crozier was 28 cm thick with a bulk salinity of 12 ppt. Frost flowers collected from this ice had a bulk salinity of 62 ppt. Young sea ice collected on 2011 October 24 from Lewis Bay was 13.5 cm thick with a bulk salinity of 8 ppt. Frost flowers collected from this ice had a bulk salinity of 78 ppt. The temperature at the ice surface on October 24 was  $-6.1^{\circ}\text{C}$  and, at the top of the frost flowers, –

10.4 °C; the air temperature was –14.7 °C. Temperature data are not available for 2011 October 22.

#### **A.2.3.2 Community composition**

The supraglacial samples (Taylor Glacier, Wilson Piedmont, and the snow sample) formed a cluster separate from the marine samples in an analysis of the 50 most abundant OTUs across all samples, although there was significant variation in community composition between these samples (Fig. 2.2). The 3.0 µm fractions for Taylor Glacier and the snow sample were dominated by the Cyanobacterium *Oscillatoria limnetica* (TG = 30.0 %, snow = 37.7 %), an unidentified Gammaproteobacterium (TG = 8.6 %, snow = 2.7 %), and *Saprospirales* sp. (TG = 11.4 %, snow = 1.0 %). The 1.0 µm fraction from Wilson Piedmont Glacier was similar, although these OTUs formed a smaller fraction of the overall community, and *Bradyrhizobiales* sp. was the second most abundant OTU (24.3 %) followed by *Sphingomonadales* (13.9 %). In the 0.2 µm fraction from Wilson Piedmont Glacier there were no *O. limnetica*; the dominant genera were *Sphingomonadales* (49.9 %) and *Bradyrhizobiales* (7.3 %). An alternate 3.0 µm filter for Wilson Piedmont, and alternate 1.0 and 0.2 µm filters for Taylor Glacier and the snow sample, were not available or generated poor sequence libraries.

With some exceptions the frost flower samples formed a distinct cluster among the marine samples, as did the seawater and young sea ice samples, however, the compositional differences between these groups were slight. The most abundant OTU among the marine samples classified as a member of the Deltaproteobacterial Sva0853 clade (SG: 0.006 %, FF: 5.9 %, YI: 9.9 %, SW: 15.9 %), the next most abundant was a member of the Gammaproteobacterial SUP05 clade (SG: 0 %, FF: 6.6 %, YI: 6.4 %, SW: 6.0 %), with considerable numbers of Gammaproteobacteria ZA2333c (SG: 0.0 %, FF: 2.2 %, YI: 5.1 %, SW: 4.2 %), an unclassified

Gammaproteobacterial genus (Gamma0013) (SG: 0 %, FF: 7.4 %, YI: 1.7 %, SW: 0.6 %),  
Gammaproteobacteria Arctic96B-1 (SG: 0.006 %, FF: 2.4 %, YI: 4.6 %, SW: 3.1 %),  
Alphaproteobacteria *Consitiales* (SG: 0 %, FF: 3.1 %, YI: 2.9 %, SW: 2.7 %), and  
Flavobacteriales Arctic97A-17 (SG: 0 %, FF: 1.4 %, YI: 2.8 %, SW: 4.3 %).

Between the frost flower samples there was considerable variation, with one 0.2 µm filter from 2011 October 22 as a clear outlier. Although this sample contained significant SUP05 (4.4 %) and unclassified Gammaproteobacterial contigs (16.5 %), as did the other marine samples, it also contained a large number of *Oscillatoria limnetica* contigs (14.1 %). Lower numbers of *O. limnetica* contigs (0.03 % to 3.4 %) were found in the other frost flower samples collected on 2011 October 22, with only a single contig found in the frost flower samples from 2011 October 24. Two Gammaproteobacterial OTUs were elevated in the frost flowers relative to the other marine samples (Gamma0013; FF: 7.4 %, YI: 1.7 %, SW: 0.61 %, Gamma0029; FF: 2.9 %, YI: 0.18 %, SW: 0.01 %). Gamma0013 was most abundant in the Lewis Bay frost flower samples (11.3 % and 8.1 %), but was present in some abundance in the Cape Crozier samples (2.8 %). Gamma29 was much more abundant in the Lewis Bay samples (0.5 % and 8.0 %) than the Cape Crozier samples (0.03 %).

The deeper subsampling identified four less abundant OTUs that were shared between the supraglacial and marine environments; Introasporangiaceae (SG: 0.73 %, FF: 0.09 %, YI: 0.12 %, SW: 0.06 %), an unclassified Cyanobacterium (Cyano00118) (SG: 0.09 %, FF: 0.17 %, YI: 0.007 %, SW: 0 %), an unclassified Saprospirales (SG: 0.09 %, FF: 0.003 %, YI: 0 %, SW: 0 %), and an unclassified Actinobacteridae (SG: 0.21 %, FF: 0.08 %, YI: 0.02 %, SW: 0.01 %).

Four OTUs of interest were investigated more closely: the three poorly classified OTUs, Cyano00118, Gamma0013 and Gamma0029, and *Oscillatoria limnetica* (Cyano0005).

Phylogenetic placement of Cyano00118 placed it within the Cyanobacterial genus *Anabaena* (mean posterior probability = 1). Phylogenetic placement of the OTU associated with Cyano0005 placed it within the genus *Pseudanabaena* (mean posterior probability = 1); the reference *O. limnetica* sequence also placed with this genus. The Gamma0013 OTU placed with a Gammaproteobacterial clade consisting of the genera *Thiohalobacter*, *Thiopfundum*, *Thiohalophilus*, and *Thioalkalispira* (mean posterior probability = 1). Gamma0029 placed with the Gammaproteobacterial genus *Rugamonas* (mean posterior probability = 1).

#### **A2.4 Discussion**

In this analysis we sought to identify whether there was overlap in the bacterial communities present at the surface of young sea ice and glacial ice. Although these environments differ dramatically in their nutrient regimes, carbon concentrations, and bulk salinity, they have some key similarities, including high incident radiation and low temperature. Sampling occurred during the Austral spring, when the young sea ice and supraglacial environments in the vicinity of McMurdo Sound were experiencing approximately 22 hours of daylight. We did not measure the temperature of the glacial surface, but air temperature at McMurdo Station at the time of sampling was  $-16.2^{\circ}\text{C}$  for Wilson Piedmont Glacier and  $-12.5^{\circ}\text{C}$  for Taylor Glacier. These values fall well below the values for the young sea ice environment ( $-6.1^{\circ}\text{C}$  at the ice surface) which, due to the thin nature of young sea ice, was not well insulated from the underlying warm seawater. Unknown however, is the microscale temperature distribution at the glacial ice surface. During the polar spring local warming around dark particles at the glacial surface lead to the development of cryoconite holes and a diverse microbial community (Christner, Kvitko II, & Reeve, 2003; Edwards et al., 2010). Prior to

establishing cryoconite holes these particles, warmed by sunlight, could generate local warming and temperatures similar to the young sea ice surface.

