Microbial Ecology, Biogeochemistry, and Signatures of Life at Low Temperature in Arctic Thermokarst Lake Sediments and High Sierra Snow Fields

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Biochemistry

by
Paula B. Matheus Carnevali

Dr. Alison Murray/Dissertation Advisor

May, 2015
We recommend that the dissertation prepared under our supervision by

PAULA B. MATHEUS CARNEVALI

Entitled

Microbial Ecology, Biogeochemistry, And Signatures Of Life At Low Temperature In Arctic Thermokarst Lake Sediments And High Sierra Snow Fields

be accepted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Alison E. Murray, Ph.D., Advisor
Gary Blomquist, Ph.D., Committee Member
David Shintani, Ph.D., Committee Member
Henry Sun, Ph.D., Committee Member
Kevin Hand, Ph.D., Committee Member
Jerry Qualls, Ph.D., Graduate School Representative
David W. Zeh, Ph. D., Dean, Graduate School

May, 2015
Abstract

The subjects of biological methane (CH$_4$) production and microbial community diversity and structure in extreme cold environments were at the core of this dissertation. CH$_4$ is a potent greenhouse gas that contributes to the warming of the planet and has chemical properties that allow its detection by instruments that can go into space. In this sense, CH$_4$ is also a biosignature, because its presence could be indicative of the presence of life. Mechanisms that lead to the formation of this gas on Earth include reactions mediated by biological enzymes that evolved early in life’s evolutionary time. Therefore, microbial communities composed of different metabolic guilds that act in synergy to decompose organic matter are key elements of understanding biological CH$_4$ production. Two chapters of this dissertation were devoted to studying the potential for biological CH$_4$ production from arctic thermokarst lake sediments. Thermokarst lakes are abundant in the area of continuous permafrost on the Coastal Plain on the North Slope of Alaska, and they are sensitive to climate warming. In Chapter 2 we detected biological CH$_4$ production from sediments of two lakes by studying the isotopic composition of C in CH$_4$ ($\delta^{13}$C$_{CH_4}$) and its concentration in porewaters, and by conducting sediment incubation experiments with no substrate additions. At CH$_4$ concentrations no higher than $\sim$ 2.2 µmoles CH$_4$ g$^{-1}$ dry sediment, an average $\delta^{13}$C$_{CH_4}$ signature $\sim$ 72 $\delta/^{00}$ in the surface sediment intervals of Siqulukaq Lake (Siq) was indicative of biogenic CH$_4$, while a $\delta^{13}$C$_{CH_4}$ signature $\sim$ 51 $\delta/^{00}$ indicated that in Sukok Lake (Suk) there was a mixed signature in the range of thermogenic CH$_4$. We also determined that the highest amount of CH$_4$ produced in the lab at 10 °C ($\sim$ 7 µmoles CH$_4$ g$^{-1}$ dry sediment) occurred in Siq, the lake where we detected four times more organic carbon than in the other two sites (SukS and SukB). These findings were supported
by high amounts in Siq sediments of archaeal lipid biomarkers and by copies of the mcrA gene, which is specific to the pathway of methanogenesis. Furthermore, we determined that the geochemical environment of these lakes was different, with higher concentrations of dissolved iron (Fe) in Siq (51.6 - 855 µM) than in SukB (0.1 - 482 µM) and SukS (33.8 - 141 µM); higher concentrations of sulfate (SO₄²⁻) in SukB (0 - 752 µM) and SukS (0 - 329 µM) than in Siq (0.0 - 23.9 µM); and similar concentrations of nitrate (NO₃⁻) at all sites (0.0 - 213 µM) except for a couple of depths in SukB where NO₃⁻ was much higher (up to 2.3 mM).

These differences in sediment geochemistry led us to investigate what the diversity and structure of the bacterial and archaeal assemblages was in the sediments of these two lakes. In Chapter 3 we sequenced the small subunit (SSU) ribosomal RNA (rRNA) gene and the SSU rRNA molecule of Bacteria and Archaea by Sanger and iTag sequencing (Illumina MiSeq). iTag sequence analyses showed that the microbial assemblage was highly diverse (maximum Simpson's reciprocal index of 148) and that there was a difference in operational taxonomic units (OTUs) predicted for DNA and aRNA sequences between Siq, SukB, and SukS, as reflected by non-metric multidimensional scaling (NMDS). OTUs related to taxa that are capable of organic matter degradation were found among the OTUs comprising ≥ 1% of the aRNA sequences, which were taken as a proxy for members of the community that may have been potentially active in the sediments at the time of sampling. These taxa included members of the phyla Actinobacteria and Bacteroidetes and represented up to 10% of the sequences found in all the sites. There was an important presence (> 5% of the sequences) of sulfate reducing bacteria (SRB) affiliated with the orders Syntrophobacterales and Desulfobacterales at SukB. OTUs at this site were also related to Gammaproteobacteria of the order Pseudomonadales, and to a few groups of unicellular Cyanobacteria. In Siq, OTUs related to the phylum Nitrospirae comprised ~ 5%
of the sequences, and the same group comprised ~2.5% in SukB. Further investigation of the functional potential of the community using phylogenetic investigation of communities by reconstruction of unobserved state (PICRUSt) supported the potential for microbial mediated transformations of nitrogen and sulfur, and for CH_4 cycling. One of the most noteworthy findings was that we observed distinct differences in the distribution of Archaea between the two lakes, in which methanogenic Archaea were mostly found in Siq. Through phylogenetic inference of all archaeal sequences we determined that archaea in these arctic thermokarst lakes are evolutionarily related to other archaea in the Arctic. As an appendix to this work, we also conducted a culturing effort to enrich for methanogenic Archaea adapted to this cold environment. This effort led to enrichment of several methanogens in co-culture with bacteria, augmenting the short list of cold-adapted methanogens in culture.

The last chapter of this dissertation focused on an entirely different kind of cold tolerant, ice loving microbial community – those inhabiting high altitude snow packs. In these habitats, a variety of microorganisms including unicellular algae, bacteria, and fungi can be found under freezing conditions during the seasonal snowpack. Here, we aimed to uncover the diversity and structure of the microbial community in summer snowpacks where conspicuous red algae were present. We also aimed to understand the mechanisms that drive community assembly in this ephemeral habitat. Snow samples were collected from Mt. Conness (CNS) in the eastern Sierra Nevada of California (as our primary study site), and two other geographically distant sites: the Pacific Crest Trail (PCT) and Wheeler Peak (WP). Chemical analyses of inorganic and organic nutrients and Chlorophyll a (Chl_a) were conducted in CNS, and the snow microbiota in samples from all three sites was studied by iTag sequencing of the SSU rRNA gene of the three domains of life. Detailed analysis of
the resampled data set of bacterial and archaean iTag sequences indicated that the bacterial assemblage had a unique structure. In addition to commonly observed ‘singletons’ (OTUs with just 1 sequence) that dominate NGS datasets, we found OTUs that were only present at one site and comprised a high number of sequences (‘singleton OTUs’). Interestingly, these OTUs were mostly affiliated with two different families of Bacteroidetes in the genera Polaromonas and Herminiimonas. Furthermore, even though the bacterial assemblage was up to 80% dissimilar between sites at the OTU level, there was a resemblance in taxonomic composition at the Order level among all three sites, and a few ‘core’ OTUs affiliated with the Phyla Bacteroidetes and Actinobacteria were found in all samples. Variation in structure of the abundant OTUs between sites (and depths) was reflected in the proportion of sequences related to two families of Betaproteobacteria (the Comamonadaceae and the Oxalobacteraceae) and groups of Bacteroidetes. To elucidate possible mechanisms for microbial community assembly and to explain bacterial assemblage structure we tested two hypotheses: (i) that the presence of snow algae determined bacterial assemblage structure, and (ii) that environmental factors affected the structure. A Spearman’s rank correlation between chlorophyll a (Chla; as a proxy to snow algae) and bacterial diversity, indicated that bacterial diversity decreased with Chla concentration. This finding was consistent with three different kinds of Speraman’s rank correlations (positive, logarithmic, and negative) between specific algal OTUs (from a parallel data set) and specific bacterial OTUs, including OTUs that were present in all three sites. A relationship between environmental factors and bacterial OTUs was found for only two of the environmental factors studied (DOC and PN). Further testing using multivariate statistical approaches may allow us to formulate possible explanations to how geographically distant regions share similar microbial community
composition at this high taxonomic level, and determine what mechanisms drive community assembly in these snow packs.

Together, the chapters of this dissertation allowed me to address first order questions concerning life in extremely cold environments that have yet to be well-studied, and that are particularly sensitive to climate warming. Gaining a better understanding of biogeochemical cycling in these ecosystems, may allow us to predict the impact of environmental change on the resident microbial communities, and ultimately how will that affect the cycling of carbon and other major nutrients. Furthermore, these ecosystems are of great interest, because they may be used as analogs to study biosignatures relevant to the search for life in other icy worlds in the Solar System.
Acknowledgements

There are not enough words to express my gratitude to everyone who made this dissertation possible, but I want them to know how grateful I am.

Alison, my Ph.D. advisor, who along this process became a true mentor. I believe she will have a lasting positive impact on my career as a scientist for years to come. From her I learned how science is done with the highest levels of professionalism, and for that I respect and admire her. I thank her for pushing me past my limits, because now I know what I am truly capable of. I thank her for giving me unimaginable and countless opportunities to learn and grow, and most importantly, for having the patience to let me do it. She made me stronger, and for that I will always be grateful.

Kevin Hand, who led the NASA Astrobiology of Icy Worlds Investigation 3 team. I thank him for being a great leader, for creating amazing opportunities for all of us, and for inspiring me to dream even bigger. His hard work has already made my dreams come true!

Megan Rorhssen, a true warrior and collaborator. I thank her for letting me learn from her, for giving me the strength and courage to make things happen, and for working by my side when we were already exhausted. Mostly, I thank her for being my friend.

Dan Berisford, who helped me overcome several of the technical challenges of sampling the arctic lakes, and even taught me how to use a drill. His friendship and generous support made the most difficult things look better and brighter to me.
Andy Klesh and John Leitchy, I thank them for all the support they provided in the field, even at the expense of their own sleep. Their work and dedication is admirable, and it was most inspiring to me!

John Priscu, I thank him for showing me what many years of experience in polar field work looks like, and for giving me a new goal to reach. I also thank him for all of his support, and for always keeping things interesting!

Pamela Santibanez, Alex Michaud, Yonqin Liu, Ralph Lorenz, and everyone whose participation in field work made it possible to do this research. Special thanks to Heather Adams who organized the first few field campaigns, and started to uncover what was interesting about the different arctic lakes we surveyed.

Jeremy Dodsworth and Brian Hedlund, who taught me everything I know about anaerobic work, and provided enormous amounts of support that made the culturing effort possible. I especially thank Jeremy for being a true teacher, he was always there for me, patient and caring, and I will always be grateful for everything I learned from him.

Everyone who was part of the Murray Lab during my tenure as a Ph.D. student. I thank them for the technical help, moral support, and their friendship: Vivian Peng, Ema Kuhn, Gary Trubl, Olivia Rassuchine, Emily Ulrich, Michael Blayberg, Robert Read, Charlotte Tyler, Protima Wagh, and the newest additions to the lab, Mary Highman and Samantha Mayer.

Chris Fritsen and his group, especially Jeramie Memmott, for all the help and support he provided. I thank Chris for giving me great opportunities to learn!

DRI staff, who were always helpful: Barbara Jackson, Cindy Littlefield, Petra Bartella, Russ Bergin, Wendy Long, Ryan Coots, Maria Vasquez, Melissa McCoy, Martha McRae, Denisse
Hamond, Paul Neely, John Karlas, Glenn Wilson, Mitch Goldfin, Lisa Wable, John Ford, James Rios, Melanie Scott, Wayne, Mike, Jeff...and the list goes on!

Other faculty and researchers at DRI who were there when I needed them most: John Mejia, Mary Miller, Hans Moosemuller, Duane Moser, Sophie Baker, and Bill Coulomb. Also, Jay Arnone and Alan Gertler, for providing other opportunities to develop my career as a scientist.

Grad students at DRI who were friends and companions on this journey: Kathy Bywaters, KC King, Megan Johnson, Eric Wirthlin, and Clint Davis.

My committee members, Gary Blomquist, Jerry Qualls, David Shintani, Henry Sun, and Kevin Hand, who were very supportive of me, and gave me encouragement and great advice along the way.

Dr. Glenn Miller, who was most helpful with GC work, and was always there when I needed help.

Everyone at BASC and UMIAQ, for keeping us safe in the field, and for providing all the logistics support and good stories.

DEES directors over the years, who supported Alison and made it possible for me to be at DRI all these years. Thanks to DRI for its financial support, and for being an institution that trains students under the highest of expectations, so that they leave prepared for the challenges of being a scientist.

The NASA Astrobiology Institute, which provided the funding for this work through the Jet Propulsion Lab, and all the other members of the Astrobiology of Icy Worlds Projects.
My friends and family, who have supported me unconditionally over the years, and always give me a reason to want to reach for the stars.

Leo, my best friend, husband, and partner, the only person who knows all the sacrifices we have made to reach this goal, and whose support has made reaching this point possible. Simply put, I could not have done this without him; he kept me going when I felt like giving up, he was the ever present light that guided my journey, and his unconditional love and support never cease to amaze me. I thank him for being in my life all of these years.
Table of Contents

Chapter 1 .......................................................................................................................................................... 1
General Introduction ........................................................................................................................................ 1

1. How to search for life? Methane as a biosignature ........................................................... 5
2. Methanogenesis ................................................................................................................................. 8
3. How is biogenic CH₄ distinguished from abiogenic CH₄? .............................................. 13
4. Methane as a greenhouse gas ................................................................................................. 15
5. Thermokarst lakes ..................................................................................................................... 18
6. Microbial life at the limits: examples from two cold environments ............................ 23
7. Overall objectives ...................................................................................................................... 29
References ............................................................................................................................................... 31

Chapter 2 .................................................................................................................................................... 37
Methane sources in arctic thermokarst lake sediments on the North Slope of Alaska ........ 37

Abstract ..................................................................................................................................................... 38

1. Introduction .......................................................................................................................................... 39
2. Methods .............................................................................................................................................. 41
2.1. Sampling sites .............................................................................................................................. 41
2.2. Sediment geochemistry ............................................................................................................. 44
2.2.1. Oxygen (O₂) microelectrode profiles ............................................................................. 44
2.2.2. Pore water chemistry ........................................................................................................ 46
2.3. Sediment CH₄ ............................................................................................................................ 48
2.3.1. CH₄ concentrations and stable carbon isotope analyses .............................................. 48
2.3.2. CH₄ production experiment ........................................................................................... 49
2.4. Quantitative PCR (qPCR) of the methyl-coenzyme reductase alpha subunit gene (mcrA) .......................................................................................................................... 51
2.5. Organic carbon and lipid biomarkers analyses ........................................................................ 52
2.5.1. Sediment texture, total carbon, and inorganic carbon content .................................. 52
2.5.2. Lipid biomarker analyses ............................................................................................... 53
3. Results ................................................................................................................................................... 54
Abstract
Archaea in sediments of Alaskan Arctic thermokarst lakes
Diversity, structure, functional potential of the microbial community, and phylogeny of

Chapter 3
Supplementary References
Supplementary Tables and Figures
References
Acknowledgements

1. Introduction
2. Methods
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.4.</td>
<td>Nucleic acids extraction and next generation sequencing</td>
<td>161</td>
</tr>
<tr>
<td>2.5.</td>
<td>iTag sequences analyses and statistics</td>
<td>162</td>
</tr>
<tr>
<td>3.</td>
<td>Results</td>
<td>163</td>
</tr>
<tr>
<td>3.1.</td>
<td>Cell counts</td>
<td>163</td>
</tr>
<tr>
<td>3.2.</td>
<td>Chemical analyses</td>
<td>164</td>
</tr>
<tr>
<td>3.3.</td>
<td>Microbial community structure and diversity based on iTag sequences</td>
<td>167</td>
</tr>
<tr>
<td>3.4.</td>
<td>Factors that drive community assembly</td>
<td>175</td>
</tr>
<tr>
<td>4.</td>
<td>Discussion</td>
<td>178</td>
</tr>
<tr>
<td>4.1.</td>
<td>Cell counts from Mt. Conness</td>
<td>178</td>
</tr>
<tr>
<td>4.2.</td>
<td>Snowpack dynamics and abiotic parameters related to the microbial community</td>
<td>180</td>
</tr>
<tr>
<td>4.3.</td>
<td>Diversity of the microbial community in comparison to other places</td>
<td>182</td>
</tr>
<tr>
<td>4.4.</td>
<td>Theory of community assemblage and sources of microorganisms</td>
<td>183</td>
</tr>
<tr>
<td>4.5.</td>
<td>Future directions and closing remarks</td>
<td>186</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td></td>
<td>188</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>189</td>
</tr>
<tr>
<td>Supplementary Tables and Figures</td>
<td></td>
<td>192</td>
</tr>
<tr>
<td>Supplementary Information</td>
<td></td>
<td>200</td>
</tr>
<tr>
<td>S1.</td>
<td>Supplementary Methods</td>
<td>200</td>
</tr>
<tr>
<td>S1.1</td>
<td>Amplification of the SSU rRNA gene of Bacteria, Archaea and Eukarya</td>
<td>200</td>
</tr>
<tr>
<td>S1.2</td>
<td>Diversity profiling of CNS samples using DGGE</td>
<td>201</td>
</tr>
<tr>
<td>S2.</td>
<td>Supplementary Results</td>
<td>202</td>
</tr>
<tr>
<td>Supplementary References</td>
<td></td>
<td>206</td>
</tr>
<tr>
<td>Chapter 5</td>
<td></td>
<td>207</td>
</tr>
<tr>
<td>General Conclusions</td>
<td></td>
<td>207</td>
</tr>
<tr>
<td>1.</td>
<td>Positive and negative feedback loops as a property of self-regulating systems</td>
<td>208</td>
</tr>
<tr>
<td>2.</td>
<td>Microbial communities in cold environments and their global distribution</td>
<td>211</td>
</tr>
<tr>
<td>2.1.</td>
<td>Microbial communities in thermokarst lake sediments</td>
<td>211</td>
</tr>
<tr>
<td>2.2.</td>
<td>Microbial communities in snow samples</td>
<td>214</td>
</tr>
<tr>
<td>3.</td>
<td>Astrobiological implications of the study of life in cold environments on Earth</td>
<td>216</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>217</td>
</tr>
<tr>
<td>Appendix</td>
<td></td>
<td>218</td>
</tr>
</tbody>
</table>
List of Tables

Table 1.1  Some cold-adapted methanogens in culture ........................................24

Table 2.1  Sampling sites, dates, and analyses performed in each core .........................45

Table S2.1  CH₄ concentration and CH₄ production rates reported in the literature for arctic permafrost samples ..................................................................................................................85

Table S2.2  Free hydrocarbon biomarker extracted from Siqlukaq and Sukok sediments ... 86

Table S2.3  Lipid biomarker ratios from catalytic hydropyrolysis hydrocarbon products ... 87

Table 3.1  iTag sequence statistics ..................................................................................107

Table 3.2  Alpha diversity of samples subsampled to 28,000 sequences per group ........109

Table S3.1  Tukey’s test comparing the log transformed abundances of phyla ..............143

Table S4.1  Alpha diversity statistics of the snow samples .............................................192

Table A1  Summary of relevant enrichment cultures from arctic lake sediments ...........219
List of Figures

Figure 1. 1 Two of the main pathways of methanogenesis and their common steps........ 12
Figure 1. 2 Satellite image of the western portion of the Coastal Plain in the North Slope
of Alaska .................................................................................................................... 18
Figure 1. 3 Model of a thermokarst lake with open water in the summer and frozen
water in the early winter......................................................................................... 22

Figure 2. 1 Landsat 7 image of the Arctic Coastal Plane near Barrow, AK......................... 43
Figure 2. 2 Oxygen microprofiles measured at 100 µm resolution from Siqlukaq, Sukok B,
and Sukok S .................................................................................................................. 55
Figure 2. 3 Pore water chemistry profiles from replicate sediment cores ....................... 58
Figure 2. 4 Methane concentration, δ¹³C₅H₄ and δ¹³C²O² in arctic thermokarst lakes...... 60
Figure 2. 5 CH₄ production from sediment incubations at 2 temperatures using in situ
organic matter............................................................................................................. 62
Figure 2. 6 October 2011 sediment depth profiles of (A) total organic carbon, (B)
archaeol concentration, and (C) archaeol to n-C₃₁ ratios ...................................... 66
Figure S2.1 Methane concentration and δ¹³C₅H₄ from all the cores collected for these
analyses ....................................................................................................................... 88

Figure 3.1a NMDS analysis of DNA iTag sequences......................................................... 110
Figure 3.1b NMDS analysis of aRNA iTag sequences...................................................... 111
Figure 3.2 Relative abundance of bootstrap resampled OTUs observed with ≥ 1%
relative abundance in either the aRNA or the RNA datasets............................... 113
Figure 3.3 Archaeal OTUs that were observed with > 95% confidence over all the
resamplings ............................................................................................................... 117
Figure 3.4a Maximum likelihood phylogenetic tree of SSU rRNA gene Sanger sequences
related to Euryarchaeota.......................................................................................... 118-119
Figure 3.4b Maximum likelihood phylogenetic tree of SSU rRNA gene Sanger sequences
related to other Archaea phyla.............................................................................. 119
Figure S4.1 Abundance distribution of the 'singleton OTUs' comprising < 1 % of the sequences .................................................................193
Figure S4.2 Relative abundance of the 'core' OTUs that comprised ≥ 1 % of the sequences and that were present in all the samples .................................................................194
Figure S4.3 Intra-sample rarefaction curves for the observed number of OTUs ..........195
Figure S4.4 Spearman's rank correlation between chlorophyll a and the inverse Simpson's diversity index ...............................................................................................................196
Figure S4.5 Spearman's rank correlations between algal OTUs and bacterial OTUs .....197
Figure S4.6 Snow sampling sites .................................................................................198
Figure S4.7 Wheeler Peak (WP) chlorophyll a concentration in the vertical profile ......199
Figure S4.8a DGGE of PCR-amplified fragments using general primers for Bacteria .....202
Figure S4.8b Dendrogram of phylotypes observed by DGGE analysis of the SSU rRNA gene of Bacteria .................................................................................................................203
Figure S4.9a DGGE of PCR-amplified fragments using general primers for Eukarya ...204
Figure S4.9b Dendrogram of phylotypes observed by DGGE analysis of the SSU rRNA gene of Eukarya .................................................................................................................205
Chapter 1

General Introduction
There are two underlying themes of my dissertation research: signatures of life in cold environments that could potentially be used to search for life elsewhere, and the importance of microbial communities involved in carbon cycling as it influences the greenhouse gas inventory in our planet. These topics are related to the goals described in the Astrobiology Roadmap, a document created by the NASA Astrobiology Institute to guide the design and execution of research in the context of Astrobiology, which also provides a framework for my dissertation.

