Microbial Community Composition of Lake Sediment in the High Arctic

by Brittany M. Berdy

B.A. in Biology, Skidmore College
M.S. in Biology, Northeastern University

A dissertation submitted to

The Faculty of
the College of Science of
Northeastern University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

April 20, 2017

Dissertation directed by

Slava Epstein
College Distinguished Professor
Acknowledgements

I would like to thank my advisor, Dr. Slava Epstein, who has helped me grow as a person and scientist, and who has always been there to provide helpful guidance and advice, and supported through many wonderful learning opportunities.

I would also like to thank my committee members for their valuable feedback, discussions, and advice: Dr. Veronica Godoy-Carter, Dr. Win Chai, Dr. Kim Lewis, and Dr. Karen E. Nelson.

A special thank you to the J. Craig Venter Institute and our collaborators Chris Dupont, Manny Torralba, Alex Richter, and Drishti Kaul; you were instrumental throughout the entire project, particularly all genomic data and analyses. Thank you for always making yourself available to answer my questions and provide advice.

We could have never done this project without the entire CH2M Hill Polar Services Team, whose expertise, planning, and advice allowed for travel to Greenland.

The entire Epstein Lab, past, visiting, and present members, have all been amazing: encouraging me and providing both emotional and scientific support. I want to extend my thanks to all members, with a special mention of Dawoon Jung, Maria Sizova, Paul Muller, Sebastian Doerfert, Eva Wunschel, Yoshi Aoi, Alina Eilers, Svenja Grobe, Charlie Fox, Alexander Sousa, Claire Williams, Ethan Edson, Josh Timmons, and Thao Nyugen who were all active members of the Greenland Project, culminating with this dissertation. I also want to thank David Stanczyk, Violetta Medik, and Ellen Taylor for their continued support in and outside of the lab.

I want to thank all my friends, who have put up with me during this stressful time, and have never left my side, continued to encourage me, and have truly become my family. I especially want to thank those whom I met in my travels to Thule, Greenland. I also want to thank those who helped review my thesis: Dan Monroe, Bijaya Sharma, Amanda Hazen, David and Aimee Ward, and Harrison Wolnick.

My entire family has been amazing at supporting me during this time, and each and every one of them deserves my gratitude. Thank you to my sisters, Sherry and Kim, and my brothers-in-law, for your encouragement. And to Paige Regan, who has been like a sister to me my entire life, thank you for always making me giggle!
A very special thank you to my wonderful boyfriend, Joe Baker, who has always believed in me and been by my side, even during the most stressful times, and provided me with unconditional love and support.

And lastly, I want to thank the three people I owe everything to, the people who have been absolutely instrumental in my successes both in science and life: my parents Alice and Jack Berdy, and James Mosel. I could have never made it this far without your unconditional love and unwavering support. You have fielded more frantic phone calls and late night panic attacks than I can possibly count. James, you have been by my side for the last 12 years, and have made me a stronger, better person. You will always be my family. Mom and Dad, you are my best friends, my heroes, and I owe everything to you. I promise to work the rest of my life to make you both proud, every day. I love you both! Mom, Dad, and James you believed in me in moments when I didn’t believe in myself, and you are the reason I have always, and will always continue to strive to be my best self.
Abstract of Dissertation

The holy grail of microbial ecology would be to know what species are present and active in a community, what functions they are performing, and at what point in time. In recent years it has become clear that environmental meta-omic approaches are essential to gain insight into microbial communities; however the importance of cultivation cannot be overlooked. High quality reference genomes are necessary for the interpretation of meta-omic data, and these can only come from pure cultures. In order to contribute to the ever-growing body of work investigating microbial ecology, and to assess current methods commonly used, we studied the composition of a microbial community within lake sediment in the High Arctic (Thule, Greenland) using a multi-faceted approach. My specific aims were to:

1. Utilize cultivation-independent molecular approaches to define microbial community structure and identify potentially active organisms
2. Utilize culture-dependent approaches to create a comprehensive culture collection
3. Analyze and synthesize the data obtained from Aims 1 and 2 and assess the biological relevance of cultured organisms

The community appeared relatively rich and stable over time based upon a 16S rRNA gene survey. However, analysis of the expressed 16S rRNA genes indicated that activity of operational taxonomic units (OTUs) within the sediment community was dynamic, even across very small distances. Additionally, the most active members were not represented in any database to date.

Despite increasing advancements in microbial cultivation, our results emphasize the need to develop new tools and techniques. A total of 1173 strains were cultivated, but only a small fraction was detected as active (<7%), or even present in the community (<12%) based on 16S gene surveys. We conclude that the variation in OTU activity between replicate samples, and the low abundance and activity of OTUs cultured, are a result of microheterogeneity in spatial
distribution within microbial communities. Thus, samples taken even millimeters apart were actually from potentially different communities. The results of this dissertation highlight the increasing need to study microbial communities on a scale relevant to microbial life.
Table of Contents

Acknowledgements ................................................................................................................................. ii
Abstract of Dissertation ............................................................................................................................ iv
List of Figures ........................................................................................................................................... viii
List of Tables ........................................................................................................................................... xi
List of Abbreviations ............................................................................................................................... xii
Chapter 1. Introduction .............................................................................................................................. 1
  1.1 Microbiology Through Cultivation ................................................................................................. 2
  1.2 The Genomic Approach .................................................................................................................. 3
  1.3 A Synergistic Approach .................................................................................................................. 5
  1.4 Dissertation Aims ........................................................................................................................... 6
Chapter 2. Investigating the Lake Community using Culture-Independent Approaches ................. 8
  2.1 Introduction ....................................................................................................................................... 9
  2.2 Methods .......................................................................................................................................... 11
  2.3 Results and Discussion .................................................................................................................... 14
    2.3.1 Taxonomic Composition Overview by 16S rDNA sequencing .................................................. 14
    2.3.2. Comparison of Taxa Present According to site and date .......................................................... 17
    2.3.3 16S rRNA Sequencing and Taxonomic Activity ................................................................. 21
    2.3.4 Patterns of Activity as Determined by 16S rRNA Sequencing .............................................. 24
    2.3.5 Unclassified OTUs: Phantom Taxa and Future Experimentation ........................................... 27
    2.3.6 Comparing OTU abundance (rDNA libraries) and Activity (rRNA libraries): Are the Most Rare
        the Most active? .............................................................................................................................. 29
    2.3.7 The Importance of Microheterogeneity ..................................................................................... 33
Chapter 3. Culturing the Lake Community ............................................................................................... 35
  3.1 Introduction ....................................................................................................................................... 36
  3.2 Cultivation-Dependent Methods .................................................................................................... 39
    3.2.1 Sampling Procedure for cultivation .......................................................................................... 39
    3.2.2. Cultivation conditions ............................................................................................................ 41
    3.2.3 Anaerobic and enrichment cultivation ..................................................................................... 41
    3.2.4 Standard Cultivation ................................................................................................................ 42
    3.2.5 In situ cultivation ....................................................................................................................... 42
    3.2.6 Harvesting Growth from In Situ Devices and Laboratory Cultivation ..................................... 46
    3.2.7 Isolation, Identification, and Downstream Analysis of Cultivated Species .............................. 46
  3.3 Results and Discussion ..................................................................................................................... 48
    3.3.1 Overview of Cultivated Microorganisms ................................................................................ 48
    3.3.2 Comparison of Standard, Anaerobic, and In Situ Cultivation Approaches ............................. 53
    3.3.3 Comprehensive Comparison of In Situ Cultivation Methods ................................................... 59
    3.3.4 Novelty of Isolates .................................................................................................................... 66
Chapter 4. Biological Relevance of Culture Isolates: Comparing Culture-independent and Culture-dependent Data ................................................................. 69
4.1 Cultivation of OTUs detected using Culture-Independent Methods ................................ 70
4.2 Cultivation of Phantom Taxa ................................................................................ 72
4.3 Abundance and Activity of Cultured Isolates ............................................................ 73
4.4 Final Conclusions .................................................................................................. 77
References .................................................................................................................. 80
List of Figures

Figure 1. Microbial community composition as determined by a 16S rDNA gene survey. Taxonomy was determined at the phylum level. Data from all libraries and time points was averaged to generate percentages. Inset: breakdown of Proteobacteria at the class level. 

Figure 2. Beta diversity analysis for lake sediment samples (DNA libraries) from the High Arctic. Principal coordinates analysis (PCoA) of Bray Curtis distances was focused on the 29 sediment samples from DNA libraries. Red = RL1, Blue = RL2. The amount of variability explained by all three principal coordinates (PCs) is shown on the axes. (Kaul and Dupont)

Figure 3. Comprehensive taxonomic breakdown of RL1 and RL2 samples with respect to different dates. Taxa abundance values for top ten most prevalent taxa at the class level for different time points are indicated. The remaining taxa are pooled into an additional taxon labeled “Other.” (Kaul and Dupont)

Figure 4. Comprehensive taxonomic breakdown of Rich Lake 1 samples with respect to different dates. Taxa abundance values for top ten most prevalent taxa at the class level for different time points for 16 rDNA samples from Rich Lake 1. The remaining taxa are pooled into an additional taxon labeled “Other.” (Kaul and Dupont)

Figure 5. Microbial community composition as determined by 16S rRNA sequencing. Taxonomy was determined at the phylum level averaging all libraries.

Figure 6. Unclassified sequences cluster within three main OTUs in the 16S rRNA libraries. 25% of all rRNA sequences were unclassified. Out of these sequences, rRNA_OTU_113 accounts for 10% of the total sequences detected and 41% of all unclassified OTU sequences.

Figure 7. Beta diversity analysis for the lake sediment samples from High Arctic. Principal coordinates analysis (PCoA) of Bray Curtis distances was done for the sediment samples from a lake in Thule, Greenland. The DNA Libraries are in red (right) and the RNA libraries (left) in blue. There are 29 samples from DNA libraries and 16 samples from RNA libraries. The amount of variability explained by all three principal coordinates (PCs) is shown on the axes. (Kaul and Dupont)

Figure 8. Comprehensive taxonomic breakdown of Rich Lake 1 samples with respect to different dates. Taxa abundance values for top ten most prevalent taxa at the order level for different time points for six rRNA samples from Rich Lake 1. The remaining taxa are pooled into an additional taxon labeled “Other”. Unclassified eukaryotic sequences have been removed from analysis (Dupont and Kaul).

Figure 9. Comprehensive taxonomic breakdown of Rich Lake 2 samples with respect to different dates. Taxa abundance values for top ten most prevalent taxa at the order level for different time points for 10 rRNA samples from Rich Lake 2. The remaining taxa are pooled into an additional taxon labeled “Other”. Unclassified eukaryotic sequences have been removed from analysis (Dupont and Kaul).

Figure 10. Abundance of the 17 active OTUs at Rich Lake 1. Seventeen OTUs were active at RL1. The percent each OTU contributed to rDNA (blue) and rRNA (red) libraries. Most OTUs were rarely detected in rDNA libraries, indicating low abundance. Phantom OTUs detected only in RNA libraries are labeled as rRNA_OTU_#.
Figure 11. Abundance of the 16 active OTUs at Rich Lake 2. Sixteen OTUs were active at RL2. The percent each OTU contributed to rDNA (blue) and rRNA (red) libraries. Most OTUs were rarely detected in rDNA libraries, indicating low abundance. Phantom OTUs detected only in RNA libraries are labeled as rRNA_OTU_#.32

Figure 12. Concept of ipore device. Figure from (Tandogan 2016).39

Figure 13. Design and application of the (A) diffusion chamber and (B) microbial trap. Figure from (Epstein et al. 2010).43

Figure 14. Schematic of a growth chamber in a FPMT. Figure from (Jung et al. 2013).44

Figure 15. Schematic of the itip. Figure from (Jung et al. 2014).45

Figure 16. Photograph of an ipore device. Photo from (Tandogan 2016).45

Figure 17. Overlap of OTUs cultured by each approach. N=number of total isolates cultivated by that method. Note: all in situ methods were treated as one approach, resulting in a total of 722 isolates.54

Figure 18. Phylum break down for standard compared to in situ cultivation methods. All values were normalized by dividing the number of isolates in a phylum cultured using a specific approach by the total number of isolates in that phylum.55

Figure 19. Distribution of recovered isolates, by order, for in situ, standard, and anaerobic methods. Percent indicates the proportion of isolates obtained within that order. All values were normalized by dividing the number of isolates in a phylum cultured using a specific approach by the total number of isolates in that phylum.56

Figure 20. Venn diagram depicting the overlap of OTUs cultivated by in situ methods. Numbers within Venn diagram indicate OTUs unique to/shared between approaches. N=total number of OTUs cultivated using the given approach.60

Figure 21. Phyla cultured using in situ approaches. Taxonomic preferences can be observed at the phylum level. Firmicutes and eukaryotes were only cultivated by two methods (each), while Proteobacteria were well represented in all approaches. All values were normalized by dividing the number of isolates in a phylum cultured using a specific approach by the total number of isolates in that phylum.61

Figure 22. Orders cultivated via in situ methods. Taxonomic preferences can be see, such as exclusive cultivation of Neisseriales by the itip. Orders without any bars were cultivated by standard approaches only. DC=diffusion chamber, T=trap, FP=filter plate, I=itip, IP=ipore. All values were normalized by dividing the number of isolates in a phylum cultured using a specific approach by the total number of isolates in that phylum.61

Figure 23. Percent of isolate collection (by method) that matched to a LCA within the indicated percent similarity.67

Figure 24. OTUs common to all methodologies. N=total number of OTUs detected via each approach. All OTUs detected at least one time were counted to include “rare” taxa. OTUs were clustered at 97% sequence similarity. rDNA and rRNA library sequences were ~300 bp. Cultured isolate sequences were on average 772 bp.71

Figure 25. Abundance of cultured OTUs at Rich Lake 1. 151 cultivated OTUs were also detected as present in rDNA libraries at RL1. None of the cultivated OTUs comprised more than 3.5% of any rDNA library.74

Figure 26. Abundance of cultured OTUs at Rich Lake 2. 148 cultivated OTUs were also detected as present in DNA/RNA libraries at RL2. No cultured OTU comprised more than 5% of any rDNA library, except rDNA_OTU_28 during June 27 (12%).75
Figure 27. Cultured OTUs active at Rich Lake 1. A total of 55 cultivated OTUs were also deemed active at RL1, and OTU contributed, at most, 3% to sequences from rRNA libraries.

Figure 28. Cultivated OTUs active at Rich Lake 2. A total of 45 active OTUs were cultivated at RL2. Sequences from cultivated OTUs contributed, at most, 7% to sequences from rRNA library.
List of Tables

Table 1. Sediment samples collected for 16S rDNA and rRNA surveys. Each sample was processed as one library, and all libraries underwent quality control. Those that passed control, as indicated in the table, were used for downstream analyses.................................................. 12

Table 2. Summary of isolates cultivated by each method. Date of device set up = date when in situ devices were placed in the environment. Date of collection = date the sediment was collected and plated or the date the in situ device was processed. *Note that due to windy conditions, a 1-gallon Tupperware container filled with a layer of sediment was used for incubation of the filter plate and itip devices for a portion of in situ cultivation. ................. 40

Table 3. Phylum level breakdown of cultivated isolates from the summer of 2014. Proteobacteria are further divided into classes. .............................................................................. 48

Table 4. Order level breakdown of cultivated isolates (N=24). ............................................................. 50

Table 5. Summary of isolates cultivated, as broken down by genus (N=89). ...................................... 51

Table 6. Diversity indices of standard and in situ collections. Calculations by PAST3 software (Hammer et al. 2001). ........................................................................................................ 54

Table 7. Diversity indices of in situ approaches. Calculations performed using Past3 software (Hammer et al. 2001). .......................................................................................................................... 60

Table 8. Summary of cultivation preferences for different taxonomic divisions. ......................... 65

Table 9. Novelty Ratios. Fungal isolates excluded from calculations ..................................................... 68
List of Abbreviations

16S rDNA – 16S rRNA gene
16S rRNA – cDNA of 16S rDNA gene
bp – base pairs
CO₂ – carbon dioxide
DC – diffusion chamber
DNA – deoxyribonucleic acid
FISH – fluorescent in situ hybridization
FPMT/FP – filter plate microbial trap/filter plate
g – gram
H₂ – hydrogen
I – itip
IP – ipore
JCVI – J. Craig Venter Institute
LCA – least common ancestor
MAR – microautoradiography
ml – milliliter
mm – millimeter
NA – nutrient agar
NGS – next generation sequencing
OTU – operation taxonomic unit
PBS – phosphate buffered saline
PC – principal component
PCoA – principal coordinate analysis
PCR – polymerase chain reaction
PDMS - polydimethylsiloxane
PVDF – polyvinylidene difluoride
RL1 – Rich Lake 1
RL2 – Rich Lake 2
RNA – ribonucleic acid
rRNA – ribosomal RNA
SE – soil extract
SINA – SILVA incremental aligner
T – trap
μl – microliter
Chapter 1. Introduction
1.1 Microbiology Through Cultivation

Microbes have survived, and thrived, on our planet for at least 3.5 billion years, and yet we know surprisingly little about these organisms. The field of microbiology is relatively new. Antoni van Leeuwenhoek was the first to visualize bacteria through a hand built microscope in 1675. It took until the late 19th century for scientists to develop and employ universal cultivation techniques, as we know them today (i.e. nutrient media in gelling agents solidified within petri dishes). Cultivation techniques have continued to advance, but we are still greatly limited by “uncultivability” of the majority of microbial species.