We identified only one OTU, Cyano0015, of the *Pseudanabaena*, that was a major member of both the young sea ice and supraglacial microbial communities. In the young sea ice environment this clade was most abundant in older frost flowers growing on relatively thick young sea ice. Algal mats from the numerous McMurdo Dry Valley lakes and McMurdo Ice Shelf ponds, where the genus has been observed previously (de los Ríos, Ascaso, Wierzychos, Fernández-Valiente, & Quesada, 2004; Jungblut et al., 2005), are a likely source for these *Pseudanabaena*. Mat material is known to detach and become frozen into lake ice, where it can be ablated and transported by wind (Parker, Simmons, Wharton, Seaburg, & Love, 1982). Under this model the older young sea ice surfaces will have had more time to accumulate wind-borne material. While the slow growth of psychrophilic bacteria has been observed at temperatures to  $-12^{\circ}\text{C}$  (Breezee, Cady, & Staley, 2004; Wells & Deming, 2006), the lower temperature limit for Cyanobacterial growth is generally considered to be much higher (Tang, Tremblay, & Vincent, 1997). Psychrophilic Cyanobacteria, however, have been observed to grow at  $2^{\circ}\text{C}$  (Nadeau & Castenholz, 2000), the lowest tested temperature that we could find in the literature. Growth at  $-6^{\circ}\text{C}$  may be unlikely for these bacteria, but psychrophilic Cyanobacteria at the young sea ice surface could still play an ecological role at this temperature by exuding DOC and exopolymers (EPS). These carbon sources would aid the growth of heterotrophic bacteria at the ice surface or within saline snow (Ewert et al., 2013).

Unlike the Cyanobacteria, the Gammaproteobacterial clade contains some of the most cold-active psychrophiles known (Breezee et al., 2004; Wells & Deming, 2006), though most known Gammaproteobacterial strains are mesophiles. The marine samples contained a large

number of Gammaproteobacterial OTUs, consistent with other analyses of late-winter polar marine microbial communities (Alonso-Saez, Galand, Casamayor, Pedros-Alio, & Bertilsson, 2010; Grzymski et al., 2012). During the polar winter photosynthesis in the water column reaches a minimum and the system becomes oligotrophic. These conditions favor bacteria capable of an autotrophic lifestyle (Alonso-Saez et al., 2010; Grzymski et al., 2012), such as the sulfur-oxidizing Gammaproteobacterial clades SUP05 (Marshall & Morris, 2012), the closely related clade Arctic96B-1, and (putatively) the OMZ associated clade ZA2333c (Wright, Konwar, & Hallam, 2012). The abundance of Sva0853 and SUP05, typically associated with deeper waters (Shi, Tyson, Eppley, & DeLong, 2010; Kim et al., 2013), suggests upwelling or the colonization of carbon-poor surface waters by deep clades during winter. Although all of these clades were present within frost flowers, they were present at a lower abundance than in seawater.

In contrast to these clades the Gamma0013 and Gamma0029 OTUs were elevated in frost flowers relative to the underlying sea ice or seawater. Gamma0029 placed with *Rugamonas*, a pigmented aquatic bacterium (Austin & Moss, 1986) for which little additional information is available to aid an ecological interpretation. The placement of Gamma0013 with the alkaliphilic and halophilic sulfur-oxidizing genera *Thiohalobacter*, *Thiopfundum*, *Thiohalophilus*, and *Thioalkalispira* is more suggestive. These genera are not associated in the literature with the marine environment. Although some of these genera are typically found in alkaline hypersaline lakes (Sorokin, Sjollem, & Kuenen, 2002; Sorokin, Kovaleva, Tourova, & Muyzer, 2010), transport to the young sea ice surface from lakes in the McMurdo Dry Valleys seems unlikely as these genera are completely absent from the supraglacial samples. Strains from these genera are associated with hydrothermal systems, however (Takai et al., 2009; Mori et al., 2011); thus one



possible source for these bacteria is the nearby volcano Mt. Erebus. The microbial community composition of hot and cold hydrothermally influenced environments around Mt. Erebus is almost entirely unknown. Soo et al. (2009) used a shallow 16S rRNA gene clone library to evaluate the microbial community composition of hot soils on Tramway Ridge, Mt. Erebus, but did not identify genera related to those observed at the young sea ice surface. We could not find any study that has evaluated microbial community composition of cold hydrothermally altered environments on Mt. Erebus (e.g. frozen fumaroles, melt water streams) or that has applied a quantitative deep-sequencing method to any environment on Mt. Erebus.

There should be opportunities for material from Mt. Erebus to reach the supraglacial environment in the McMurdo Dry Valleys. The absence of Gamma0013 in supraglacial samples suggests a short lifetime for these bacteria at the glacier surface. An inadequate energy source (relatively oligotrophic and little reduced sulfur), low pH, low ion concentration, and high UV could all select against these strains. The pH of brines at the surface of young sea ice is presumed to be only slightly below that of seawater, due to the precipitation of calcite and ikaite at relatively warm temperatures (Papadimitriou, Kennedy, Kattner, Dieckmann, & Thomas, 2004), but higher than the acidic (and very low volume) liquid phase expected for a glacier ice surface sourced from meteoric water. Ion concentrations at the sea ice surface should favor moderate halophiles, along with (presumed) higher concentrations of organic substrate and reduced sulfur.

## **Acknowledgments**

This work was supported by NSF DPP award 1043265 to JWD and the Walters Endowed Professorship. JSB was supported by an NSF IGERT Fellowship through the University of Washington's Astrobiology Program.