Astrobiology seeks to answer some of the most fundamental questions that humankind has yet to solve, what is the origin of life on Earth? and what is the extent of life in the universe? To answer these questions we must first establish what forms of life exist on Earth and what the limits are for life as we know it. Life is defined as a sum of characteristic features that include the presence of a membrane, metabolic activity, and most importantly, the ability to reproduce. Studying how these features have co-evolved with our planet may be the key to understanding what we need to look for in other planets, and it is among the goals outlined by the Astrobiology Roadmap (Des Marais et al., 2008).

My dissertation was consistent with goals 4, 6, and 7 as described in the Astrobiology roadmap. Goal 6 outlines the need to understand how microbial metabolic activities will respond to environmental change in light of increasing evidence of human-induced change of geochemical cycles and their associated microbial communities (Des Marais et al., 2008). Chapter 2 characterizes the geochemical environment of arctic thermokarst lake sediments that are known to
produce methane (CH$_4$), a potent greenhouse gas critical in carbon (C) cycling. Rates of biological CH$_4$ formation (methanogenesis) in lake sediments were obtained, which could be used in current estimates of CH$_4$ emissions from the Arctic. The effect of a temperature increase on methanogenesis was also addressed (relevant to goal 4 in the Astrobiology Roadmap too), and a baseline for prediction of scenarios of future change was provided.

Goal 4 in the Astrobiology Roadmap recommends to perform comparative genomic studies of microbial diversity, to understand the metabolic requirements of these microbes; to study how microorganisms affect the cycling of essential nutrients (e.g., C, N, S, O), and what sort of record is left by their metabolic activities on Earth (Des Marais et al., 2008). Chapter 3 addressed the issues of microbial diversity in thermokarst lake sediments, the metabolic potential of the microbial community inhabiting these sediments, and the likely involvement of these microbes in geochemical cycling.

Goal 7 in the Astrobiology Roadmap is to identify biosignatures that are indicative of past or present life on Earth, and that can be used to detect life in extraterrestrial samples or that could be measured remotely (Des Marais et al., 2008). Biosignatures are defined as features that can emerge from life, and include atmospheric gases and molecular structures that can be found on a planet’s surface (Pilcher, 2003; Seager et al., 2012). Overall, my dissertation work looked at CH$_4$ as a biosignature that results from a chain of microbial metabolic activities codependent on each other, and at photosynthetic pigments that are found in microbial life forms
and that are easily detected by remote-sensing technologies. Additionally, I believe that a molecular marker used in this work to investigate microbial community diversity and structure \textit{i.e.}, the SSU rRNA gene could be considered another kind of biosignature.

It was this goal of identifying biosignatures that incentivized the overall project of which my dissertation was only a small part of. Our investigation within the Astrobiology of Icy Worlds project, led by the NASA Jet Propulsion Lab (JPL, Pasadena, CA), aimed to study the detectability of spectral signatures beyond photosynthetic pigments that are of potential relevance to the search for life in icy worlds of the Solar System. Studying CH$_4$ emissions from arctic thermokarst lakes in Alaska, allowed our collaborators at JPL to work on the issue of detectability, and allowed us to determine what the potential for biological CH$_4$ production from these lakes was (Chapter 2 & 3). Chapter 4 was motivated by the same theme of detectability, in this case of spectral signatures in summer snow fields, potentially laden with algal biomass that can be remotely sensed.

In the three main chapters of the dissertation, I addressed first order questions concerning life and associated processes in habitats that are sensitive to climate warming and relevant analogs to other Icy Worlds in our Solar System. This work contributes to the fundamental understanding of microbial life in high latitude and high altitude environments where freezing temperatures are prevalent and life experiences environmental extremes.
The next few sections in this general introduction provide background information concerning: (1) why CH$_4$ is of so much interest to us as a signature of life, (2) describe how CH$_4$ is produced by microorganisms, (3) what other sources of CH$_4$ are there (or its pitfalls as a biosignature), (4) how important CH$_4$ is in the context of the greenhouse gas inventory, (5) arctic lakes as an example of ecosystem that are major sources of CH$_4$ on Earth, and (6) what kind of life thrives in this and other kinds of cold environments. This chapter ends with the main goals of my dissertation research.

1. **How to search for life? Methane as a biosignature**

The transmission spectra of certain atmospheric gases are potentially detectable using remote-sensing technologies, a desirable feature of biosignatures. What determines the detectability of atmospheric gases is a combination of factors that affect their absolute concentration in the atmosphere, including the accumulation time, global mixing with other gases, and their atmospheric lifetime, which is partially affected by radiation from the parent star (Seager, 2013). Mixing processes require time from the initial step in the chain of metabolic reactions that lead to gas formation, to allow for extensive accumulation in the atmosphere, this last step being limited by the rate of destruction under atmospheric conditions. Extensive atmospheric accumulation is essential for detectability because biosignature gases could be locally produced and therefore undetectable in a given atmospheric mix.
Oxygen (O\textsubscript{2}) is currently considered the dominant biosignature on Earth, under the assumptions that biosignature gases result from specific steps along metabolic pathways (as the byproduct of a set of redox reactions), and that they accumulate at a rate faster than their destruction (McKay, 2011; Seager et al., 2012). Oxygen makes up 21% of the atmosphere and it is the byproduct of photosynthesis, a ubiquitous metabolic process that captures the energy of sunlight and enables the formation of cellular components. Our current understanding is that O\textsubscript{2} started to accumulate in the atmosphere after the Great Oxidation Event ~2.4 billion years ago (Ga), when oxygenic photosynthesis became a widespread process among photosynthetic bacteria. Enough time passed to allow for atmospheric accumulation, which also meant a change in the redox state of the planet from reducing to oxidizing. However, under certain planetary conditions, for instance, planets with high atmospheric concentrations of water vapor, O\textsubscript{2} levels could increase due to photodissociation instead of photosynthesis (Seager, 2014).

Therefore, one important aspect of the use of atmospheric gases to infer the presence of life is the understanding of their environmental or planetary context (Segura et al., 2007).

Similarly, the presence of CH\textsubscript{4} in an atmosphere must be interpreted in its planetary context, because some gases that are byproducts from metabolic reactions could also have alternative sources (Seager, 2014). For instance, even though ~95% of atmospheric CH\textsubscript{4} on Earth is microbially produced, there are other sources of CH\textsubscript{4} that do not involve biology. CH\textsubscript{4} can be formed by break down of buried organic
matter at high temperature and pressure (thermogenesis), or as a result of serpentinization reactions in hydrothermal vents (e.g. Lost City Hydrothermal Field in the mid-Atlantic ocean). During serpentinization, ferrous iron-rich rocks are oxidized by hot water, thus reducing the water, and producing hydrogen (H₂), which in the presence of CO₂ would form CH₄ independently from biology (Schrenk et al., 2013). The trace amount of CH₄ recently detected by the Curiosity rover on Mars would be another good example of alternative sources of CH₄. CH₄ there seems to be generated by a combination of mechanisms that may include photochemical degradation of exogenous organics, release from clathrates, or perhaps biology (Webster et al., 2015; Atreya et al., 2007).

Furthermore, as with other biosignature gases, CH₄ is not easily detected remotely using current spectroscopy-based technologies. At the current mixing ratio in Earth's atmosphere the absorption band of CH₄ at 7.6 µm would be overshadowed by the strong water absorption at 6 µm, making remote detection very difficult (Pilcher, 2003). CH₄ in Mars was measured in situ using tunable laser absorption in a closed sample cell (Mahaffy et al., 2012) and was detected at the 3.3-µm band in the infrared (IR) spectrum. In places like Europa, an icy moon of Jupiter, the search for CH₄ may be even harder, because CH₄ would have to be detected in ice using the 2.27-2.46 µm bands in the near-IR spectrum, where CH₄ shows weaker features (Hand et al., 2010).

Finally, the search for biosignature gases in the solar system and in remote exoplanets is constrained by the parameters that define a “habitable zone” around
the parent star(s). On Earth, when we think of life, we think of water, and it is precisely water that determines the potential for life in a given planet, because liquid water only occurs in a certain temperature range and a given pressure (McKay, 2011). Without water, biology as we know it would not exist, O\textsubscript{2} would not be produced by photosynthesis, and perhaps CH\textsubscript{4} would still be the dominant gas in the atmosphere. That is why an atmospheric biosignature gas would be in complete chemical disequilibrium with the atmosphere, \textit{i.e.}, it would be several orders of magnitude higher than if generated by photochemical disequilibrium (Seager, 2014), even though that also means that it could be a false positive. This challenge has lead Segura (2005) to propose the use of both O\textsubscript{2} and CH\textsubscript{4} together as biosignatures, with the caveat that these atmospheric gases could only be found at detectable concentrations in UV radiation environments lower than Earth’s radiation environment (Seager, 2014).

2. **Methanogenesis**

Before Earth’s atmosphere became oxidized, it was full of CH\textsubscript{4}. Methanogenesis is the process in which microorganisms of the domain Archaea obtain minimal amounts of energy for growth (as low as - 3.2 KJ/reaction under specific environmental conditions, see below) and release CH\textsubscript{4} as a result, and it is one of the most ancient forms of microbial metabolism. Methanogenic Archaea are among the oldest microorganisms in the tree of life, and methanogenesis seems to have originated between 4.11 Ga (3.31-4.49 Ga) and 3.78 Ga (3.05-4.16 Ga) (Battistuzzi et
That means that Archaea may have arisen between the Hadean and the Archaean, during the time of the “Faint Young Sun Paradox”, when the sun was 20-25% less bright than today but the Earth was kept warm by greenhouse gases (e.g., CO$_2$ and CH$_4$) that may have been formed by microorganisms like methanogenic Archaea (Kasting & Ono, 2006).

In modern times, to obtain substrates for methanogenesis, organic matter must be first broken down by primary fermenters, including bacteria and fungi that are able to decompose complex organic molecules such as cellulose and lignin. The products of this step are short-chain fatty acids (e.g., acetate and formate), some H$_2$ and some CO$_2$, which can be utilized by methanogens and other kinds of bacteria in anaerobic reactions. The remaining partially decomposed organic matter with three or more carbon atoms (e.g., butyrate and propionate) is then taken by secondary fermenters that make more acetate, H$_2$ and CO$_2$. However, this step is only possible at low concentrations of H$_2$ (< 1 μM) and formate (< 100 μM) that would enable regeneration of the NAD$^+$ consumed during acetate formation by fermentation. If the bacteria carrying out this fermentation require the presence of methanogens to keep the levels of H$_2$ and formate low, the relationship between these microorganisms is considered to be a form of obligate syntrophy, otherwise it would be considered a form of non-obligate interspecies electron transfer (Whitman et al., 2006).

An example of obligate syntrophy between a bacterium and an archaeaon that may be relevant in the context of this dissertation research follows.
Syntrophomonas sp., a fatty acid-oxidizing bacterium oxidizes butyrate to acetate and H₂ under thermodynamically unfavorable conditions (ΔG⁰' = + 48.2 KJ/reaction), which means that it cannot grow on its own; but in the presence of hydrogenotrophic methanogens that use this H₂ (or formate) to reduce CO₂ (ΔG⁰' = -130.7 KJ/reaction), the combined reactions will be thermodynamically favorable (ΔG⁰' = -82.5 KJ) (Madigan et al., 2014) and inhibition of fatty acid oxidation by end-product formation can be prevented.

Another form of methanogenesis, acetoclastic methanogenesis, is a disproportionation reaction in which the methyl C of acetate acts as the electron acceptor in the oxidation of the carboxyl C to CO₂, turning it into CH₄. Similarly, a variety a C₁ compounds such as methanol, trimethylamine, dimethylsulfide, and some alcohols can be disproportionated to form CH₄. The standard change in free energy of some of these reactions is enough to make 1 ATP (G⁰ = -31.8 KJ/mol) or less per mole of CH₄, especially under environmental conditions which tend to include lower concentration of substrates. For instance, the free energy change under standard conditions (ΔG; 1 M solutes, 1 atm gases, 25 °C) for the reduction of CO₂ with H₂ to CH₄ is -136 KJ/reaction, and -31 KJ/reaction for the conversion from acetate; but in a typical anoxic freshwater environment the concentrations of substrates are much lower (1 mM fatty acids, 20 mM HCO₃⁻, 10 μM glucose, 10⁻⁴ atm H₂) and the free energy change for hydrogenotrophic methanogenesis is -3.2 KJ/reaction and -24.7 KJ/reaction for acetoclastic methanogenesis (Zinder, 1984).
The different pathways for methanogenesis involve several coenzymes and cofactors (see Ferry & Kastead, 2007) for a complete review of the biochemistry, but two of those coenzymes are common to the different pathways: coenzyme M and coenzyme B (Fig. 1.1). Coenzyme M (HS-CoM) accepts the methyl group from acetate in acetoclastic methanogenesis (CH$_3$-S-CoM) and acts as the substrate for a key enzyme in all methanogenic pathways, the methyl coenzyme M reductase (Mcr), which catalyzes the reductive demethylation of CH$_3$-S-CoM to CH$_4$ and the subsequent reduction of coenzyme B (HS-CoB). The resulting heterodisulfide CoMS-SCoB is then reduced by a heterodisulfide reductase to the active sulfhydryl forms of the coenzymes, generating an electrochemical proton gradient that involves an A$_1$-A$_0$-type ATP synthase and driving ATP synthesis. However, the electrons participating in this reduction originate from the carbonyl group of acetate in the acetoclastic pathway, or from the oxidation of H$_2$ or formate in the hydrogenotrophic pathway (Ferry, 2010). It is important to note that there are two mechanisms by which methanogens conserve energy: a sodium motive force associated to hydrogenotrophic methanogenesis, during transfer of the methyl group from methanopterin (one of the coenzymes that functions as a C$_1$ carrier) to CoM; and a proton motive force during both acetoclastic and hydrogenotrophic methanogenesis, linked to the reaction catalyzed by the heterodisulfide reductase (Ferry, 2007; Madigan, 2014).
Figure 1.1 Two of the main pathways of methanogenesis and their common steps. In the CO$_2$-reduction pathway reduced ferredoxin (Ferredoxin-red) donates electrons to CO$_2$ in a reaction catalyzed by an enzyme containing methanofuran, the first C$_1$ carrier (^) in this pathway. In the next step, a tetrahydromethanopterin (methanopterin^) coenzyme receives the formyl group from methanofuran, and the formyl group undergoes dehydration to methenyl, then reduction to methylene coupled to F$_{420}$ (^electron donor) and to methyl, with electrons from H$_2$. Coenzyme M (CoM-SH^) methyltransferase relocates the methyl group from methanopterin to CoM, the first common step between hydrogenotrophic and acetoclastic methanogenesis. The last step, leading to CH$_4$ formation involves the key enzyme methylreductase, which is a highly complex system (containing subunit α coded by the mcrA gene) that interacts with coenzyme B (HS-CoB^) and cofactor F$_{430}$. The resultant disulfide complex (CoM-S-S-CoB) is regenerated with electrons from H$_2$, the same way oxidized ferredoxin (Ferredoxin-ox) is regenerated. The first two steps during acetoclastic methanogenesis convert acetate to acetyl-CoA. Then, the carbon monoxide dehydrogenase/acetyl-CoA synthase complex catalyze the cleavage of the C-C and C-S bonds, the transfer of the methyl group to methanopterin^, and the oxidation of the carbonyl group to CO$_2$ while reducing ferredoxin-ox. This last set of reactions provides the electrons to reduce CoM-S-CH$_3$ to CH$_4$ (Whitman et al., 2006; Ferry & Kastead, 2007; Ferry, 2010).
Methanogenic Archaea can be outcompeted by other microorganisms like sulfate-reducing bacteria or iron-reducing bacteria, which in comparison carry out more energy efficient reactions. The main issue is that many of these microorganisms can use the same electron donors (by products of organic matter decomposition) for their respective metabolic pathways. For example, Group II marine sulfate-reducing bacteria (complete oxidizers that are able to oxidize acetate and other long chain fatty) can obtain energy from acetate oxidation (except Desulfo bacter sp.): CH$_3$COO$^- +$ SO$_4^{2-} + 3H^+ \rightarrow 2$ CO$_2 +$ H$_2$S + 2H$_2$O ($\Delta G^\circ = -57.5$ KJ/reaction) (Madigan et al., 2014). Another group of organisms that can use acetate as an electron donor belong to the family Geobacteraceae in the Deltaproteobacteria. Members of the genera Geobacter, Desulfuromonas, and Desulfuromusa can oxidize acetate during dissimilative ferric iron reduction (Madigan et al., 2014).

3. **How is biogenic CH$_4$ distinguished from abiogenic CH$_4$?**

CH$_4$ can also be formed by alternative mechanisms that are abiogenic. Biology, more specifically, enzymes, discriminate against substrates with heavier isotopic compositions in their atoms, a process known as isotopic fractionation. In the case of the substrates for methanogenesis, compounds with the lighter carbon isotope ($^{12}$C) would be strongly preferred by microbial enzymes over compounds with the heavier carbon isotope ($^{13}$C), resulting in lighter CH$_4$ relative to the parent substrates. We can use $\delta^{13}$C, that is the $^{13}$C to $^{12}$C ratio compared to the Pee Dee
belemnite standard (a 65- to 150-million year old limestone formation and the geologic standard for stable carbon isotope analysis) to determine the origin of the CH$_4$. $^{13}$C-depleted CH$_4$ is indicative of biological CH$_4$ production, while $^{13}$C-enriched CH$_4$ is associated to thermogenic CH$_4$ formation (Whiticar, 1999).

However, to confirm the origin of CH$_4$ it is useful to look at the stable isotopes of both C and H atoms in CH$_4$. The same way the carbon in CH$_4$ originates from the substrates for methanogenesis, some of the hydrogen atoms in CH$_4$ are derived from the water and substrates available to methanogens (Whiticar, 1999; Boone, 2000), and there is a preference for the lighter isotope ($^1$H) over deuterium ($^2$H). The $\delta^{13}$C$_{CH4}$ can be plotted against the $^2$H to $^1$H ratio ($\delta$D) relative to the standard mean ocean water (SMOW standard) in a CD-diagram to identify the source of CH$_4$ (Whiticar, 1999). Furthermore, in combination with molecular or compositional data on short-chain hydrocarbons ($C_2$-$C_4$), the origin of CH$_4$ can be determined more accurately. This is especially useful when the isotopic signature of CH$_4$ is borderline thermogenic, because there is overlap with the $\delta$D$_{CH4}$ of certain kinds of biogenic CH$_4$ (Whiticar, 1999). Noticeably, the following factors affect the isotopic composition of CH$_4$: the isotopic fractionation in the parent organic matter, which is ultimately determined by the isotopic composition of CO$_2$; kinetic isotope effects that determine the isotopic fractionation of CH$_4$ relative to the parent substrates, and in the case of atmospheric CH$_4$ sources, it is also important to consider the effect of CH$_4$ oxidation on isotopic fractionation (Whiticar, 1999; Whiticar, 2000).
Additional information can be gleaned from the $\delta_{CH_4}$ and the source of hydrogen, because the aceticlastic pathway uses three hydrogens from the methyl groups and one from water, while the hydrogenotrophic pathway extracts all the hydrogen from water or protons (Whiticar, 1999). Moreover, one can use the isotopic fractionation of C in CO$_2$ to determine the isotope separation factor ($\epsilon_C$) between $\delta^{13}C_{CO_2}$ and $\delta^{13}C_{CH_4}$, which is relative to the pathway of CH$_4$ formation (Whiticar, 1999).