The phenomenon of uncultivability was first noted in 1898, when Heinrich Winterberg observed the number of microbial cells in his samples was inconsistent with the number of colonies formed on nutrient media (Winterberg 1898). In 1911, it was again noted that the number of microscopically observed cells outnumbered cultivable cells by almost 150 times (Amann 1911). This phenomenon was continuously observed, became widely recognized, and eventually was termed the “great plate count anomaly” (Jannasch & Jones 1959; Staley & Konopka 1985). With the advent of the rRNA approach (Woese 1987; Olsen et al. 1986), the extent of uncultured diversity became clear (Michael S Rappé & Giovannoni 2003; Giovannoni et al. 1990; Uroz et al. 2013; Venter et al. 2004; Ward et al. 1990). In 1998 only 13 out of 36 known phyla had cultivated representatives (Hugenholtz et al. 1998), and the realization of this “uncultivated majority” led to a resurgence of cultivation efforts, with attempts to break convention and devise new methods to bridge the gap. These methods included single-cell and high-throughput strategies (Connon & Giovannoni 2002; Rappé et al. 2002; Nichols et al. 2010), increased incubation length and decreased nutrient concentration (Davis et al. 2005), and

1 The term “uncultivable microorganisms” is widely used but may be misleading: since they do grow in nature, they are better defined as “yet-to-be-cultivated.” It is in this latter sense, and solely for the sake of brevity, that we use the term “uncultivable” or “uncultured” here.
simulation of the natural environment (Aoi et al. 2009; Stevenson et al. 2004; Ferrari et al. 2005). By 2003, 26 of 52 known phyla had cultured representatives (Rappé & Giovannoni 2003), and currently, a new view of the tree of life indicates there may be as many as 92 named bacterial phyla, with at least half lacking cultured representatives (Hug et al. 2016).

In 1993 Button et al. proposed dilution culture, incubating bacteria with low concentration of nutrients, as a method to study marine bacteria (Button et al. 1993). Connon & Giovannoni utilized this approach, which increased microbial recovery over traditional cultivation approaches, and importantly facilitated the isolation of four unique cell lineages (Connon & Giovannoni 2002). Around the same time, the idea of in situ cultivation was born with the development of the diffusion chamber (Kaeberlein, Lewis, Epstein, et al. 2002). This method has been further expanded upon, and research has demonstrated that in situ cultivation can increase microbial recovery up to 40% (Bollmann et al. 2010), and the richness and novelty of recovered isolates is often greater than that obtained via standard cultivation (Kaeberlein, Lewis, Epstein, et al. 2002; Gavrish et al. 2008; Bollmann et al. 2010; Bollmann et al. 2007; Aoi et al. 2009). These advancements have led to exciting new discoveries, including the discovery of Teixobactin, the first member of a new class of peptidoglycan synthesis inhibitors. Teixobactin is produced by a previously uncultured bacteria that was recovered by in situ approaches (Ling et al. 2015).

1.2 The Genomic Approach

The advent of cultivation-independent approaches has rapidly reshaped the way we view the microbiological world. DNA (deoxyribonucleic acid) sequencing technology advanced in 1977 as Sanger introduced the use of dideoxynucleotides as chain terminators, a process now commonly referred to as Sanger sequencing (Sanger et al. 1977). Around the same time, Dr.
Carl Woese proposed the use of ribosomal (r)RNA (ribonucleic acid) genes as molecular markers (Woese & Fox 1977), and pioneered the use of 16S rRNA gene sequencing to determine relatedness between organisms and to understand bacterial evolution (Woese 1987).

As molecular techniques developed, the extent of uncultivated diversity became evident (Giovannoni et al. 1990). Metagenomics, defined as the examination of a collection of all genomes from a specific environment, was first proposed in 1998 (Handelsman et al. 1998), and quickly grew in popularity starting in the mid-2000’s. For example, whole genome shotgun sequencing was used to assess the level of ocean biodiversity using the Sargasso Sea by Venter et al. in 2004 (Venter et al. 2004). This pioneer study was performed using Sanger sequencing approaches coupled with capillary electrophoresis allowing for high throughput genetic analysis. Sequencing technology continued to improve and in 2004, Roche commercialized sequencing-by-synthesis (Margulies et al. 2005).

Using the small ribosomal subunit gene for bacterial taxonomy and phylogeny has become the standard, and is used in microbial ecology to assess the makeup of bacterial and archaeal community structure. The 16S rRNA gene is present in all prokaryotes, and its functionality over time has not changed. Therefore, random sequence changes are more likely consequences of evolution and a good measure of time and relatedness, rather than functional changes. Furthermore, the gene (averaging 1500 base pairs (bp)) is long enough to be used for bioinformatics purposes (Janda & Abbott 2007). However, despite the usefulness of this approach, there are problems with sequencing and relying on the 16S rRNA gene for taxonomic classification and community diversity analyses.

All cells, whether active, dying, dead, or dormant, possess DNA that can be extracted and sequenced, and therefore presence does not equal activity. RNA is less stable compared to DNA,
and many metabolically active bacteria have a higher number of ribosomes than dormant cells. Thus, RNA is a potential tool for assessing metabolic activity (Moeseneder et al. 2005). It has been shown that bacterial growth rate correlates with cellular RNA content (DeLong et al. 1989; Poulsen L. K. et al. 1993), and therefore using reverse-transcribed 16S rRNA should allow for assessment of relative activity. An increasing number of studies are now assessing the presence of expressed 16S RNA to analyze the active microbial population within a community (Lay et al. 2013; Klein et al. 2016; Murray et al. 2012). While targeting RNA has proved a useful tool for detecting activity, it should be noted that RNA isolation, reverse transcription, and amplification steps might also contribute biases and select for/against certain species, as with DNA-based approaches. Advances in the field of “omics” are continuing, and studies are also incorporating the use of proteomics, metabolomics, lipidomics, etc. (Bundy et al. 2009; Baker & Dick 2013; Ritchie et al. 2015).

1.3 A Synergistic Approach
Currently most microbiology studies are performed using either culture-independent or –dependent methods, often separately within different labs specializing in one or the other. Studies focused on culture-dependent techniques, as discussed above, face many challenges as a consequence of uncultivability. Culture-independent approaches, centering on meta-omics, use various extraction, amplification, sequencing, and bioinformatic methods to bypass cultivation, but numerous biases still exist at each step. Each method has unique limitations, so both need to be used synergistically to provide the most useful insight. For culture-independent information to be useful, traditional cultivation methods are required to produce reference genomes to enable interpretation of the ‘omic’ data. Furthermore, cultivation is necessary to study an organism’s biology, or to utilize the organism’s natural products for human benefit (e.g., antibiotic
production). Cultivation-dependent and -independent approaches are complementary, and when combined may overcome the caveats of each other. To enhance the field of microbial ecology, we must advance tools and methodologies for both of these approaches to allow access to microbes in their natural environment.

1.4 Dissertation Aims
This dissertation aimed to use a multidisciplinary approach to examine a microbial community by utilizing multiple culture and molecular genomic methods. By combining traditional cultivation with advanced approaches developed in our lab, and by collaborating with experts in metagenomics, we studied a sedimentary lake community from the High Arctic throughout the summer of 2014. Our goal was not only to study the community, but also to gain an understanding of how these approaches complement each other. My specific aims were to:

1. Utilize cultivation-independent molecular approaches to define microbial community structure and identify potentially active organisms
2. Utilize culture-dependent approaches to create a comprehensive culture collection
3. Analyze and synthesize data obtained from Aims 1 and 2 to assess the biological relevance of cultured organisms

To achieve Aim 1 and bypass cultivation limitations, the microbial community composition was assessed by sequencing the 16S rRNA gene loci (rDNA\(^2\)) directly from sediment samples. To determine the active fraction of the community we analyzed expressed 16S RNA by sequencing cDNA of the 16S rRNA gene (herein referred to as rRNA). To achieve Aim 2, both traditional cultivation methods and in situ approaches were utilized. No studies to date have combined such substantial cultivation efforts as employed here with genomic approaches. Direct comparison of the datasets obtained from Aims 1 and 2 allowed for evaluation of the biological relevance of cultured organisms.

---

\(^2\)To differentiate between the 16S rRNA gene data and the analysis of the expressed 16S rRNA, we use the terms rDNA and rRNA, respectively.
Using data from the 16S rDNA libraries we were able to identify a tens of thousands of operational taxonomic units (OTUs; clustered at 97% sequence similarity), with consistent presence detected throughout the season. However, only a small fraction of these OTUs were active, and the active OTUs differed greatly between sites, dates, and even among replicates. Furthermore, many of the active OTUs were not detected in the rDNA libraries, and more importantly not found in any reference database.

We successfully developed a vast cultivation library consisting of over a thousand strains, but observed limited overlap between OTUs cultured and those detected from DNA/RNA libraries. We also noted minimal overlap between the OTUs cultured via different methods, indicating that no single approach was sufficient to access most organisms in the community.

The limited overlap between cultivation-dependent and independent data was an unexpected finding, and we interpret this to be a result of important biogeochemical differences at the microscale level, leading to “microheterogeneity” in microbial spatial distribution. Therefore, all samples obtained, including replicate samples taken millimeters (mm) apart, might have come from substantially different communities. This dissertation demonstrates the importance of utilizing multiple approaches synergistically within microbial ecology, and highlights the need to develop tools and methods to study community composition and activities on a scale relevant to microbial life.
Chapter 2. Investigating the Lake Community using Culture-Independent Approaches
2.1 Introduction

The field of microbiology has advanced over the years: beginning with the first reports of microbes by Leeuwenhoek, and followed by the influential discoveries of the fathers of modern microbiology: Ferdinand Cohn, Louis Pasteur and Robert Koch. The development of standard approaches for microbial cultivation, such as the use of nutrient media for isolation, coupled with advancements in microscopy, allowed easily cultivable organisms to be readily cultured and studied (Escobar-Zepeda et al. 2015). However, it became increasingly clear that the number of organisms scientists were observing using the microscope did not match the number of colonies they would form on Petri dishes (Staley & Konopka 1985).

In the late 1970s, Carl Woese proposed the idea of using rRNA genes as molecular markers for classification (Woese & Fox 1977). This idea, combined with Sanger automated sequencing (Sanger et al. 1977), completely transformed the classification and study of microorganisms. As molecular techniques, such as the polymerase chain reaction (PCR), continued to advance, the extent of uncultivated diversity became illuminated (Giovannoni et al. 1990).

Following Julian Davis’ pioneering ideas (Handelsman et al. 1998), metagenomics became one of the key tools in microbiology. Whole genome shotgun sequencing was used to assess the level of ocean biodiversity using the Sargasso Sea as a model (Venter et al. 2004). Sequencing technology rapidly developed to handle ever increasing endeavors, lead by Roche commercializing pyrosequencing in 2005 (Margulies et al. 2005). Since then, sequencing technology and bioinformatic pipelines have continued to improve, allowing unprecedented insight into the microbial world.

Since Woese advocated the use of the 16S rRNA gene for bacterial taxonomy and phylogeny, it has become the standard to describe community composition. This approach is
incredibly useful, but it is important to note the caveats. The 16S rRNA gene can be present in multiple copy numbers (Větrovský & Baldrian 2013; Wooley et al. 2010; Moeseneder et al. 2005), and the DNA sequences of these copies can differ even within the same organism. This variation may lead to the overestimation of microbial community richness. The phylogenetic resolution power of the 16S rRNA gene can be low, which is particularly important in regard to the short reads generated by many next generation sequencing (NGS) methods. Even 16S rRNA gene sequences much longer in length may be fraught with taxonomic classification uncertainties. Proper taxonomic identification is limited by the quality of sequences in the existing databases, which are widely different in how they are curated. Furthermore, culture-independent microbial community analysis involves DNA extraction and PCR-aided amplification of a gene of interest (often 16S rRNA gene). It is well established that multiple biases exist with both DNA extraction and PCR procedures (Suzuki & Giovannoni 1996; von Wintzingerode et al. 1997; Aird et al. 2011; Janda & Abbott 2007). In addition, species classified as identical by 16S rRNA gene sequences can actually have drastically different genome sizes (Perna et al. 2001); this could lead to an underestimation of richness in a community. Therefore, while in this study we focus on targeted surveys of the 16S rDNA gene to analyze the microbes in polar lake sediment, we must consider the limitations of this approach.

The vast majority of studies characterizing microbial communities using culture-independent methods, including those conducted in polar environments, have been based on the 16S rRNA gene sequencing approach. 16S data does not give specific insight into the activity within the community, but rather provides a look into the microbial composition, including populations of dead, dormant, and active bacteria and archaea. It is becoming more common to also assess the presence of expressed 16S rRNA to gather information on potential metabolic
activity, as there is evidence that ribosomes are more abundant in active cells compared to
dormant cells (Fegatella et al. 1998; Blazewicz et al. 2013; Kerkhof & Kemp 1999). Studies use
a combined 16S DNA and RNA survey approach to look at the total bacterial community
composition compared to its potentially metabolically active fraction (Lay et al. 2013;
number of studies that employ this combined approach is still relatively small, especially in
extreme environments such as the Arctic (Lay et al. 2013; Stibal et al. 2015). Here, we use both
the 16S rRNA gene\textsuperscript{3} and the cDNA from the 16S rRNA to describe microbes present, and active,
within the lake sediment.

\subsection{2.2 Methods}
*Collaborators (Karen Nelson, Chris Dupont, Alex Ritcher, Manolito Torralba and Drishti Kaul)
at the J. Craig Venter Institute (JCVI) performed all sequencing and data processing related to
culture-independent methods, provided valuable and critical insight into data analysis, and
provided figures where indicated.

Sample Collection

\textit{Site location.} Sediment samples were collected from the upper (oxic) layer of a man-made lake
in Northwest Greenland, outside of Thule Airbase (N 76°32.659' W 68°27.458'). This lake was
chosen based on a preliminary survey of a multiple locations around Thule Airbase in 2013. This
particular lake was easily accessible and harbored a microbial community of moderate richness
(data not shown). Compared to other sample sites, there was more vegetation within the
surrounding soil, and thus we termed it the “Rich Lake.” For this study, two sample sites within
the lake were chosen, designated as Rich Lake 1 (RL1) and Rich Lake 2 (RL2). Markers (plastic
pipes dug into the ground) were placed at both sites to ensure continuous sampling from the

\textsuperscript{3} For simplicity, to distinguish between 16S rRNA gene libraries and libraries obtained from cDNA, we use the terms 16S rDNA
and 16S rRNA, respectively
same location throughout the season. The sites were 30 feet from each other, about 15 cm from the waters edge, with 1-3 mm of water above the sediment. Sediment samples were collected at various time points during the summer of 2014 (Table 1). The temperature and pH of the sediment was measured periodically and stayed essentially unchanged throughout the season: 10°C with a pH of 6.8.

Table 1. Sediment samples collected for 16S rDNA and rRNA surveys. Each sample was processed as one library, and all libraries underwent quality control. Those that passed control, as indicated in the table, were used for downstream analyses.