## References

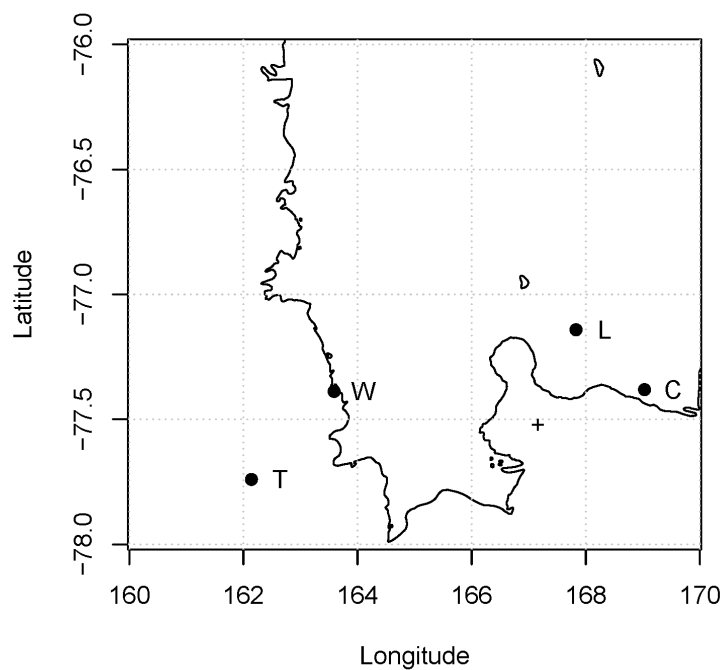
- Alonso-Saez, L., Galand, P. E., Casamayor, E. O., Pedros-Alio, C., & Bertilsson, S. (2010). High bicarbonate assimilation in the dark by Arctic bacteria. *ISME J*, 4(12), 1581–1590.
- Austin, D., & Moss, M. (1986). Numerical Taxonomy of Red-pigmented Bacteria Isolated from a Lowland River, with the Description of a New Taxon, *Rugamonas rubra* gen. nov., sp. nov. *J. Gen. Microbiol.*, 132(7), 1899–1909.
- Beaudon, E., & Moore, J. (2009). Frost flower chemical signature in winter snow on Vestfonna ice cap (Nordaustlandet, Svalbard). *The Cryos. Discuss.*, 3(1), 159–180.
- Bowman, J. S., Berthiaume, C. T., Armbrust, E. V., & Deming, J. W. (2014). Metagenomic analysis of frost flowers and young sea ice reveals genes involved in key biogeochemical processes. *FEMS Microb. Ecol.*, Available online in advance of print.
- Bowman, J. S., & Deming, J. W. (2010). Elevated bacterial abundance and exopolymers in saline frost flowers with implications for atmospheric chemistry and microbial dispersal. *Geophys. Res. Lett.*, 37(L13501).
- Bowman, J. S., Larose, C., Vogel, T. M., & Deming, J. W. (2013). Selective occurrence of Rhizobiales in frost flowers on the surface of young sea ice near Barrow, Alaska and distribution in the polar marine rare biosphere. *Environ. Microbiol. Rep.*, 5, 575–582. doi: 10.1111/1758-2229.12047
- Breezee, J., Cady, N., & Staley, J. T. (2004). Subfreezing growth of the sea ice bacterium “*Psychromonas ingrahamii*”. *Microb Ecol*, 47(3), 300–304.
- Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., . . . Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci.*, 108(Supplement 1), 4516–4522.
- Christner, B. C., Kvitko II, B. H., & Reeve, J. N. (2003). Molecular identification of bacteria and eukarya inhabiting an Antarctic cryoconite hole. *Extremophiles*, 7(3), 177–183.
- Cole, J. R., Chai, B., Farris, R. J., Wang, Q., A. S. Kulam-Syed-Mohideen, D. M. M., Bandela, A. M., . . . Tiedje, J. M. (2007). The ribosomal database project (RDP-II): introducing myRDP space and quality controlled public data. *Nuc. Acids Res.*, 35, D169–D172. doi: 10.1093/nar/gkl889
- Cox, M. P., Peterson, D. A., & Biggs, P. J. (2010). SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data. *BMC Bioinformatics*, 11(1), 485.
- de los Ríos, A., Ascaso, C., Wierzos, J., Fernández-Valiente, E., & Quesada, A. (2004). Microstructural characterization of cyanobacterial mats from the McMurdo Ice Shelf, Antarctica. *Appl. Environ. Microbiol.*, 70(1), 569–580.

- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., . . . Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.*, 72(7), 5069–5072. doi: 10.1128/aem.03006-05
- Domine, F., Sarapani, R., Lanniello, A., & Beine, H. J. (2004). The origin of sea salt in snow on Arctic sea ice and in coastal regions. *Atmos. Chem. Phys. Discuss.*, 4, 4737–4776.
- Edwards, A., Anesio, A. M., Rassner, S. M., Sattler, B., Hubbard, B., Perkins, W. T., . . . Griffith, G. W. (2010). Possible interactions between bacterial diversity, microbial activity and supraglacial hydrology of cryoconite holes in Svalbard. *ISME J*, 5(1), 150–160.
- Eronen-Rasimus, E., Kaartokallio, H., Lyra, C., Autio, R., Kuosa, H., Dieckmann, G. S., & Thomas, D. N. (2014). Bacterial community dynamics and activity in relation to dissolved organic matter availability during sea ice formation in a mesocosm experiment. *MicrobiologyOpen*, 3(1), 139–156.
- Ewert, M., Carpenter, S., Colangelo-Lillis, J., & Deming, J. (2013). Bacterial and extracellular polysaccharide content of brine-wetted snow over Arctic winter first-year sea ice. *J. Geophys. Res-Oceans*, 118, 726–735.
- Grzyski, J., Riesenfeld, C., Williams, T., Dussaq, A., Ducklow, H., Erickson, M., . . . Murray, A. (2012). A metagenomic assessment of winter and summer bacterioplankton from Antarctic Peninsula coastal surface waters. *ISME J*, 6(10), 1901–1915.
- Han, M., & Zmasek, C. (2009). phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics*, 10(1), 356.
- Jungblut, A.-D., Hawes, I., Mountfort, D., Hitzfeld, B., Dietrich, D. R., Burns, B. P., & Neilan, B. A. (2005). Diversity within cyanobacterial mat communities in variable salinity meltwater ponds of McMurdo Ice Shelf, Antarctica. *Environ. Microbiol.*, 7(4), 519–529. doi: 10.1111/j.1462-2920.2005.00717.x
- Kellogg, C. T. E., & Deming, J. W. (2009). Comparison of free-living, suspended particle, and aggregate-associated bacterial and archaeal communities in the Laptev Sea. *Aquat. Microb. Ecol.*, 57(1), 1–18. doi: 10.3354/ame01317
- Kim, J.-G., Park, S.-J., Quan, Z.-X., Jung, M.-Y., Cha, I.-T., Kim, S.-J., . . . Lee, S.-H. (2013). Unveiling abundance and distribution of planktonic Bacteria and Archaea in a polynya in Amundsen Sea, Antarctica. *Environ. Microbiol.*, 16(6), 1566–1578.
- Marion, G., Farren, R., & Komrowski, A. (1999). Alternative pathways for seawater freezing. *Cold Reg. Sci. Technol.*, 29(3), 259–266.
- Marshall, K. T., & Morris, R. M. (2012). Isolation of an aerobic sulfur oxidizer from the SUP05/Arctic96BD-19 clade. *ISME J*, 7(2), 452–455.