4. **Methane as a greenhouse gas**

There is an equally important reason to look at CH$_4$ in Earth’s atmosphere besides its astrobiological relevance. CH$_4$ is the third main contributor to the greenhouse effect, but it is $\sim 30$ times more potent as a greenhouse gas compared to CO$_2$ over a 100-year time horizon (Lelieveld, 1998; IPCC, 2001). Moreover, CH$_4$ is rising at a rate of 0.4 % yr$^{-1}$ (Kirschke et al., 2013) although the Intergovernmental Panel for Climate Change (IPCC) estimated a total source of 598 Tg CH$_4$ yr$^{-1}$ in 2001, and later adjusted that value to 582 Tg CH$_4$ yr$^{-1}$ in 2007 (IPCC, 2007) due to changes in CH$_4$ sinks over time. These sources of CH$_4$ are classified as natural (wetlands, termites, ocean, hydrates, geological sources, wild animals and wild fires) or anthropogenic sources (energy, coal mining, gas and oil industry, landfills and waste, ruminants, rice agriculture, biomass burning, and vegetation). Thus, both kinds of CH$_4$ sources can be of biogenic (microbial) or non-biogenic origin. In particular, natural gases
can be formed by microorganisms, thermogenic processes, hydrothermal and geothermal activity, and in the mantle (Whiticar, 2000).

A distinction must be made between current biological CH\textsubscript{4} formation and biological CH\textsubscript{4} formation that happened in the past. The \textsuperscript{14}C content of CH\textsubscript{4} can be used to determine the age of the C source, because only modern organic carbon that is < 35,000 yrs old contains \textsuperscript{14}C (radiocarbon). Furthermore, the age of the C source is a useful proxy to detect thermogenic CH\textsubscript{4}, which is formed from \textsuperscript{14}C-depleted organic carbon (Kvenvolden & Rogers, 2005). Based on the isotopic composition of C in the atmosphere (reviewed by Stevens & Wahlen, 2000; Whiticar, 2000) biogenic or young CH\textsubscript{4} contributes the most to the CH\textsubscript{4} budget (Kvenvolden & Rogers, 2005). However, a high degree of uncertainty in the total CH\textsubscript{4} budget is reflected in the most current estimates for global CH\textsubscript{4} emissions that attribute > 70 % of the total CH\textsubscript{4} sources to modern biochemical sources including: wildfires (~ 3 Tg CH\textsubscript{4} yr\textsuperscript{-1}), oceans (~ 9 Tg CH\textsubscript{4} yr\textsuperscript{-1}), wild animals (~ 15 Tg CH\textsubscript{4} yr\textsuperscript{-1}), termites (~ 22 Tg CH\textsubscript{4} yr\textsuperscript{-1}), and wetlands (~ 175 Tg CH\textsubscript{4} yr\textsuperscript{-1}); and place the geologic sources at ~ 30 % of the total CH\textsubscript{4} sources (~ 175 Tg CH\textsubscript{4} yr\textsuperscript{-1}) (IPCC, 2007; Etiope, 2008). According to these estimates, geologic CH\textsubscript{4} sources comprise ~ 53 Tg CH\textsubscript{4} yr\textsuperscript{-1} and include emissions from volcanoes (< 1 Tg CH\textsubscript{4} yr\textsuperscript{-1}), geothermal sources (2.5-6.3 Tg CH\textsubscript{4} yr\textsuperscript{-1}), mud volcanoes (6-9 or 10-20 Tg CH\textsubscript{4} yr\textsuperscript{-1}), seeps (3-4 Tg CH\textsubscript{4} yr\textsuperscript{-1}), microseepage (10-25 Tg CH\textsubscript{4} yr\textsuperscript{-1}), marine seepage (~ 20 Tg CH\textsubscript{4} yr\textsuperscript{-1}), and serpentinization (unknown Tg CH\textsubscript{4} yr\textsuperscript{-1}). However, process-based modeling (bottom-up estimates) place geologic CH\textsubscript{4} estimates at 60 Tg CH\textsubscript{4} yr\textsuperscript{-1}; and inverse
modeling based on concentration, isotopic atmospheric composition, aircraft and satellite, spatially distributed and uninterrupted observations (top-down estimates) place geologic CH₄ estimates at 80 Tg CH₄ yr⁻¹ (Etiope, 2012). Additionally, anthropogenic fossil emissions are estimated to be ~ 90 Tg CH₄ yr⁻¹ (Denman et al., 2007), which leaves a residual fossil emission source (~ 32 Tg CH₄ yr⁻¹) yet to be identified (Etiope et al., 2008). Each one of these sources of CH₄ is regulated by mechanisms of CH₄ consumption or by CH₄ sinks that ultimately determine the net CH₄ flux to the atmosphere.

This work was focused on a particular kind of natural CH₄ source, northern wetlands, which have been estimated to contribute < 6 to 40 Tg CH₄ yr⁻¹ (IPCC, 2001; Zhuang et al., 2006). Furthermore, IPCC (2007) established that Polar Regions and perennially frozen environments are the most vulnerable to climate change. For instance, the North Slope of Alaska, located entirely above the Arctic Circle (Kittel et al., 2011) has experienced a temperature increase of 0.5 °C decade⁻¹ from 1966 to 1995 (Serreze et al., 2000), and the mean annual temperature in Barrow, the northernmost town in Alaska, has increased 2.4 °C decade⁻¹ from 1949 to 2008 based on the historical record (ACRC, 2008). This trend of increasing temperatures have impacted permafrost temperatures in the Arctic by > 3 °C (Zhang et al., 2003; Osterkamp, 2007) affecting the landscape in different ways and specifically impacting CH₄ emissions from thermokarst lakes (Rouse et al., 1997). Current estimates that rely on bottom-up estimates suggest that the Arctic contributes 15-50 Tg CH₄ yr⁻¹ to the total CH₄ budget, or 3-9 % of the net source (excluding CH₄ sinks;
Denman et al., 2007; McGuire et al., 2009). Top-down estimates include a 15-35 Tg CH$_4$ yr$^{-1}$ contribution from thermokarst lakes (Walter et al., 2006), of which 0.72 Tg CH$_4$ yr$^{-1}$ correspond to Alaska lakes emissions (Sepulveda-Jauregui et al., 2014), increasing the contribution from the Arctic to 32-112 Tg CH$_4$ yr$^{-1}$ (McGuire et al., 2009).

Figure 1. Satellite image of the western portion of the Coastal Plain in the North Slope of Alaska, showing terrain occupied by thermokarst lakes and drained lake basins.

5. Thermokarst lakes

Permafrost is defined as ground that remains frozen ($\leq$ 0 °C) for a minimum of two years (Davis, 2000). Permafrost underlies ~ 25 % of Earth’s terrestrial surface, and recent estimates suggest that there are between 1400 and 1850 Gt of organic carbon
stored in arctic soil stocks (McGuire et al., 2009; Tarnocai et al., 2009; Graham et al., 2012). The top layer of permafrost may thaw seasonally and it is thus known as 'the active layer', with depths that range from 30-60 cm at high latitudes (e.g., Point Barrow, Alaska at 71° 23’ N, 156° 28’ W) to 300 cm at lower latitudes (e.g., Anchorage, Alaska at 61° 13’ N, 149° 54′ W) (Brewer, 1958; Péwé, 1982; Washburn, 1980). The bottom of the active layer is known as 'the permafrost table', and the permafrost itself can be hundreds of meters deep, particularly in regions of continuous permafrost that comprise the northern halves of Canada and Alaska, the northern two-thirds of Greenland, and grand parts of Siberia, China and Tibet (Davis, 2000). Depending on the amount of ice in permafrost and its spatial extent, permafrost can be thick and continuous, discontinuous, sporadic, or isolated (Péwé, 1975; Péwé, 1983).

Permafrost is generally found at high latitudes (lowlands, highlands, intra- or inter-mountane depressions) including shallow continental shelves under the polar seas (Davis, 2000), or high altitudes at mid-latitudes (mountains, highlands, plateaus) (Péwé, 1983; Heginbottom & Dubreuil, 1993; Davis, 2000; Brown et al., 1997). In Siberia, Alaska, and western Canada permafrost occurs in unglaciated regions with ice-rich deposits (ice wedges, ice lenses, and pingo ice cores) or in previously glaciated regions with remnant ice in the ground (Grosse et al., 2013). During the last deglaciation, in the transition from the Pleistocene to the Holocene, temperatures warmed, inducing a form or permafrost degradation that led to thermokarst lake formation in lowland regions with ice-rich unconsolidated
sediment deposits (Zimov et al., 1997; Walter Anthony et al., 2014). Thermokarst lakes are mostly found in Arctic, sub-Arctic, and Boreal regions (Lehner & Doll, 2004; Grosse et al., 2013) and Grosse et al. (2013) estimated a total thermokarst lake and pond area in the pan-Arctic between 250,000 Km² and 380,000 Km².

In northern permafrost, both localized disturbances to the ground’s thermal regime and large climatic changes promote thawing of massive ice wedges in the frozen ground (French, 1976). Competing forces such as climate, hydrology, geology, and short-term ecosystem alterations may affect the thermal stability of permafrost (Jorgenson et al., 2010; Grosse et al., 2013), inducing ice wedge melting. Melted ice-wedges coalescence through superficial ponds and the water that accumulates on top of the permafrost increases water-sediment interface temperatures up to 10 °C above the mean annual air temperatures, due to water properties such as low albedo, absorption of long wave radiation, and a higher heat storage capacity in comparison to ice and dry ground (Walter Anthony et al., 2014). A positive feedback loop is then created in which the warmer temperatures promote even more permafrost degradation and lake bottom expansion, to the point where water column depth exceeds winter ice formation enhancing the growth of the so called thaw bulb (Grosse et al., 2013). This process may be also enhanced by substrates with higher thermal diffusivity like ice-rich mineral substrates, according to a modeling scenario by West and Plug (2008). Lateral expansion of the lakes is aided by thermal and mechanical erosion of the shorelines as described by Grosse et al. (2013), with lake sizes ranging between 100 m and 15 Km in diameter, and lake
depths of up to 25 m depending on the permafrost region. For instance, in the Coastal Plain of the North Slope of Alaska there are shallow ice wedges in the permafrost and lakes in this area tend to be shallow (< 3 m deep), but in Siberia and the Northwestern Canadian territories, there are deeper and larger ice wedges and massive glacial ice, and the lakes are much deeper (Grosse et al., 2013).

In the Coastal Plain of the North Slope of Alaska, an area that extends 100 Km to the South from the Arctic Ocean (Morrissey & Livingston, 1992), an estimated 50-70 % of the region is occupied by thermokarst lakes and drained lake basins. Thermokarst lakes in this region present an elliptical shape, and are oriented perpendicular to the prevailing wind direction according to the most widely accepted, but still controversial hypothesis. Wind-induced waves create circulation cells in the ice-free period of the year, causing lateral shoreline erosion 50 ° from the wave approach, deposition of sediments in other shores, and the formation of littoral shelves. Therefore, this hypothesis suggests that the prevailing wind direction would dictate the orientation of the long axis of the lakes and enhance thermokarst erosion at the ends (Carson & Hussey, 1962).

Thermokarst lakes on the North Slope are frozen for ~ 9 months of the year (Phelps et al., 1998) and some lakes freeze all the way to the bottom due to shallow depths. During the summer months, mean daily temperatures above freezing favor lake thaw in late spring, and continuous sunlight reaches the water column over 24 hr periods. During this time of the year CH$_4$ is released from the sediments into the water column where it may be oxidized to CO$_2$, and then released to the atmosphere.
An estimated 0.5 to 11.3 g CH$_4$ m$^{-2}$ from high-latitude lakes and ponds is contributed to the atmospheric CH$_4$ budget (Whalen et al., 1990; Bartlett et al., 1992; Roulet et al., 1997) during the ice-free season, but there is also evidence of a CH$_4$ pulse during spring thaw from lakes on the North Slope (Phelps, 1998). Thermokarst lakes have played different roles on CH$_4$ emissions over geologic time. After formation during the Holocene, lakes in Siberia and Alaska may have shifted from carbon sources to sinks (Walter Anthony, 2014), but currently they are considered to be CH$_4$ sources.
6. Microbial life at the limits: examples from two cold environments

Microorganisms that inhabit permafrost and thermokarst lakes must be able to cope with a variety of environmental stressors such as low temperatures, low water activity, high levels of radiation, and freeze-thaw cycles. However, these arctic ecosystems are just one kind of extreme cold environment where microbial communities thrive; other examples of cold environments are glaciers, ice sheets, marine sediments, and snow. Snowpacks are also of interest for my dissertation research, because they represent a contrasting ecosystem that imposes a different set of stressors on microbial life adapted to the cold.

Methanogenic Archaea are unique in the microbial world because of their ability to produce CH\(_4\). The majority of methanogens in culture were isolated from warm environments like wetlands and rice-paddies; but despite their ability to thrive in warm environments, methanogens have also colonized extremely cold places. Psychrophilic and psychrotrophic methanogens not only survive in these conditions, but are also able to maintain an active metabolism and to reproduce. I dedicated two chapters of my dissertation research to study the activity of methanogens in arctic thermokarst lakes, and their role in the microbial community inhabiting the sediments of these lakes. My ultimate goal was to further the understanding of cold-adapted methanogens and their impact on the greenhouse gas budget of our planet. For the same reason, I attempted to expand the short list of cold-adapted methanogens in culture (Table 1.1; and see appendix), which in the
future would allow us to learn more about their genetic adaptations to different environmental stresses, and what we could look for in other icy bodies of the solar system.

Table 1.1 Some cold-adapted methanogens in culture

<table>
<thead>
<tr>
<th>Scientific name</th>
<th>Temperature preference</th>
<th>Place of origin</th>
<th>Substrate</th>
<th>Reference</th>
</tr>
</thead>
</table>

(Franzmann et al., 1997; Franzmann et al., 1992; Chong et al., 2002; Singh et al., 2005; Kendall et al., 2007; Krivushin et al., 2010)

Some environments that are favorable for methanogenesis occur in cold climates (northern wetlands in Siberia, Alaska, Canada, Scandinavia), and are dominated by permafrost, where seasonal variation in ground temperature determines the availability of pore water (Wagner, 2008). However, methanogens can be metabolically active in permafrost (Mikucki et al., 2011) when certain
conditions are met. For instance, the freezing point of water can be depressed during freezing of pure water because of solute exclusion, which allows the presence of thin films of water at temperatures as low as -20 °C (Price, 2000; Bakermans, 2008). Incorporation of radiolabeled acetate at ~20 °C over a 550-day period as well as microbial growth at -10 °C has been detected in permafrost sediments (Rivkina et al., 2000). These findings suggest that microbial life is metabolically active in permafrost at subzero temperatures, and that it is able to continue the cycling of nutrients under this conditions (Mikucki et al., 2011), though at very slow reactions rates.

Of all the environmental factors that can be considered extreme (i.e., departure from optimal conditions), low water activity due to freezing temperatures is perhaps one of the most stressful to microorganisms. To be active at subzero temperatures microbes have to deal with limited amounts of water (the solvent for enzymes), and the resulting decrease in catalysis and transport; the decrease in membrane fluidity, ice crystal formation, and the destabilization of important molecules like DNA and RNA (Cavicchioli et al., 2000; Bakermans, 2008). Some of the adaptations to this type of stress include modification of enzyme kinetics to enable high enzymatic activity at low temperatures; synthesis of glycoproteins and peptides that prevent freezing by dropping the freezing point of water inside the cell; increased membrane fluidity by increasing the degree of unsaturation of fatty acids, modifying branching patterns, and shortening fatty acid chain length; regulation of ion channel permeability; synthesis of cold-shock proteins (Csps) that
stabilize the mRNA to allow protein synthesis, and synthesis of cold acclimation proteins (Caps) that are synthesized during prolonged exposure to cold temperatures (Pikuta et al., 2007).

Methanogens are just one example of the kind of microorganisms that are adapted to live in cold environments, in conditions of freezing temperatures and freeze-thaw cycles, low water activity, high or low levels of salinity or pH, high levels of UVB radiation during the summer, and complete darkness during the winter (Hoover et al., 2002; Pikuta et al., 2007). In snow-covered regions in the Arctic, the Antarctic, and in high altitude environments, microorganisms belonging to several different phyla from the three domains of life must also adapt to conditions of low water activity due to low temperature. One of the most studied eukaryotic microorganisms in snow is the unicellular alga *Chlamydomonas nivalis*, which produces pigmented red cells (cysts and zygotes) giving this coloration to the snow. *C. nivalis* uses these pigments (flavonoids and astaxanthins) as antioxidants, an adaptation strategy that makes it resistant to high levels of UVA and UVC radiation (Hoover et al., 2002; Pikuta et al., 2007). As a consequence, the red biomass in snow changes the albedo, which induces snow melt and allows other microorganisms to thrive because of increased water availability (Pikuta et al., 2007). This ‘red snow’ commonly found in the Sierra Nevada of California is a signature of life occupying a contrasting habitat of interest for my dissertation.

The very high radiation levels the algae experience at high altitudes decrease with snow depth, however, snowpack temperatures at depth can be well below the
constant 0-1 °C characteristic of wet snow (Remias, 2012). Wet snow mostly occurs during the spring melt, when the snowfall from the previous winter starts to thaw, but persist in a water-logged state until the summer, favoring snow algae development (Remias, 2012). Unicellular algae like *C. nivalis* require higher amounts of liquid water than bacteria in permafrost because of the larger cell sizes (10-30 µm), and old summer snow allows the formation of a continuous water film around macroscopic snow crystals that is the ideal microhabitat for the algae (Jones, 2001; Remias, 2012).

The life cycle of *C. nivalis* as described by Remias (2012) is tied to annual environmental changes in snowpack, reflecting adaptation to the harsh conditions of this cold environment. In the cyst phase, cells are dormant, *i.e.*, non-motile, stop reproducing, preserve energy and nutrients, rearrange their chloroplast, produce lipids to decrease the water content in the cell, and accumulate UV-protecting pigments. All of these changes allow the algae to withstand conditions of high radiation, low nutrient availability, and freeze-thaw cycles during the summer. Some discrepancies remain in the understanding of cyst physiology, as Remias (2005) determined that mature cysts of *C. nivalis* from the Alps were photosynthetically active and produced detectable levels of oxygen, indicating at least low levels of metabolism. In the spring of the next season, cells germinate (reproduce asexually) and release four-to-eight daughter cells into the snowpack that has accumulated over the winter. In the early summer, the flagellated algal cells that emerge from cyst forms can propel upwards seeking adequate light conditions for sexual
reproduction and growth. At this stage, these cells are most sensitive to
environmental conditions; therefore, they remain deep enough in the snowpack to
be protected from high irradiation and freeze-thaw cycles (Remias, 2012). When
temperatures rise, cellular growth and metabolism ensues, and the nutrients can be
exhausted. Flagellates then turn into vegetative cysts or generative zygotes, losing
motility and the ability to reproduce. In this state, the algae become more resistant
to the harsher environmental conditions associated to the snowmelt period
(Remias, 2012). Likewise, concurrent irradiation levels on the surface of the
snowpack become increasingly high stimulating carotenoid production and
chloroplast size reduction so that the cells finally enter the dormant state that lasts
until next season (Remias et al., 2005).

During their flagellated phase, snow algae are photosynthesizing, which
makes them primary producers and CO₂ sinks under favorable conditions (Painter
et al., 2001; Williams et al., 2009). Additionally, they create conditions that promote
the growth of heterotrophic bacteria. These two types of metabolism
(photoautotrophic and chemoheterotrophic) are different from
chemolithoautotrophic metabolism common to several species of cold-adapted
methanogens, which is less favorable in cold environments (Hoover et al.; Pikuta et
al., 2007). Nevertheless, the microbial community associated with snow algae has
only been described in a few studies (Segawa et al., 2005; Liu et al., 2009; Zhang et
al., 2010). Much less attention has been paid to these communities in comparison to
microbial communities in permafrost (Zinger et al., 2009; Larose et al., 2010;
Harding et al., 2011; Hamilton et al., 2013; Moller et al., 2013)) or other icy environments including glacial ice, subglacial systems, lake ice, sea ice, and ice-sheet environments (Mikucki et al., 2011).

An interesting characteristic of algae in snow is that they constitute a detectable biosignature due to carotenoid pigmentation. Painter (2001) have used imaging spectroscopy to analyze algal spectral features enabling detection by remote-sensing which can be calibrated to determine algal concentrations. Unlike methanogens, snow algae can be detected directly using spectroscopy-based techniques. While spectroscopy alone would not tell us whether methanogens are present, a combination of spectroscopic and spectrometric techniques can detect the metabolic byproducts formed by methanogens. However, the likelihood of encountering evolutionarily advanced photosynthetic microorganisms elsewhere in the Solar System is questionable, whereas the presence of a primitive form of metabolism like methanogenesis is much more likely, particularly in Mars.