<table>
<thead>
<tr>
<th>Date</th>
<th>DNA Libraries</th>
<th>Failed Libraries</th>
<th>% Successful DNA Libraries</th>
<th>RNA Libraries</th>
<th>Failed Libraries</th>
<th>% Successful RNA Libraries</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-Jun</td>
<td>2</td>
<td>2</td>
<td>50%</td>
<td>4</td>
<td>2</td>
<td>67%</td>
</tr>
<tr>
<td>27-Jun</td>
<td>6</td>
<td>0</td>
<td>100%</td>
<td>4</td>
<td>2</td>
<td>67%</td>
</tr>
<tr>
<td>14-Jul</td>
<td>6</td>
<td>0</td>
<td>100%</td>
<td>5</td>
<td>1</td>
<td>83%</td>
</tr>
<tr>
<td>15-Jul</td>
<td>2</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>4-Aug</td>
<td>2</td>
<td>1</td>
<td>67%</td>
<td>3</td>
<td>3</td>
<td>50%</td>
</tr>
<tr>
<td>14-Aug</td>
<td>5</td>
<td>0</td>
<td>100%</td>
<td>0</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>20-Aug</td>
<td>4</td>
<td>2</td>
<td>67%</td>
<td>0</td>
<td>6</td>
<td>0%</td>
</tr>
<tr>
<td>6-Oct</td>
<td>2</td>
<td>4</td>
<td>33%</td>
<td>0</td>
<td>5</td>
<td>0%</td>
</tr>
<tr>
<td>Total</td>
<td>29</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sampling procedure for DNA and RNA

Using a sterile teaspoon, the uppermost 2-3 mm of sediment was placed in a 50 ml Falcon tube. Approximately 5 grams (g) of sediment was collected for RNA analysis and 10 g was collected for DNA library preparations. For every time point, at both RL1 and RL2, at least three closely spaced samples (1-3 cm apart) were collected as biological replicates for each RNA and DNA quantification/analysis. Sediment for DNA analysis was frozen at -20°C within minutes of collection. The samples for RNA analysis were placed in a falcon tube, and at least 10-15 ml of AllProtect Tissue Reagent (Qiagen Catalog # 76405) was slowly added. The sample was mixed thoroughly to ensure the RNA buffer penetrated the sample and RNA was stabilized.
throughout. Both DNA and RNA samples were stored at -20°C, shipped to Northeastern University frozen, and stored at -80°C until RNA extraction.

DNA extraction, PCR, library preparation, and sequencing

Sediment samples were transported at -20°C to JCVI for DNA analysis (Table 1). Following the manufacturer’s specifications, total DNA was extracted using PowerSoil DNA Isolation Kit (MO Bio Catalog #12888). Extracted DNA was stored at -80°C until further processing. JCVI prepared 16S rRNA gene libraries using adaptor and barcode ligated primers that targeted the V1-V3 regions of the 16S gene. Amplicons were purified using the Qiagen PCR Purification Kit (Catalogue #28106), quantified using Sybr Gold, normalized, and pooled to ensure equimolar quantities per sample. The 16S rRNA gene library was sequenced using Illumina MiSEQ and the V3 chemistry kit 2x300bp format following the manufacturer’s recommendations.

RNA extraction, PCR, library preparation, and sequencing

RNA from sediment samples (Table 1) was extracted using the MoBio Total RNA PowerSoil Extraction Kit (Catalogue # 12866-25). Approximately 2-4 g of the collected sediment was used as starting material, and RNA extraction was performed according to the manufacturer’s specifications. RNA was eluted in 100 µl of elution buffer. The 16S rRNA cDNA library was prepared using ScriptSeq Complete Gold (Epidemiology) – Low Input Kit (Illumina catalog BEP1224) according to manufacturer’s recommendations for cDNA generation and library preparation. rRNA depletion was not performed to allow for analysis of 16S rDNA gene transcripts. Reverse ScriptSeq index primers were used to barcode each library. MiSEQ
sequencing was completed using standard dual index sequencing protocol as specified by Illumina.

**Sequencing analysis**

Sequence data were deconvolved using sample barcodes to identify sequences from each sampled time point (Table 1). Barcode, primer, and adaptor sequences were trimmed using JCVI’s in-house pipelines. Chimeric sequences were identified using the ChimeraSlayer program (Haas et al. 2011), and removed prior to downstream analysis. OTUs were generated using the Uparse pipeline (Edgar 2013) at 97% nucleotide identity. Taxonomic assignments for the OTUs were performed using mother classifying.seqs (Schloss et al. 2009) and the Silva (SSURef_NR99_119) database (Quast et al. 2013; Pruesse et al. 2007). The OTU membership of sequences was used to construct a sample-OTU count matrix. Counts were normalized based on the total number of reads per library. Low abundance OTUs (observed counts <5) were removed as potential artifacts. All calculations of alpha and beta diversity were conducted in QIIME. Bray Curtis distance matrices were created from the sample-OTU table and used for principal coordinate analysis (PCoA) of beta-diversity (Caporaso et al. 2010).

**2.3 Results and Discussion**

**2.3.1 Taxonomic Composition Overview by 16S rDNA sequencing**

To assess community composition, DNA libraries were constructed from 36 sediment samples, followed by targeted sequencing of the 16S rRNA genes. Twenty-nine libraries passed quality control, totaling 2,785,845 reads obtained from all libraries. The average number of reads per library was 96,063. OTUs were generated at 97% sequence similarity, and the longest representative sequence of each OTU was assigned taxonomic identification based on the closest
relative in the Silva v.117 database (Quast et al. 2013; Pruesse et al. 2007). We identified a total of 20,321 OTUs; omitting those encountered fewer than five times lowered the final count to 13,113 OTUs. Low abundance reads were omitted to ensure potential sequencing artifacts are removed from downstream data analysis.

The total community composition as determined by combining all samples is shown in Figure 1. Here we note that the known biases of DNA extraction protocols and PCR methods (Martin-Laurent et al. 2001; Forney et al. 2004) make the richness and abundance estimates conservative (von Wintzingerode et al. 1997).

Figure 1. Microbial community composition as determined by a 16S rDNA gene survey. Taxonomy was determined at the phylum level. Data from all libraries and time points was averaged to generate percentages. Inset: breakdown of Proteobacteria at the class level.
The 16S rDNA gene phylotypes were dominated by Proteobacteria (37%), followed by Bacteroidetes (13%), Cyanobacteria (11%), and Actinobacteria (10%) (Figure 1). Within the Proteobacteria, Alpha- and Betaproteobacteria classes were most abundant, comprising 50% and 28% of Proteobacterial sequences, respectively (19% and 28% of the total) (Figure 1 inset). Thirteen percent of Proteobacterial sequences were classified as Deltaproteobacteria (5% of all 16S rDNA sequences), and 8% (3% of total) were identified as Gammaproteobacteria. The abundance of Proteobacteria corresponds to a previously reported figure (39%) from a Canadian High Arctic permafrost sample (Steven et al. 2007). In contrast to our study, the majority of Proteobacteria phylotypes from permafrost were related to Deltaproteobacteria. Steven et al. also found many more Actinobacteria sequences; this may be due to differences in sediment types. The community composition of Thule Lake sediment is consistent with that of another study from Arctic snow and fresh water, which found Proteobacteria to dominate snow sites, and Bacteroidetes prevalent in the freshwater samples (Møller et al. 2013), supporting our results on the presence and dominance of these taxa in the Arctic environment.

Lay et al. conducted a metagenomic survey of the microbial sediment community of a hypersaline, subzero, perennial spring in the High Arctic, and detected a significant portion of sequences tracing back to the Bacteroidetes phylum (13.3%) (Lay et al. 2013), as also found in our study. In contrast, their study identified significantly less Proteobacteria (only 6.6% of their community).

We also noted that, in our library, a large portion of sequences (11%) was related to Cyanobacteria. This is not surprising since Cyanobacteria are typically photosynthetic, and our samples were collected from the top layer of sediment where light penetration is still possible. Furthermore, our samples were collected during the summer, when there is constant sunlight,
potentially explaining the prominence of photosynthesizers. However, it is important to note that among Cyanobacteria detected, about 40% were attributed to chloroplast DNA, such as from diatoms and other eukaryotes. Chloroplast contamination is known to result from the use of universal bacterial primers in the V1-V3 region (used here) matching the 16S rRNA gene of the chloroplast (Hanshew et al. 2013). Excluding chloroplast sequences, Cyanobacteria comprise 6% of total sequences. Both chloroplast DNA (Møller et al. 2013) and 16S rDNA from Cyanobacteria (Steven et al. 2013; Lay et al. 2013) have been detected in Arctic environments by other 16S rDNA pyrosequencing studies.

2.3.2. Comparison of Taxa Present According to site and date

DNA libraries were constructed from samples obtained at both RL 1 and RL 2. Thus, we were interested in exploratory analysis of multidimensional scaling of OTU abundance. We used PCoA to visualize trends in beta diversity (among different sample libraries) using Bray-Curtis distances. We found that microbial communities sampled from the same site were slightly more similar to each other than samples from different sites (Figure 2). For example, DNA libraries from RL1 were more similar to each other than to samples from RL2. However, this trend was not particularly strong or significant, as all three principal components explained only 34.19% of variability within the data. No other trends were observed in the DNA libraries via this analysis.
Beta diversity analysis for lake sediment samples (DNA libraries) from the High Arctic. Principal coordinates analysis (PCoA) of Bray Curtis distances was focused on the 29 sediment samples from DNA libraries. Red = RL1, Blue = RL2. The amount of variability explained by all three principal coordinates (PCs) is shown on the axes. (Kaul and Dupont)

Graphical representation of the taxonomic breakdown of sequences detected throughout the season at each site is presented in Figure 3. It can be seen that community composition was relatively similar and consistent between replicate samples, within sites, and throughout the season. Note, the data presented here was used qualitatively to assess large-scale changes in taxonomic groups, as the 16S rDNA libraries allow limited inferences into the actual numerical abundances.

At both sites, we observed a predominance of Alphaproteobacteria. At the RL1 site, we noted a cyanobacterial bloom during the middle of July (Figure 3). This is in contrast with the RL2 site, where we observed a bimodal pattern of two cyanobacterial blooms, first in mid-July and then again in late August. At RL1, corresponding to the mid-season increase in cyanobacterial abundance, there was a decrease in the presence of Deltaproteobacteria. At RL2 this trend was much less pronounced. At both sites, Verrucomicrobia peaked mid season, July 14-15.
Figure 3. Comprehensive taxonomic breakdown of RL1 and RL2 samples with respect to different dates. Taxa abundance values for top ten most prevalent taxa at the class level for different time points are indicated. The remaining taxa are pooled into an additional taxon labeled “Other.” (Kaul and Dupont)

Figure 3 displays the 10 most prevalent classes common to both RL1 and RL2. When community analysis was performed separately for each site, we noted that Anaerolineae was one of the most prevalent classes at RL1, primarily detected in the beginning of June and end of August and October, with very few sequences found in July (Figure 4). Anaerolineae was not highly detected at RL2, but rather Flavobacteriia were more dominant (data not shown).
Figure 4. Comprehensive taxonomic breakdown of Rich Lake 1 samples with respect to different dates. Taxa abundance values for top ten most prevalent taxa at the class level for different time points for 16 rDNA samples from Rich Lake 1. The remaining taxa are pooled into an additional taxon labeled “Other.” (Kaul and Dupont)

Future studies should incorporate metadata for each sample, including but not limited to carbon, nitrogen, and phosphorous levels. Metadata is crucial to connect an organism’s presence in the community with any possible functional roles or to hypothesize reasons for seasonal trends. For example, Cyanobacteria are known to survive in extreme environments, and have been found in polar regions, dominating biomass and productivity of freshwater ecosystems (Lionard et al. 2012). They play important roles in photosynthesis, as well as nitrogen and carbon fixation, and also possess the ability to produce noxious compounds that can influence growth of neighboring organisms (Berman-Frank et al. 2003; Lay et al. 2013). To determine if the blooms detected here contribute to the change of community structure through the season and impact the immediate environment, further biogeochemical data are needed. Importantly though, the presence of cyanobacterial 16S rRNA gene sequences in our libraries does not indicate activity. A previous study of an Arctic metagenome found 66.3% of dormancy genes were of
cyanobacterial origin (Lay et al. 2013; Jones & Lennon 2010), indicating that the OTUs detected in our study may be representative of dormant cyanobacteria. Ideally, future studies of Rich Lake sediment would incorporate metagenomic analysis to couple with the metadata.

2.3.3 16S rRNA Sequencing and Taxonomic Activity

To determine the active fraction of the microbial sediment community, we assessed the presence of expressed 16S RNA by sequencing its cDNA. The sequencing success of the rRNA libraries was low; only 16 of 35 rRNA libraries passed quality control, limiting downstream analysis of microbial activity in the 2014 dataset. We note that the ability of AllProtect Buffer, used here, to preserve RNA can be negatively affected by a sediments’ humic acid, potentially influencing the results of reverse transcription of RNA (Daniel 2005; Wang et al. 2009; Mettel et al. 2010).

We obtained 268,663 reads from all libraries, with an average of 17,910 reads per library. It is well documented that not all organisms present are active, which may explain the difference between total reads obtained from our rDNA libraries (2 million) and rRNA libraries (couple hundred-thousand).

Our data suggest that the most active groups were Proteobacteria (23%), followed by Bacteroidetes (14%), and Cyanobacteria (12%) (Figure 5). These same taxonomic groups were determined as abundant based on analysis of the 16S rDNA gene data (Figure 1). Uniquely characteristic of the 16S rRNA library data was that 25% of all rRNA sequences did not match any known organisms, and had to be placed into an “unclassified” category (Figure 5).
Unclassified sequences formed 116 distinct OTUs when clustered at 97% similarity. The longest representative sequence from each of the 21 most abundant unclassified OTUs was checked against available sequences in the Silva database (Pruesse et al. 2007), and all sequences were deemed to be truly novel. The majority of these sequences (70%) fell into three OTUs, with almost half of them (41%) forming a single OTU: rRNA_OTU_113. The latter may thus be an active and important group within our community; in fact 10% of all rRNA sequences obtained were classified within this OTU (Figure 6). Unexpectedly, this OTU was not detected in any of the 16S rDNA gene libraries. This OTU will be further discussed in section 2.3.5, p. 27-29.
Figure 6. Unclassified sequences cluster within three main OTUs in the 16S rRNA libraries. 25% of all rRNA sequences were unclassified. Out of these sequences, rRNA_OTU_113 accounts for 10% of the total sequences detected and 41% of all unclassified OTU sequences.

PCoA of Bray Curtis distances was performed for just the rRNA libraries, but provided no meaningful data, as fewer OTUs were detected from 16S rRNA data than from 16S rDNA data (not shown). PCoA was then performed on all DNA and RNA libraries, and we found that the active OTUs detected in RNA libraries were more dynamic than those present in DNA libraries (Figure 7). DNA library data clustered together, with minimal variability noted along principal component 2 (PC2). In contrast, the OTUs within RNA libraries were highly variable across both PC1 and PC2, indicating temporal and spatial variability in their activity. The total amount of variability explained by the first three principal components is 49.2%, which is more robust than we calculated for the rDNA samples alone (Figure 2). The majority of the variability can be explained across PC1, which accounted for 39% of the variability. Exploratory analysis
led us to the conclusion that RNA libraries, and thus patterns of activity across all samples, were potentially very dynamic – in contrast to the DNA-only data.

![Figure 7. Beta diversity analysis for the lake sediment samples from High Arctic. Principal coordinates analysis (PCoA) of Bray Curtis distances was done for the sediment samples from a lake in Thule, Greenland. The DNA libraries are in red (right) and the RNA libraries (left) in blue. There are 29 samples from DNA libraries and 16 samples from RNA libraries. The amount of variability explained by all three principal coordinates (PCs) is shown on the axes. (Kaul and Dupont)](image)

2.3.4 Patterns of Activity as Determined by 16S rRNA Sequencing

As indicated by exploratory analysis using PCoA (Figure 7), the activity of OTUs between rRNA libraries was more variable than the presence of OTUs between DNA libraries. We found that the most active taxa were those deemed unclassified. The remaining active taxa varied greatly between all samples (Figure 8, Figure 9). The variability between replicate samples was particularly striking, as they were meant to serve as biological replicates, and sampled only centimeters apart (Figure 8, Figure 9).
For example, from both samples taken, unclassified taxa were the most active on August 4 at RL2 (Figure 9). However, Sphingobacteriales were the second most active taxa from sample 1, and Burkholderiales also showed some activity. In contrast, the orders Bacillales and Rhizobiales were most active in the second sample; no significant activity was attributed to the Sphingobacteriales order (Figure 9). We observed considerable differences of the most active taxa between available biological replicates (i.e. July 14 at both sites, 2 Figure 8, Figure 9).

Figure 8. Comprehensive taxonomic breakdown of Rich Lake 1 samples with respect to different dates. Taxa abundance values for top ten most prevalent taxa at the order level for different time points for six rRNA samples from Rich Lake 1. The remaining taxa are pooled into an additional taxon labeled “Other”. Unclassified eukaryotic sequences have been removed from analysis (Dupont and Kaul).
Figure 9. Comprehensive taxonomic breakdown of Rich Lake 2 samples with respect to different dates. Taxa abundance values for top ten most prevalent taxa at the order level for different time points for 10 rRNA samples from Rich Lake 2. The remaining taxa are pooled into an additional taxon labeled “Other”. Unclassified eukaryotic sequences have been removed from analysis (Dupont and Kaul).