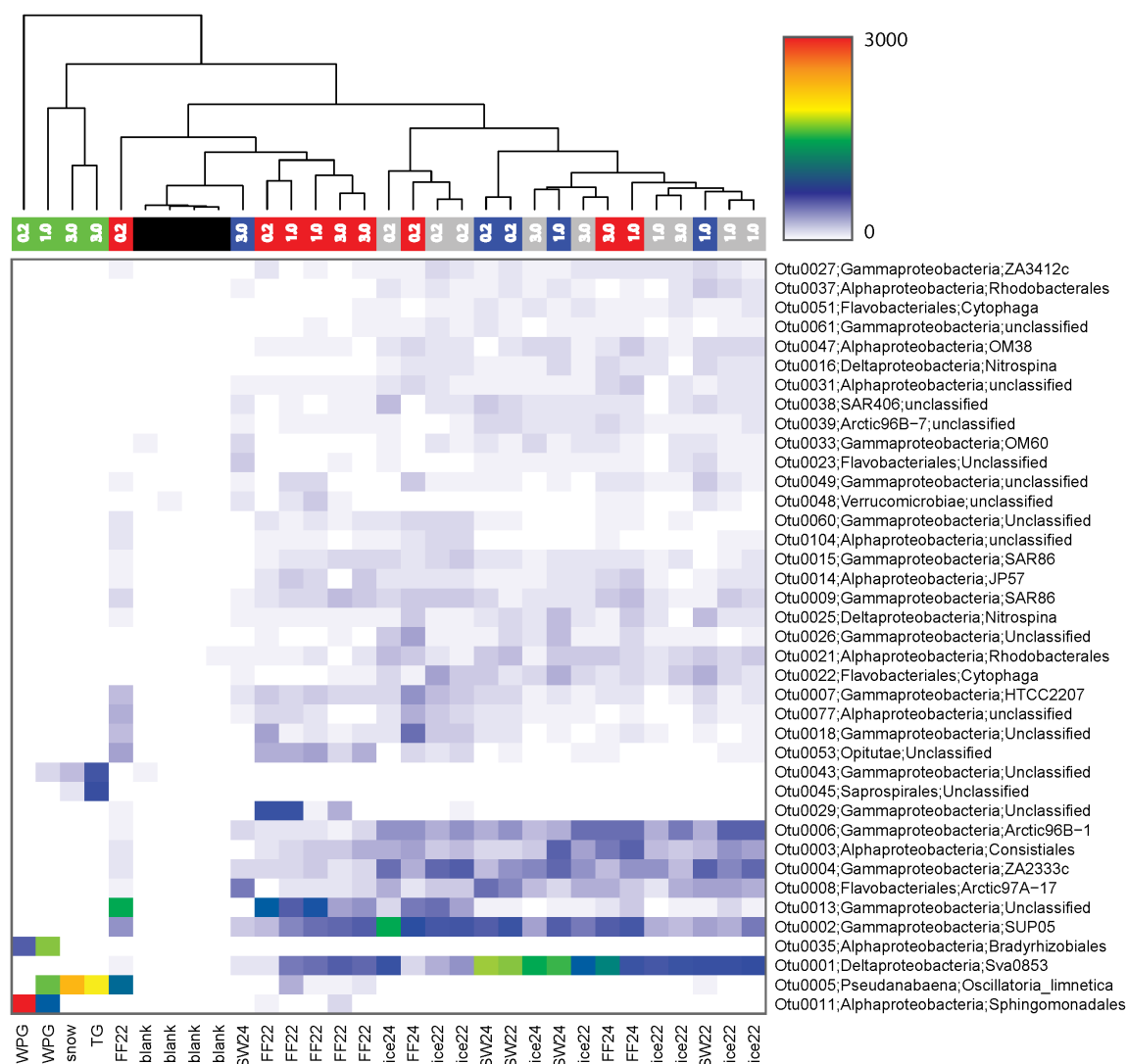
- Matsen, F., Kodner, R., & Armbrust, E. V. (2010). pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11(1), 538.
- Mori, K., Suzuki, K.-i., Urabe, T., Sugihara, M., Tanaka, K., Hamada, M., & Hanada, S. (2011). *Thiopfundum hispidum* sp. nov., an obligately chemolithoautotrophic sulfur-oxidizing gammaproteobacterium isolated from the hydrothermal field on Suiyo Seamount, and proposal of Thioalkalspiraceae fam. nov. in the order Chromatiales. *Int. J. Syst. Evol. Micr.*, 61(10), 2412–2418.
- Nadeau, T.-L., & Castenholz, R. W. (2000). Characterization of psychrophilic Oscillatorians (Cyanobacteria) from Antarctic meltwater ponds. *J. Phycol.*, 36(5), 914–923. doi: 10.1046/j.1529-8817.2000.99201.x
- Papadimitriou, S., Kennedy, H., Kattner, G., Dieckmann, G., & Thomas, D. (2004). Experimental evidence for carbonate precipitation and CO<sub>2</sub> degassing during sea ice formation. *Geochimica et Cosmochimica Acta*, 68(8), 1749–1761.
- Parker, B. C., Simmons, G. M., Wharton, R. A., Seaburg, K. G., & Love, F. G. (1982). Removal of organic and inorganic matter from antarctic lakes by aerial escape of bluegreen algal mats. *J. Phycol.*, 18(1), 72–78.
- Pasteris, D. R., McConnell, J. R., Das, S. B., Criscitiello, A. S., Evans, M. J., Maselli, O. J., . . . Layman, L. (2014). Seasonally resolved ice core records from West Antarctica indicate a sea ice source of sea salt aerosol and a biomass burning source of ammonium. *J. Geophys. Res-Atmos.*, Available online in advance of print, 2013JD020720. doi: 10.1002/2013jd020720
- Price, M., Dehal, P., & Arkin, A. (2010). FastTree 2 - Approximate maximum likelihood trees for large alignments. *PLOS one*, 5(3), e9490.
- Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., & Glockner, F. O. (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl. Acids Res.*, 35(21), 7188–7196. doi: 10.1093/nar/gkm864
- Rankin, A. M., Auld, V., & Wolff, E. W. (2000). Frost flowers as a source of fractionated sea salt aerosol in the polar regions. *Geophys. Res. Lett.*, 27(21), 3469–3472. doi: 10.1029/2000gl011771
- Rankin, A. M., Wolff, E. W., & Martin, S. (2002). Frost flowers: Implications for tropospheric chemistry and ice core interpretation. *J. Geophys. Res.*, 107, 4683–4695. doi: 10.1029/2002jd002492
- Roscoe, H. K., Brooks, B., Jackson, A., Smith, M., Walker, S., Obbard, R. W., & Wolff, E. W. (2011). Frost flowers in the laboratory: Growth, characteristics, aerosol, and the underlying sea ice. *J. Geophys. Res-Atmos.*, 116, D12301.

- Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., . . . Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.*, 75(23), 7537–7541. doi: 10.1128/aem.01541-09
- Shi, Y., Tyson, G. W., Eppley, J. M., & DeLong, E. F. (2010). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *ISME J.*, 5(6), 999–1013.
- Soo, R. M., Wood, S. A., Grzyski, J. J., McDonald, I. R., & Cary, S. C. (2009). Microbial biodiversity of thermophilic communities in hot mineral soils of Tramway Ridge, Mount Erebus, Antarctica. *Environ. Microbiol.*, 11(3), 715–728. doi: 10.1111/j.1462-2920.2009.01859.x
- Sorokin, D. Y., Kovaleva, O. L., Tourova, T. P., & Muyzer, G. (2010). Thiohalobacter thiocyanaticus gen. nov., sp. nov., a moderately halophilic, sulfur-oxidizing gammaproteobacterium from hypersaline lakes, that utilizes thiocyanate. *Int. J. Syst. Evol. Micro.*, 60(2), 444–450.
- Sorokin, D. Y., Sjollem, K. A., & Kuenen, J. G. (2002). Thioalkalispira microaerophila gen. nov., sp. nov., a novel lithoautotrophic, sulfur-oxidizing bacterium from a soda lake. *Int. J. Syst. Evol. Micro.*, 52(6), 2175–2182.
- Takai, K., Miyazaki, M., Hirayama, H., Nakagawa, S., Querellou, J., & Godfroy, A. (2009). Isolation and physiological characterization of two novel, piezophilic, thermophilic chemolithoautotrophs from a deep-sea hydrothermal vent chimney. *Environ. Microbiol.*, 11(8), 1983–1997.
- Tang, E. P. Y., Tremblay, R., & Vincent, W. F. (1997). Cyanobacterial dominance of polar freshwater ecosystems: Are high-latitude mat-formers adapted to low temperature? *J. Phycol.*, 33(2), 171–181. doi: 10.1111/j.0022-3646.1997.00171.x
- Team, R. C. (2012). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Wagenbach, D., Ducroz, F., Mulvaney, R., Keck, L., Minikin, A., Legrand, M., . . . Wolff, E. W. (1998). Sea-salt aerosol in coastal Antarctic regions. *J. Geophys. Res.*, 103(D9), 10961–10974. doi: 10.1029/97jd01804
- Wells, L. E., & Deming, J. W. (2006). Characterization of a cold-active bacteriophage on two psychrophilic marine hosts. *Aquat. Microb. Ecol.*(45), 15–29.
- Wright, J. J., Konwar, K. M., & Hallam, S. J. (2012). Microbial ecology of expanding oxygen minimum zones. [10.1038/nrmicro2778]. *Nat. Rev. Micro.*, 10(6), 381–394.