7. **Overall objectives**

My dissertation aimed to study microbial life in two different kinds of cold environments with astrobiological relevance: the sediments of thermokarst lakes in permafrost at high-latitudes, and snow fields at high-altitude environments. Understanding how life thrives in these environments will aid our efforts to search for life in the Solar System, especially in places like the permafrost regions in Mars and the icy surface of Europa. The microbial communities inhabiting these cold
environments are also very relevant for the global cycling of C, and with the current trend of climate change, we must fully understand their role in nutrients cycling and how they may respond to change.

With these goals in mind my dissertation aimed to:

1. Determine the sources of CH$_4$ in sediments of arctic thermokarst lakes in the North Slope of Alaska, and what geochemical factors favor biological CH$_4$ production from these lakes.

2. Describe the diversity and structure of the microbial community that supports organic matter decomposition leading to methanogenesis, and to analyze how organic carbon affects microbial community composition.

3. Uncover the microbial communities associated with snow algae in the snow fields of Mt. Conness and on the Pacific Crest Trail near Mt. Lincoln in the Sierra Nevada of California, and on Wheeler Peak in the Great Basin National Park in Nevada.
References


Chapter 2

Methane sources in arctic thermokarst lake sediments on the North Slope of Alaska

Co-authorship statement


PM, MR and AEM designed the sampling approach and the experiments. PM, MR, AEM, DB, HA, JP, AM, KH collected samples for all of the analyses. PM executed the CH$_4$ production experiment, qPCR, and analyzed the data generated from CH$_4$ production experiment, CH$_4$ concentration, $\delta^{13}$C$_{CH_4}$ and $\delta^{13}$C$_{CO_2}$, pore water chemistry of anions and metals, and qPCR under the supervision of AEM. OR helped with CH$_4$ concentration and $\delta^{13}$C analyses, and with DNA extractions. MR and MW performed analyses for TOC and all of the lipid biomarkers under the supervision of GL. AM and JP performed O$_2$ analyses. HA and JP performed limnological analyses. PM, MR and AEM wrote this chapter in collaboration with GL, JP, AM, KH, DB and HA. AEM, JP, GL and KH secured the funding for this research.

The contents of this chapter were published in:

*Geobiology Volume 13, Issue 2, March 2015, Pages: 181-197*
Abstract

The permafrost on the North Slope of Alaska is densely populated by shallow lakes that result from thermokarst erosion. These lakes release methane (CH$_4$) derived from a combination of ancient thermogenic pools and contemporary biogenic production. Despite the potential importance of CH$_4$ as a greenhouse gas, the contribution of biogenic CH$_4$ production in arctic thermokarst lakes in Alaska is not currently well understood. To further advance our knowledge of CH$_4$ dynamics in these lakes, we focused our study on (i) the potential for microbial CH$_4$ production in lake sediments, (ii) the role of sediment geochemistry in controlling biogenic CH$_4$ production, and (iii) the temperature dependence of this process. Sediment cores were collected from one site in Siqlukaq Lake and two sites in Sukok Lake in late October to early November. Analyses of pore water geochemistry, sedimentary organic matter and lipid biomarkers, stable carbon isotopes, results from CH$_4$ production experiments, and copy number of a methanogenic pathway-specific gene (mcrA) indicated the existence of different sources of CH$_4$ in each of the lakes chosen for the study. Analysis of this integrated data set revealed that there is biological CH$_4$ production in Siqlukaq at moderate levels, while the very low levels of CH$_4$ detected in Sukok had a mixed origin, with little to no biological CH$_4$ production. Furthermore, methanogenic Archaea exhibited temperature dependent use of in situ substrates for methanogenesis, and the amount of CH$_4$ produced was directly related to the amount of labile organic matter in the sediments. This study constitutes an
important first step in better understanding the actual contribution of biogenic CH$_4$ from thermokarst lakes on the coastal plain of Alaska to the current CH$_4$ budgets.

1. Introduction

Thermokarst lakes, resulting from ground ice melting in permafrost regions (Howard & Prescott, 1973; French, 1976; Kokelj & Jorgenson, 2013) may be significant contributors to the global CH$_4$ budget (Oechel et al., 1993; Phelps et al., 1998; Walter et al., 2007; Schuur et al., 2009). A large proportion of this CH$_4$ may be derived from ancient thermogenic CH$_4$ trapped deep within or under the permafrost (referred to as the "cryosphere cap", Walter-Anthony et al., 2012). However, the large amount of organic matter stored in the thaw layer (talik) between the water column and the permafrost table (Ling & Zhang, 2003; Pedersen et al., 2011; Parsekian et al., 2013), either from interglacial or contemporary photosynthesis, also serves as a significant source of carbon for in situ methanogenesis. Identifying and deconvolving the production sources of CH$_4$ will improve our ability to generate accurate predictions about the changing climate in the Arctic.

The North Slope of Alaska has extensive continuous permafrost (~ 60-75 % ice by volume and ~ 400 m deep; Sellmann et al., 1975; Hinkel et al., 2003; Jorgenson et al., 2008), and is occupied by thousands of shallow (~ 2-3 m deep), relic (i.e., drained), and contemporary thermokarst lakes (Hussey & Michelson, 1966; Frohn et al., 2005; Jorgenson & Shur, 2007) that are ice-covered or frozen to the ground for at least 9 months of the year. Northern Alaska lake sediments may be
gradually eroded through the lake thaw cycle and preferentially re-deposited at the upwind and downwind lake margins (Carson & Hussey, 1962; Hinkel et al., 2003). Relative contributions from allochthonous or autochthonous sources of organic matter have not been well studied in the sediments of this area. Allochthonous organic matter can be transported via fluvial or aeolian processes, and derive from modern active layer soils or Pleistocene-aged terrigenous organic matter from permafrost (Repenning, 1983). Autochthonous organic matter can be produced by present-day lacustrine autotrophs (Ramlal et al., 1994; Hecky & Hesslein, 1995; Bonilla et al., 2005) and is potentially more labile than allochthonous organic matter.

The coastal plain in the North Slope of Alaska also contains an estimated 53 billion cubic feet of natural gas (Houseknecht et al., 2010). Radiocarbon analyses have indicated that gas seeps in the area may be sourced from the gas reservoirs at depth and/or laterally from thermogenic CH$_4$ trapped under the permafrost ice cap, rather than from present-day microbial activity within the lake, as is common in Siberia (Walter Anthony et al., 2012). A distinction between microbial and thermogenic CH$_4$ can be made by combining isotopic ratios (e.g., $\delta^{13}$C$_{CH_4}$, $\delta^{13}$C$_{CO_2}$ and $\delta$D-$CH_4$) and C2-C4 hydrocarbon ratios (Whiticar, 1999). Microbial CH$_4$ production, which $\delta^{13}$C ranges between -110 and -50 $^0$/oo (Quay et al., 1988; Whiticar, 1999), results from anaerobic decomposition of organic matter in sediments. Thermogenic CH$_4$ has a range between -52 and -20 $^0$/oo (Whiticar, 1999; Judd, 2000; Kvenvolden & Rogers, 2005) and it is generated at subsurface depths between 1 and 4 Km by
decomposition of residual organic matter under high pressure and temperature, during coal formation or thermal alteration of oil (Judd, 2000). $\delta^{13}C_{CH_4}$ signatures at the boundary between biogenic and thermogenic CH$_4$ could result from mixed sources, including CH$_4$ oxidation, advanced stage of parent organic matter decomposition, contributions from different methanogenic pathways, or a combination of thermogenic and biogenic signatures (Whiticar, 1999). Lastly, abiogenic CH$_4$ originates in the mantle, and it has a $\delta^{13}C$ between -45 and -50/$\text{oo}$ (Judd, 2000).

Our study focused on biological CH$_4$ production in two Alaskan thermokarst lakes. Specifically we examined: i) in situ CH$_4$ concentrations and carbon isotope compositions of CH$_4$ in sediments; ii) temperature response of methanogenesis at natural substrate levels; iii) archaeol lipid biomarkers (archaeol) and the methyl coenzyme reductase alpha subunit (mcrA) gene, which is a key enzyme in the pathway for methanogenesis; and iv) description of the substrates available for methanogenesis.

2. Methods

2.1. Sampling sites

Siqlukaq Lake (Siq) and Sukok Lake (Suk), two arctic thermokarst lakes near the town of Barrow, Alaska (Fig. 2.1), were sampled on late October to early November field campaigns. Two sites were sampled at Sukok, one near an active, submerged natural gas seep (Sukok Seep - SukS), and another about 1 km southwest from the
seep site (Sukok B site - SukB), to determine the effects of localized CH$_4$ flux on biological CH$_4$ production within these sediments. A total of 16 sediment cores were recovered for various geochemical and biological analyses over the course of four field campaigns spanning 4 years (Table 1). Sediments in both lakes lack well-defined sedimentological features, such as laminations, and the lakes possess taliks of at least 1.1 m depth (the maximum sediment thickness penetrated in coring).

Suk is located ~ 29 km south of Barrow and 12.7 km east-southeast of the mouth of Walakpa Bay, in the Walakpa gas field, a natural gas field approximately 600 m deep (the permafrost base nearby the lake is ~ 280 m; Glenn & Allen, 1992). An east-southeast trending fault occurs in the subsurface north of Sukok, however no such feature is identified beneath the lake itself (Glenn & Allen, 1992). Openings in autumn ice cover resulting from active CH$_4$ ebullition in the lake were observed in April 2010 and the late October-early November 2010-2013 field campaigns. Satellite imagery indicates that Suk consists of at least three coalesced thermokarst lakes and lies within a portion of the arctic coastal plain that has seen repeated thermokarst episodes (Fig. 2.1). Suk is approximately 4.2 km wide and 5.5 km long (not including the slightly-adjoined southern basin). At the time of sampling, total water depth for Suk was ~ 0.80-1.35 m, with ~ 0.10-0.25 m of ice, and lake water temperatures were relatively uniform with depth between 0.7 and 1.3 °C, as determined with a portable Orion 5 star multi meter (Thermo Scientific, Waltham, MA, USA).
Figure 2.1 Landsat 7 image of the Arctic Coastal Plane near Barrow, AK (L7 ETM+ SLC-on, 31 August 2000). Shaded field and hatched lines approximate subsurface boundaries of the Walakpa Gas Field (Glenn & Allen, 1992). Dots indicate sites described in the main text: Siqlukaq (Siq), and the two Sukok Lake sites are "Seep" proximal to an active, ebullient gas seep (SukS), and "B" distal from the area of active CH₄ seepage (SukB).

Siq, located 6.6 km east-northeast of the mouth of Walakpa Bay, outside of the gas field, has no visible open holes in winter lake ice due to gas ebullition. Satellite photography shows that Siq likely drains into Walakpa Bay at high-stand and shows evidence of previously higher lake levels, but a less complex hydrologic history than that of Suk (Fig. 2.1). Siq is smaller and more elongated than Suk, measuring approximately 1.0 km wide by 3.8 km long. During the sampling period, Siq was ~ 1.5-1.6 m deep, had ~ 0.15-0.25 m of ice, and lake water temperatures between 0.2 and 2.0 °C.
2.2. Sediment geochemistry

Sediment cores were retrieved from all sampling sites using a universal percussion corer (Aquatic Research Instruments), ~ 10 cm in diameter polycarbonate coring tube, and plug caps on both ends to decrease oxidation. Sediment core lengths ranged between 20 and 110 cm.

2.2.1. Oxygen (O$_2$) microelectrode profiles

Shallow sediment cores (20 cm) were collected at each sampling site (Siq13, SukB13 and SukS13, Table 1). Overlying water (4.5 cm) was left on top of the sediment to minimize atmospheric O$_2$ influence on the microprofile in the sediment (Boetius & Damm, 1998). Microelectrode O$_2$ profiles were conducted within 30 min of core collection using a Clark-style oxygen microelectrode (Unisense, Aarhus, Denmark) with a tip diameter of 500 µm. These microelectrodes respond in a linear fashion to O$_2$ concentration (Revsbech, 1989) and a two point calibration curve was used to standardize the instrument. The microelectrode was attached to a manual micromanipulator and lowered through the water and sediment column at 100 µm increments. Profiling was conducted in a darkened tent which provided a thermal barrier to prevent freezing and to maintain core temperatures (2.7-6.0 °C) during profiling.
Table 2.1 Sampling sites, dates, and analyses performed in each core. Cores were labeled using the first three letters to indicate the lake, the next letter to indicate the site (only in the case of Sukok Lake), followed by the last 2 digits of the year, and replicate cores indicated by a, b, c, and d.

<table>
<thead>
<tr>
<th>Coordinates (N, W)</th>
<th>Sampling dates</th>
<th>O₂ profile</th>
<th>Pore water chemistry</th>
<th>Total Carbon</th>
<th>Lipid Biomarkers</th>
<th>mcrA qPCR</th>
<th>[CH₄] &amp; δ¹³CCH₄</th>
<th>Porosity</th>
<th>CH₄ production</th>
<th>Sediment texture</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siquluqaq</td>
<td>71° 10.486' - 156° 53.891'</td>
<td>29-Oct-2010</td>
<td>_</td>
<td>_</td>
<td>Siq10-a</td>
<td>Siq10-a</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>71° 10.481' - 156° 53.910'</td>
<td>22-Oct-2011</td>
<td>_</td>
<td>_</td>
<td>Siq11-a</td>
<td>Siq11-a</td>
<td>Siq11-a</td>
<td>Siq11-a*</td>
<td>_</td>
<td>Siq11-b</td>
</tr>
<tr>
<td></td>
<td>71° 10.482' - 156° 53.900'</td>
<td>30-Oct-2012</td>
<td>_</td>
<td>_</td>
<td>Siq12-a*</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>Siq12-a</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>71° 10.457' - 156° 54.004'</td>
<td>02-Nov-2012</td>
<td>_</td>
<td>_</td>
<td>Siq12-c*</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>Siq12-c</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>71° 10.451' - 156° 53.872'</td>
<td>01-Nov-2013</td>
<td>Siq13-a</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>Sukok B</td>
<td>71° 04.006' - 156° 49.841'</td>
<td>24-Oct-2011</td>
<td>_</td>
<td>_</td>
<td>SukB11-a</td>
<td>SukB11-a</td>
<td>SukB11-a</td>
<td>SukB11-a</td>
<td>_</td>
<td>SukB11-b</td>
</tr>
<tr>
<td></td>
<td>71° 04.006' - 156° 49.841'</td>
<td>02-Nov-2013</td>
<td>SukB13-a</td>
<td>SukB13-b</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>SukB13-c</td>
<td>SukB13-c</td>
</tr>
<tr>
<td></td>
<td>71° 03.997' - 156° 49.918'</td>
<td>06-Nov-2013</td>
<td>_</td>
<td>_</td>
<td>SukB13-d</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td>Sukok S</td>
<td>71° 04.455' - 156° 49.250'</td>
<td>27-Oct-2010</td>
<td>_</td>
<td>_</td>
<td>SukS10-a</td>
<td>SukS10-a</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>71° 04.519' - 156° 49.208'</td>
<td>20-Oct-2011</td>
<td>_</td>
<td>_</td>
<td>SukS11-a</td>
<td>SukS11-a</td>
<td>SukS11-a</td>
<td>SukS11-a</td>
<td>_</td>
<td>SukS11-b</td>
</tr>
<tr>
<td></td>
<td>71° 04.513' - 156° 49.202'</td>
<td>31-Oct-2013</td>
<td>SukS13-a</td>
<td>SukS13-b</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
</tr>
<tr>
<td></td>
<td>71° 04.519' - 156° 49.204'</td>
<td>05-Nov-2013</td>
<td>SukS13-c</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>_</td>
<td>SukS13-d</td>
<td>SukS13-d</td>
<td>_</td>
</tr>
</tbody>
</table>

* Acetate concentration, Formate concentration, and δ¹³C CO₂ were also analyzed in these cores.
Depth-integrated aerobic O$_2$ consumption (IOC) was calculated using Fick’s second law of diffusion assuming zero-order kinetics (Nielsen et al., 1990; Rasmussen & Jørgensen, 1992). The corrected diffusion coefficient ($D_s$) was calculated by adjusting the O$_2$ diffusion coefficient in fresh water (at sediment temperature during profiling) for porosity and tortuosity, based on measured porosity values and sediment type (Broecker & Peng, 1974; Rasmussen & Jørgensen, 1992).

### 2.2.2. Pore water chemistry

Dissolved gas and chemical gradients in the sediments, were determined for deeper cores (50-70 cm) collected from Siq (Siq12-a and Siq12-c) in 2012, and from Suk (SukS13-b, SukS13-c, SukB13-b, and SukB13-d) in 2013 (Table 1). Pore waters were sampled through pre-drilled holes in the core liners (sealed from the surrounding environment until samples were taken) using Rhizons (Seeberg-Elverfeldt et al., 2005). The $\sim$ 0.15 µm porous membrane of each Rhizon was conditioned before sampling by rinsing with milli-Q water. Samples were drawn into 10 mL sterile syringes connected to the Rhizons.

Pore water samples for analysis of low molecular weight organic acids including acetate and formate, and anions SO$_4^{2-}$ and NO$_3^-$, were collected in 2012 from Siq. Samples were collected (5 mL HDPE bottles), frozen, and then transported to the Biogeochemistry Laboratory at Indiana State University. A Dionex ICS-2000 with an AS11-HC column (Sunnyvale, CA) was utilized to measure the concentration...
of each compound following Johnson et al. (2012) and Baker and Vervier (2004). Due to technical issues, acetate and formate were only measured in the surficial samples at Siq. Samples for analyses of the anions $SO_4^{2-}$ and $NO_3^-$ collected (15 mL polypropylene tubes) from Suk in 2013 were frozen, and then transported to the University of Tennessee, Knoxville. Pore waters were analyzed using a Dionex ICS-2100 RFIC fitted with an ASRS-4mm suppressor column, an AS18 analytical column, and an AG18 guard column (Sunnyvale, CA) following methods similar to Banihani et al. (2009).

Pore water samples for metals analysis collected (15 mL polypropylene tubes) from Siq and Suk, were acidified with UHP HNO$_3$ to a final concentration of 1 % HNO$_3$ (vol/vol) in the field, and stored at room temperature for < 60 days. The 5 mL samples were brought to 10 mL with 1 % UHP HNO$_3$ before analysis. A Thermo Element II high resolution inductively coupled plasma-mass spectrometer with a PFA-ST concentric Teflon nebulizer (ESI, Inc. Portland, OR) and a spray chamber of cyclonic glass (ESI, Inc. Portland, OR) in the Ultratrace Chemistry Laboratory at the Desert Research Institute in Reno, NV was used to quantify the metals. Low (LR) and medium (MR) resolutions were used as needed for isotopic separations. Standards were made from mixed stock standard from Inorganic Ventures, Inc. (Christiansburg, Virginia, USA) in a 1 % UHP HNO$_3$ matrix, and all blanks were made of 1 % UHP HNO$_3$. 
2.3. Sediment CH₄

2.3.1. CH₄ concentrations and stable carbon isotope analyses

One sediment core per sampling site was collected in 2011 (Siq11-a, SukB11-a, and SukS11-a) to determine CH₄ concentrations and stable isotope signatures. The cores were sampled on site following Riedinger et al. (2010) and Koch et al. (2009). Samples were preserved at ~ 4 °C and analyzed at the University of California Santa Barbara (UCSB), for CH₄ concentration, δ¹³C₃H₄, and δ¹³C₃O₂, following methods by Kinnaman et al. (2007), with the exceptions that the bottle headspaces were displaced with 1-5 mL degassed water containing NaCl (35 g L⁻¹), and 1 mL sample was injected onto a 250 µL sample loop for quantitation. CH₄ concentration (µmoles CH₄ g⁻¹ sediment dry weight) was estimated from the molar fraction of CH₄ in the headspace using equation 1 from D’Hondt et al. (2003), excluding the terms for porosity and sediment volume and including sediment dry weight.

To determine inter-annual variability of CH₄ concentrations and C isotopic composition, additional cores were collected from Siq in 2012 (Siq12-a and Siq12c) and from Suk (SukB13-c and SukS13-d) on November 2013. Parallel sediment plugs were collected for porosity analysis within 1.5 cm of samples for CH₄ and CO₂ analysis, and were stored at ~ 4 °C (Riedinger et al., 2010). The approximate ratio of sediment mass to volume in each syringe was obtained by measuring the total volume and wet weight of sediment. Water content was determined using dry
weight after heating at 105 °C for 24 hours. Porosity was calculated as water volume divided by wet sediment volume.

For comparison, CH$_4$ concentration was also estimated in µM following equation 1 in D'Hondt et al. (2003) and using an average porosity from depths sampled in SukB13-c and SukS13-d (data was not available for Siq), and an average sediment volume of 2.7 ± 0.7 mL. Sediment volume was estimated from the average bottle headspace, the known volume of NaCl solution, and the average known volume of the bottles with stoppers.