Since more than half of the 16S rRNA libraries were of inadequate quality, we were unable to obtain data from all replicates at every date. We successfully sequenced three out of four libraries from RL2 on June 27 and July 14 (Figure 9). The most active bacterial orders varied greatly between each replicate library, on each date. For example, on July 14 Sphingobacterales were active in only one sample. Firmicutes were also detected as active in only one sample from both June 27 and July 14 (Figure 9). Similar variation in taxa activity was observed at RL1 on July 14: in one sample Burkholderiales was the most active taxon, while Bacteroidales and Verrucomicrobiales were most active in the second sample. Desulfuromonadales and Rhodocyclales activity was detected in roughly equal proportions in the second sample, but not in the first (Figure 8). All “other” taxa showed an increase in activity during July 14th, but again with different proportions between samples.
Despite differences between replicate rRNA libraries, there were a few notable trends that could still be detected. At RL1, unclassified taxa were active throughout the season, with the most activity occurring at the beginning and end of the season; other bacterial groups predominated mid-season (Figure 8). The same bimodal pattern of activity of unclassified taxa is observed at the second Rich Lake site, however the decrease in activity manifests earlier, on June 27\textsuperscript{th} (Figure 9). At both sites we observed an increase in Burkholderiales activity, corresponding to the decrease in unclassified taxa. In general, more variation was detected between active bacterial orders throughout the entire season at RL2 compared to RL1.

Future investigations should delve into the “other” taxa, containing active OTUs that did not make it into the top 10 most predominant. Particular focus should be on the active OTUs during June 27/July 14 time points to determine if any patterns or similarities exist. Furthermore, unclassified bacteria seem to decrease activity during the increase in cyanobacterial presence as detected by rDNA analysis, despite the fact that no activity from cyanobacteria was observed. Future studies should attempt to explore this relationship. Lastly, as previously mentioned, further interrogation in the unclassified taxa is necessary for any understanding of activity or biogeochemical roles within the lake.

### 2.3.5 Unclassified OTUs: Phantom Taxa and Future Experimentation

A fairly recent observation reported from rRNA-based studies is the occurrence of phantom taxa, the term given to OTUs detected from RNA but not DNA sequencing. It is unexpected that a group would be active, when it is not even detected as present (Klein et al. 2016). Some have speculated that sample processing (e.g. cDNA synthesis from RNA vs. direct DNA extraction) may contribute to the differences detected between samples, and that RNA extraction and cDNA processing procedures may result in the introduction of erroneous base pair
changes (Van Gurp et al. 2013; Lanzén et al. 2011; Mikkonen et al. 2014). Base pair changes could lead to a RNA sequence being different enough from the DNA sequence such that they are no longer clustered into the same OTU in downstream analysis.

There is growing evidence that rare taxa, defined as taxa with low presence in 16S rRNA gene surveys, may actually be disproportionately active, relative to the abundant members in the community (Campbell et al. 2011; Hugoni et al. 2013; Hunt et al. 2013). This is so because, when sequencing the 16S rRNA gene, it may be harder to detect rare members, as there are fewer rRNA gene copies per cell (Klappenbach et al. 2001) compared to the 100s-1000s of ribosomes of active organisms, easily detectable as cDNA (Fegatella et al. 1998).

Given the fact that 10% of all detected reads, and 40% of all unclassified reads, traced back to the same phantom OTU (rDNA_OTU_113), we believe this taxon may be important to the activity of the Rich Lake community. Future experiments should use a targeted approach to confirm the presence and detect activity of this OTU, as well as the other active unclassified OTUs. Sequences obtained in this study are too short (200-300 bp) to gain further insight at this time, but could be used to design OTU specific probes targeting the 16S rRNA gene. By combining fluorescent in situ hybridization (FISH) with targeted rRNA oligonucleotide probes and microautoradiography (MAR), the in situ identities, activities, and even substrate uptake profiles for bacterial cells within complex communities can be determined (Lee et al. 1999; Ouverney & Fuhrman 1999; Torsvik & Øvreås 2002). Catalyzed reporter deposition (CARD)-FISH can also be used, and has shown success over conventional FISH approaches, in both aquatic (Pernthaler et al. 2002) and soil systems (Ferrari et al. 2006). The use of FISH or (CARD)-FISH would enable us to visualize the cellular morphology of rRNA_OTU_113, however samples would no longer be viable for cultivation (Vartoukian et al. 2010). Vartoukian
and colleagues utilized a membrane hybridization approach to locate microcolonies within mixed cultures, enabling their isolation. This method could potentially allow for cultivation of previously uncultured organisms, such as the OTUs of interest here (Vartoukian et al. 2010). Once microcolonies are isolated they can be incubated in situ for continued rounds of cultivation until domestication is possible (Nichols et al. 2008).

Another option to detect activity of OTUs of interest involves detection and selective isolation of DNA from actively growing bacteria with the use of bromodeoxyuridine (BrdU), accompanied by visualization of BrdU microbial cells. This can be done in conjunction with FISH probes specific for the OTUs of interest, thus detecting presence and activity in one step (Urbach et al. 1999; Torsvik & Øvreås 2002).

The ideal future studies would cultivate and examine the genetic potential of rRNA_OTU_113. Fluorescent activated cell sorting (FACS) has been used in conjunction with whole genome amplification (Lee et al. 2015) to obtain partial genome sequences, which then could allow for characterization of the organisms, as has been done with TM7 (Vartoukian et al. 2010; Podar et al. 2007; He et al. 2015). Using information gathered via this approach, potential targeted cultivation approaches specific to the organism of interest could be generated.

2.3.6 Comparing OTU abundance (rDNA libraries) and Activity (rRNA libraries): Are the Most Rare the Most active?

Studies combing rDNA and rRNA analyses have aided in our understanding of how microbial communities respond to environmental changes, what taxa are involved in biogeochemical processes, and how communities are assembled. In contrast to our results, several previous studies examining abundance and activity found activity over time, determined
by rRNA data, to be more constant compared to the rDNA results (Klein et al. 2016). It was also determined that activity and abundance were correlated (Klein et al. 2016; Campbell et al. 2011). However, notable exceptions present evidence that corroborates trends we observed, namely that there is a pronounced difference between the composition of species that are merely present vs. those actually active (Jones & Lennon 2010; Baldrian et al. 2012; Zhang et al. 2014). This trend has been observed in a variety of environments, such as marine (Campbell et al. 2011; Hugoni et al. 2013; Hunt et al. 2013), freshwater (Wilhelm et al. 2014), and soil systems (Gremion et al. 2003). These studies demonstrate that it is useful to incorporate rRNA surveys to link microbial community composition to ecosystem function, particularly in relatively uncharacterized environments. Furthermore, these data support the premise that different methods of observation, i.e. DNA or RNA, may portray the community in vastly different ways.

To further compare the relationship between abundance and activity within our sample site, we identified active and abundant OTUs. OTUs were considered to be active if they were detected in at least 1% or more of sequences from all rRNA libraries. Abundant OTUs were identified the same way: sequences made up more than 1% of the rDNA libraries (Campbell et al. 2011). Analysis was performed on data averaged across all dates and replicates; therefore, temporal changes or microheterogeneity examples are not discernable.

In RL1, 17 OTUs were active, and five were uniquely detected in RNA libraries only (phantom taxa). Only four OTUs were considered active and abundant (blue bars, Figure 10). The active and abundant OTUs were classified as Cyanobacteria/chloroplasts or Sphingomonadales. Consistent with previous literature (Jones & Lennon 2010; Campbell et al. 2011; Baldrian et al. 2012; Hugoni et al. 2013; Wilhelm et al. 2014; Zhang et al. 2014), the most active OTUs at RL1 were considered rare taxa. Two of the most active taxa in this analysis were
identified as Eukaryota, possibly a result of chloroplast or mitochondrial DNA contamination.

Note that eukaryotic OTUs were removed from analyses when generating figures of the 10 most active taxa (Figure 8, Figure 9) because the survey was designed to target active bacterial 16S cDNA. The other most active OTUs at RL1 were the unclassified rRNA_OTU_113, and an OTU from the order Flavobacteriales.

![Active OTUs at RL1](image)

**Figure 10. Abundance of the 17 active OTUs at Rich Lake 1.** Seventeen OTUs were active at RL1. The percent each OTU contributed to rDNA (blue) and rRNA (red) libraries. Most OTUs were rarely detected in rDNA libraries, indicating low abundance. Phantom OTUs detected only in RNA libraries are labeled as rRNA_OTU_#.

At RL2, OTUs exhibited a similar pattern as above: the most active OTUs were mostly rare, non-abundant taxa (Figure 11). A total of 16 OTUs were deemed active, and only one OTU was both active and abundant. Similar to RL1, this OTU was of the order Sphingomonadales.

Five OTUs were detected in the RNA libraries only: two were unclassified, two were possible eukaryotic contaminants, and one was related to Burkholderiales. Many OTUs affiliated with the order Burkholderiales were detected in 16S rDNA libraries, but this particularly active OTU was not found in any.
There is support in the literature for the dominance of both Cyanobacteria and Sphingomonadales in Arctic Snow (Møller et al. 2013), further providing evidence they may be important contributors of Arctic bacterial communities.

![Graph](image)

**Figure 11. Abundance of the 16 active OTUs at Rich Lake 2.** Sixteen OTUs were active at RL2. The percent each OTU contributed to rDNA (blue) and rRNA (red) libraries. Most OTUs were rarely detected in rDNA libraries, indicating low abundance. Phantom OTUs detected only in RNA libraries are labeled as rRNA_OTU_#.

While our results correspond with recent studies showing “rare” taxa are the active contributors in communities (Jones & Lennon 2010; Campbell et al. 2011; Baldrian et al. 2012; Hugoni et al. 2013; Wilhelm et al. 2014; Zhang et al. 2014), this idea only explains the discrepancies that exist between our DNA and RNA libraries. However, it does not explain the wide discrepancies in detected OTU activity between replicate samples. An alternative explanation of these differences is the idea of spatial microheterogeneity in microbial
distribution, based upon inherent differences that exist within the environment, across centimeters, millimeters and even micrometers.

2.3.7 The Importance of Microheterogeneity

We hypothesize that the differences observed between replicate samples are due to microheterogeneity in microbial distribution, such that samples taken a few millimeters to centimeters apart actually represent quite different communities. Traditionally, microbial ecology studies are designed on a “micro” scale from a scientists’ point of view (sampling centimeters or millimeters apart), but this is actually macro-scale to a microorganism (Stocker 2015). Large-scale approaches are useful and provide invaluable information regarding average composition and function, but they completely miss heterogeneity and cell-scale dynamics that appear crucial for understanding how microbes live, function, and structure the environment around them.

Small-scale variation within bacterial communities has been observed in other studies and the importance of this phenomenon is discussed in the literature (Long & Azam 2001; Kuzyakov & Blagodatskaya 2015; Seymour et al. 2000; Vos et al. 2013; Meyer-Reil 1994; Stocker 2015). It is well documented that soils and sediments are very complex, heterogeneous environments (Long & Azam 2001; Kuzyakov & Blagodatskaya 2015; Seymour et al. 2000; Pedersen et al. 2015; Meyer-Reil 1994; Vos et al. 2013). Various conditions, such as cracks and pores in sediment/sand grains/organic matter, contribute to changes in nutrient cycling (i.e. carbon, nitrogen, sulfur, metals, etc.) even across very small distances (100 um or less) (Pedersen et al. 2015). The diffusive limitation of chemicals contributes to spatial differences of biogeochemical environments on very small scales. To understand microbial life, it is imperative
that we be able to characterize these spatiotemporal niches (Vos et al. 2013; Pedersen et al. 2015; Raynaud & Nunan 2014); therefore, future studies conducted at Rich Lake must include micro-scale measurements of the environment. Micro-scale measurements can be accomplished using a wide array of nanosensors that have been developed and reviewed extensively (Pedersen et al. 2015). Furthermore, future sampling of our site should include precise measurements of the pH across small distances, as evidence indicates bacterial community changes can result from differences in pH (Fierer & Jackson 2006). Ideally transcriptomic data would be obtained, allowing insight into the active genes from each sample. This data could shed light onto functional aspects of communities, and determine if, despite different species composition, the same core ecological roles are performed, as has been observed within the human microbiome (Huttenhower et al. 2012; Turnbaugh et al. 2009). We emphasize that most studies today are still conducted on a macroscale (by comparison to the size of microbes), calling for a need for microscale studies of microbial distribution, as well as the development of tools tailored to the task.
Chapter 3. Culturing the Lake Community
3.1 Introduction

Cultivation remains an essential tool to study any organism, despite the unprecedented insight into microbial communities that meta-omic approaches have allowed. Creation of high quality reference genomes, necessary for interpretation of meta-omic data, is easier, cheaper, and more accurate when generated using a monoculture compared to single-cell or mixed culture approaches. Furthermore, cultivation is essential for testing hypotheses and determining the metabolic, biochemical, and physiological profiles of organisms (Salcher & Šimek 2016). Importantly, cultivation is also necessary to access new compounds produced by microbes that may prove useful for industry and human health, such as Teixobactin (Ling et al. 2015).

The challenge scientists face is that the majority of microbial diversity remains inaccessible, because most organisms will not grow under standard cultivation conditions (Staley & Konopka 1985; Rappé & Giovannoni 2003). Recent advances in microbial cultivation have been encouraging (Lagier et al. 2016; Browne et al. 2016), but the overall challenge of accessing previously uncultivated species remains prominent.

In 2002, our lab, in collaboration with Kim Lewis, put forth the idea of taking cultivation out of the lab, and into the environment (Kaeberlein, Lewis & Epstein 2002). This approach would take advantage of the naturally occurring growth factors and nutrients in the environment to support microbial growth. This concept formed the basis of in situ cultivation methods, and has been widely used and improved upon, beginning with the diffusion chamber (Kaeberlein, Lewis, Epstein, et al. 2002; Bollmann et al. 2007), and progressing to many other approaches, including microfluidic devices (Tandogan et al. 2014) and the high throughput ichip (Nichols et al. 2010). These methods have all proven successful in recovering organisms different from those typically cultivated using standard approaches (Epstein 2009; Tandogan et al. 2014; Aoi et
We aimed to create a culture collection representative of the sediment community, and to improve microbial recovery we employed a variety of cultivation methods, including in situ approaches, anaerobic enrichments, and traditional cultivation. In situ approaches included the diffusion chamber, trap (Figure 13), filter plate (Figure 14), itip (Figure 15), and ipore (Figure 16) (Kaeberlein, Lewis, Epstein, et al. 2002; Jung et al. 2014; Jung et al. 2013; Tandogan et al. 2014). When compared to traditional cultivation (Petri dishes), the diffusion chamber has demonstrated a 300-fold higher recovery of marine sediment microorganisms (Kaeberlein, Lewis, Epstein, et al. 2002). Other studies have demonstrated that the diversity of isolates obtained by diffusion chambers significantly exceeded that of conventional plating (Bollmann et al. 2010). Increased recovery has also been shown by the use of the trap, which successfully isolated novel Actinobacteria, fungi, and other microorganisms (Gavrish et al. 2008). The filter plate method has resulted in culture collections that are more rich, diverse, and novel than standard cultivation approaches (Jung et al. 2013), and the itip has previously been used to culture from marine sponges, resulting in isolates that differed from traditional cultivation approaches (Jung et al. 2014).

A more recent development is the ipore. Here we tested its prototype, as a proof of concept to determine if small constrictions could be used to autonomously isolate individual species in the natural environment. The same principle regarding small-scale constrictions as the primary method of isolation has been shown to autonomously isolate cells from mixtures of two different species (Tandogan et al. 2014). The ipore devices tested in Greenland were kindly provided by Nil Tandogan and Ed Goluch of Northeastern University, and thus we were able to
analyze and use the isolated bacteria to increase our culture collection. The ipore premise is the same as other *in situ* devices: incubation within the natural environment allows access to naturally occurring growth compounds and nutrients. One of the most rate-limiting steps when processing *in situ* devices is the isolation of individual species from mixed cultures, but the ipore was designed to overcome this limitation. The specific device used in this dissertation was a micro- and nano-fabricated polydimethylsiloxane (PDMS) device designed to capture a single (or a few) bacterial cell(s) from mixed populations and allow propagation of the cells within chambers to produce pure cultures with easily retrievable biomass (Figure 12, Figure 16). The ipore is placed in the environment, and microbes can enter through the main entrance and move toward nutrients and chemicals inside the growth chambers (Tandogan et al. 2014). The constrictions are designed to restrict multiple cells from entering the growth chamber, as they are narrow enough so that the cross-sectional area should only permit one single cell to enter – thus blocking the opening from additional cells (Figure 12). As the entering cell grows and divides through the constriction it will propagate within the isolation chamber. A variety of constriction channel widths and lengths were used in this particular study to try and capture an array of species from the sediment. We note that our experimentation with the ipore was exploratory in nature: we have not yet optimized the process of controlling the diameter of the constriction, which therefore varied among the devices we used.
We hypothesized that leveraging a variety of *in situ* cultivation approaches, combined with enrichment and conventional techniques, would allow us to cultivate biologically active and relevant players within our study community. However, while we were successful in creating a vast culture collection of organisms, the isolates appear to represent only a fraction of the active community, with many of our isolates not even detected as present in 16S rDNA surveys (to be discussed in the next chapter). Furthermore, no single method of cultivation proved to be sufficient to represent the cultivable organisms within the environment. Rather, each method resulted in a large number of unique OTUs, and we conclude that multiple approaches must be used in conjunction with one another to access the bulk of microbial richness. It is possible however that the lack of overlap among the culture collections may be partially due to spatial microheterogeneity in microbial distribution, as different devices would then actually be incubated within different communities.