## Tables and Figures



**Fig. A2.1.** Sample locations near McMurdo Sound, Antarctica. C: Cape Crozier sample site, L: Lewis Bay sample site, T: Taylor Glacier sample site, W: Wilson Piedmont Glacier sample site, plus sign indicates the summit of Mt. Erebus.



**Fig. A2.2.** Abundance of most dominant OTUs across samples, subsampled to a depth of 3000 contigs. The number of contigs that clustered in each OTU is given by the color in the heatmap. Rows (OTUs) are labeled according to OTU number and high and low level taxa (genus, when available). Colored band above the heatmap gives the size fraction and sample type; green: supraglacial, blue: seawater, grey: young sea ice, red: frost flower, and black: blank. Labels at the bottom of the plot give the sample type and date (day of 2011 October); WPG: Wilson Piedmont Glacier, TG: Taylor Glacier. Dendrogram at top was produced using the dist and hclust functions in R (Team, 2012).

## Appendix 3

### The Cryosphere Frost Flower Reactor for Organic Geochemistry (CRYO-FROG)

#### A3.1 Introduction

The young sea ice surface in general and frost flowers in particular are an interesting site for chemical reactions, containing high concentrations of organic and inorganic reactants (Alvarez-Aviles et al., 2008; Bowman & Deming, 2010; Thomas A. Douglas et al., 2012), abundant reaction surfaces (Domine, Taillandier, Simpson, & Severin, 2005; Obbard, Roscoe, Wolff, & Atkinson, 2009), and sunlight as a source of energy. Although there has been much speculation regarding the role of frost flowers in chemical reactions (Alvarez-Aviles et al., 2008; T. A. Douglas et al., 2005; Kaleschke et al., 2004; Kalnajs & Avallone, 2006; Rankin, Wolff, & Martin, 2002; Simpson et al., 2007), direct observations in the natural environment are lacking due to the difficult nature of isolating frost flowers as a source of reaction products. Laboratory experiments in a controlled setting are desirable to further investigate these phenomena. Most apparatus for growing frost flowers in the laboratory, however, are open to the atmosphere and inadequately controlled for detailed observations of chemical reactions (Aslam et al., 2012; Bowman & Deming, 2010; Eronen-Rasimus et al., 2014; Martin, Yu, & Drucker, 1996; Müller et al., 2013; Roscoe et al., 2011; Robert W Style & Worster, 2009). We undertook the development of a closed reaction chamber, termed CRYO-FROG, to study photochemical reactions within frost flowers and related phenomena in an environmentally controlled fashion.

#### A3.2 Methods

CRYO-FROG (Fig. A3.1) consists of an off-the-shelf commercial grade chest freezer capable of cooling to  $-30\text{ }^{\circ}\text{C}$ . A 30 L HDPE plastic chamber serves as the reaction vessel and



closely follows the design of Style (2007). The reaction chamber, heavily insulated at the sides below water level, is warmed from the bottom by a low-wattage heating coil, and it contains an ethanol filled bladder connected to an external reservoir that enables the manipulation of water column pressure during ice formation. A double paned fused silica window in the lid of the reaction vessel allows the entry of ultraviolet light from an external Hg-vapor source.

Temperature and humidity sensors in the reaction chamber monitor *in situ* conditions.

Initial designs relied on an external chiller circulating cold ethanol through copper coils in the reaction chamber to achieve the desired temperature in the reaction chamber. Preferential freeze-out of moisture in the reaction chamber onto the coils however, made it impossible to reach the temperature/humidity regime required for frost flower growth. Freeze-out was reduced by replacing the coils with a heat exchanger inside the freezer but outside the reaction chamber, although this improvement did not measurably reduce the temperature achieved by using the freezing chamber alone.

### **A3.3 Results and Discussion**

Frost flowers can form via several different mechanisms. In one mechanism atmospheric moisture, derived from the brine skim at the young ice surface or nearby open water, freezes onto a nucleation point and initiates frost flower growth. In the other mechanism, described mathematically and verified experimentally by Style and Worster (2009), a strong temperature differential between the ice surface and atmosphere enables sublimation from the ice surface and immediate deposition on a nucleation point (Fig. A3.2). Under this model the factors that influence whether frost flowers will form are relative humidity and the temperature difference between the ice surface and atmosphere. With a temperature difference of 20 °C (e.g., an atmospheric temperature of –25 °C and a surface temperature of –5 °C) frost flowers should

form with a relative humidity as low as 0.6. As the relative humidity increases the temperature differential necessary to form frost flowers decreases; at a relative humidity of 0.8 only a 15 °C difference is required for frost flower formation.

Despite the simplicity of the Style model we had difficulty inducing frost flower formation under the prescribed temperature and humidity regime. These difficulties are most likely attributable to poor temperature control and monitoring at the ice surface, a problem that could be solved through the installation of a string of thermistors and relative humidity sensors, and by coupling the heat source with a temperature sensor at the young ice surface. Currently very careful temperature manipulation is required to maintain thin ice (and thus a relatively warm ice surface).

Frost flowers were successfully formed in CRYO-FROG when the atmospheric temperature was -25 °C approximately 10 cm from the ice surface, and the relative humidity was 0.8. Failure of the humidity sensor near the end of the project prevented a more detailed study of successful frost flower growth conditions.

To evaluate how much UV light was reaching the ice surface through our light path (consisting of two fused-silica windows, an IR filter, and a mirror) we measured the incident UV light between 240 and 270 nm through each optical element, and the combined element, at a distance equivalent to a typical distance between the light source and the ice surface. Our Hg source produced  $12.0 \mu\text{W cm}^{-2}$ , the filter alone reduced this amount to  $1.0 \mu\text{W cm}^{-2}$ , the mirror alone reduced this amount to  $3.0 \mu\text{W cm}^{-2}$ , together these elements and the fused silica windows reduced this amount to  $0.38 \mu\text{W cm}^{-2}$ . Although the mirror was not the greatest absorber of UV light, replacing the mirror with a UV-reflective element would be the easiest way to increase the incident UV light at the ice surface.