2.3.2. CH$_4$ production experiment

Samples were obtained for CH$_4$ production experiments from sediment cores collected in 2011 (Siq11, SukS11, and SukB11) at the same time and within ~ 15 cm of those for in situ CH$_4$ concentration and stable carbon isotope analysis (Table 1). Sediment cores for these experiments were transported at < 4 °C to a cold room, where they were maintained at 2 °C for one month. Each core was cut in three sections of approximately equal size, subsampled inside an anaerobic chamber (3-4 % H$_2$:N$_2$ atmosphere), and mixed with approximately equal volumes of cold, sterile, anoxic, deionized water by stirring, to eliminate any gas that may have been ‘trapped’ in the sediments (Kiene & Capone, 1985). Sediment slurries in 10 mL aliquots were distributed among 18 sterile, 125 mL-serum bottles (capped with butyl rubber stoppers) per depth. Negative controls with the same water and bottles were also prepared in the anaerobic chamber. The headspace of the bottles was
exchanged with ultra-high purity N\textsubscript{2} for 5 min using a manifold with 0.2 µm filters and sterile needles.

The sediment slurries were incubated upside down at 2 °C or 10 °C for the duration of the experiment. Time zero samples were collected after 2 hr incubation. For all time points, headspace samples were collected with a gas tight syringe (Hamilton, Reno, NV, USA) following vigorous shaking, before and after autoclaving of sediment slurries. Samples were analyzed with a Mini 2 gas chromatograph (GC; Shimadzu, Kyoto, Japan) equipped with a flame ionization detector (FID), a Poropak T column, and the following settings: oven temperature 20 °C, injection port temperature 210 °C, and ultra-high purity nitrogen as carrier gas at a flow rate of 40 mL min\textsuperscript{-1}. The injection volume was 100 µL. Peak area was quantified with a Hewlett-Packard 3390A integrator. CH\textsubscript{4} standards were used for calibration. The headspace of triplicate samples was analyzed every 5 to 10 days (depending on when CH\textsubscript{4} was first detected), for a period of 25 to 60 days depending on the lake.

At the end of the experiment, the dry weight of the sediments was determined after drying for 24 h at 105 °C. The CH\textsubscript{4} concentration in the headspace (µmoles CH\textsubscript{4} g\textsuperscript{-1} of dry weight) was determined as explained in section 2.3.1. Additionally, average CH\textsubscript{4} production (accumulation) over time was estimated. Methane production rates were calculated from a linear regression of three consecutive data points. The temperature coefficient (Q\textsubscript{10}) was calculated following Duc et al. (2010).
2.4. Quantitative PCR (qPCR) of the methyl-coenzyme reductase alpha subunit gene (mcrA)

Sediment subsamples were obtained from the first 30 cm of the 2011 cores Siq11-a, SukS11-a, and SukB11-a (Table 1). Subsamples that were stored at -80 °C were later transferred to sucrose lysis buffer to preserve the integrity of the nucleic acids (SLB, 40 mM EDTA, 50 mM Tris-HCl, 0.75 M sucrose) prior to nucleic acid extractions. Community genomic DNA from surface sediment subsamples from each lake was extracted using a modified protocol for the power soil DNA isolation kit (MoBio, Carlsbad, CA, USA). Samples in SLB were thawed on ice for 45 min and centrifuged at 10,000 x g for 10 min. The supernatant was removed; ~ 0.5 g sediment subsamples (for a total weight of ~ 1 g per sample) were extracted following manufacturer’s instructions, and later quantified using a standard picogreen assay (Life Technologies, Grand Island, NY, USA).

To quantify mcrA gene fragment copy number, qPCR was carried out using the ML primer pair (Luton et al., 2002) and the following conditions: 1X SYBR Green PCR master mix, 0.1 µM of each primer, 0.1 µg.µL⁻¹ of bovine serum albumin (BSA) in a 25 µL final volume. One µL of template DNA from Siq samples was used in 2-4 replicate reactions, and 4 µL of SukS or SukB samples was used in another set of 4 replicate reactions. A standard curve was prepared using Methanocaldococcus jannaschii genomic DNA with 1 µL of 10-fold dilutions covering 5 orders of magnitude (6.4 x 10⁶ to 6.4 x 10² copies of mcrA gene assuming 1 copy of mcrA per genome), 4 replicates each. qPCR was performed using an Applied Biosystems 7500
Fast system (Life Technologies, Grand Island, NY, USA) in standard mode, and following PCR conditions by Luton et al. (2002). PCR efficiency was 75.7% and amplification of standards was linear \(r^2 = 0.993\) from \(10^2\) to \(10^6\) copies of the template per \(\mu L\). *McrA* gene copy numbers were expressed per g of wet sediment and per ng DNA. To confirm amplification specificity a melt curve analysis was performed immediately after qPCR using standard instrument settings, and agarose gel electrophoresis was used to confirm expected amplicon size.

### 2.5. Organic carbon and lipid biomarkers analyses

#### 2.5.1. Sediment texture, total carbon, and inorganic carbon content

Sediment texture was determined following a simple texture analysis chart (Thien, 1979), with the same cores (Siq11-b, SukB11-b, and SukS11-b) used for the CH\(_4\) production experiment. Total carbon and total inorganic carbon content was determined from five cores: Siq10-a, SukS10-a, Siq11-a, SukS11-a, and SukB11-a.

Frozen cores were sectioned every 5 cm to a depth of 30 cm (2011 only) and every 10 cm for the remaining length of each core. These samples were collected into furnaced (550 °C, 8 hours) glass vials with foil lids that were refrozen (-20 °C) for transportation to the University of California, Riverside. Samples were lyophilized and then analyzed using an Eltra CS-500 carbon-sulfur analyzer, yielding total carbon (TC) and total inorganic carbon (TIC). Total organic carbon (TOC) was obtained by subtraction.
2.5.2. Lipid biomarker analyses

Lipid biomarkers were extracted from lyophilized samples obtained from the same cores sampled above using a Microwave Accelerated Reaction System (CEM Corp.) with 9:1 vol/vol dichloromethane: methanol to yield total lipid extracts (TLEs). One aliquot of each TLE was derivatized with N,O-bis(trimethylsilyl)trifluoroacetamide (BSFTA) in pyridine before gas chromatography-mass spectrometry (GC-MS). A second aliquot of total extract was fractionated via solid phase extraction columns to yield neutral lipids that were also subjected to GC-MS. Catalytic hydropyrolysis (HyPy; Love et al., 2005) was applied to an additional aliquot of freeze-dried sediment. HyPy conditions were 5 wt. % molybdenum sulfide catalyst, 150 bar H2 flowing at 5 L min\(^{-1}\), and programmed temperature ramp of 100 °C min\(^{-1}\) to 250 °C, followed by ramping at 8 °C min\(^{-1}\) to 480 °C. Gas chromatography-mass spectrometry (GC-MS) analyses of freely extractable and kerogen-bound saturated hydrocarbons, and derivatized total extracts were performed with an Agilent 5973 MSD mass spectrometer interfaced to an Agilent 7890A GC, equipped with a DB-1MS capillary column (60 m x 0.32 mm, 0.25μm film) and run with He as carrier gas. The temperature program for GC-MS full scan and selected ion monitoring was 60 °C (2 min), ramp to 250 °C at 20 °C min\(^{-1}\), to 325 °C at 2 °C min\(^{-1}\), and held at 325 °C for 20 min. Lipid biomarkers were identified by comparison with published mass spectra and retention times, and quantified using a d4-\(\text{C}_{24}\)-ethylcholestane internal standard. Archaeol was quantified using m/z 130 in selected ion monitoring mode with a d14-\(\text{C}_{13}\)-para-terphenyl internal standard (m/z 244) and calculated
response factor. The response factor was obtained by comparison of detector
response areas between replicate analyses of known amounts of archaeol and d14-
p-terphenyl.

3. Results

3.1. Sediment geochemistry

3.1.1. Oxygen profiles and depth integrated aerobic O$_2$ consumption

Dissolved O$_2$ concentration at the sediment-water interface in freshly collected
cores from Siq13, SukB13, and SukS13 was 212.4, 110.0, and 6.5µmol L$^{-1}$,
respectively (Fig. 2.2). Oxygen decreased rapidly with depth in all cores, becoming
depleted by 1.0, 10.0 and 0.5 mm in the same respective cores. The depth of O$_2$
depletion typically coincided with concentration at the sediment-water interface in
the Suk cores; cores from Siq did not fit this trend indicating that O$_2$ dynamics in Siq
were different from that in Suk. Based on the shape of the O$_2$ depletion profiles, and
assuming steady state conditions, average ($\pm$ SD) rates of metabolic O$_2$ consumption
were estimated to be 503.5-892.9, 40.8 ± 14.4, and 72.6 ± 19.1 µmol O$_2$ m$^{-2}$ h$^{-1}$ in
cores from Siq, SukB, and SukS respectively.

3.1.2. Pore water chemistry

Pore water chemistry from Siq12 sediment cores was generally reproduced in
replicate cores. The average concentration of NO$_3^-$ (76.62 ± 73.18 µM) and SO$_4^{2-}$
(8.74 ± 8.47 µM) in the upper 8 cm of the two replica cores was higher than the
average concentration of NO$_3^-$ (1.40 ± 2.51 µM) and SO$_4^{2-}$ (0.96 ± 0.70 µM) below 8
The concentration of total dissolved Fe in both cores was high throughout the depth profile, reaching a maximum (854.94 µM at 20 cm in Siq12-a) between 12 and 20 cm below the sediment-water interface. Additionally, the concentration of total dissolved Mn in the pore waters was about 44-fold lower than the concentration of total dissolved Fe, but showed a similar depth profile to Fe (Figs. 3D and 3G).

Overall, the concentration of NO$_3^-$ and SO$_4^{2-}$ in Suk was higher than the concentration of these anions in Siq. Reproducibility between cores at the same site was poor. Two maxima were detected in SukB13-b (2.34 mM at 10 cm and 745.73 µM at 20 cm).
μM at 18 cm) but the concentrations were much lower (0.00-27.47 μM) in the rest of the core (sampled to 30 cm), as well as in SukB13-d (0.87-52.58 μM; Fig. 2.3B). In SukS13-b increasing NO\textsubscript{3} was detected with increasing depth (0.00-126.25 μM), but in SukS13-c, NO\textsubscript{3} peaked at 11 cm (1.07 mM) and then decreased (Fig. 2.3C). The SO\textsubscript{4}\textsuperscript{2-} concentration increased with depth in SukB13-b (0.00 μM at 2 cm to 751.68 μM at 30 cm), and in SukB13-d there was an average of 275.27 ± 112.74 μM throughout the core (sampled to 17 cm) with two maxima of ~ 450 μM (at 5 cm and 11 cm; Fig. 2.3K). The SO\textsubscript{4}\textsuperscript{2-} concentration mostly increased with depth in the first ~ 20 cm (0.00-329.29 μM) of SukS13-b. A similar pattern was observed in SukS13-c down to ~ 17 cm (0.00-296.42 μM; Fig. 2.3L).

Depth profiles of dissolved Fe and Mn in SukB13 (Figs. 3E and 3H) were somewhat similar to Siq12. The concentration of dissolved Fe in SukB increased by 300-fold from 1 cm below the surface to 12 cm below the surface, and then decreased with depth; the concentration of dissolved Mn was relatively constant throughout the profile in SukB. The concentration of these metals throughout the sediment profile in SukS (Figs. 3F and 3I) showed a lack of a curve and was lower than in Siq (dissolved Fe ~ 8-fold and Mn ~ 3-fold) and SukB (dissolved Fe ~ 5-fold and Mn ~ 2-fold).

3.2. CH\textsubscript{4} concentrations and stable carbon isotope analyses

3.2.1. In situ CH\textsubscript{4} concentration and stable carbon isotopes
Average (± SD) CH$_4$ concentration was orders of magnitude higher in the upper intervals from Siq (2.18 ± 0.24 µmoles CH$_4$ g$^{-1}$ dry sediment) than in SukB (3.12 ± 2.30 x 10$^{-4}$ µmoles CH$_4$ g$^{-1}$ dry sediment) and SukS (3.21 ± 1.44 x 10$^{-3}$ µmoles CH$_4$ g$^{-1}$ dry sediment; Fig 4A). CH$_4$ concentration decreased precipitously with depth in Siq sediments, from 0.84 µmoles CH$_4$ g$^{-1}$ dry sediment at ~ 38 cm to 0.02 µmoles CH$_4$ g$^{-1}$ sediment at 88 cm. Conversely, CH$_4$ concentration remained relatively constant throughout the SukB sediment profile (7.00 ± 4.10 x 10$^{-4}$ µmoles CH$_4$ g$^{-1}$ sediment), while CH$_4$ concentration increased with depth in the SukS sediments, with the highest amount of CH$_4$ observed at 75 cm (0.11 µmoles CH$_4$ g$^{-1}$ dry sediment). The CH$_4$ level detected in the deepest interval of SukS core was 2 orders of magnitude higher than the amount of CH$_4$ detected at a similar depth in SukB.

Methane from Siq sediments was more $^{13}$C-depleted than in SukS sediments (Fig. 2.4B). The most negative $\delta^{13}$C$_{CH_4}$ values (-76.7 to -79.2 ± 0.2 $^{0}$/00) were detected in Siq surface sediment, between 8 and 38 cm. However, the signal became less depleted in $^{13}$C at 58 cm (-64.3 ± 1.6 $^{0}$/00), and even less depleted at 88 cm (-43.3 ± 6 $^{0}$/00). The $\delta^{13}$C$_{CH_4}$ values in SukS sediments were less depleted at 35 cm (-55.5 ± 1.4 $^{0}$/00) and at 55 cm (-47.3 ± 0.2 $^{0}$/00) than those for Siq. The least depleted $\delta^{13}$C$_{CH_4}$ value in SukS was observed at 75 cm (-43.4 ± 0.2 $^{0}$/00). The surface CH$_4$ concentrations for SukS sediments along with the entire depth profile in SukB were insufficient for the analysis of isotopes.
3.2.2. CH4 production experiment

Biological CH4 production was observed in sediments from Siq11 at both temperatures (2 and 10 °C), and the upper sediments from SukB11 at 10 °C (Fig. 2.5). The highest amount of CH4 produced was 7.4 ± 1.2 µmoles CH4 g⁻¹ dry weight (n=3) in the 2-20 cm interval of Siq sediments at 10 °C, and 3.7 ± 0.0 µmoles CH4 g⁻¹ dry weight (n=2) at 2 °C after 25 days of incubation (Fig. 2.5A). The next deeper interval of Siq sediment (26-47 cm) produced 0.9 ± 0.2 µmoles CH4 g⁻¹ dry weight (n=6) at 10 °C, and 0.2 ± 0.0 µmoles CH4 g⁻¹ dry weight (n=6) at 2 °C (Fig. 2.5B). CH4 production from the 2-30 cm SukB-11 sediment interval was not detected until day 10 of incubation at 10 °C. After 50 days of incubation, 0.3 ± 0.1 µmoles CH4 g⁻¹ dry weight (n=6) were produced at 10 °C and < 0.1 µmoles CH4 g⁻¹ dry weight (n=6) were produced at 2 °C in this SukB upper interval (Fig. 2.5D).

Upper Siq and SukB sediments produced more CH4 than any other interval of sediment sampled. CH4 production < 0.1 µmoles CH4 g⁻¹ dry weight was observed in the deepest interval from Siq (49-76 cm; Fig. 2.5C), and no CH4 production was observed in the deeper intervals from SukB (34-62 cm and 64-90 cm) during the incubation period (Figs. 5E and 5F). Furthermore, no CH4 production was observed from the SukS sediments, although very small amounts of CH4 were detected at all depths sampled.
Figure 2.4 Methane concentration, $\delta^{13}C_{CH_4}$ and $\delta^{13}C_{CO_2}$ in arctic thermokarst lakes. (A) $CH_4$ concentration per gram of dry sediment (Siq11-a, SukB11-a, and SukS11-a). (B) $\delta^{13}C_{CH_4}$ (Siq11-a and SukS11-a) and $\delta^{13}C_{CO_2}$ (Siq11-a) in pore waters. The approximate boundary between biogenic and thermogenic $CH_4$ (Whiticar, 1999) is indicated with a dashed line for visual convenience. The error for the average $CH_4$ concentration (analytical replicates) was ± 2 %. The error for the average $\delta^{13}C_{CH_4}$ (analytical replicates) was ± 0.2 °/oo, except at 58 cm (± 1.6 °/oo) and 88 cm (± 6.0 °/oo) in Siq, and at 55 cm (± 1.4 °/oo) in SukS. For comparison to other data sets, $CH_4$ concentration in pore waters as µM was estimated in Section 4.5 (main text).

Temperature had a distinct influence on the rate of $CH_4$ production (obtained as the slope of a linear regression of 3 consecutive data points). $CH_4$ was produced
at a rate of 2.2 µmoles CH₄ day⁻¹ g⁻¹ dry weight at 10 °C and at a rate of 0.8 µmoles CH₄ day⁻¹ g⁻¹ dry weight at 2 °C in the upper Siq sediments. In the next deeper interval of Siq sediment, CH₄ was produced at a much lower rate: 0.3 µmoles CH₄ day⁻¹ g⁻¹ dry weight at 10 °C and < 0.1 µmoles CH₄ day⁻¹ g⁻¹ dry weight at 2 °C. The temperature coefficient (Q₁₀) in Siq was 3.7 for the upper sediment interval, and 9.5 for the next interval down indicating different metabolic responses of the microbial community through the sediment core. The rate of CH₄ production in the upper sediments of SukB could only be estimated at 10 °C (0.9 µmoles CH₄ day⁻¹ g⁻¹ dry weight), given that CH₄ did not show a linear increase over time at 2 °C, hence, no Q₁₀ value was computed.

### 3.3. Proxies for Methanogen abundance

The mcrA gene was detected in the surface sediments from the 3 sites by qPCR amplification, but only Siq11 gene copy numbers were at or above a conservative limit of detection that was established for this assay (100 copies mcrA µL⁻¹). The highest mcrA gene copy number detected was 1.9 x 10⁴ mcrA copies g⁻¹ sediment (5.4 x 10² mcrA copies ng⁻¹ DNA) for Siq samples between ~ 14-15 cm below the surface, while the copy number between ~ 6-7 cm was 6.4 x 10³ mcrA copies g⁻¹ sediment (1.3 x 10² mcrA copies ng⁻¹ DNA), and between 23-24 cm was 5.0 x 10³ mcrA copies g⁻¹ sediment (4.7 x 10² mcrA copies ng⁻¹ DNA).
CH$_4$ production from sediment incubations at 2 temperatures using *in situ* organic matter. CH$_4$ production at 10 °C (black) and at 2 °C (white) at 3 different depths. Siq11-b (A) 2-20 cm, (B) 26-47, and (C) 49-76 cm. SukB11-b (D) 2-30 cm, (E) 34-62 cm, and (F) 64-90 cm. Note the different scales. CH$_4$ production was an average of 2 to 6 replica samples (see methods). Soil texture from the upper and lower layer of each sediment depth is described in each panel. CH$_4$ was detected in some incubations from SukS, but no pattern of CH$_4$ production was observed.

Archaeol was highest in Siq where two maxima were encountered: 45 and 37 µg g$^{-1}$ sediment at ~12.5 cm and ~17.5 cm, but was lower outside these intervals.
(Fig. 2.6B). Absolute archaeol concentration was greater in the distal seep site, SukB, beginning at 0.1 µg g\(^{-1}\) sediment at 7.0 cm, reaching a maximum of 1.4 µg g\(^{-1}\) sediment at ~ 26.5 cm, and then below detection for the lowermost portion of the studied samples. In the SukS11 core, the absolute archaeol concentration was slightly elevated in the upper ~ 23.5 cm but consistently low (≤ 0.07 µg g\(^{-1}\) sediment) throughout the entire 1 m core (2-3 orders of magnitude lower than at Siq and SukB). The archaeol to n-C\(_{31}\) ratio (Fig. 2.6C) was consistently higher in Siq (average 0.10 ± 0.05) and SukB (average 0.08 ± 0.06) than in the SukS site (0.01 ± 0.01).

3.4. Sediment properties, organic matter content and composition

Total organic carbon (TOC) measurements varied widely among and within cores analyzed in the three sites (Fig. 2.6A). TOC was highest in the upper 40 cm of Siq11 sediment (avg. 14.9 wt. %), then decreased to an average of 2.1 wt. % at depths greater than 40 cm below the sediment-water interface. The Siq11 sediment core had an overall average of 7.2 ± 2.9 wt. % TOC. SukB11 exhibited relatively low and consistent carbon contents down core (overall average 1.9 ± 3.7 wt. %), with the exception of a high TOC interval around 23 cm sediment depth (12.9 wt. %). Excluding this interval, the SukB11 core had an average of 0.9 ± 0.5 wt. % TOC. SukS sediments contained the lowest amounts of organic carbon, averaging 0.5 ± 0.3 wt. % over the entire SukS11 sediment core. Sediment TOC profiles from the 2011 samples reported here were consistent with those from a similar sample set collected in 2010 at the same locations (Table S2). Sediment texture analyses of
2011 sediment cores indicated that SukS and Siq have similar grain size profiles, with a dominance of clay and silt (90-95%) in the upper portion of the core, and 60-80% sand in the lower portions. SukB has the opposite profile, with >70% sand in the uppermost portion, and 90% silt and clay in the remainder of the core. SukB sediments contained carbonate, which was most likely detrital in origin.