### 3.2 Cultivation-Dependent Methods

#### 3.2.1 Sampling Procedure for cultivation

Samples for standard cultivation were collected at three points throughout the season (Table 2). Sediment samples were collected in the same manner as described for the DNA
samples: using a sterile teaspoon, the uppermost 2-3 mm of sediment was placed in a 50 ml Falcon tube. Samples were immediately brought to the lab, and all samples (RL1 & RL2) were mixed together and vortexed. Serial dilutions were made using 1X PBS, starting with the undiluted sample through 10⁻⁵ dilutions. These dilutions were used for the standard cultivation and diffusion chamber experiments.

Table 2. Summary of isolates cultivated by each method. Date of device set up = date when in situ devices were placed in the environment. Date of collection = date the sediment was collected and plated or the date the in situ device was processed. *Note that due to windy conditions, a 1-gallon Tupperware container filled with a layer of sediment was used for incubation of the filter plate and itip devices for a portion of in situ cultivation.

<table>
<thead>
<tr>
<th>Method</th>
<th>Date of Collection*</th>
<th>Date Colonies Picked</th>
<th>Colonies Attempted for Isolation</th>
<th>Isolates Cultured, Sequenced &amp; Stored</th>
<th>Percent Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anaerobic (all) Ipose</td>
<td>16-Jul and 6-Aug</td>
<td>216</td>
<td>133</td>
<td>62%</td>
<td></td>
</tr>
<tr>
<td>Standard Cultivation Total</td>
<td>Sample Collection</td>
<td>569</td>
<td>318</td>
<td>56%</td>
<td></td>
</tr>
<tr>
<td>Standard Cultivation 1</td>
<td>11-Jun</td>
<td>14-Jun</td>
<td>168</td>
<td>113</td>
<td>67%</td>
</tr>
<tr>
<td>Standard Cultivation 2</td>
<td>8-Jul</td>
<td>13-Aug</td>
<td>160</td>
<td>93</td>
<td>58%</td>
</tr>
<tr>
<td>Standard Cultivation 3</td>
<td>22-Jul</td>
<td>25-Sep</td>
<td>241</td>
<td>112</td>
<td>46%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>Date of device set up</th>
<th>Device Retrieval</th>
<th>Colonies Picked</th>
<th>Isolates Cultured, Sequenced &amp; Stored</th>
<th>Percent Success</th>
</tr>
</thead>
<tbody>
<tr>
<td>In situ 1 total</td>
<td>502</td>
<td>371</td>
<td>1025</td>
<td>619</td>
<td>74%</td>
</tr>
<tr>
<td>Diffusion Chamber 1</td>
<td>5-Jun</td>
<td>26-Jun</td>
<td>1-Aug</td>
<td>225</td>
<td>73%</td>
</tr>
<tr>
<td>Trap 1</td>
<td>5-Jun</td>
<td>26-Jun</td>
<td>1-Aug</td>
<td>21</td>
<td>90%</td>
</tr>
<tr>
<td>Filter Plate 1</td>
<td>5-Jun</td>
<td>27-Jun</td>
<td>2-Aug</td>
<td>141</td>
<td>78%</td>
</tr>
<tr>
<td>Itip 1</td>
<td>5-Jun</td>
<td>27-Jun</td>
<td>2-Aug</td>
<td>115</td>
<td>67%</td>
</tr>
<tr>
<td>In situ 2 total</td>
<td>110</td>
<td>72</td>
<td>413</td>
<td>176</td>
<td>65%</td>
</tr>
<tr>
<td>Diffusion Chamber 2</td>
<td>30-Jun</td>
<td>15-Jul</td>
<td>18-Aug</td>
<td>86</td>
<td>67%</td>
</tr>
<tr>
<td>Trap 2</td>
<td>30-Jun</td>
<td>15-Jul</td>
<td>18-Aug</td>
<td>24</td>
<td>58%</td>
</tr>
<tr>
<td>In Situ 3 total</td>
<td></td>
<td></td>
<td>413</td>
<td>176</td>
<td>43%</td>
</tr>
<tr>
<td>Diffusion Chamber 3</td>
<td>29-Jul</td>
<td>8-Aug</td>
<td>15-Sep</td>
<td>260</td>
<td>48%</td>
</tr>
<tr>
<td>Trap 3</td>
<td>29-Jul</td>
<td>8-Aug</td>
<td>15-Sep</td>
<td>139</td>
<td>35%</td>
</tr>
<tr>
<td>Filter Plate 3</td>
<td>29-Jul</td>
<td>8-Aug</td>
<td>15-Sep</td>
<td>12</td>
<td>17%</td>
</tr>
<tr>
<td>Itip 3</td>
<td>29-Jul</td>
<td>8-Aug</td>
<td>15-Sep</td>
<td>2</td>
<td>50%</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td></td>
<td>1940</td>
<td>1173</td>
<td></td>
</tr>
</tbody>
</table>

40
**3.2.2. Cultivation conditions**

Three cultivation media were used: R2A, 1:100 Nutrient Agar (1:100 NA) and Soil Extract Agar (SE). R2A was made following the manufacturer recommendations (BD, Difco). A 1:100 dilution of Nutrient Agar was made using 0.8 g/L Nutrient Broth (Difco) and Bacto Technical Agar (15 g/L; Difco). Sediment from the lake was mixed with DI water and sterilized at 121°C and 15 PSI for one hour. The solution was allowed to sediment and the supernatant was collected. For SE agar, Bacto Technical Agar (15 g/L) was added to the sediment and autoclaved.

The average temperature of the lake throughout the entire sampling campaign was 10°C, however the sediment experiences colder temperatures during other parts of the season (such as 2°C). To simulate the natural conditions of the lake, all cultures were incubated at both 0-2°C and 10°C.

**3.2.3 Anaerobic and enrichment cultivation**

Maria Sizova performed anaerobic and enrichment cultivation. Hungate test tubes were filled with 10 ml liquid mineral media (per liter: 2 g K2HPO4, 1.5 g NaH2PO4, 1 g (NH4)2SO4, 0.1 g CaCl2, 0.4 g MgSO4, 0.01 g Na-EDTA, 1x10-3 g FeCl3.6H2O, 2x10-4 g KI, 2x10-4 g CoCl2.6H2O, 8x10-4 g MnCl2.4H2O, 8x10-4 g ZnSO4, 1x10-4 g H3BO3, 1x10-4 g Na2MoO4.2H2O, 1x10-4 g CuCl2 and 2x10-4 g NiCl2.6H2O) and supplemented with various carbon and energy sources to enrich for different bacterial groups.

To enrich for methanogens we used anaerobic mineral media with hydrogen (H2) and carbon dioxide (CO2) gas as energy and carbon sources, with and without reducing agent (sodium-citrate). For cultivation of aerobic methanotrophs we used aerobically prepared mineral medium with added methane gas. Aerobic hydrogen oxidizers were enriched for using mineral
media with H\textsubscript{2} and CO\textsubscript{2} gas. R2 liquid medium was prepared and organisms were incubated under low oxygen/anaerobic culture conditions to select for chemoorganotrophic organisms. We also enriched for anoxic phototrophs using three conditions: 1) H\textsubscript{2} and CO\textsubscript{2} gas for cyanobacteria 2) iron added to media with CO\textsubscript{2} gas for iron oxidizers, and 3) sulfate with H\textsubscript{2}, and CO\textsubscript{2} gases for green and purple bacteria. All samples were incubated in the presence of light.

In addition to enrichment cultivation, a portion of petri dishes from the processing of standard and diffusion chamber methods were anaerobically incubated under 95% nitrogen and 5% carbon dioxide in anaerobic boxes at room temperature.

\textbf{3.2.4 Standard Cultivation.}

One hundred microliters of each serial dilution, described above, was plated on each media (R2A, 1:100 Nutrient Agar, SE agar) in triplicate. Plates were divided equally between two incubation temperatures (2\textdegree C or 10\textdegree C) for 3-8 weeks (Table 2). Following incubation, plates were individually examined and dilutions resulting in single colonies were selected. Biomass from single colonies was lifted off the plate with a toothpick, restreaked, and incubated at the same temperature as the parent plate. Colonies were picked so that, to the best of our ability, we cultured a representative of each phenotype that was visible (to the naked eye). Lastly, we examined plates with denser growth for any additional unique phenotypes, which were isolated as above.

\textbf{3.2.5 In situ cultivation}

\textit{Diffusion chamber}

A 0.03 \textmu m pore-size polycarbonate membrane was affixed to the bottom of a stainless steel O-ring (diffusion chamber) using silicone glue, to allow exchange of chemicals but restrict
movement of cells (Figure 13A). Serially diluted sediment samples, as used for standard cultivation, were mixed with warm agar and used as the inoculum for our devices. Roughly 3 ml of agar-sediment mix was pipetted into the diffusion chamber. Following solidification of the agar, the top of the device was sealed with another 0.03 µm pore-size membrane using silicone glue. Sealed chambers were incubated in situ just below the surface of the sediment.

**Trap**

We used a microbial trap to enrich for filamentous, chain forming, and motile organisms (Figure 13B). The trap was constructed in the same way as the diffusion chamber, except the inoculum was warm sterile agar, and the bottom polycarbonate membrane pore size was larger (0.4 µm), allowing some organisms to move/grow through the larger pores and colonize the sterile space inside the device. A 0.03 µm membrane was used on top to seal the device.

![Figure 13. Design and application of the (A) diffusion chamber and (B) microbial trap. Figure from (Epstein et al. 2010).](image)
**Filter plate microbial trap (FPMT)**

A FPMT base plate contained 96 wells each serving as a small growth chamber. The bottom of each well was fitted with a hydrophilic polyvinylidene fluoride (PVDF) membrane with a 0.45 µm pore size (Figure 14). Wells were filled with agar and the entire device was placed on top of sediment to allow direct contact of the membrane with the target environment.

![Diagram of FPMT](image)

**Figure 14. Schematic of a growth chamber in a FPMT.** Figure from (Jung et al. 2013).

**Tip**

The lower portion of a sterilized standard yellow 200 µl pipette tip was filled with acid-washed glass beads to prevent invasion of larger organisms (Figure 15). Sterilized media containing 0.7% agar was added above the glass beads. The size of the glass beads varied, ranging from 60-100 µm in diameter to 150-212 µm in diameter. Organisms could enter the device through the narrow opening, and the opposite end (larger opening) was sealed with waterproof silicone adhesive.
Figure 15. Schematic of the itip. Figure from (Jung et al. 2014).

**Ipore**

The theory, design and proof of concept for this device has been published previously (Tandogan et al. 2014; Tandogan 2016). The devices (Figure 16) were kindly provided by the Goluch lab, and were deployed in Greenland to test the ability to isolate individual species within the natural environment. Ipore devices were placed in the sediment just beneath the surface. A variety of constriction channel widths and lengths were used to try and capture an array of different types of species from the environment. The optimal constriction size was not determined or relevant for the purposes of this dissertation.

Figure 16. Photograph of an ipore device. Photo from (Tandogan 2016).
3.2.6 Harvesting Growth from In Situ Devices and Laboratory Cultivation.

After 2-3 weeks of incubation (Table 2) in situ cultivation devices were collected, and grown biomass was harvested in a uniform manner. Devices were aseptically disassembled with a sterile blade; the agar containing microorganisms was carefully removed using sterile loops, added to sterile water or media, homogenized, and vortexed. This mixture was used as an inoculum for serial dilutions, which were spread on agar plates. For ipore devices, we used a sterile toothpick to remove the contents of wells and streak directly on agar plates. Three media types were used: R2A, 1:100 nutrient agar, and soil extract agar and cultures were grown at 0°C or 10°C. Following incubation, plates were processed as described in standard cultivation methods.

3.2.7 Isolation, Identification, and Downstream Analysis of Cultivated Species

Sealed petri dishes were transported at 0°C to Northeastern University and immediately returned to their original cultivation temperature (0-2°C or 10°C). Isolates were passaged multiple times on either 1% NA, 10% NA, or R2A, until determined pure by macroscopic and microscopic visualization. Pure isolates were archived in glycerol stocks (20% glycerol) at -80°C.

Taxonomic identification was performed by sequencing the 16S rRNA gene. Biomass from a colony was picked with a sterile toothpick and homogenized with DNA-grade water for colony PCR. One microliter (µl) of homogenate was used as a template for PCR-enabled 16S rRNA gene amplification. Alternatively, isolates were transferred to liquid R2A media. After growth, 1 ml of culture was pelleted and washed. One microliter of the supernatant directly above the pellet was used as the template. PCR-aided amplification of the 16S rRNA gene was performed using the 27F (AGA GTT TGA TCC TGG CTC AG) and 1492R (GGT TAC CTT
GTT ACG ACT T) primers (Lane 1991), and a HotStarTaq system (Qiagen, Catalog #203445). PCR conditions were as follows: initial denaturation for 15 min at 95°C, followed by 20 cycles of 1 minute 95°C, 1 minute 55°C and 1 minute at 72°C. PCR products were purified and sequenced commercially (Macrogen or Genewiz) by fluorescent dye terminator sequencing using 27F primers. As a result of poor quality sequences, some isolates were sequenced again with the use of 1492R primer. A total of 1173 isolates were sequenced.

Raw “.abi” files downloaded from Macrogen were quality controlled to remove primers and manually trimmed. The average sequence length was 772 base pairs (standard deviation = 129). OTUs were generated using the Uparse pipeline and isolates were clustered at 97% sequence similarity (Edgar 2013). Taxonomic assignment was performed using mother classifying.seqs (Schloss et al. 2009) and the Silva (SSURef_NR99_119) database (Quast et al. 2013; Pruesse et al. 2007).

The culture-independent 16S sequencing rDNA and rRNA data was mapped to the cultivated OTUs using UPARSE. Note that using the Silva v. 119 database for taxonomic identification did not allow every isolate to be classified at genus level with high confidence. If every individual isolate within an OTU was taxonomically identified within the same genus (with more than 80% confidence), that OTU was deemed a member of that genus for the purposes of further analysis (i.e. Chapter 4). When genus level identification was not possible, the OTU was identified using the next taxonomic ranking. Taxonomic grouping was used to assess what types of organisms preferred different cultivation methods. Diversity of culture collections was compared using Past v3.14 (Hammer et al. 2001).

In order to assess novelty of isolates, the SILVA Incremental Aligner (SINA v1.2.11) was used to align and taxonomically classify the 16S rRNA gene sequences from cultured
isolates (Pruesse et al. 2012). Using the Search and Classify feature, each sequence was classified with the last common ancestor (LCA) and percent similarity was noted.

3.3 Results and Discussion

3.3.1 Overview of Cultivated Microorganisms

A total of 1173 isolates were cultivated and, utilizing PCR-aided amplification of the 16S rRNA gene, taxonomically identified using the Silva v.117 database (Quast et al. 2013; Pruesse et al. 2007). The bulk of isolates classified within four bacterial phyla (Table 3), however eight isolates were eukaryotic fungi; these isolates were characterized by sequencing the ITS domain. All eukaryotic isolates belonged to the supergroup Opisthokonta. The 16S rRNA gene sequence from one isolate did not match any known domain. The sequence from this isolate was 525 base pairs, and was most closely related to the published strain Sphinogomonas faeni MA-olki(T) (69.19% similarity, determined by EZ-Taxon (Yoon et al. 2016) Database Version 2017.01). We are currently in the process of obtaining a full-length 16S rRNA sequence and plan to pursue whole genome sequencing in collaboration with JCVI.

Table 3. Phylum level break down of cultivated isolates from the summer of 2014. Proteobacteria are further divided into classes.