## **Acknowledgments**

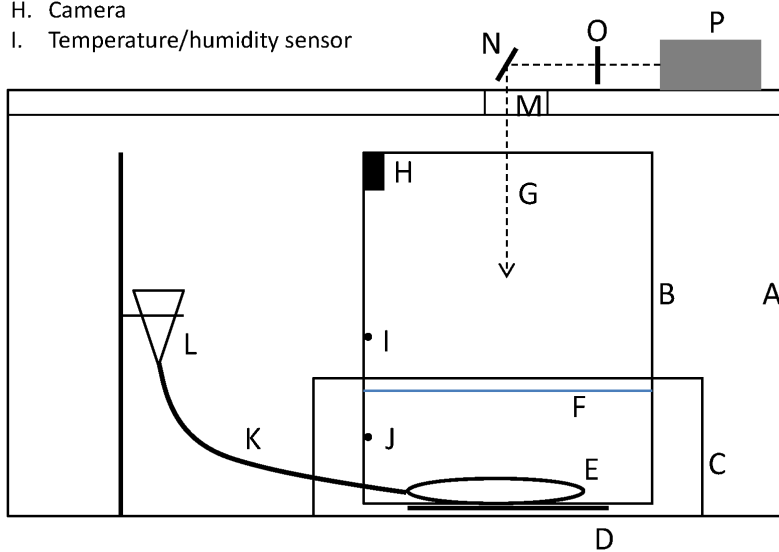
This work was supported by a NASA Astrobiology Institute DDF grant to JWD, through the NAI Virtual Planetary Laboratory node. JSB was supported by an NSF IGERT Fellowship through the University of Washington's Astrobiology Program.

## References

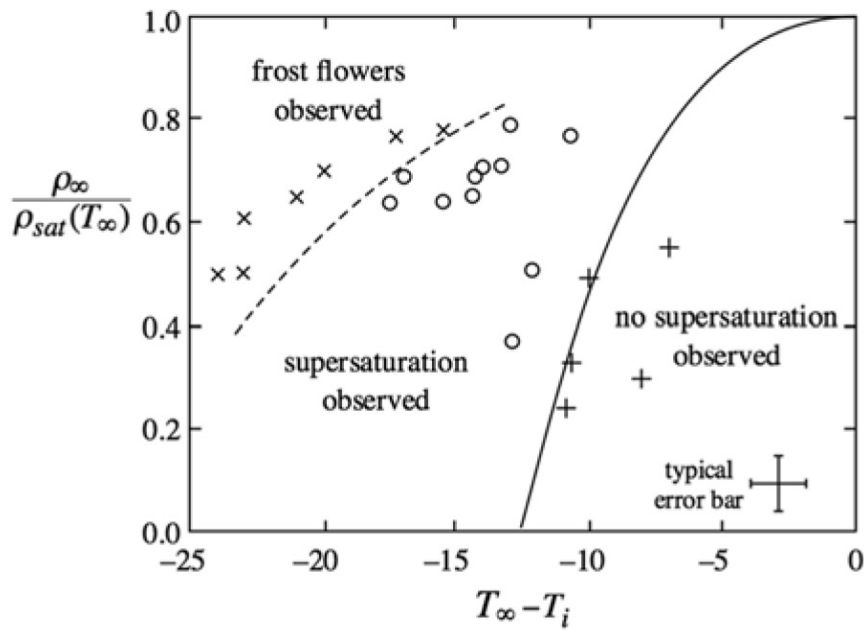
- Alvarez-Aviles, L., Simpson, W. R., Douglas, T. A., Sturm, M., Perovich, D., & Domine, F. (2008). Frost flower chemical composition during growth and its implications for aerosol production and bromine activation. *J. Geophys. Res.*, *113*. doi: 10.1029/2008jd010277
- Aslam, S., Underwood, G., Kaartokallio, H., Norman, L., Autio, R., Fischer, M., . . . Thomas, D. (2012). Dissolved extracellular polymeric substances (dEPS) dynamics and bacterial growth during sea ice formation in an ice tank study. *Pol. Biol.*, *35*(5), 661–676. doi: 10.1007/s00300-011-1112-0
- Bowman, J. S., & Deming, J. W. (2010). Elevated bacterial abundance and exopolymers in saline frost flowers with implications for atmospheric chemistry and microbial dispersal. *Geophys. Res. Lett.*, *37*(L13501).
- Domine, F., Taillandier, A. S., Simpson, W. R., & Severin, K. (2005). Specific surface area, density and microstructure of frost flowers. *Geophys. Res. Lett.*, *32*. doi: 10.1029/2005gl023245
- Douglas, T. A., Domine, F., Barret, M., Anastasio, C., Beine, H. J., Bottenheim, J., . . . Steffen, A. (2012). Frost flowers growing in the Arctic ocean-atmosphere-sea ice-snow interface: 1. Chemical composition. *J. Geophys. Res.*, *117*, D00R09. doi: 10.1029/2011jd016460
- Douglas, T. A., Sturm, M., Simpson, W. R., Brooks, S., Lindberg, S. E., & Perovich, D. K. (2005). Elevated mercury measured in snow and frost flowers near Arctic sea ice leads. *Geophys. Res. Lett.*, *32*. doi: 10.1029/2004gl022132
- Eronen-Rasimus, E., Kaartokallio, H., Lyra, C., Autio, R., Kuosa, H., Dieckmann, G. S., & Thomas, D. N. (2014). Bacterial community dynamics and activity in relation to dissolved organic matter availability during sea ice formation in a mesocosm experiment. *MicrobiologyOpen*.
- Kaleschke, L., Richter, A., Burrows, J., Afe, O., Heygster, G., Notholt, J., . . . Jacobi, H. W. (2004). Frost flowers on sea ice as a source of sea salt and their influence on tropospheric halogen chemistry. *Geophys. Res. Lett.*, *31*. doi: 10.1029/2004gl020655
- Kalnajs, L. E., & Avallone, L. M. (2006). Frost flower influence on springtime boundary-layer ozone depletion events and atmospheric bromine levels. *Geophys. Res. Lett.*, *33*. doi: 10.1029/2006gl025809
- Martin, S., Yu, Y., & Drucker, R. (1996). The temperature dependence of frost flower growth on laboratory sea ice and the effect of the flowers on infrared observations of the surface. *J. Geophys. Res.*, *101*. doi: 10.1029/96jc00208
- Müller, S., Vähätalo, A. V., Stedmon, C. A., Granskog, M. A., Norman, L., Aslam, S. N., . . . Thomas, D. N. (2013). Selective incorporation of dissolved organic matter (DOM) during sea ice formation. *Mar. Chem.*, *155*, 148–157.

- Obbard, R. W., Roscoe, H. K., Wolff, E. W., & Atkinson, H. M. (2009). Frost flower surface area and chemistry as a function of salinity and temperature. *J. Geophys. Res.*, *114*(D20), D20305. doi: 10.1029/2009jd012481
- Rankin, A. M., Wolff, E. W., & Martin, S. (2002). Frost flowers: Implications for tropospheric chemistry and ice core interpretation. *J. Geophys. Res.*, *107*. doi: 10.1029/2002jd002492
- Roscoe, H. K., Brooks, B., Jackson, A., Smith, M., Walker, S., Obbard, R. W., & Wolff, E. W. (2011). Frost flowers in the laboratory: Growth, characteristics, aerosol, and the underlying sea ice. *J. Geophys. Res-Atmos.*, *116*, D12301.
- Simpson, W. R., Carlson, D., Hönninger, G., Douglas, T. A., Sturm, M., Perovich, D., & Platt, U. (2007). First-year sea-ice contact predicts bromine monoxide (BrO) levels at Barrow, Alaska better than potential frost flower contact. *Atmos. Chem. Phys.*, *7*(3), 621-627. doi: 10.5194/acp-7-621-2007
- Style, R. W. (2007). *The formation and evolution of frost flowers and related phenomena*. (PhD), St. Catharine's College, Cambridge.
- Style, R. W., & Worster, M. G. (2009). Frost flower formation on sea ice and lake ice. *Geophys. Res. Lett.*, *36*(11).