Organic matter sources and composition were evaluated through analysis of saturated hydrocarbons, BSTFA-derivatized total lipid extracts, and catalytic hydropyrolysis products (functionalized free- and kerogen-bound lipids that have been converted to hydrocarbons). The proportions and amounts of these compounds differed from site to site and with depth at a given site (Tables S2 and S3). Both the extractable and the kerogen-bound hydrocarbons in Siq10 sediments were dominated by intermediate chain length \( n \)-alkanes. The most abundant \( n \)-alkane in free hydrocarbons was \( n \)-C\(_{23} \) and in the kerogen-bound fraction was \( n \)-C\(_{24} \). The proportion of \( n \)-alkanes derived from aquatic plants (\( P_{aq} \) values, average 0.77 for the free hydrocarbons) was consistent with this intermediate chain length. These samples also contained low short to long \( n \)-alkane ratios, and relatively low sterane to hopane ratios (free sterane to hopane average 0.03, kerogen-bound sterane to hopane average 2.27). SukS10 organic matter showed a pronounced difference between free hydrocarbons and kerogen-bound hydrocarbons. The most abundant \( n \)-alkane in the free fraction was \( n \)-C\(_{31} \), but the most abundant \( n \)-alkane in the bound-fraction was \( n \)-C\(_{16} \). The short to long \( n \)-alkane ratio average was 0.02 in the free fraction, and the short to long \( n \)-alkane ratio average in the bound fraction
was 7.35. Free \( n \)-alkanes also had lower \( P_{aq} \) values (average 0.18) in this lake. Carbon preference indices (CPI) for both lakes showed strong odd-over-even predominance in free \( n \)-alkanes (from decarboxylation of free fatty acids) and even-over-odd in bound \( n \)-alkanes (from reduction of functionalized lipids).

4. Discussion

Methane emissions have received extensive attention in numerous environments (Wagner et al., 2007; Liu et al., 2013; Negandhi et al., 2013), where as much as 80-90% of the atmospheric \( \text{CH}_4 \) is microbially derived (Whiticar, 1999). Furthermore, atmospheric \( \text{CH}_4 \) is a potent greenhouse gas that is currently rising (Hoehler & Alperin, 2014). Considering that the Arctic is highly sensitive to climate change (Kittel et al., 2011), accurate estimates of \( \text{CH}_4 \) emissions are of the utmost importance for the global \( \text{CH}_4 \) budget. Methane has also been recognized as a biosignature for life beyond Earth (Tazaz et al., 2013). In this section, we discuss the implications of our findings for North American Arctic \( \text{CH}_4 \) budgets.

4.1. \( \text{CH}_4 \) sources and sediment biogeochemistry in Siqulukaq Lake

4.1.1. Sediment \( \text{CH}_4 \) profiles

Detection of \( \text{CH}_4 \) (highest between 8-18 cm below the surface) in Siq was consistent with rapid \( \text{O}_2 \) depletion (\( \sim 1 \) mm, Fig. 2.2) and pore water biogeochemistry (Fig. 2.3).
Figure 2.6 October 2011 sediment depth profiles of (A) total organic carbon, (B) archaeol concentration, and (C) archaeol to \( n-C_{31} \) ratios. Note differences in archaeol scales among sites.

The \( O_2 \) concentration in the water-sediment interface was within the range of concentrations detected in other shallow arctic lakes (Whalen et al., 2013), but \( O_2 \) was depleted in Siq at a shallower depth. High amounts of surficial TOC (Fig. 2.6) in combination with a silty clay loam, suggested that Siq sediments would become anoxic at shallow sediment depths. In fact, within the first 1 mm in the sediment profile \( O_2 \) was consumed at a rate higher than that computed in the other sediment cores we studied. This rate is at the low end of those observed in a eutrophic lake (Lake Zug, Switzerland; Maerki et al., 2009), and it is similar to the rate measured in a meso-eutrophic lake (Lake Vechten, The Netherlands; Sweerts et al., 1991). Moreover, \( SO_4^{2-} \) and \( NO_3^- \) were consumed within the first 10 cm below the
sediment-water interface, while dissolved Fe and Mn were chemically reduced, indicating thermodynamic conditions favorable for methanogenesis deeper in the profile. At depths > 48 cm, CH₄ concentration decreased in concert with TOC values. Overall, a positive correlation was found between the in situ CH₄ concentration and the TOC content in the sediments ($r^2 = 0.80$, $p = 0.008$), partly explaining the decrease in biological CH₄ production with depth.

### 4.1.2. $\delta^{13}$C<sub>CH₄</sub>

The stable isotope signature of C in CH₄ ($\delta^{13}$C<sub>CH₄</sub>) from Siq sediment suggested a biogenic source (Fig. 2.4 and Fig. S2). The most depleted $\delta^{13}$C<sub>CH₄</sub> values (-79.2 to -57.6 ± 0.2 ₀/₀₀) fell within the range of $\delta^{13}$C<sub>CH₄</sub> values recorded in the literature for biogenic CH₄ production in other arctic studies (Quay et al., 1988; Walter et al., 2008). Only the deepest sediment layer (~ 88 cm) in Siq showed a borderline thermogenic (c.f. Walter Anthony et al., 2012) signal. At this depth, however, substrate depletion due to extensive organic matter decomposition (Pedersen et al., 2011) could result in biogenic CH₄ with heavier $\delta^{13}$C<sub>CH₄</sub> values (Whiticar, 1999).

Biological CH₄ production pathways may be inferred by estimating the isotope separation factor ($\varepsilon_C$) between $\delta^{13}$C<sub>CO₂</sub> and $\delta^{13}$C<sub>CH₄</sub> and the apparent C fractionation factor ($\alpha_C$) (Whiticar, 1999). For the upper 38 cm of Siq sediments, we calculated $\varepsilon_C$ between 60.4 and 67.1, and $\alpha_C$ between 1.065 and 1.073. This corresponds to CH₄ production by CO₂ reduction (Whiticar, 1999), and is in accordance to the values observed by Walter et al. (2008) for ebullient CH₄ sources.
in Siberian lakes. Alternative pathways, such as acetate fermentation, have been found to occur in freshwater sediments that are rich in organic carbon. Although acetate was present in the first few centimeters of Siq sediment (0.3-0.6 µM), it is possible that other forms of anaerobic respiration outcompeted methanogenesis in the use of acetate, leaving CO₂ reduction as the dominant pathway. For instance, the similarity between the dissolved Fe (Fig. 2.3D) and CH₄ (Fig. 2.4A) profiles is not currently understood in these lakes, though could be explained by a syntrophic relationship between iron reducing bacteria and methanogens (Zhou et al., 2014), in which iron-reducing bacteria oxidize acetate to CO₂, and the CO₂ is reduced by methanogenic Archaea to CH₄ (via the hydrogenotrophic pathway). Despite the little we know about iron reduction in the sediments of these lakes, this process has been connected to ecosystem respiration in drained-lake basins on the coastal plain of Alaska (Lipson et al., 2013).

4.1.3. CH₄ production

Methane was biologically produced from organic matter present in the first ~ 47 cm of Siq11 sediment at 2 °C and at 10 °C. Coincidentally, the amounts of CH₄ produced at 2 °C were similar to pore water CH₄ levels, indicating that CH₄ present in the sediments was most likely derived from in situ, present-day methanogenesis. The rate of CH₄ production at 10 °C in the upper sediments of Siq was in the range of the CH₄ production rates reported for the 0.4-9.0 cm sediment depth of three shallow (4.1-6.7 m) arctic lakes in the Alaskan Foothill (Bretz & Whalen, 2014).
These results also suggest that CH$_4$ was produced in a temperature-dependent fashion. This kind of temperature dependence is expected for biological reactions where enzymes are involved (Hochachka & Somero, 1973), and corroborates the temperature dependence of methanogenesis at the microbial community-level recently shown by Yvon-Durocher et al. (2014). As observed in surface (2-20 cm) Siq sediment, a 10 °C temperature rise can be correlated to a 2.5 to 3.5-fold increase in CH$_4$ production (Conrad & Schutz, 1988); however the 3-fold increase in the rate of CH$_4$ production in the next deeper (26-47 cm) Siq interval, supports the idea that decomposing structurally complex, aromatic molecules requires higher activation energies, causing enzymatic reactions to be more sensitive to temperature (Mikan et al., 2002; Davidson & Janssens, 2006). Below the methanogenic zone in Siq sediment, older, more recalcitrant organic matter may be found.

4.2. CH$_4$ sources and sediment biogeochemistry in Sukok Lake

4.2.1. Sediment CH$_4$ profiles

CH$_4$ was not retained in the sediments of Sukok Lake. The low CH$_4$ concentrations in both SukB and SukS sediments (Fig. 2.4) were surprising, considering the proximity of the SukS sediments to an ebullient gas seep (Fig. 2.1). SukS sediments transitioned from silty clay at the upper intervals, to a combination of sandy clay loam and loamy sand in the deeper intervals, perhaps reflecting current- or ebullition-induced winnowing that facilitated CH$_4$ channeling through the seep. Additionally, lower TOC contents in Suk compared to Siq (by a factor of ~ 4) may
have indirectly affected O\textsubscript{2} levels in the sediments, supporting lower O\textsubscript{2} consumption rates in SukB and in SukS versus Siq. TOC concentration was slightly higher in SukS than in SukB, and O\textsubscript{2} concentration was lower at the SukS water-sediment interface, but comparable in magnitude to another arctic lake in the Alaskan Foothill (Bretz & Whalen, 2014). The lower TOC content at both Suk sites may also explain the higher concentrations of alternative electron acceptors (NO\textsubscript{3}\textsuperscript{-} and SO\textsubscript{4}\textsuperscript{2-}) in the sediments of this lake, and establishes the conditions for other biogeochemical transformations to take place in the upper layers of SukB and SukS (\textit{i.e.}, anaerobic methane oxidation).

4.2.2. \(\delta^{13}\text{C}_{\text{CH}_4}\)

Physical mixing of CH\textsubscript{4} from different sources (\textit{e.g.} microbial and thermogenic) could explain the “transitional” isotope signature observed in SukS. The primary source of CH\textsubscript{4} in the deeper sediment layers was consistent with thermogenic CH\textsubscript{4}, though there may also be some biological CH\textsubscript{4} production in discrete layers of these sediments. Considering that Sukok Lake is located in a gas field, and that the sample from SukS was collected nearby an active CH\textsubscript{4} seep, a thermogenic CH\textsubscript{4} signal is plausible. Additional measures of C\textsubscript{2}-C\textsubscript{4} hydrocarbon ratios or \(\delta\text{D-CH}_4\) could be used to further substantiate this observation. Also, methanogenesis could be thermodynamically constrained by high CH\textsubscript{4} partial pressures (like the ones observed at the nearby CH\textsubscript{4} seep) in combination with low concentrations of methanogenic substrates (Chong et al., 2002), or by high redox potentials. Data from the CH\textsubscript{4} production experiment with SukS samples however ruled out the possibility
of thermodynamic inhibition imposed by CH$_4$ itself. Note that the $\delta^{13}$C$_{CH4}$ for SukB samples could not be determined due to insufficient CH$_4$ levels.

4.2.3. CH$_4$ production

Biological CH$_4$ production was not observed from any SukS11 interval at 2 or 10 °C, which is consistent with the isotopically heavy $\delta^{13}$C$_{CH4}$ in SukS11. However, CH$_4$ production at 10 °C was detected in the upper SukB11 sediments at a rate 23-fold lower than Siq11. Temperature had a weaker influence on CH$_4$ production in Suk than in Siq. According to Davidson and Janssens (2006), when substrate is abundant, temperature increases affect the maximum reaction rate ($V_{max}$), but under substrate-limiting conditions the substrate concentration at which the reaction rate equals $V_{max} [2 (K_m)]^{-1}$ also increases with temperature, leading to a lower apparent temperature dependence. Therefore, the lag phase of CH$_4$ production observed in SukB surface sediments, may be consistent with the observation that organic matter in Suk has a higher vascular plant input and thus, may not only be less abundant, but also less labile than in Siq.

The fact that there was more TOC in the surface sediments of SukS than in SukB, but there was no CH$_4$ production in SukS at 10 °C indicates that biological CH$_4$ production in SukS sediments may be subject to other environmental constraints. Alternatively, longer incubation times or substrate additions could lead to CH$_4$ production from SukS, as reported in other ecosystems (e.g. Tibetan plateau lakes; Liu et al., 2013).
4.3. **Proxies for Methanogen abundance as a control on CH$_4$ concentration in thermokarst lake sediments**

Copy numbers of the *mcrA* gene in lake sediments were used to study the relationship between methanogen biomass and CH$_4$ concentration in sediment pore waters (Colwell et al., 2008; Freitag & Prosser, 2009; Liu et al., 2011). Assuming the presence of 1 copy of *mcrA* gene per genome of methanogenic Archaea, the copy number of the *mcrA* gene may be proportional to the number of cells with potential for CH$_4$ production (Luton et al., 2002; Steinberg & Regan, 2008) or consumption, because *mcrA* genes are also found in anaerobic CH$_4$ oxidizers (Raghoebarsing et al., 2006; Beal et al., 2009; Ettwig et al., 2010).

In Siq surface sediment the *mcrA* gene copy numbers were low but detectable, while *mcrA* copy numbers in Suk were below the limits of detection established for the assay at both sites. This pattern may at least partly explain the observed difference in CH$_4$ concentration between the 2 lakes and is a good approximation of relative differences between the lakes. In comparison to *mcrA* copy numbers found in a sample of active layer (permafrost) from the Canadian High Arctic (Yergeau et al., 2010), the *mcrA* copy numbers detected in Siq11 were 1 to 2 orders of magnitude higher. Also, our results are quite comparable to two active layer samples from the Western Canadian Arctic, where *mcrA* copies g$^{-1}$ wet soil were between $10^3$ and $10^6$ (Frank-Fahle et al., 2014), though our highest copy number was $1.86 \times 10^4$ *mcrA* copies g$^{-1}$ sediment in Siq.
Methanogen/archaeal biomass was corroborated by the detection of archaeol in the lake sediments (McCartney et al. 2013). Archaeol is an isoprenoid membrane lipid produced by Archaea that has been attributed in freshwater sediments and peats to methanogenic Archaea, (Pancost et al., 2011). The presence of archaeol may reflect active shallow sedimentary Archaea (Parkes et al., 2007), or preserved dead microbial biomass (Pancost et al., 2011; Bischoff et al., 2013). In either case, not only was the concentration of archaeol much higher in Siq than in SukB and in SukS, but it also tracked the trends of mcrA copy numbers per gram of sediment, amounts of TOC, and CH₄ concentrations observed in the depth profile of Siq, supporting the use of this compound as a proxy for methanogenic Archaea in lacustrine environments. Compared to the concentrations of archaeol detected in an area of continuous permafrost in Siberia (Bischoff et al., 2013), where CH₄ has been measured in situ and in incubations at 10 °C (Table S1), archaeol concentrations detected in the upper intervals of Siq were three orders of magnitude higher, and the concentrations detected in Suk were within the range of concentrations or slightly higher.

Archaeol abundance was also normalized to the C₃₁ n-alkane abundance, a lipid likely to derive primarily from diagenesis of land plant leaf waxes, to correct for terrigenous inputs and to serve as a means of assessing which lake had the highest in situ archaeol production. Archaeol to n-C₃₁ ratio was consistently higher in Siq and SukB than in the SukS site, indicating higher contributions from lake sedimentary Archaea, above baseline values transported in by soil lipid inputs.
4.4. Amount of organic matter, sources, and its relevance to CH$_4$ production

Within and among the studied thermokarst lakes, organic matter content and composition was heterogeneous. Unlike Siberian permafrost where the main source of organic matter is from the Pleistocene (Zimov et al., 1997), lacustrine sedimentary organic matter in these lakes appears to have a contribution from ongoing in situ primary production, as observed in other lakes of the North Slope of Alaska (Bretz & Whalen, 2014). Although ubiquitous in the environment and of moderately low source specificity, $n$-alkane chain length has been shown to differ between vascular plant waxes (typically odd carbon number $n$-alkanes greater than C$_{22}$; Killops & Killops, 2005) and microbial lipids (approximately C$_{16}$-C$_{24}$; reviewed in Meyers & Ishiwatari, 1993). Additionally, $n$-alkane chain length may reflect the proportion of submerged versus emergent or terrigenous macrophytes, in which submerged macrophytes produce larger proportions of shorter (C$_{23}$, C$_{25}$) $n$-alkanes (Ficken et al., 2000).

The dominance of intermediate chain length $n$-alkanes in Siq sediments indicated a substantial organic contribution from a mixture of aquatic microbial lipids (bacteria) and algae, and terrestrial plants and soils. Siq P$_{aq}$ values were consistent with a source from submerged macrophytes or microbial primary producers. Sterane to hopane ratios reflected the relative contributions of plants and algae versus hopanoid-producing bacteria, with a larger proportion of eukaryotic material present in the bound fraction.
Suk organic matter showed a greater contribution from allochthonous, perhaps more resistant or reworked organic matter. Free hydrocarbons in SukS were dominated by terrigenous leaf waxes (larger amounts of long-chain n-alkanes with high CPI; Tables S2 and S3) and trace amounts of mature diastereoisomers of steranes and hopanes, indicating a small contribution from petroleum at this site. 

$P_{aq}$ values for the free n-alkanes in SukS sediments also reflected supply from terrigenous plants including bryophytes (e.g., *Pogonatum* sp. which produces low amounts of C$_{25}$, C$_{27}$, and C$_{23}$ n-alkanes; Haas, 1982), while the longer chain n-alkanes more likely derived from graminoids (e.g., *Arctophila fulva*, *Carex* spp. and *Eriophorum* spp. which produce predominantly C$_{27}$, C$_{31}$, and C$_{29}$ n-alkanes; Oros et al., 2006; Ronkainen et al., 2013). However, high sterane to hopane ratios (with C$_{29}$ steranes the most abundant) and elevated short-chain n-alkanes in the kerogen-bound hydrocarbons showed significant inputs of likely microbially-derived n-alkanes. These findings from bound organics from the kerogen phase most likely reflect additional, though proportionately less abundant inputs from C$_{29}$ steroid-producing algae or macrophytes (e.g. chlorophytes) in comparison to Siq.

Microbial CH$_4$ production rates in lake sediments can be controlled by the amount of dissolved or total organic carbon (Kelly & Chynoweth, 1981; Avery et al., 2003; Bergman et al., 2000; Liu et al., 2011). In this study, organic matter composition and proxies for methanogen abundance also seemed to influence CH$_4$ production. For instance, the maximum amount of organic carbon observed in SukB sediments was comparable to that of the methanogenic zone in Siq, however organic
carbon in Suk was more recalcitrant than in Siq, and archaeol concentrations in Suk were much lower, perhaps explaining the low to non-existent CH$_4$ production in this lake. Moreover, in Siq sediments TOC in excess of 10 wt. % extended to 42 cm, but archaeol concentrations diminished below ~ 22 cm, suggesting depletion of readily available fermentation products with depth as well. TOC content may not be used as sole predictor of short term CH$_4$ production in these lakes; the lability of sedimentary organic matter influences the amount and composition of substrates ultimately available for methanogenesis.

4.5. **Implications for CH$_4$ production from permafrost in the North American Arctic**

In comparison to permafrost samples from other arctic environments, the amount of CH$_4$ produced in incubations of Siq and SukB sediments was 2 to 3 orders of magnitude higher; while the amount of CH$_4$ accumulated in the sediments of the lakes was within the range of CH$_4$ concentrations detected in other arctic locations (Table S1). Note that we have not taken seasonal or inter-annual variation in this study (beyond dissolved CH$_4$ concentrations and the carbon isotopes), which may introduce some degree of uncertainty upon our observations. Assuming an average porosity of 0.28 ± 0.07, an average headspace volume of 0.009 ± 0.001 L, and an average sediment volume of 0.003 ± 0.001 L, CH$_4$ concentrations detected in Siq11 ranged between ~121.40 µM and ~4.89 mM. Conversely, CH$_4$ concentrations in SukB11 ranged between ~0.75 µM and ~5.80 µM, and CH$_4$ concentrations in SukS11 ranged between ~12.62 µM and ~405.12 µM. These concentrations are comparable
to CH$_4$ concentrations detected 60 miles to the south in Qalluuraq Lake, which has active gas seeps, although the concentrations in the first ~ 30 cm sediments of Siq11 were ~ 2.5 X higher than in Qalluuraq Lake (He et al., 2012). Moreover, CH$_4$ concentrations in Siq exceeded by approximately an order of magnitude the maximum CH$_4$ concentration detected in another shallow (6 m) arctic lake (GTH 112) in the Alaskan Foothills, which displayed a similar trend in the pore water CH$_4$ profile (Bretz & Whalen, 2014).