<table>
<thead>
<tr>
<th>Phylum</th>
<th>Number of Isolates</th>
<th>Percent of total collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proteobacteria</td>
<td>682</td>
<td>58.142%</td>
</tr>
<tr>
<td>Alpha-</td>
<td>193</td>
<td>16.454% (28% of Proteobacteria)</td>
</tr>
<tr>
<td>Beta-</td>
<td>309</td>
<td>26.343% (45% of Proteobacteria)</td>
</tr>
<tr>
<td>Gamma-</td>
<td>180</td>
<td>15.345% (26% of Proteobacteria)</td>
</tr>
<tr>
<td>Bacteroidetes</td>
<td>251</td>
<td>21.398%</td>
</tr>
<tr>
<td>Actinobacteria</td>
<td>220</td>
<td>18.755%</td>
</tr>
<tr>
<td>Firmicutes</td>
<td>11</td>
<td>0.938%</td>
</tr>
<tr>
<td>Fungi – Eukaryota</td>
<td>8</td>
<td>0.682%</td>
</tr>
<tr>
<td>Unclassified</td>
<td>1</td>
<td>0.085%</td>
</tr>
<tr>
<td>Total</td>
<td>1173</td>
<td>100%</td>
</tr>
</tbody>
</table>

Proteobacterial isolates were the most abundant in our collection; more than half of isolates (58%) were classified within this phylum (Table 3). The predominance of Proteobacteria
in culture collections, from the Arctic (and Antarctic) environment, is expected based on published literature (Steven et al. 2007; Møller et al. 2013; Xuezheng et al. 2014). We cultivated three lineages from Proteobacteria: Alpha-, Beta-, and Gammaproteobacteria. Most isolates were classified as Betaproteobacteria (45%), which have been found to dominate freshwater systems (Kirchman 2002), such as our lake. The remaining Proteobacterial isolates were equally divided between the Alphaproteobacteria (28%) and Gammaproteobacteria (26%) classes (Table 3). We also frequently encountered members of the phylum Bacteroidetes and Actinobacteria, comprising 21% and 18% of isolates, respectively. In addition, we cultivated 11 isolates from the Phylum Firmicutes. Our collection is consistent with other Arctic cultivation studies (Steven et al. 2007; Møller et al. 2013; Pearce et al. 2003; Xuezheng et al. 2014).

The majority of bacterial isolates belonged to the following three orders: Burkholderiales, Flavobacteriales, and Micrococcales (25.92%, 19.69%, and 17.05% of the total collection, respectively, Table 4). Møller and colleagues conducted a study of Arctic snow, and faced trouble cultivating organisms from Burkholderiales, in contrast with our study (Møller et al. 2013). Flavobacteriales are within the Cytophaga-Flavobacterium cluster, and can commonly be found in Arctic environments, including sea ice (Staley & Gosink 1999), cold marine surface waters (Wells & Deming 2003), and ice covered freshwater lakes (Møller et al. 2013). These organisms may play an important role as heterotrophs in aquatic environments (Kirchman 2002), particularly in marine and sea-ice locations. We also encountered a number of Pseudomonadales and Sphingomonadales isolates. Flavobacteriales, Pseudomonadales, and Sphingomonadales have been observed within Arctic (and Antarctic, (Shivaji et al. 2011) environments using culture-dependent (Steven et al. 2007) and independent methods (Pearce et al. 2003; Møller et al. 2013; Tazi et al. 2014; Ravenschlag et al. 2001; Lay et al. 2013; Tian et al. 2009).
Table 4. Order level breakdown of cultivated isolates (N=24).

<table>
<thead>
<tr>
<th>Order</th>
<th>Total</th>
<th>Percent total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burkholderiales</td>
<td>304</td>
<td>25.92%</td>
</tr>
<tr>
<td>Flavobacteriales</td>
<td>231</td>
<td>19.69%</td>
</tr>
<tr>
<td>Micrococcales</td>
<td>200</td>
<td>17.05%</td>
</tr>
<tr>
<td>Pseudomonadales</td>
<td>128</td>
<td>10.91%</td>
</tr>
<tr>
<td>Sphingomonadales</td>
<td>85</td>
<td>7.25%</td>
</tr>
<tr>
<td>Caulobacteriales</td>
<td>78</td>
<td>6.65%</td>
</tr>
<tr>
<td>Enterobacteriales</td>
<td>29</td>
<td>2.47%</td>
</tr>
<tr>
<td>Rhizobiales</td>
<td>22</td>
<td>1.88%</td>
</tr>
<tr>
<td>Propionibacteriales</td>
<td>18</td>
<td>1.53%</td>
</tr>
<tr>
<td>Aeromonadales</td>
<td>17</td>
<td>1.45%</td>
</tr>
<tr>
<td>Sphingobacteriales</td>
<td>11</td>
<td>0.94%</td>
</tr>
<tr>
<td>Cytophagales</td>
<td>9</td>
<td>0.77%</td>
</tr>
<tr>
<td>Lactobacillales</td>
<td>9</td>
<td>0.77%</td>
</tr>
<tr>
<td>Fungi</td>
<td>8</td>
<td>0.68%</td>
</tr>
<tr>
<td>Neisseriales</td>
<td>5</td>
<td>0.43%</td>
</tr>
<tr>
<td>Chromatiales</td>
<td>4</td>
<td>0.34%</td>
</tr>
<tr>
<td>Rhodobacteriales</td>
<td>4</td>
<td>0.34%</td>
</tr>
<tr>
<td>Rhodospirillales</td>
<td>4</td>
<td>0.34%</td>
</tr>
<tr>
<td>Xanthomonadales</td>
<td>2</td>
<td>0.17%</td>
</tr>
<tr>
<td>Bacillales</td>
<td>1</td>
<td>0.09%</td>
</tr>
<tr>
<td>Clostridiales</td>
<td>1</td>
<td>0.09%</td>
</tr>
<tr>
<td>Solirubrobacteriales</td>
<td>1</td>
<td>0.09%</td>
</tr>
<tr>
<td>Streptomycetales</td>
<td>1</td>
<td>0.09%</td>
</tr>
<tr>
<td>Unknown</td>
<td>1</td>
<td>0.09%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1173</strong></td>
<td><strong>100.00%</strong></td>
</tr>
</tbody>
</table>

Our isolates represented 89 genera, which clustered into 183 OTUs at 97% sequence similarity (Table 5). Two genera appear especially abundant, each comprising ≥10% of our entire collection: *Flavobacteria* (19%) and *Pseudomonas* (10%). We cultured 128 organisms from the Pseudomonadales order of which 121 (94.5%) were identified as *Pseudomonas* species that clustered into eight distinct OTUs. The genus *Flavobacterium* also was the main component of the Flavobacteriales order (227 out of 231 isolates, 98.2%). There was more intra-generic diversity among Flavobacterial isolates, which clustered into 27 unique OTUs. The data from the
*Pseudomonas* and *Flavobacteria* isolates from the summer of 2014 correspond well to cultivation data obtained during the summer of 2013. In 2013 our lab performed a cultivation survey from multiple locales around Thule Airbase (unpublished). The 2013 culture collection was dominated by *Pseudomonas* (66% of isolates) and *Flavobacteria* (17%) isolates. As with species cultivated in 2014, we noted more intra-generic diversity of *Flavobacteria* (21 OTUs) compared to *Pseudomonas* (6 OTUs). Marked variability within the Flavobacterial clade has been documented before (Gómez-Pereira et al. 2010).

Further investigation into the Flavobacteria cultured here may provide useful information regarding biogeochemical cycles in the Arctic, as the Cytophaga-Flavobacterium-Bacteroides (CFB) clade are major constituents of sea ice, which makes up approximately 7% of the earth's surface (Staley & Gosink 1999). Future studies should include whole genome sequencing of representative isolates from each OTU. Additional sampling from sediments for transcriptomic analysis would allow us to infer the activity and functional role of this genus within the Arctic lake communities.

Table 5. Summary of isolates cultivated, as broken down by genus (N=89).

<table>
<thead>
<tr>
<th>Genus (#OTUs)</th>
<th>Sum of Total Isolates</th>
<th>Percent of Total</th>
<th>Genus</th>
<th>Sum of Total Isolates</th>
<th>Percent of Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flavobacterium (27)</td>
<td>227</td>
<td>19.352%</td>
<td>Psychrobacter</td>
<td>4</td>
<td>0.341%</td>
</tr>
<tr>
<td>Pseudomonas (8)</td>
<td>121</td>
<td>10.315%</td>
<td>Rheinheimera</td>
<td>4</td>
<td>0.341%</td>
</tr>
<tr>
<td>Sphingomonas (4)</td>
<td>73</td>
<td>6.223%</td>
<td>Acinetobacter</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>Cryobacterium (4)</td>
<td>66</td>
<td>5.627%</td>
<td>Actinotalea</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>Arthrobacter (6)</td>
<td>57</td>
<td>4.859%</td>
<td>Algoriphagus</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>Polaromonas (3)</td>
<td>52</td>
<td>4.433%</td>
<td>Devosia</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>Janthinobacterium (1)</td>
<td>49</td>
<td>4.177%</td>
<td>Dikarya</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>Massilia (4)</td>
<td>44</td>
<td>3.751%</td>
<td>Mucilaginibacter</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>Brevundimonas (6)</td>
<td>43</td>
<td>3.666%</td>
<td>Noviherbaspirillum</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>Microbacteriaceae genus (3)</td>
<td>43</td>
<td>3.666%</td>
<td>Phenyllobacterium</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>Caulobacter (2)</td>
<td>32</td>
<td>2.728%</td>
<td>Rathayibacter</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>Herminiimonas (1)</td>
<td>29</td>
<td>2.472%</td>
<td>Sandarakinorhavirus</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>Comamonadaceae genus (3)</td>
<td>23</td>
<td>1.961%</td>
<td>Sphingomonadaceae genus</td>
<td>3</td>
<td>0.256%</td>
</tr>
<tr>
<td>genus</td>
<td>species</td>
<td>number</td>
<td>percentage</td>
<td>species</td>
<td>number</td>
</tr>
<tr>
<td>------------------------</td>
<td>------------------</td>
<td>--------</td>
<td>------------</td>
<td>------------------</td>
<td>--------</td>
</tr>
<tr>
<td>Serratia</td>
<td></td>
<td>20</td>
<td>1.705%</td>
<td>Subtercola</td>
<td>3</td>
</tr>
<tr>
<td>Oxalobacteraceae</td>
<td></td>
<td>19</td>
<td>1.620%</td>
<td>Trichococcus</td>
<td>3</td>
</tr>
<tr>
<td>Aeromonas (1)</td>
<td></td>
<td>17</td>
<td>1.449%</td>
<td>Undibacterium</td>
<td>3</td>
</tr>
<tr>
<td>Albidiferax (1)</td>
<td></td>
<td>16</td>
<td>1.364%</td>
<td>Comamonadaceae</td>
<td>2</td>
</tr>
<tr>
<td>Paucibacter (1)</td>
<td></td>
<td>16</td>
<td>1.364%</td>
<td>Dyadobacter</td>
<td>2</td>
</tr>
<tr>
<td>Comamonadaceae</td>
<td></td>
<td>13</td>
<td>1.108%</td>
<td>Kocuria</td>
<td>2</td>
</tr>
<tr>
<td>(Albidiferax) (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duganella (1)</td>
<td></td>
<td>13</td>
<td>1.108%</td>
<td>Phycicoccus</td>
<td>2</td>
</tr>
<tr>
<td>Bosea (1)</td>
<td></td>
<td>12</td>
<td>1.023%</td>
<td>Sanguibacter</td>
<td>2</td>
</tr>
<tr>
<td>Hydrogenophaga</td>
<td></td>
<td>9</td>
<td>0.767%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nocardioides</td>
<td></td>
<td>8</td>
<td>0.682%</td>
<td>Acetobacteraceae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>genus</td>
<td></td>
</tr>
<tr>
<td>Yersinia</td>
<td></td>
<td>8</td>
<td>0.682%</td>
<td>Actimicrobium</td>
<td>1</td>
</tr>
<tr>
<td>Comamonadaceae</td>
<td></td>
<td>7</td>
<td>0.597%</td>
<td>Blastomonas</td>
<td>1</td>
</tr>
<tr>
<td>(Variovorax)</td>
<td></td>
<td></td>
<td></td>
<td>Clostridium</td>
<td>1</td>
</tr>
<tr>
<td>Pedobacter</td>
<td></td>
<td>7</td>
<td>0.597%</td>
<td>Curtobacterium</td>
<td>1</td>
</tr>
<tr>
<td>Carnobacterium</td>
<td></td>
<td>6</td>
<td>0.512%</td>
<td>Dothideomycetes</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Dikarya)</td>
<td></td>
</tr>
<tr>
<td>Marisediminicola</td>
<td></td>
<td>6</td>
<td>0.512%</td>
<td>Enterobacteraceae</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>genus</td>
<td></td>
</tr>
<tr>
<td>Microbacterium</td>
<td></td>
<td>6</td>
<td>0.512%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aeromicrobium</td>
<td></td>
<td>5</td>
<td>0.426%</td>
<td>Ferruginibacter</td>
<td>1</td>
</tr>
<tr>
<td>Isodobacter</td>
<td></td>
<td>5</td>
<td>0.426%</td>
<td>Janibacter</td>
<td>1</td>
</tr>
<tr>
<td>Rhizobium</td>
<td></td>
<td>5</td>
<td>0.426%</td>
<td>Leifsonia</td>
<td>1</td>
</tr>
<tr>
<td>Agaricomycotina</td>
<td></td>
<td>4</td>
<td>0.341%</td>
<td>Leptothrix</td>
<td>1</td>
</tr>
<tr>
<td>(Dikarya)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chryseobacterium</td>
<td></td>
<td>4</td>
<td>0.341%</td>
<td>Leucobacter</td>
<td>1</td>
</tr>
<tr>
<td>Hymenobacter</td>
<td></td>
<td>4</td>
<td>0.341%</td>
<td>Luteimonas</td>
<td>1</td>
</tr>
<tr>
<td>Nocardoides</td>
<td></td>
<td>4</td>
<td>0.341%</td>
<td>Marmorica</td>
<td>1</td>
</tr>
<tr>
<td>Novosphingobium</td>
<td></td>
<td>4</td>
<td>0.341%</td>
<td>Methylobacterium</td>
<td>1</td>
</tr>
<tr>
<td>Pseudorhodobacter</td>
<td></td>
<td>4</td>
<td>0.341%</td>
<td>Microbacteriaceae</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Micrococcus</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Mycetocola</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Oxalobacteraceae</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Patulibacter</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pigmentiphaga</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Porphyrobacter</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Reyranella</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Roseomonas</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Roseomonas</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Staphylococcus</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Streptomyces</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Tardiphaga</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Thermomonas</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Unknown</td>
<td>1</td>
</tr>
</tbody>
</table>
3.3.2 Comparison of Standard, Anaerobic, and In Situ Cultivation Approaches

To increase microbial recovery beyond that which can typically be achieved via standard cultivation, we additionally employed multiple in situ methods. Furthermore, since preliminary data from a culture-independent survey of the 16S rRNA gene, performed in 2013, indicated the presence of anaerobic organisms (unpublished), and in collaboration with Maria Sizova, we additionally incorporated the use of anaerobic enrichments to capture potential anaerobes.

We observed minimal overlap among culture collections obtained by different approaches: only 21 of 183 OTUs were common among all three techniques (Figure 17). Both standard and in situ cultivation resulted in large culture libraries: 58 and 56 unique OTUs respectively, and an additional 50 OTUs encountered by both methods. The recovery of similar numbers of OTUs using both approaches was surprising since we cultivated twice as many isolates compared to standard cultivation. Results from previous studies that utilized in situ approaches found richer culture collections compared to standard cultivation (Jung et al. 2013; Bollmann et al. 2007; Jung et al. 2014). In contrast, our in situ isolate collection was not significantly more rich or diverse than the standard cultivation collection (Table 6).
Figure 17. Overlap of OTUs cultured by each approach. N= number of total isolates cultivated by that method. Note: all in situ methods were treated as one approach, resulting in a total of 722 isolates. Circles are scaled to represent percent of overlap.

Table 6. Diversity indices of standard and in situ collections. Calculations by PAST3 software (Hammer et al. 2001).