- |                                |  |
|--------------------------------|--|
| A. Chest freezer               | J. Temperature sensor                  |
| B. Reaction chamber            | K. Bladder feed line                   |
| C. Insulation                  | L. Height-adjustable ethanol reservoir |
| D. Heating element             | M. Fused-silica port in freezer lid    |
| E. Ethanol bladder             | N. UV reflective mirror                |
| F. Water level                 | O. Infrared filter (heat trap)         |
| G. Light path                  | P. Hg-vapor light source               |
| H. Camera                      |  |
| I. Temperature/humidity sensor |  |



**Fig. A3.1. CRYO-FROG schematic.**



**Fig. A3.2. Frost flower growth conditions.** Taken from Style and Worster (2009) and reproduced with permission. The figure describes empirical observations (x's, crosses, circles) and model results (dotted and solid line). Frost flowers growth conditions were predicted as shown by the dotted line; observed frost flower growth conditions are shown by the x's. Y-axis is relative humidity; x-axis is the temperature difference between the atmosphere and the ice surface.

## Appendix 4

### Curriculum vitae (as of August 13, 2014)

#### Jeff Bowman

Lamont-Doherty Earth Observatory  
PO Box 1000  
Palisades NY, USA 10964-8000  
bowmanjs@ldeo.columbia.edu  
+1 425 753 3735

<b>Current Position:</b>	<b>Postdoctoral Fellow.</b> Lamont-Doherty Earth Observatory, Columbia University. 2014 to present. <b>Research Associate.</b> Blue Marble Space Institute of Science. 2013 to present.
<b>Degrees Awarded:</b>	<b>Doctor of Philosophy, Oceanography and Astrobiology.</b> University of Washington, 2014. <b>Master of Science, Oceanography.</b> University of Washington, 2010. <b>Bachelor of Science, Biological Oceanography.</b> University of Washington, 2008. <b>Associate of Science with Honors.</b> Bellevue Community College, 2006.
<b>Additional Education:</b>	<b>University of Hawaii Astrobiology Winter School.</b> University of Hawaii, 2011. <b>Program on Climate Change Summer Institute.</b> University of Washington, 2010. <b>Biological Sequence Analysis.</b> Danish Technical University, 2010. <b>Nordic-NASA Astrobiology Summer School.</b> University of Iceland, 2009.
<b>Research Experience:</b>	<b>Postdoctoral Research Scientist.</b> November 2014 to present. Department of Biology and Paleo Environment, Lamont-Doherty Earth Observatory, Columbia University. Supervisor: Hugh Ducklow, PhD, Professor. <b>Graduate Research Assistant.</b> September 2008 to August 2014. School of Oceanography and Astrobiology Program, University of Washington. Supervisor: Jody Deming, PhD, Professor. <b>Undergraduate Research Assistant.</b> August 2006 to September 2008. School of Oceanography, University of Washington. Supervisor: Julian P. Sachs, PhD, Associate Professor.
<b>Field Experience:</b>	<b>Laurence M. Gould, LMG1401.</b> January-February 2014. <b>Barrow, Alaska.</b> May 2013. <b>McMurdo, Antarctica.</b> August-November 2011.



**Barrow, Alaska.** March-April 2011.  
**Barrow, Alaska.** April 2010.  
**Barrow, Alaska.** February 2010.  
**Oden, LOMROGII.** August-September 2010.  
**Barrow, Alaska.** April 2009.  
**Cargill Saltworks.** June 2008.  
**Canada Salt Lakes.** June 2007.

**Publications:**

- Bowman, JS and Deming, JW. **Alkane hydroxylase genes in psychrophile genomes and the potential for cold active catalysis.** Submitted to *BMC Genomics*.
- Bowman, JS and Deming, JW. **Amino acid preferences in the proteomes of psychrophilic bacteria: Is serine the answer to enzyme adaptation to low temperature and high salinity?** Submitted to *BMC Genomics*.
- Miller L, Bowman JS, Brown K, Collins ER, Else B, Ewert M, Fransson A, Fripiat F, Gosselin M, Lannuzel D, Meiners K, Michel C, Nishioka J, Nomura D, Papadimitriou S, Russell L, Sørensen L, Thomas D, Tison JL, A van Leeuwe M, Vancoppenolle M, Wolff E, Zhou J. **Methods for biogeochemical studies of sea ice: Where we are and where we are going.** Submitted to *Elementa: Science of the Anthropocene*.
- Ehn J, Pucko M, Rysgaard S, Deming J, Bowman JS, Papakyriakou T, Galley R, Sogaard D. **Frost flower on young sea ice: The climatic, chemical, and microbial significance of an emerging ice type.** Submitted to *JGR Atmospheres*.
- Hauptmann A, Stibal M, Baelum J, Sicheritz-Pontén T, Brunak S, Bowman JS, Hansen L, Jacobsen C, Blom N. **Bacterial diversity in snow on North Pole ice floes.** Available online in advance of print.
- Bowman JS, Berthiaume C, Armbrust V, Deming JW. **The genetic potential for key biogeochemical processes in Arctic frost flowers and young sea ice revealed by metagenomic analysis.** *FEMS Microbiol Ecol*, 2014. 89: 376–387. doi: 10.1111/1574-6941.12331.
- Bowman JS, Larose C, Vogel T, Deming JW. **Selective occurrence of Rhizobiales on the surface of young sea ice near Barrow, Alaska and distribution in the polar marine rare biosphere.** *EMIR*, 2013. 5: 575–582. doi: 10.1111/1758-2229.12047.
- Stüeken EE, Anderson RE, Bowman JS, Brazelton WJ, Colangelo-Lillis J, Goldman AD, Som SM, Baross JA. **Did life originate in a global chemical reactor?** *Geobiology*, 2013. 11: 101–126. doi: 10.1111/gbi.12025.
- Wietz M, Mansson M, Bowman JS, Blom N, Ng Y, Gram L. **Wide distribution of closely related, antibiotic-producing strains throughout the Arctic Ocean.** *Appl Env Microbiol*, 2012. 78:6. doi:10.1128/AEM.07096-11.
- Bowman JS, Rasmussen S, Blom N, Deming JW, Rysgaard S, Sicheritz-Pontén T. **Microbial community structure of Arctic multiyear sea ice and surface seawater by 454 sequencing of the 16S RNA gene.** *ISME J*, 2011. 6. 11–20.

doi:10.1038/ismej.2011.76.

Bowman JS and Deming JW. ***Elevated bacterial abundance and exopolymers in saline frost flowers and implications for atmospheric chemistry and microbial dispersal.*** *Geo Res Let*, 2010. 37. L13501. doi:10.1029/2010GL043020.

Bowman JS and Sachs J. ***Chemical and physical properties of some saline lakes in Alberta and Saskatchewan.*** *Saline Systems*, 2008. 4:3.

#### **Conferences and Presentations:**

Bowman JS, Deming JW, Collins ER. ***Increased rates of horizontal gene transfer in psychrophilic genomes and potential links to the Phanerozoic climate record.*** Oral presentation. August, 2014. ISME15, Seoul, S. Korea.

Bowman JS, Deming JW. ***Prevalence of horizontal gene transfer among cold adapted microbes and implications for crude oil bioremediation in the polar environment.*** Poster presentation. September, 2013. Polar and Alpine Microbiology Conference, Big Sky, MT.