Our study demonstrated that methanogenic Archaea present in Alaska’s North Slope thermokarst lakes are able to use in situ substrates for methanogenesis in a temperature-dependent fashion, and that the amount of CH$_4$ produced is proportional to the mcrA copy number, the concentration of archaeol, and the amount of labile organic matter in the sediments. These findings are particularly important when considering possible scenarios of climate change (Yvon-Durocher et al., 2014). The effect of increasing the temperature by 8 °C (from 2 to 10 °C) on CH$_4$ production rates was substantial for Siq. Currently, the largest source of CH$_4$ in this region of the North Slope of Alaska is release of thermogenic CH$_4$ (Walter Anthony et al., 2012). In scenarios of warming climate, our data lead us to contend that biological CH$_4$ production may play a larger role in CH$_4$ emissions in the future, though here, we have not considered CH$_4$ sinks (e.g. methanotrophy), which will also likely respond in parallel with temperature (Lofton et al., 2014).

A point of caution is that this study focused on the interior of the lakes, although it is possible CH$_4$ production varies throughout the lakes, and may be...
strongest at organic-rich thermokarst lake margins. Therefore, future estimates of CH₄ emissions should comprise spatial characterization and include the organic-rich shelf area, to be an adequate predictor of CH₄ release from Alaskan thermokarst lakes. CH₄ emission estimates would also benefit from including annual components of the production cycle to account for the impact of temperature shifts. With these cautions in mind, this study constitutes an important first step in determining the contribution of biogenic CH₄ to CH₄ budgets in the changing Alaskan arctic environment in proximal, yet contrasting thermokarst lake ecosystems.

Acknowledgements

We especially thank A. Klesh, J. Leichty, and P. Santibañez, for assistance in the field; Katey Walter-Anthony for sharing her observations about the study area; and Frank Löffler for analyses of pore waters in the University of Tennessee. We are very grateful to N. Riedinger, J. Memmott, G. Miller, E. Ulrich, M. Miller, G. Trubl, J. Dodsworth, and J. Qualls for invaluable technical support. Likewise, we appreciate the efforts of the Barrow Arctic Science Consortium (BASC) and the UMIAQ Corporation in Barrow, AK for providing logistical support and insight into the local region. Special thanks to the anonymous reviewers of the manuscript for their comments and suggestions, and to Life Technologies for use of the Applied Biosystems 7500 Fast system to conduct qPCR. PMC was supported in part by the Division of Earth and Ecosystem Sciences, DRI. Funding for ABM was provided in part by NSF IGERT Program in Geobiological Systems (DGE 0654336). KPH and DB
acknowledge support through the Jet Propulsion Laboratory (JPL), California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA). Financial support for this work was provided in part by the NASA Astrobiology Institute, Astrobiology of Icy Worlds program at JPL, and a NASA Astrobiology Science and Technology for Exploring Planets (ASTEP) award (Project Narvak, NNN13D036T). Support from these programs is gratefully acknowledged.

References


biodiversity of methane cycling in subsurface marine sediments (Skagerrak, Denmark). *Environmental Microbiology*, 9, 1146-1161.


eutrophic lake (Lake Vechten, the Netherlands). *Limnology and Oceanography*, **36**, 1124-1133.


### Supplementary Tables and Figures

#### Table S2.1 CH₄ concentration and CH₄ production rates reported in the literature for arctic permafrost samples.

<table>
<thead>
<tr>
<th>Location</th>
<th>Type of sample</th>
<th>Depth (cm)</th>
<th>% Total Organic Carbon (mg g⁻¹)</th>
<th>CH₄ (µmol g⁻¹)</th>
<th>CH₄ production rate (µmoles CH₄·g⁻¹·day⁻¹)</th>
<th>Annotated citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lena Delta, Siberia</td>
<td>Holocene permafrost deposits</td>
<td>0-150</td>
<td>4.2 % (42)</td>
<td>0.282</td>
<td>&lt; ~ 0.0036 at 5 °C</td>
<td>Wagner et al. (2007). Sampled in the summer of 2001. TOC is the median for the specific time unit (Unit I).</td>
</tr>
<tr>
<td>Lena Delta, Siberia</td>
<td>Low-centered ice-wedge permafrost polygon</td>
<td>0-100</td>
<td>&lt; ~ 4-18 % (40-180)</td>
<td>0.004-3.600</td>
<td>&lt; ~ 0.0022 at 10 °C</td>
<td>Bischoff et al. (2013). Sampled in the summer of 2002.</td>
</tr>
<tr>
<td>Laptev Sea coast</td>
<td>Polygon center, Samoylov Island</td>
<td>0-40</td>
<td>3.6-18.3 % (36-183)</td>
<td>0.150-542.0</td>
<td>&lt; ~ 0.0144 at 5 °C</td>
<td>Ganzert et al. (2007).</td>
</tr>
<tr>
<td></td>
<td>Floodplain, Samoylov Island</td>
<td>0-52</td>
<td>0.8-3.1 % (8-31)</td>
<td>0.004-0.411</td>
<td>≤ ~ 0.0144</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Submarine permafrost sediments</td>
<td>40.5-63.4 (mbsf)</td>
<td>0.03-87 % (0.3-87)</td>
<td>0.000-0.284</td>
<td></td>
<td>Koch et al. (2009). Sampled in the spring of 2005.</td>
</tr>
<tr>
<td>Cold foot, Alaska</td>
<td>Permafrost-protected soil</td>
<td>35-80</td>
<td>0.063 ± 0.041-0.087 ± 0.07</td>
<td>~ 0.001-0.010 at 5 °C</td>
<td></td>
<td>Waldrop et al. (2010). Sampled in the spring of 2007. Location North of the Arctic Circle at 67° N, 150° W. Organic-mineral active layer between 35-50 cm, and continuous mineral permafrost between 65-80 cm.</td>
</tr>
<tr>
<td>Hess Creek, Alaska</td>
<td>Permafrost-protected soils</td>
<td>35-100</td>
<td>0.46 ± 0.031-0.72 ± 0.28</td>
<td>~ 0.010-0.100 at 5 °C</td>
<td></td>
<td>Waldrop et al. (2010). Lowland south of Yukon River, below Arctic Circle at 65° N, 149° W. Samples composed of organic active layer and organic permafrost.</td>
</tr>
<tr>
<td>Herschel Island, Canadian Arctic</td>
<td>Drained lake, low centered polygon</td>
<td>0-35</td>
<td>23-28 % (230-280)</td>
<td>0.0336 -0.924 at 10 °C</td>
<td></td>
<td>Barbier et al. (2012). Samples collected in the summer of 2010. Location at 69° N, 138° W. Active layer (0-36 cm). No CH₄ production was detected between 0-5 cm.</td>
</tr>
</tbody>
</table>
Table S2.2 Free hydrocarbon biomarker extracted from Siqlukaq and Sukok sediments.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Depth (cm)</th>
<th>TOC (wt. %)</th>
<th>Sterane: Hopane</th>
<th>Short:Long n-alkanes</th>
<th>ACL (odd)</th>
<th>Most abundant n-alkane</th>
<th>CPI</th>
<th>Paq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siq10-a</td>
<td>6.5</td>
<td>14.23</td>
<td>0.01</td>
<td>0.13</td>
<td>27</td>
<td>23</td>
<td>3.06</td>
<td>0.77</td>
</tr>
<tr>
<td></td>
<td>21.5</td>
<td>4.84</td>
<td>0.02</td>
<td>0.04</td>
<td>27</td>
<td>23</td>
<td>2.45</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>38.0</td>
<td>0.89</td>
<td>0.05</td>
<td>0.01</td>
<td>27</td>
<td>27</td>
<td>3.33</td>
<td>0.72</td>
</tr>
<tr>
<td>SukS10-a</td>
<td>5.0</td>
<td>1.03</td>
<td>0.29</td>
<td>0.01</td>
<td>29</td>
<td>31</td>
<td>9.94</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>15.0</td>
<td>1.17</td>
<td>0.27</td>
<td>0.01</td>
<td>29</td>
<td>31</td>
<td>13.31</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>25.0</td>
<td>0.97</td>
<td>0.30</td>
<td>0.02</td>
<td>30</td>
<td>31</td>
<td>13.41</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>35.0</td>
<td>1.11</td>
<td>0.43</td>
<td>0.01</td>
<td>29</td>
<td>31</td>
<td>12.84</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>45.0</td>
<td>1.06</td>
<td>0.34</td>
<td>0.03</td>
<td>30</td>
<td>31</td>
<td>11.24</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>55.0</td>
<td>0.77</td>
<td>0.36</td>
<td>0.00</td>
<td>30</td>
<td>31</td>
<td>14.30</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>65.0</td>
<td>1.62</td>
<td>0.35</td>
<td>0.01</td>
<td>30</td>
<td>31</td>
<td>14.94</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>75.0</td>
<td>0.80</td>
<td>0.44</td>
<td>0.06</td>
<td>28</td>
<td>27</td>
<td>5.05</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>85.0</td>
<td>1.82</td>
<td>0.33</td>
<td>0.02</td>
<td>30</td>
<td>31</td>
<td>14.40</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>99.0</td>
<td>1.04</td>
<td>0.36</td>
<td>0.00</td>
<td>29</td>
<td>31</td>
<td>14.28</td>
<td>0.13</td>
</tr>
</tbody>
</table>

1TOC, Total Organic Carbon
2Calculated as: all isomers of C_{27}-C_{29} steranes/all isomers of C_{29}-C_{35} desmethylhopanes
3Calculated as: (C_{16}-C_{20})/(C_{27}-C_{31})
4Average chain length calculated as: (25*area C_{25} + 27*area C_{27} + 29*area C_{29} + 31*area C_{31} + 33*area C_{33})/(area C_{25} + area C_{27} + area C_{29} + area C_{31} + area C_{33})
5Carbon preference index calculated as: (0.5)*(((C_{25} + C_{27} + C_{29} + C_{31} + C_{33})/(C_{24} + C_{26} + C_{28} + C_{30} + C_{32} + C_{34}))*((C_{25} + C_{27} + C_{29} + C_{31} + C_{33})/(C_{26} + C_{28} + C_{30} + C_{32} + C_{34})))
6Proportion of n-alkanes derived from aquatic plants calculated as (C_{23} + C_{25})/(C_{23} + C_{25} + C_{29} + C_{31})
<table>
<thead>
<tr>
<th>Sample</th>
<th>Depth (cm)</th>
<th>Sterane: Hopane²</th>
<th>Short:Long n-alkanes³</th>
<th>ACL (even)⁴</th>
<th>Most abundant n-alkane</th>
<th>CPI⁵</th>
<th>Pₐq⁶</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siq10-a</td>
<td>6.5</td>
<td>2.4</td>
<td>0.59</td>
<td>27</td>
<td>24</td>
<td>0.54</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>21.5</td>
<td>1.3</td>
<td>0.56</td>
<td>27</td>
<td>24</td>
<td>0.42</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>38.0</td>
<td>3.1</td>
<td>1.13</td>
<td>27</td>
<td>24</td>
<td>0.34</td>
<td>0.67</td>
</tr>
<tr>
<td>SukS10-a</td>
<td>5.0</td>
<td>25.8</td>
<td>5.73</td>
<td>25</td>
<td>24</td>
<td>0.29</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>15.0</td>
<td>17.9</td>
<td>3.42</td>
<td>26</td>
<td>24</td>
<td>0.28</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>25.0</td>
<td>18.0</td>
<td>6.73</td>
<td>25</td>
<td>24</td>
<td>0.32</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>35.0</td>
<td>8.8</td>
<td>4.02</td>
<td>26</td>
<td>24</td>
<td>0.40</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>45.0</td>
<td>31.0</td>
<td>3.95</td>
<td>26</td>
<td>24</td>
<td>0.36</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>55.0</td>
<td>14.8</td>
<td>13.12</td>
<td>25</td>
<td>18</td>
<td>0.39</td>
<td>0.96</td>
</tr>
<tr>
<td></td>
<td>65.0</td>
<td>30.1</td>
<td>2.60</td>
<td>26</td>
<td>24</td>
<td>0.40</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>75.0</td>
<td>6.3</td>
<td>16.83</td>
<td>25</td>
<td>20</td>
<td>0.49</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>85.0</td>
<td>46.3</td>
<td>7.01</td>
<td>25</td>
<td>24</td>
<td>0.41</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td>99.0</td>
<td>57.2</td>
<td>10.11</td>
<td>25</td>
<td>18</td>
<td>0.36</td>
<td>0.93</td>
</tr>
</tbody>
</table>

¹TOC, Total Organic Carbon
²Calculated as: all isomers of C₂₇-C₂₉ steranes/all isomers of C₂₉-C₃₅ desmethylhopanes
³Calculated as: (C₁₆-C₂₀)/(C₂₇-C₃₁)
⁴Calculated as: (24*area C₂₄ + 26*area C₂₆ + 28*area C₂₈ + 30*area C₃₀ + 32*area C₃₂)/(area C₂₄ + area C₂₆ + area C₂₈ + area C₃₀ + area C₃₂)
⁵Carbon preference index calculated as: (0.5)*((C₂₅ + C₂₇ + C₂₉ + C₃₁ + C₃₃)/(C₂₄ + C₂₆ + C₂₈ + C₃₀ + C₃₂)+(C₂₅ + C₂₇ + C₂₉ + C₃₁ + C₃₃)/(C₂₆ + C₂₈ + C₃₀ + C₃₂ + C₃₄))
⁶Proportion of n-alkanes derived from aquatic plants calculated as (C₂₃ + C₂₅)/(C₂₃ + C₂₅ + C₂₉ + C₃₁).
Figure S2.1 Methane concentration and $\delta^{13}$C$_{CH_4}$ from all the cores collected for these analyses. (A-C) CH$_4$ concentration and (D,E) $\delta^{13}$C$_{CH_4}$ in pore waters. Note the different x-axis scales. $\delta^{13}$C$_{CH_4}$ could not be measured for Sukok lake samples collected in 2013 due to CH$_4$ concentrations too low for the analysis. The approximate boundary between biogenic and thermogenic (Whiticar, 1999) is indicated with a dashed line. The error for the average CH$_4$ concentration (analytical replicates) was ± 2 %. The error for the average $\delta^{13}$C$_{CH_4}$ (analytical replicates) was ± 0.2 $^\circ$/oo, except: 58 cm (± 1.6 $^\circ$/oo) and 88 cm (~ ± 6.0 $^\circ$/oo) in Siq11-a; 25 cm (± 0.4 $^\circ$/oo) and 45 cm (± 0.8 $^\circ$/oo) in Siq12a; 5 cm (± 0.4 $^\circ$/oo), 25 cm (± 0.3 $^\circ$/oo), and 69 cm (~ ± 8.1 $^\circ$/oo) in Siq12-c; and 55 cm (± 1.4 $^\circ$/oo) in SukS.
**Supplementary References**


Chapter 3

Diversity, structure, functional potential of the microbial community, and phylogeny of Archaea in sediments of Alaskan Arctic thermokarst lakes

Co-authorship statement

P. B. Matheus Carnevali, C. Herbold, K. P. Hand, and A. E. Murray

PM and AM collected samples and designed the experiments. PM executed the experiments and data analyses with mentoring by AM. CH performed statistical analyses in R. PM wrote the chapter under the supervision of AM. AM and KH secured support for the research program and AM secured support for the NGS.
Abstract

Thermokarst lakes sediments provide ideal conditions for organic matter degradation through fermentation and respiration processes, which ultimately lead to the production of greenhouse gases such as CO₂ and CH₄. Microorganisms inhabiting the sediments are the engine behind some of these processes, and their metabolic potential determines the fate of important nutrients in the environment. Two lakes (Siqlukaq Lake – Siq, and Sukok Lake – Suk) on the Coastal Plain of Alaska contain 0.2 to 15 wt. % of total organic carbon (TOC) in their sediments, and the lake with higher TOC content (Siq) exhibited higher concentrations and a higher potential for biological CH₄ production (Chapter 2). Likewise, there are different geochemical gradients in the sediments of Siq and Suk; and in one site in Suk, there was an active thermogenic CH₄ seep at the time of sampling. Our goals for this study were to understand the effect that the sediment geochemistry (and vice versa) may have on the diversity and structure of the active microbial community (Bacteria and Archaea), and to determine the phylogeny of Archaea inhabiting these arctic thermokarst lake sediments. To study the microbial community as represented by the small subunit (SSU) ribosomal RNA (rRNA) gene and the SSU rRNA, we used a combination of molecular approaches, including MiSeq next-generation sequencing (iTag sequencing), archaeal variable region 3 (V3) SSU rRNA gene denaturing gradient gel electrophoresis (DGGE), and archaeal V3 SSU rRNA gene cloning. We focused on the analyses of rRNA iTag sequences as proxies for the potentially active
microorganisms in the sediments at the time of sampling, which allowed us to investigate the functional potential of the community as well. The findings from iTag sequencing indicated that the microbial community in the sediments was highly diverse (Inverse Simpson's reciprocal index 33.4-148.6), and that it was composed of 20 abundant taxa of Bacteria (≥ 1% of amplified RNA – aRNA or RNA sequences or more) across three depths surveyed from cores collected in Siq and two sites in Suk (SukB and SukS). These phyla encompassed groups capable of organic matter degradation, sulfate reduction, nitrogen cycling, and photosynthesis. A deeper look into the functional potential of the community using phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt) supported these findings. The archaeal assemblage was found to be portioned between lakes, and methanogenic Archaea were mostly found in the sediments of Siq and were almost entirely absent in Suk. To know the status of the microbial community in these of climate-sensitive ecosystems is essential, because it will allow for predictive changes in community structure that could lead to changes in greenhouse gas emissions.

1. Introduction

Arctic regions are among the most climate-sensitive ecosystems on the planet (McGuire et al., 2009; Kittel et al., 2011; Graham et al., 2012), particularly because of the high amount of partially degraded organic matter that is stored in permafrost (frozen ground), both in continental and oceanic territories (Zimov et al., 2006).
Thermokarst lakes occupy an estimated 250,000 to 380,000 km$^2$ area in the pan-Arctic (Grosse et al., 2013), and are perhaps even more sensitive to warming than permafrost itself. The large heat storage capacity of the water column combined with long term heat flux to the sediments, maintains lake bottoms above freeze temperatures and result in expansion of the thaw bulb underlying the lake bottom (Grosse et al., 2013).

The thaw bulb underneath thermokarst lakes contains organic matter that was stored in permafrost, in addition to fresh organic matter produced in the water column during the long days of summer. These two sources of carbon can fuel microbial processes such as anaerobic respiration, fermentation, and methanogenesis. The sediments of two lakes in the area (Siqlukaq Lake – Siq, and Sukok Lake – Suk) contain 0.2 to 15 wt. % of total organic carbon (TOC), and the lake with higher TOC content (Siq) exhibited a higher potential for biological CH$_4$ production (Chapter 2; Matheus Carnevali et al., 2015). Biological CH$_4$ production occurs as a result of these microbial activities in arctic thermokarst lake sediments on the Coastal Plain of Alaska. Therefore, we hypothesized that the microbial community in Siq would be different from the microbial community in SukB and SukS, considering the differences in total organic carbon (TOC) content detected in these sediments.

Microbial communities in arctic rivers, wetlands, permafrost, and permafrost-affected soils have been studied with DNA profiling techniques and clone libraries of the SSU rRNA gene (Høj et al., 2005; Galand et al., 2006; Galand et
al., 2008; Steven et al., 2007; Barbier et al., 2012). Next generation sequencing (NGS) technologies that provide the ability to parallelize the analysis of multiple samples, and to sequence at a higher sequencing depth at a faster pace, are being used more frequently as their cost decreases rapidly. In arctic environments, the use of NGS to study microbial communities has been limited to the Arctic Ocean (Galand et al., 2009; Kirchman et al., 2014), peat soils (Tveit et al., 2013), permafrost (Yergeau et al., 2010), permafrost-affected soils (Frank-Fahle et al., 2014), and wetlands (Serkebaeva et al., 2013; Stoeva et al., 2014). Only recently, Neghandi (2013) studied CH$_4$ sources in thaw ponds located in the Canadian High Arctic.

Most studies about thermokarst lakes in the North Slope of Alaska have focused on the limnology of the lakes (Howard & Prescott, 1973), on CH$_4$ fluxes (Morrissey & Livingston, 1992; Phelps et al., 1998), CH$_4$ cycling (Bretz & Whalen, 2014; Lofton et al., 2014), and other aspects. A few studies have looked at microbial communities in the sediments of thermokarst lakes on the North Slope of Alaska (He et al., 2012a; He et al., 2012b; He et al., 2012c). To know the status of the microbial community in this kind of climate-sensitive ecosystems is of utmost importance, because it allows for informed predictive changes in community structure and function that could lead to changes in greenhouse gas emissions to be estimated. Thus, knowing the microbial community ultimately improves our ability to estimate current and future greenhouse gas emissions in the Arctic. To advance our understanding of these lake ecosystems, we investigated the diversity, composition,
and functional potential of the microbial communities inhabiting two lakes on the Coastal Plain of Alaska.

Our goal was to study the bacterial and archaeal assemblages in the sediments of thermokarst lakes in light of the differences in geochemistry between lakes. A combination of molecular approaches, including MiSeq next-generation sequencing (iTag sequencing) of the small subunit (SSU) ribosomal RNA (rRNA) gene and the SSU rRNA was used. Ribosomal RNA sequences were taken as a proxy for potentially active microorganisms in the sediments at the time of sampling (Putkinen et al., 2009), which allowed us to investigate the functional potential of the community as well. Furthermore, the phylogeny of Archaea in these arctic lakes was determined.