<table>
<thead>
<tr>
<th></th>
<th>Standard (N=318)</th>
<th>In Situ (N=722)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxa_S</td>
<td>110</td>
<td>114</td>
</tr>
<tr>
<td>Individuals</td>
<td>318</td>
<td>722</td>
</tr>
<tr>
<td>Dominance_D</td>
<td>0.03</td>
<td>0.02883</td>
</tr>
<tr>
<td>Simpson_1-D</td>
<td>0.97</td>
<td>0.9712</td>
</tr>
<tr>
<td>Shannon_H</td>
<td>4.117</td>
<td>3.966</td>
</tr>
<tr>
<td>Evenness_e^H/S</td>
<td>0.558</td>
<td>0.4627</td>
</tr>
<tr>
<td>Chao-1</td>
<td>173.8</td>
<td>202.4</td>
</tr>
</tbody>
</table>

Utilizing anaerobic approaches, we cultivated nine unique OTUs from the following genera: *Yersinia, Clostridium, Carnobacterium, Mucilaginibacter*, and *Flavobacterium*; we also isolated one fungus (class: Dothideomycetes). All of these genera are either facultative or
obligate anaerobes, explaining their presence in the anaerobic enrichment cultures. In line with our findings is that *Carnobacteria* species are known to withstand freezing temperatures (Leisner et al. 2007), and some strains of *Mucilaginibacter* have been previously detected in High Arctic tundra (Jiang et al. 2012).

While representatives of most taxonomic divisions using anaerobic methods, over 40% of Firmicutes cultivated were obtained by this approach (Figure 18). We also noted that most eukaryotic organisms were isolated by standard approaches (5 of 8), while Proteobacterial isolates were predominantly grown via *in situ* methods. We term this tendency of the given approach to preferentially cultivate representatives of some taxonomic divisions as “cultivation preference.” Cultivation preferences observed throughout our study are summarized in Table 8 at the end of this section.

![Phylum breakdown for standard compared to in situ cultivation methods](image)

**Figure 18. Phylum break down for standard compared to *in situ* cultivation methods.** All values were normalized by dividing the number of isolates in a phylum cultured using a specific approach by the total number of isolates in that phylum.

Cultivation preferences were increasingly noticeable as isolates were grouped into taxonomic divisions (i.e. order level, Figure 19). It is important to note there is no way to
accurately correlate the number of organisms cultured directly to their abundance in nature since isolates were not selected randomly but by phenotype. The number of isolates from a given taxon is used as a proxy for its abundance, with the underlying assumption that the more abundant an organism is, the more likely we are to encounter it multiple times.

Figure 19. Distribution of recovered isolates, by order, for in situ, standard, and anaerobic methods. Percent indicates the proportion of isolates obtained within that order. All values were normalized by dividing the number of isolates in a phylum cultured using a specific approach by the total number of isolates in that phylum.

**Taxonomy of isolates cultured mostly (or only) via standard cultivation**

Representatives from the orders Xanthomonadales (two isolates), Streptomycetales (one isolate), and Cytophagales (nine isolates, six OTUs) were isolated exclusively by traditional
cultivation. Note that taxonomy of Cytophagales is quite complex and controversial. The families within this order have faced a number of phylogenetic regroupings since the establishment of the genus Cytophaga in the 1940s (DeLong et al. 2014). Many bacteria previously thought to be of the order Cytophagales have since been reclassified into various genera within the family Flavobacteriaceae. The nine isolates identified as members of Cytophagales were classified using the SINA pipeline and SILVA database, which currently maintains the original Cytophagales classification for many organisms. According to the GenBank database (NCBI Resource Coordinators, 2016), our isolates had been reclassified to the order Flavobacteriales. Other members of this order were isolated via in situ approaches, but these six OTUs were unique to standard cultivation.

**Taxonomy of isolates cultured mostly or entirely by in situ cultivation**

Isolates belonging to the orders Solirubrobacterales (1 isolate), Chromatiales (4 isolates), and Bacillales (1 isolate) were cultivated exclusively by the in situ approaches. The singular Bacillales isolate was most closely related to *Staphylococcus aureus* and was obtained by the ipore method. It is unclear if this isolate was a human contaminant or a naturally occurring strain, since related sequences were detected in both the 16S rDNA and rRNA libraries, potentially indicating its prevalence and activity in nature.

The four cultured organisms from Chromatiales clustered into two OTUs, both most closely related to the genus *Rheinheimera*. This genus is relatively new, as it was first documented in 2002 (Brettar et al. 2002), and has been amended four times, most recently in 2012. As of 2015 there were 16 species within this genus (Kumar et al. 2015). *Rheinheimera*
species have not been studied extensively, and may require growth very specific factors or nutrients, which could conceivably be provided by the use of in situ cultivation devices.

**Multiple cultivation approaches are necessary to culture the microbial community**

We propose that neither standard nor in situ approaches alone are sufficient to access the representative members of a community. Møller and colleagues used a pre-incubation approach, using culture media directly from the sample site to provide unknown growth factors (Møller et al. 2013), and this resulted in a culture collection distinct from that obtained via standard methods. In the present study, each approach used resulted in different collections, and the collections were equally rich (Table 6), indicating both standard and in situ cultivation are necessary to increase microbial recovery.

Furthermore, we isolated most orders by a combination of approaches (Figure 19), but still observed that some methods exhibited “cultivation preferences”. For example Caulobacterales and Burkholderiales isolates were obtained primarily via in situ approaches (70% and 74% of all isolates of that order, respectively). Provided cultivation preferences exist, and the culture collections obtained by different methods vary greatly, we propose multiple approaches must be used in conjunction with one another.

Microheterogeneity can also explain the dissimilarity between collections obtained by standard and in situ approaches. For example, we cultivated only one Streptomycetales isolate and two Xanthomonadales isolates using standard cultivation. If they were simply fastidious organisms, we would anticipate in situ approaches to improve cultivability, however we did not culture any isolates from these orders using in situ approaches. Alternatively, microheterogeneity
may contribute to spatial distribution of these species such that they were only present in some samples. Therefore, it was simply by chance we did not encounter them by *in situ* approaches.

### 3.3.3 Comprehensive Comparison of *In Situ* Cultivation Methods

To our knowledge, this study represents the largest culture collection generated during a single cultivation survey and is the first to compare multiple *in situ* approaches. We expected some variation between culture collections obtained by various *in situ* methods, as each device was designed to overcome a limitation or target a specific group of organisms. The trap selects for filamentous, motile, and *Actino*-type organisms (Gavrish et al. 2008). The trap device can be easily overgrown by fast-growing species, and the filter plate was designed to overcome this limitation. It functions like a trap, but each contains 96 individual small wells to prevent overgrowth (Jung et al. 2013). The itip was initially designed to cultivate microorganisms associated with marine sponges, and therefore had a smaller area for microbial entry, compared to the large flat surface of the filter plate and other devices (Jung et al. 2014). The ipore prototype used here highly selects for motile bacteria because the only way they can reach growth chambers is through a maze of microfluidic channels (Figure 12) (Tandogan et al. 2014; Tandogan 2016). The diffusion chamber does not rely on motility, as it is inoculated prior to *in situ* incubation. However, multiple cells are placed in the single diffusion chamber, enabling fast growing or metabolically competitive species to easily outcompete others (Freilich et al. 2011).

As expected, we found the majority of isolates cultivated by the ipore can be motile, however, three isolates recovered by the ipore were not. Of these, two were Actinobacteria, and thus may be filamentous, allowing them to reach the entrance of the device and continue to grow with branching mycelium.
The richest and most diverse culture collections resulted by use of the diffusion chamber and ipore devices (Table 7). Interestingly, we observed minimal overlap of cultivated OTUs by *in situ* approaches, with only one OTU common to all methods (Figure 20). As expected, we noted method specific “cultivation preferences” that were more noticeable as taxonomy level decreases from Phylum (Figure 21) to Order (Figure 22). We also observed some genus specific patterns.

**Table 7. Diversity indices of *in situ* approaches.** Calculations performed using Past3 software (Hammer et al. 2001).

<table>
<thead>
<tr>
<th></th>
<th>DC</th>
<th>T</th>
<th>FP</th>
<th>I</th>
<th>IP</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Taxa_S</strong></td>
<td>66</td>
<td>27</td>
<td>19</td>
<td>25</td>
<td>46</td>
</tr>
<tr>
<td><strong>Individuals</strong></td>
<td>347</td>
<td>82</td>
<td>112</td>
<td>78</td>
<td>103</td>
</tr>
<tr>
<td><strong>Dominance_D</strong></td>
<td>0.04633</td>
<td>0.06336</td>
<td>0.1629</td>
<td>0.1913</td>
<td>0.05345</td>
</tr>
<tr>
<td><strong>Simpson_1-D</strong></td>
<td>0.9537</td>
<td>0.9366</td>
<td>0.8371</td>
<td>0.8087</td>
<td>0.9466</td>
</tr>
<tr>
<td><strong>Shannon_H</strong></td>
<td>3.536</td>
<td>3.006</td>
<td>2.22</td>
<td>2.405</td>
<td>3.414</td>
</tr>
<tr>
<td><strong>Evenness_e^H/S</strong></td>
<td>0.5202</td>
<td>0.7486</td>
<td>0.4845</td>
<td>0.4433</td>
<td>0.6604</td>
</tr>
<tr>
<td><strong>Chao-1</strong></td>
<td>89</td>
<td>36</td>
<td>31</td>
<td>65</td>
<td>73.08</td>
</tr>
</tbody>
</table>

**Figure 20. Venn diagram depicting the overlap of OTUs cultivated by *in situ* methods.** Numbers within Venn diagram indicate OTUs unique to/shared between approaches. N = total number of OTUs cultured using the given approach.
Figure 21. Phyla cultured using in situ approaches. Taxonomic preferences can be observed at the phylum level. Firmicutes and eukaryotes were only cultivated by two methods (each), while Proteobacteria were well represented in all approaches. All values were normalized by dividing the number of isolates in a phylum cultured using a specific approach by the total number of isolates in that phylum.

Figure 22. Orders cultivated via in situ methods. Taxonomic preferences can be seen, such as exclusive cultivation of Neisseriales by the itip. Orders without any bars were cultivated by standard approaches only. DC = diffusion chamber, T = trap, FP = filter plate, I = itip, IP = ipore. All values were normalized by dividing the number of isolates in a phylum cultured using a specific approach by the total number of isolates in that phylum.
**Eukaryotic Isolate Trends**

All eukaryotic organisms in our culture collection were isolated via the diffusion chamber or itip approaches, most likely a result of the large cell size of eukaryotes compared to bacteria. Eukaryotic cells may be too large to enter through the small opening of the ipore or to penetrate the small pore sized polycarbonate membranes (filter plate and trap). Previous studies using the itip have also found culture collections to contain many fungal isolates (Jung et al. 2014).

**Phylum Firmicutes**

The only phylum in our collection that we were unable to grow via the diffusion chamber was Firmicutes (Figure 21). Another study that compared cultivation methods cultured fewer Firmicutes by the diffusion chamber (one isolate) compared to standard cultivation (five isolates), (Bollmann et al. 2010), indicating the diffusion chamber may not be ideal for this phylum. In this study we cultivated four isolates from this phylum by either the trap or ipore devices (Figure 21). One of these was the *Staphylococcus* species described above. The other three Firmicutes were related to the genus *Carnobacterium*, and all clustered together. This genus is known to inhabit polar regions (Leisner et al. 2007), and possesses motility that would allow access to the ipore.

**Order-Level Patterns**

Sphingomonadales activity was detected in the 16S rRNA libraries, and we were able to successfully cultivate members from this order. Most isolates were cultured via the diffusion chamber (56%) and standard cultivation (27%) approaches; one isolate was cultivated by the itip,
four were grown in a trap, and the filter plate grew none. The majority of isolates cultivated from this order clustered to two OTUs, both most closely related to *Sphingomonas* species.

The outer cell layer of Sphingomonadales organisms, including *Sphingomonas* species, possess sphingoglycolipids, contributing to slight hydrophobicity of the layer (Rosenberg et al. 2014a). The slightly hydrophobic membrane may make it harder for these species to penetrate devices with hydrophilic membranes, such as the trap and filter plate. Diffusion chamber and standard cultivation methods involve direct plating of cells on/in the media, and thus eliminate exposure to a hydrophilic membrane. Furthermore, we cultivated *Sphingomonas* via the itip, which does not utilize a membrane and could easily be colonized due to the motility of some of these species (Rosenberg et al. 2014a).

Isolates of the order Flavobacteriales were commonly encountered by most methods used except by ipore and itip devices (<2%). Most isolates (227/231) were related to the genus *Flavobacteria*. This genus is known to exhibit a gliding motility that requires respiratory metabolism (i.e. O2). Once oxygen levels decrease, bacterial movement completely stops (DeLong et al. 2014), as can be seen during motility testing using a wet mount. Therefore, only cultivation approaches that allow for continued oxygen exposure would also allow motility. The main openings of the itip and ipore device serve as the source of oxygen: the channels in the ipore are surrounded by PDMS, and the itip is constructed from of plastic; both materials prevent diffusion of oxygen from the environment. Once an organism entered these openings, it is possible that the oxygen level decreased in the pipette tip or channels of the ipore preventing further colonization. Many other taxa were cultivated by itip and ipore approaches, potentially because their motility allowed them to outcompete *Flavobacterium* species. All other cultivation
methods that required motility also had constant oxygen exposure (trap, filter plate) or involved direct plating (diffusion chamber, standard cultivation).

About 80% of Caulobacterales isolates were cultivated via the filter plate or the ipore device. The filter plate may have selected for these organisms due to their slow-growth rate (Rosenberg et al. 2014a). The smaller growth compartments would eliminate faster growing competitors. In addition, the individual chambers of the ipore device may have provided similar space for growth without competition. This order also has a two-part life cycle consisting of prosthecate and swarmer cells. The PDMS walls of the ipore’s main entry may have served as a base for the prosthecate cell, which produces flagellated swarming cells that can swim through channels toward constrictions (Rosenberg et al. 2014a).

Aeromonadales isolates comprised 1% of our collection and were cultivated exclusively by the diffusion chamber (aerobic and anaerobic). Some species of this order are known to be facultatively anaerobic, thus subculturing the diffusion chamber-grown material under anaerobic conditions would favor their growth. Standard cultivation involves directly placing cells from the environment into a petri dish, which may have shocked the cells if they prefer microaerophilic conditions; a round of in situ cultivation in the diffusion chamber, prior to growth in the lab, may have resulted in domestication and increased recovery of these species (Rosenberg et al. 2014b; Bollmann et al. 2007). Furthermore, some species lack motility, and therefore would be unable to access all devices favoring motile strains.

Genus Level Patterns

Consistent with our preliminary study in 2013, we cultivated many organisms from the Pseudomonas genus, which was also the second most prominent genus in our collection (Table
The majority of isolates were grown via the itip method. *Pseudomonas* species are known to be motile and growing quickly, allowing them to easily colonize the itip (Ramos et al. 2015; Ramos & Filloux 2010). In contrast, a previous study using the itip in marine sponges found that this genus was rarely encountered (Jung et al. 2014).

Only one OTU was cultivated from the genus *Herminimonas* but it was comprised of 29 isolates all cultivated by the filter plate. Organisms from this genus are able to survive at very low temperatures. Some of these species reproduce optimally at 30°C, with a doubling time of 4 hours (Lang et al. 2007). Others, however, require extremely long term (months) incubation for recovery (Loveland-Curtze et al. 2009). In the present study, we incubated organisms at potentially suboptimal temperatures between 0 and 10°C. In addition, *Herminimonas glaciei* is described as an ultramicrobacterium (Loveland-Curtze et al. 2009). The extremely small size, and inherently slow growth rate of some species in this genus, would make the filter plate ideal for its cultivation, as it prevents overgrowth by and thus competition from larger, faster growing species.

Table 8. Summary of cultivation preferences for different taxonomic divisions.