Bowman JS, Berthiaume CT, Armbrust VE, Deming JW. ***Metagenomic analysis of an unexpected community of widely distributed Rhizobiales in Arctic frost flowers.*** Poster Presentation. March, 2013. Gordon Research Conference on Polar Marine Science, Ventura, CA.

Bowman JS and Deming JW. ***The distribution and cold adapted character of putative hydrocarbon degradation genes among the genomes of cold adapted microbes.*** Oral presentation. January, 2013. Gulf Oil and Ecosystems Conference, New Orleans, LA.

Deming JW, Bowman JS, Ewert M, Collins RE. ***Microbial life at the boundary between sea ice and atmosphere.*** Poster presentation. April, 2012. IPY, Montreal, Quebec, Canada.

Bowman JS, Larose C, Vogel T, Deming JW. ***Evidence for strong selective enrichment of bacteria within frost flowers on the surface of Arctic sea ice.*** Oral presentation. February, 2012. Ocean Sciences Meeting, Salt Lake City, UT.

Bowman JS, Schmidt BE, Blankenship DD. ***Life in Ice: Defining the habitability of Europa.*** Poster Presentation. October, 2011. AAS Division of Planetary Science Annual Meeting, Nantes, France.

Bowman JS, Chan KYK, Durkin C, Hennon G, Smith D, Sullivan B. ***Is diversity related to service provision across an ecosystem? An estuarine case study.*** Oral presentation. September, 2011. World Conference on Marine Biodiversity, Aberdeen, Scotland.

Bowman JS and Deming JW. ***Integrating sequence based biology with biogeochemical and physical studies.*** Poster presentation, June 2011. OASIS Science Meeting, Telluride, CO.

Bowman JS, Rasmussen S, Blom N, Deming JW, Thomas SP. ***Determining microbial community structure in MYI with mid-depth sequencing.*** Oral presentation, March 2011. Gordon Research Seminar on Polar Marine Science, Ventura, CA.

Bowman JS, Rasmussen S, Blom N, Deming JW, Thomas SP. ***Determining***

***microbial community structure in MYI with mid-depth sequencing.*** Poster presentation, March 2011. Gordon Research Conference on Polar Marine Science, Ventura, CA

Bowman JS. Planets and Pedagogy: ***An astrobiology course for science educators.*** Poster presentation, January 2011. UH NAI Winter School, Hilo, HI.

Bowman JS and Schmidt B. ***Putting the biology back in astrobiology for EJSM: How ice penetrating radar can define key habitat parameters.*** Poster presentation, December 2010. American Geophysical Union Annual Meeting, San Francisco, CA.

Bowman JS. ***A possible role for surface ice features in the origin of life.*** Oral presentation, June 2010. ABGRADCON, Tällberg, Sweden.

Bowman JS and Deming JW. ***A novel method for measuring microbial respiration at very low temperatures.*** Poster presentation, April 2010. Astrobiology Science Conference, Houston, TX.

Bowman JS. ***Implications of a hydrated gel on the sea surface microlayer for prebiotic chemistry and the origin of life.*** Poster presentation, January 2010. Gordon Research Conference on the Origin of Life, Galveston, TX.

Bowman JS and Deming JW. ***Elevated bacterial abundance in laboratory grown and naturally occurring frost flowers under late winter conditions.*** Poster presentation, December 2009. American Geophysical Union Annual Meeting, San Francisco, CA.

Bowman JS. ***Extracellular polymers produced in abundance by microorganisms in saline ice: A potential biomarker for mid-IR spectroscopic observations of Europa?*** Oral presentation, July 2009. ABGRADCON, Seattle, WA.

Bowman JS and Deming J. ***Marine frost flowers as analogues for extraterrestrial environments of very low temperature and low water activity.*** Poster presentation, June 2009. NASA-Nordic Astrobiology Summer Program, Reykjavik, Iceland.

Bowman JS and Deming J. ***Microbial responses to sea ice conditions encountered during an arctic winter.*** Poster presentation, November 2008. IMPETUS, St. Petersburg, Russia.

Bowman JS and Sachs J. ***Developing an indicator of salinity from membrane lipids in halophilic microorganisms.*** Oral presentation, June 2008. American Society of Limnologists and Oceanographers Summer Meeting, St. John's, Newfoundland.

**Teaching:** **Seminar Leader**, University of Washington. *OCEAN 599, Bioinformatics Seminar.* 2012.

**Instructor**, University of Washington. *BEDUC 592, Fundamentals of Astrobiology for Science Educators.* 2010.

**Teaching Assistant**, University of Washington. *ASTBIO 115, Introduction to Astrobiology.* 2009.

**Professional Service:** **Chair**, Gordon Research Seminar Polar Marine Science. March 2013, Ventura, CA.  
**Reviewer**, *Applied and Environmental Microbiology*  
**Reviewer**, *Microbial Biotechnology*  
**Reviewer**, *JGR Atmospheres*

**Outreach:** **Science Advisor**. *Wild Canada*. May, 2013.  
**Science Advisor**. The Weather Channel. *Earth's Strangest Weather: Arctic Blooms*. March, 2013.  
**National Science Bowl Question Reviewer**. December 2011.  
**Invited Presenter, Ballard High School**. October 2010.  
**National Science Bowl Question Writer**. August 2010, December 2010, May 2011, June 2013.  
**Invited Presenter**, Novozymes. June 2010.  
**Science Communication Fellow**, Pacific Science Center. 2010.  
**Invited Panelist**, Seattle Central Community College. September 2009.  
**Invited Presenter**, Garfield High School. September 2009.  
**Presenter**, Theodore Jacobsen Observatory. September 2007 to present.  
**Scientist Like Me**, Pacific Science Center. June 2009 to September 2009.  
**Invited Presenter**, University of Washington Department of Education. December 2009.

**Contract Work:** **North Central School District Math-Science Partnership**. Instructor. Wenatchee, WA. August 2011.  
**North Central School District Math-Science Partnership**. Instructor. Wenatchee, WA. August 2010.

**Military Service:** **2<sup>nd</sup> Airborne Ranger Battalion**. Infantry mortar team leader, E5 (last position held). Ft. Lewis, WA. August 1999 to December 2003.

**Awards:** **EPA STAR Fellowship**. June 2011.  
**Moore Foundation**. March 2010.  
**Soffen Memorial Fund**. December 2009  
**Allen Library Research Award**. June 2008.  
**National Science Foundation IGERT Fellowship**. September 2008-June 2011.  
**National Science Foundation Graduate Research Fellowship Honorable Mention**. April 2008.  
**Washington Sea Grant Science Writing Fellowship**. April, 2008.  
**Levinson Emerging Scholar**. November 2007.

**Mary Gates Scholar.** November 2007.

**Dean's List.** Winter 2007, Spring 2007, Fall 2007, Spring 2008. College of Ocean and Fisheries Science, University of Washington.

**Societies:**

**International Society of Microbial Ecologists.** 2014.

**American Association for the Advancement of Science.** 2009.

**American Geophysical Union.** 2009.

**Association of Polar Early Career Scientists.** 2008.

**American Society of Limnologists and Oceanographers.** 2008.

**Phi Sigma Theta.** 2007.

**Phi Theta Kappa.** 2004.