<table>
<thead>
<tr>
<th>Group</th>
<th>Most effective method for Isolation</th>
<th>Least effective method for isolation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Firmicutes</td>
<td>Anaerobic, <em>In situ</em>: trap, ipore</td>
<td>Diffusion chamber</td>
</tr>
<tr>
<td>Eukaryotes</td>
<td>Standard, <em>In situ</em>: diffusion chamber, itip</td>
<td></td>
</tr>
<tr>
<td>Proteobacteria</td>
<td><em>In situ</em></td>
<td></td>
</tr>
<tr>
<td>Xanthomonadales</td>
<td>Standard</td>
<td></td>
</tr>
<tr>
<td>Streptomycetales</td>
<td>Standard</td>
<td></td>
</tr>
<tr>
<td>Cytophagales</td>
<td>Standard</td>
<td></td>
</tr>
<tr>
<td>Solirubrobacterales</td>
<td><em>In situ</em></td>
<td></td>
</tr>
<tr>
<td>Chromatiiales</td>
<td><em>In situ</em></td>
<td></td>
</tr>
<tr>
<td>Bacillales</td>
<td><em>In situ</em></td>
<td></td>
</tr>
<tr>
<td>Caulobacterales</td>
<td><em>In situ</em></td>
<td></td>
</tr>
<tr>
<td>Burkholderiales</td>
<td><em>In situ</em></td>
<td></td>
</tr>
<tr>
<td>Sphingomonadales</td>
<td><em>In situ</em>: diffusion chamber</td>
<td></td>
</tr>
<tr>
<td>Flavobacteriales</td>
<td></td>
<td>Ipore, itip</td>
</tr>
<tr>
<td>Caulobacterales</td>
<td><em>In situ</em>: filter plate</td>
<td></td>
</tr>
<tr>
<td>Neisseriales</td>
<td><em>In situ</em>: itip</td>
<td></td>
</tr>
<tr>
<td>Aeromonadales</td>
<td><em>In situ</em>: diffusion chamber</td>
<td></td>
</tr>
<tr>
<td><em>Pseudomonas</em></td>
<td><em>In situ</em>: itip</td>
<td></td>
</tr>
<tr>
<td><em>Herminimonas</em></td>
<td><em>In situ</em>: filter plate</td>
<td></td>
</tr>
</tbody>
</table>
Specific growth requirements and life strategies of species may partly explain the minimal overlap of OTUs cultivated by different methods. However, it was surprising that the majority of OTUs detected were exclusive to a single cultivation method, and only one OTU was common among all five approaches. We propose the lack of OTU overlap is a result of two different factors. First, as discussed above, the method or device used will inherently select for certain species. Second, we believe this to be further compounded by small-scale biogeochemical differences within the environment that result in spatial heterogeneity in microbial distribution, as discussed in Chapter 2. Therefore, it is possible that in situ devices allow for isolation of those microbes that are actively growing in the locations where the devices are placed. Owing to the size of the devices, these locations are always centimeters apart.

### 3.3.4 Novelty of Isolates

In the past, in situ approaches have all proven successful in recovering organisms different from, and more novel than, those typically cultivated using standard methods (Kaeberlein, Lewis, Epstein, et al. 2002; Gavrish et al. 2008; Bollmann et al. 2010; Bollmann et al. 2007; Aoi et al. 2009; Rappé et al. 2002; Ben-Dov et al. 2009; Jung et al. 2014; Jung et al. 2013). Novelty of cultivated OTUs was assessed using the SINA sequence aligner (Pruesse et al. 2012) and comparing the representative sequence to the last common ancestor (LCA) (Figure 23). Most OTUs had a 98-99% similar sequence match to a LCA, indicating our isolates may fall into already recognized and cultivated genera or species.
Figure 23. Percent of isolate collection (by method) that matched to a LCA within the indicated percent similarity.

We note that there are known limitations to using the 16S rRNA gene as the only means of classification. In this study, we use it only to implicate potential novelty. It is standard practice to consider a sequence similarity of less than 97% as a new species, but the significance of greater than 97% similarity scores is unclear, and further phenotypic and genomic analysis would be needed for conclusive results (Janda & Abbott 2007).

A novelty ratio, used to determine the likelihood of cultivation a novel organism, was calculated for each method based on the cultivation effort (Table 9). The ratio was calculated by dividing the number of novel isolates (<97% sequence similarity) by the total number of isolates per each method (results in percent).
Table 9. Novelty Ratios. Fungal isolates excluded from calculations.

<table>
<thead>
<tr>
<th></th>
<th>Standard Cultivation</th>
<th>Ipore</th>
<th>Anaerobic</th>
<th>Trap</th>
<th>Diffusion Chamber</th>
<th>Itip</th>
<th>Filter plate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of novel isolates</td>
<td>28</td>
<td>7</td>
<td>8</td>
<td>2</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Ratio (Percent of collection)</td>
<td>9%</td>
<td>7%</td>
<td>6%</td>
<td>2%</td>
<td>1%</td>
<td>1%</td>
<td>1%</td>
</tr>
</tbody>
</table>

Using the novelty ratio as a method of comparison, we found standard cultivation to be the most effective at isolating new species (novelty ratio 9%, Table 9).

Table 9. The ipore and anaerobic cultivation approaches had similar ratios: 7% and 6% respectively (Table 9, Figure 23). All other cultivation methods were less effective, with novelty ratios of 2% or less. We did succeed in isolation of at least one novel organism using every method. Overall we conclude that the ipore may be the best method to access previously uncultured organisms: 7% of the ipore isolates were <97% similar to the LCA, and the novelty ratio was the second largest. Also worthy of noting, the ipore and the diffusion chamber approaches both resulted in the largest number of unique OTUs per method (Table 7, Figure 20).
Chapter 4. Biological Relevance of Culture Isolates: Comparing Culture-independent and Culture-dependent Data
4.1 Cultivation of OTUs detected using Culture-Independent Methods

To our knowledge, this is the first study to generate such a vast culture collection (over 1000 organisms) from a single survey and to synthesize both culture-independent and culture-dependent data to assess the presence and activity of cultivated isolates. For cultivated isolates, the PCR-amplified 16S rDNA gene product was sequenced using Sanger sequencing, resulting in longer sequences (~800 base pairs) than those generated MiSeq for the culture-independent dataset. To allow for comparison of isolate sequences to those generated from MiSeq, we first generated isolate OTUs at 100% sequence similarity, and then each OTU was assigned a representative sequence (the longest sequence in the cluster). Using the representative sequences, another alignment was performed and all OTUs were clustered once again, but at 97% to maintain consistency with protocols used for generating OTUs for culture-independent libraries. We then performed a multiple alignment of OTUs cultivated with those generated from rDNA and rRNA libraries. Resulting analysis allowed us to determine common OTUs detected by all methods (Figure 24).

It should be noted that, while most isolates’ sequences were determined using the 27F primer, a small fraction of isolates were sequenced again using the 1492R primer due to poor sequencing quality. The sequences were manually trimmed and the reverse complement was used for initial isolate OTU (100%) alignments. When performing multiple alignments with rDNA/rRNA data, the reverse complemented sequence would not overlap with the V1-V3 regions. The lack of overlap could result in the appearance of OTUs successfully cultured but missed by sequencing approaches. However, most cultivated organisms matched at least one other strain in our collection at 99-100% sequence similarity, and we used the longest representative sequence for all alignments. Therefore, the reverse complement sequences were
rarely included as they were usually shorter than sequences from the forward direction. Within our entire collection, sequencing using the reverse primer potentially impacted only 13 isolates.

**Figure 24. OTUs common to all methodologies.** N = total number of OTUs detected via each approach. All OTUs detected at least one time were counted to include “rare” taxa. OTUs were clustered at 97% sequence similarity. rDNA and rRNA library sequences were ~300 bp. Cultured isolate sequences were on average 772 bp.

*Note multiple rDNA/rRNA OTUs can align with a single cultured OTU, which can inflate the number of common OTUs. For example, representative sequences from the following culture-independent OTUs (rDNA_OTU_10322, rDNA_OTU_16734, rDNA_OTU_183, and rDNA_OTU_3689) align with OTU 69 from our culture collection. Therefore, four rDNA OTUs = one cultured OTU, inflating OTUs that overlap between rDNA and cultivation by four.*

We cultivated 61 OTUs that were also detected in rDNA and rRNA libraries (Figure 24), indicating isolates from these 61 OTUs were likely present and active at Rich Lake. In total, our 16S rDNA gene survey detected over 17,000 OTUs, but only 157 (0.9%) were cultivated, corresponding with the commonly cited figure that only 1% of bacteria are actually cultivable (Solden et al. 2016). In addition, we cultivated 96 OTUs that were detected in rDNA libraries, but not rRNA libraries, indicating these OTUs may represent dormant organisms in the sediment, despite their cultivability. This is consistent with the knowledge that many cells in the environment are not active contributors to the community, but rather in a dormant state (Stevenson 1978).
We successfully cultivated 5% of all active OTUs (70 of 1395). These results indicate that standard methods, coupled with multiple in situ approaches, may be effective at targeting at least a small fraction of the active, and biologically relevant, members of the community. Future studies should integrate meta-data from cultivated isolates (i.e. date of isolation) with abundance and activity data from individual DNA/RNA libraries to assess if the cultivation of an OTU is directly associated with increases of detectable activity.

4.2 Cultivation of Phantom Taxa

Out of 1395 OTUs deemed active based on 16S rRNA data, 218 were considered “phantom taxa,” – potentially rare but active taxa not detected in any 16S rDNA gene library (Klein et al. 2016). This concept is discussed more thoroughly in Chapter 2, and may be a result of insufficient DNA sampling coverage (Kamke et al. 2010; Klein et al. 2016). We successfully cultivated eight of these phantom taxa, although most were represented by only one or two isolates. Interestingly, one OTU (rRNA_OTU_80) contained 55 isolates related to the Oxalobacteraceae family with 100% sequence similarity. Specifically, many of the 16S rDNA gene sequences of these isolates were 100% similar to members in the genus Janthinobacterium. We cultivated representatives of this OTU throughout the season by every cultivation method. However 58% (38/55) were grown via the diffusion chamber. Eleven isolates were cultured by a variety of other in situ methods, five by standard cultivation, and one by anaerobic enrichment. These results highlight the potential for in situ methods to culture rare, active organisms.

Intriguingly, we cultivated 50 OTUs that were not detected in any DNA or RNA libraries (Figure 24). Eight of these OTUs represented fungal sequences, identified via the ITS region, and therefore would not have been detected from our 16S culture-independent surveys. The remaining 42 OTUs fell into the following orders: Burkholderiales*, Pseudomonadales,
Caulobacterales, Propionibacterales*, Flavobacterales*, Neisseriales, Cytophagales,
Sphingobacterales*, Lactobacillales, Sphingomonadales, Micrococcales*, Enterobacterales,
Bacillales, Rhizobiales, and Rhodospirillales*. Asterisks indicate that the representative
text of that OTUs was most closely related to an “uncultured bacterium” when checked
against the Silva database, and thus may be very interesting for further characterization. Future
studies should include whole genome sequencing of these isolates and further investigation into
the specific culture approaches that most commonly resulted in cultivation of “phantom taxa.”

4.3 Abundance and Activity of Cultured Isolates

In an ideal scenario, we aim to cultivate the majority of microbes in the sediment
community, both those present and active. However, integrated data analysis revealed that the
majority of our isolates clustered with OTUs that were not abundant or active at Rich Lake.
Specifically, out of over 20,000 OTUs, 151 OTUs cultivated were also detected at RL1, and 55
of those were also considered active. At RL2, we cultivated 148 OTUs common to rDNA/rRNA
libraries, and 45 were considered active.

The OTUs in our collection comprised, at most, 3.5% of any rDNA library at RL1
(Figure 25), and most OTUs constituted less than 1% of the community. We did note that the
abundance, while low, remained consistent between samples over time: OTUs that were detected
in one library with “high” abundance (~2%) were detected with similar abundance throughout
the season. At RL2 we also observed that all cultivated OTUs were detected in low, but
consistent, abundance (Figure 26), most comprised less than 4% of any rDNA library. One
unique observation at RL2 was the presence of rDNA_OTU_28 on June 27; this OTU
contributed over 12% of sequences detected in the rDNA library from this sample. Isolates
cultivated from rDNA_OTU_28 were most closely related to the genus Sphingomonas. This
observation corresponds to rRNA data from June 27, where the Sphingomonadales order was first detected as active (Figure 9). Consistency in detectable abundance may lend support the idea that “everything is everywhere,” commonly cited to why geography does not impact microbial distribution (Meyer-Reil 1994; O’Malley 2008).

**Figure 25. Abundance of cultured OTUs at Rich Lake 1.** 151 cultivated OTUs were also detected as present in rDNA libraries at RL1. None of the cultivated OTUs comprised more than 3.5% of any rDNA library.
The abundance of cultured OTUs was low and stable over time, which may indicate cultured organisms are from “rare” taxa. As previously described, “rare” taxa may actually be the most active and relevant to the community function (Campbell et al. 2011; Hugoni et al. 2013; Hunt et al. 2013). In contrast to abundance, by examining the 16S rRNA data of cultivated OTUs, we found that activity varied greatly between samples, and even replicates (Figure 27, Figure 28). Consistent with patterns observed for abundance, OTUs that were active also comprised only a small percentage (3-7%) of rRNA libraries.
Figure 27. Cultured OTUs active at Rich Lake 1. A total of 55 cultivated OTUs were also deemed active at RL1, and OTU contributed, at most, 3% to sequences from rRNA libraries.
We detected two OTUs (rDNA_OTU_786 and rDNA_OTU_962) cultivated during every time frame at RL1. Furthermore, at RL2, rDNA_OTU_786 exhibited detectable activity in all rRNA libraries. Representative sequences from the latter OTUs both match to the genus Pseudomonas. We consistently detected activity from another OTU, rDNA_OTU_126, at RL2. This OTU was related to the order Burkholderiales, which notably was found to be both present and active in all culture-independent samples (rDNA and rRNA libraries). Therefore, we conclude that the order Burkholderiales and the genus Pseudomonas may be important to the Rich Lake community, and these isolates from our collection should be the focus of future studies and whole genome sequencing.

4.4 Final Conclusions
Microbiology is typically studied from a culture-independent or culture-dependent perspective, but rarely are the two approaches combined. In this study we combined these
methods to examine lake sediment and assessed the overlap of information gathered using different approaches. All aims were achieved in this dissertation, and by integrating the data we found that we were able to successfully culture some rare and biologically relevant organisms. Most intriguingly, we cultured “phantom taxa”. However, despite large-scale efforts and extensive use of \textit{in situ} approaches, we did not cultivate the most abundant or active OTUs and no single cultivation approach was sufficient to access these OTUs. Further investigations are needed to validate the reasons and mechanisms behind cultivation approaches that did not succeed in isolating the most active species. When sampling, either for sequencing or for cultivation studies, we may be using sediment that, in fact, contains many differently structured microbial communities, and we propose that this explains the results obtained here.

Small-scale biogeochemical differences across small-scale distances may be the most important and understudied aspect of microbial ecology. Diffusive limitation, especially as a result of microscale features of particulate matter, can result in significant spatial heterogeneity of the environment. This heterogeneity may be critical, as it can determine which organisms are most active and where they are located. When examining a single gram of soil, a typical amount used for sample collection and sequencing, there may be $10^{10-11}$ microorganisms, but the actual “space” occupied can be as little as 1% of the total volume (Pedersen et al. 2015). If one gram of soil was taken, and divided into three parts for DNA, RNA, and cultivation, similar to this experiment, the resulting libraries and collections might differ drastically, depending on where the bulk 1% was located.

The idea of microbial hot spots was first noted in 1987 when a study found that 85% of overall denitrification took place in 0.0015% of a sample soil core (Parkin 1987). This idea of “microbial hot spots” has seen a resurgence in the literature in recent years (Seymour et al. 2000;
Kuzyakov & Blagodatskaya 2015; Long & Azam 2001; Blagodatskaya & Kuzyakov 2013; Pedersen et al. 2015). A recent study has shown that while a microbial hotspot may make up only 0.2-5% of soil volume, that same area may be responsible for most processes measured at a larger scale, such as organic matter degradation and carbon cycling (Kuzyakov & Blagodatskaya 2015). This concept relates to microheterogeneity, and thus provides further explanation of our data. Unless we were culturing organisms directly from a hot spot, we likely cultivated transient or dormant bacteria, explaining why they were not prominent members of the active RNA community.

The results of this dissertation highlight the need for multi-pronged approaches to study microbial ecology including multiple cultivation methods, coupled with culture-independent approaches. We must start using smaller scales (100 um scale, for example, (Stocker 2015)) and sampling biogeochemical parameters coupled with our cultivation efforts in order to determine what contributes most to microbial community structure. Furthermore, as microscale heterogeneity may be a driving force behind microbial community structure, we emphasize the need to develop tools that allow study at a micro-or nano-scale. These tools should be sensitive, *in situ*, minimally invasive to the environment, and economically accessible. As early as 1976 the evidence of small microbial colonies surrounded by different immediate-microenvironments (Hattori et al. 1976) has been known, and yet microscale techniques have not been widely adopted and are still not standard practice within microbiology, despite many nanoscale developments in other areas of research, such as technology and engineering. While developing the tools to study these organisms is critical, the larger challenge may be popularization of microscale techniques and approaches amongst environmental researchers.
References


Kuzyakov, Y. & Blagodatskaya, E., 2015. Microbial hotspots and hot moments in soil: Concept


Pruesse, E. et al., 2007. SILVA: A comprehensive online resource for quality checked and


Schloss, P.D. et al., 2009. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and
Environmental Microbiology, 75(23), pp.7537–7541.
Steven, B. et al., 2013. High Bacterial Diversity of Biological Soil Crusts in Water Tracks over Permafrost in the High Arctic Polar Desert. PLoS ONE, 8(8).


Wilhelm, L. et al., 2014. Rare but active taxa contribute to community dynamics of benthic biofilms in glacier-fed streams. *Environmental Microbiology*, 16(8), pp.2514–2524.