2. **Methods**

2.1. **Sampling sites and samples collection**

Artic thermokarst lakes nearby the town of Barrow, Alaska, were sampled during three different field campaigns (April and Oct 2010 and Oct 2011). Sediment cores were collected using a universal percussion corer (Aquatic Research Instruments) and stored frozen at -80 °C until subsampling. Samples for the primary iTag sequencing analyses described in this study were collected in 2011 from Siqluakaq Lake (Siq11; 71° 10.481’ N, 156° 53.910’ W), and from two sites in Sukok Lake: a site nearby an open natural gas seep (SukS11; 71° 04.519’ N, 156° 49.208’ W) and another site ~ 1 km southwest from the seep (Sukok B site - SukB; 71° 04.006’ N,
Lakes Ikroavik (Ikv10; 71° 14.036’ N, 156° 38.349’ W), Siqlukaq (Siq10; 71° 10.486’ N, 156° 53.891’ W), Sukok (Sukok Seep site – SukS10; 71° 04.455’ N, 156° 49.250’ W), and Qalluuraq Lake (70° 22.672’ N, 157° 20.927’ W; 60 miles to the south) had been also sampled in 2010. The 2010 samples were used to survey the archaeal diversity by profiling the V3 region of the SSU rRNA gene with denaturing gradient gel electrophoresis (DGGE), and to determine the phylogeny of Archaea in these lakes.

2.2. Nucleic acid extraction for iTag sequencing of the SSU rRNA gene and SSU rRNA of Bacteria and Archaea

Over the course of this study three DNA extraction protocols (one described in this section and two described in section 2.3) and one RNA extraction protocol were executed. Efforts were made to optimize DNA extraction procedures to ensure amplifiability of the DNA. Extracts were used for different downstream analyses and were not intercompared.

To study the microbial community diversity and structure and the ‘potentially active’ members of the community, genomic DNA and total RNA were extracted from three sections (U-upper, M-middle, L-lower) of the first 30 cm of the 2011 cores (Siq11, SukS11, and SukB11). Sediment subsamples that had been stored at -80 °C and later transferred to sucrose lysis buffer (SLB, 40 mM EDTA, 50 mM Tris-HCl, 0.75 M sucrose) to better preserve the integrity of the nucleic acids, were used to extract community genomic DNA as previously described in (Chapter 2) (Matheus Carnevali et al., 2015). In brief, samples in SLB were thawed on ice for
45 min and centrifuged at 10,000 x g for 10 min. The supernatant was removed; 2 x
~ 0.5 g sediment subsamples from each depth interval (for a total weight of ~ 1 g
per sample) were extracted using a Power Soil DNA Isolation Kit (MoBio, Carlsbad,
CA, USA) following manufacturer’s instructions, and later quantified using the
Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Grand Island, NY, USA).

To ultimately obtain total RNA, nucleic acids were also extracted from each
section of Siq11, SukS11, and SukB11 using a modified RNA PowerSoil Total RNA
Isolation Kit (MoBio, Carlsbad, CA, USA). Two grams of sediment (in duplicate) were
bead beat with MoBio’s proprietary solutions and phenol:chloroform:isoamyl
alcohol (pH 6.5-7.0) by vortexing at maximum speed for 15 min. The organic and the
aqueous phases were separated by centrifugation at 1,040 x g for 24 min, followed
by a secondary precipitation step at 4 °C for 10 min to further purify the aqueous
phase, and another centrifugation step under the same conditions. The resulting
supernatant was split in 1.7 ml Eppendorf tubes and the extracted nucleic acids
were precipitated in isopropanol at room temperature for 30 min, followed by cold
centrifugation at 13,000 x g for 15 min, and a clean-up step. This last step involved
70 % ethanol at room temperature, cold centrifugation at 13,000 x g for 15 min, air-
drying of the pellets (or using a Speed Vac Vacuum Concentrator at medium
temperature), and final suspension of the total RNA in 45-50 µL of nuclease-free
water (Life Technologies, Carlsbad, CA, USA). All reagents, disposable supplies and
instruments used for this protocol were RNAsse-free or treated with RNaseZap.
decontamination solution (Life Technologies, Carlsbad, CA, USA). Reagent controls were also processed to create reagent extraction controls.

In order to isolate the total rRNA, nucleic acid extracts were treated two to three times with 4 U RNase-free DNase I (Life Technologies, Carlsbad, CA, USA) at 37 °C in 100 µL reaction volumes for 30 min. The enzyme was later extracted with phenol:chloroforma:isoamyl alcohol (pH 6.5-7.0). Total RNA was precipitated at -20 °C overnight and suspended in 25 µL of nuclease-free water. These extracts were quantified using a Quant-iT RiboGreen RNA Assay Kit (Life Technologies, Carlsbad, CA, USA).

2.3. Reverse transcription of the SSU rRNA gene and RNA amplification

To determine if the total RNA could reverse transcribed to complementary DNA (cDNA) in sufficient amounts for sequencing, reverse transcription of the total RNA was directed with a SSU rRNA-specific primer (806R) (Caporaso et al., 2010). Total RNA was incubated with the dNTPs (Qiagen, Valencia, CA, USA) for 5 min at 65 °C, transferred to 4 °C, and then incubated at 55 °C for 60 min with 200 U of SuperScript III Reverse Transcriptase (Life Technologies, Carlsbad, CA, USA) in a 20 µL reaction, following manufacturer’s instructions. Some of the total RNA samples (Siq11-U, -M, and -L) were diluted 1/10 before the reverse transcription step to prevent downstream reactions inhibition, due to presence of humic and fulvic acids in the extracts. Final products were checked on an agarose gel. cDNA was quantified using a Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Grand Island, NY, USA).
Reagent controls were subjected to DNase treatments and reverse transcription with no subsequent amplification.

Only Siq11-U, Siq11-M, SukB11-U, SukB11-M, and SukB11-L samples could be reverse transcribed to cDNA directly from the RNA in sufficient amounts for sequencing. The remaining samples (Siq11-L; SukS11-U, SukS11-M, and SukS11-L) had to be first amplified using the MessageAmp II aRNA Amplification Kit (Life Technologies, Carlsbad, CA, USA) according to manufacturer’s instructions. Additionally, RNA obtained from Siq11-U, Siq11-M, SukB11-U, SukB11-M, and SukB11-L samples was amplified for comparisons of cDNA derived from amplified RNA (aRNA) from all samples. The resulting aRNA was quantified using a Quant-iT RiboGreen RNA Assay Kit (Life Technologies, Carlsbad, CA, USA), and reverse transcribed to cDNA with primer 806R (Caporaso et al., 2010). DNA and cDNA from both RNA and aRNA samples were sent to the Joint Genome Institute (JGI, Walnut Creek, CA) for library preparation and paired-end (2 x 250 bp) MiSeq Illumina sequencing of the variable region 4 (V4) of the SSU rRNA using primers 515F (5’-GTGCCAGCMGCCGCGGTAA-3’) and 806R (5’-GGACTACHVGGGTWTCTAAT-3’) following Caporaso et al. (2010).

### 2.4. DNA extraction for DGGE profiling and cloning of the SSU rRNA gene of Archaea

To profile the SSU rRNA gene of Archaea using DGGE and to make clone libraries, 1 g sediment subsamples from 2010 cores were collected every 10 cm from the top ~50 cm of Siq10, the top ~10-50 cm of Ikv10, and the 2-106 cm length of the SukS10
core. Community genomic DNA was extracted using a phenol-chloroform method (Massana, 1997) and quantified with a NanoDrop. To generate a second round of clone libraries of the SSU rRNA gene of Archaea from cores obtained in 2011, DNA was extracted from ~ 1 g sediment approximately every 10 cm of the first 30 cm of each core (Siq11, SukS11, and SukB11). A Power Soil DNA isolation kit (MoBio, Carlsbad, CA, USA) was utilized for genomic DNA extractions following manufacturer’s instructions.

2.5. **Archaeal PCR and denaturing gradient gel electrophoresis analysis of 2010 samples**

In order to get a rapid snapshot of the archaeal assemblage diversity and structure, primers targeting the archaeal SSU rRNA gene: GC_347F (5’-CGCCCGCCGCACCCCCCGCCGCCCCTCCGCGCCGCACGCCAGCGCAGGCGMGA-3’); and UNIV519R (Muyzer et al., 1993) were used to amplify the variable region 3 (V3) with the following reaction mix: 1X Standard Taq Buffer (New England BioLabs, Ipswich, MA, USA), 2 mM MgCl₂, 0.2 mM of each dNTP (Qiagen, Valencia, CA, USA), 0.5 µM of each primer, and 1.25 U Taq DNA polymerase (New England BioLabs, Ipswich, MA, USA) in a final volume of 50 µL. Thermocycler conditions were: 94 °C for 5 min, followed by 10 cycles of denaturation at 94 °C for 45 sec, touch-down primer annealing with the temperature decreasing from 65 °C to 55 °C (1 °C per cycle) for 30 sec, extension at 72 °C for 30 sec; 18 cycles of denaturation at 94 °C for 30 sec, primer annealing at 55 °C for 30 sec, extension at 72 °C for 30 sec, and a final extension step at 72 °C for 10 min. Amplification products were precipitated overnight at -20 °C with two volumes
of absolute ethanol and 1/10 of the volume of 3 M sodium acetate pH 5. After centrifugation in cold for 30 min, the DNA was washed with 70% ethanol, centrifuged for 10-15 min, air-dried, or dried in a DNA speed vac (Savant), and suspended in 18 µl of sterile DI water. Two to seven PCR reaction products were pooled to reach 300 ng of DNA per sample. A 30 to 65 % gradient of denaturants (7 M urea and 40% deionized formamide in 100% concentration of denaturants) was used. Electrophoresis was run at 60 °C for 16 hrs at 62 V (Murray et al., 1996). Analyses of the banding patterns resulting from DGGE were done with the GelComparer II software (Applied Maths).

2.6. Clone libraries construction and analyses

To study the phylogeny of Archaea in these thermokarst lake sediments, 5 µL of each extract obtained in 2010 from the different sediment intervals were pooled for amplification of the SSU rRNA gene using primers 109F (Großkopf et al., 1998) and 915R (Stahl & Amann, 1991) using the following reaction mix: 1X Standard Taq Buffer (New England BioLabs, Ipswich, MA, USA), 2 mM MgCl₂, 0.2 mM of each dNTP (Qiagen, Valencia, CA, USA), 0.5 µM of each primer, and 1U Taq DNA polymerase (New England BioLabs, Ipswich, MA, USA) in a final volume of 20 µL. Thermocycler conditions were: 94 °C for 5 min, followed by 28 cycles of denaturation at 94 °C for 1 min, primer annealing at 52 °C for 1 min, extension at 72 °C for 1 min 30 sec, and a final extension step at 72 °C for 6 min.
The amplification products (bp) were gel purified using a QIAquick Gel Extraction Kit (Qiagen, Valencia, CA, USA), pooled, and cloned using the TOPO TA cloning kit (Life Technologies, Carlsbad, CA, USA) according to manufacturer’s recommendations. Plasmids were extracted with the Plasmid Miniprep Kit (EMD Millipore, Temecula, CA, USA) and screened using a restriction digest as follows: 25 U EcoRI (Life Technologies, Carlsbad, CA, USA), 1X Buffer H (Life Technologies, Carlsbad, CA, USA), 0.1 mg mL⁻¹ bovine serum albumin (Promega, Sunnyvale, CA, USA), and 1 µL of plasmid DNA at 37 °C for 1 hr. The enzyme was deactivated at 65 °C for 20 min. Insert size was checked in 1 % agarose gels stained with GelRed (Biotium, Hayward, CA, USA). Plasmids with the desired insert were sequenced in one direction with the ABI BigDye Terminator Cycle Sequencing Ready Reaction Kit V3.1 using vector primer T7 on a ABI3730 DNA Analyzer (Life Technologies, Carlsbad, CA, USA).

In 2011, the SSU rRNA gene was amplified using primers 20F-958R (DeLong, 1992) using the following reaction mix: 1X Ampli Taq Buffer II (Life Technologies, Carlsbad, CA, USA), 2.5 mM MgCl₂ (Life Technologies, Carlsbad, CA, USA), 0.2 mM of each dNTP (Qiagen, Valencia, CA, USA), 0.5 µM of each primer, and 1.25 U AmpliTaq DNA polymerase (Life Technologies, Carlsbad, CA, USA) in a final volume of 25 µL. Thermocycler conditions were: 94 °C for 10 min, followed by 35 cycles of denaturation at 94 °C for 1 min, primer annealing at 55 °C for 1 min, extension at 72 °C for 1 min, and a final extension step at 72 °C for 7 min. The amplification products from each depth were gel purified using a QIAquick Gel Extraction Kit (Qiagen,
Valencia, CA, USA), pooled with other amplification products from the same depth, and cloned using the TOPO TA cloning kit (Life Technologies, Carlsbad, CA, USA) according to manufacturer’s recommendations. Plasmids were extracted with a Qiagen 3000 BioRobot and a R.E.A.L prep (Qiagen, Valencia, CA, USA), quantified with the Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies, Carlsbad, CA, USA), then sequenced in one direction with the ABI BigDye Terminator Cycle Sequencing Ready Reaction Kit V3.1 using vector primer T7 on a ABI3730 DNA Analyzer (Life Technologies, Carlsbad, CA, USA).

Only high quality Sanger sequences from both libraries were used for sequence analyses, including trimming with BioEdit version 7.2.5, and chimera check with Decipher (Wright et al., 2012). To study the evolutionary relatedness between thermokarst sediment Archaea and known Archaea lineages, the SSU rRNA gene sequences were aligned with their closest matches from the GenBank, several sequences from related cultivated Archaea, and other relevant sequences from the Silva database. Sequences were aligned against the Archaea variability profile using Silva’s SINA aligner (Pruesse et al., 2012). Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 6 (Tamura et al., 2013).

2.7. iTag sequence processing and statistical analyses

JGI initially processed the sequences to remove contaminants, barcodes, and primers. The quality of these sequences was evaluated using the FastQC software (Babraham Bioinformatics). Trimmomatic (Bolger et al., 2014) was then used in the
single end mode to filter the low quality sequences, and 3,834,911 (85.8 %) sequences out of 4,469,505 sequences passed the quality filtering step. Sequences were then separated into forward and reverse reads with an in-house script and used to form contigs with Mothur (Kozich et al., 2013), which was also the platform used to remove chimeras, classify the sequences against the Greengenes gg_13_8_99 taxonomy, and to cluster them into operational taxonomic units (OTUs) at a distance of 0.03. These sequences were used to create an OTU database containing OTU abundance, a representative sequence, and the consensus taxonomy (Kozich et al., 2013). Mothur was also used to do alpha-diversity analyses on a resampled OTU table. Beta-diversity analyses were done after bootstrap resampling using R (R Core Development Team, 2015) and methods and custom scripts developed by Herbold et al. (2014). To ensure that the data was free of sequencing error, OTUs with 5 sequences or less among all of the samples (DNA, aRNA, and RNA) were removed prior to resampling. To determine between site similarity for DNA and aRNA sequences, a Mann-Whitney test was used to test pairwise comparisons of the Bray-Curtis distances between sites, and the significant thresholds were defined by the Benjamini-Hochberg procedure with a false discovery rate of 0.05 in each sample. To determine whether the abundance of any aRNA-OTU at any site was greater or less than the abundance of any aRNA-OTU at any site, abundances were log-transformed and compared across sites at the phylum-level (or the class level for Proteobacteria) using a Tukey’s test, where phyla and site were interacting factors
and depth was the independent variable. The median $p$-value was obtained by repeating the Tukey’s test for all resamplings.

2.8. Functional profiling

To determine the functional potential of the community represented in the aRNA samples, as a proxy for ‘potentially active’ microorganisms (Putkinen et al., 2009) in the lake sediments at the time of sampling, phylogenetic investigation of communities by reconstruction of unobserved states (PICRUSt) (Langille et al., 2013) was implemented through the Hunttenhover lab modules in the Galaxy server. Sequences were reclassified with Mothur against the Greengenes gg_13_5_99 taxonomy and the OTU table was formatted to match PICRUSt software requirements. In simple terms, PICRUSt predicts community functional composition by associations of SSU rRNA gene sequence data (corrected by copy number) with reference genomes based on relatedness (80 %). The Kyoto Encyclopedia of Gene and Genomes (KEGG) Orthology (KO) was used in this study to annotate the table of predicted gene family counts per sample. Statistical analyses of differences between predicted gene content was performed using Statistical Analysis of Metagenomic (and other) Profiles, STAMP (Parks & Beiko, 2010; Parks et al., 2014).

3. Results

3.1. $\alpha$-diversity and $\beta$-diversity of the microbial assemblage (Bacteria and Archaea)
A total of 1,213,686 filtered sequences comprising DNA, aRNA and RNA were used to analyze the diversity and structure of the microbial community. Out of this total, 290,351 sequences were unique, with an average of 62,297 ± 7,068 DNA sequences (n=9), an average of 47,783 ± 4,817 aRNA sequences (n=9), and average of 44,783 ± 8,699 RNA sequences (n=5) (Table 3.1). The sequences were clustered into 39,244 operational taxonomic units (OTUs) at a distance of 0.03. DNA sequences formed an average of 4,325 ± 1,116 OTUs, of which 2,928 ± 838 OTUs were doubletons and singletons; the aRNA sequences formed an average of 3,832 ± 1,119 OTUs including 2,434 ± 787 doubletons and singletons; and the RNA sequences formed an average of 4,096 ± 1,043 OTUs including 2,649 ± 959 doubletons and singletons. Therefore, a great majority of the OTUs were comprised of rare sequences.

After 1000 iterations of subsampling to less than the lowest number of sequences (28,000) in a given sample, iTag sequence coverage, richness and diversity were estimated (Table 3.2). Good’s coverage (Good, 1953) indicated sampling over 92 %. Simpson’s diversity index indicates the probability that any two individuals of the same species would be randomly drawn from a community. The index is inversely proportional to diversity, but the reciprocal of the index rises as the community becomes more even (Magurran, 2003). Therefore, Simpson’s reciprocal index (Table 3.2) suggested that the samples were highly diverse (33-132 for DNA, 53-223 for aRNA, and 68-149 for RNA). Diversity among the DNA iTags was similarly high (104-133) in the upper and middle intervals of Siq and SukB, and it decreased with depth. The diversity of RNA iTags was higher in Siq than in SukB,
Table 3.1 iTag sequence statistics. Data presented here was not normalized to correct for iTag pool sizes. Sediment intervals correspond to three sections in the first 30 cm of sediment: upper (U), middle (M), and lower (L).

<table>
<thead>
<tr>
<th>Lake</th>
<th>Sediment Interval</th>
<th>Number of filtered sequences (1,213,686)</th>
<th>Number of OTUs at 0.03 distance (singletons and doubletons removed)</th>
<th>Number of singleton or doubleton OTUs at 0.03 distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siq11</td>
<td>U</td>
<td>DNA 66,997 aRNA 46,989 RNA 51,281</td>
<td>DNA 1,669 aRNA 1,681 RNA 1,645</td>
<td>DNA 4,152 aRNA 3,406 RNA 3,954</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>71,199 42,585 48,254</td>
<td>1,766 1,346 1,463</td>
<td>3,997 2,260 3,064</td>
</tr>
<tr>
<td></td>
<td>LB</td>
<td>67,282 47,140 ------</td>
<td>1,443 1,685 ------</td>
<td>3,102 2,961 ------</td>
</tr>
<tr>
<td>SukB11</td>
<td>U</td>
<td>63,393 46,861 47,983</td>
<td>1,564 1,251 1,341</td>
<td>3,204 2,324 2,767</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>63,461 48,210 29,397</td>
<td>1,449 1,279 1,111</td>
<td>3,061 2,056 1,951</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>66,139 41,519 46,050</td>
<td>1,163 934 1,673</td>
<td>2,305 1,072 1,509</td>
</tr>
<tr>
<td>SukS11</td>
<td>U</td>
<td>58,683 47,850 ------</td>
<td>1,099 1,440 ------</td>
<td>2,156 1,966 ------</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>48,521 50,672 ------</td>
<td>1,533 2,045 ------</td>
<td>2,805 3,634 ------</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>54,999 58,221 ------</td>
<td>890 916 ------</td>
<td>1,566 2,198 ------</td>
</tr>
</tbody>
</table>
and the highest diversity for both was found in the middle interval. Among the aRNA iTags the diversity in SukS-M was the highest of all (223).

For β-diversity analyses, the read depth between different samples was normalized by bootstrap resampling (500 bootstraps). A non-metric multidimensional scaling (NMDS) ordination analysis of Bray-Curtis dissimilarity with 20 resamplings for each sample (Fig. 3.1) revealed the separation between DNA sequences from all the three sites (Fig. 3.1a). However, this pattern changed for the aRNA sequences, because the NMDS analysis (Fig. 3.1b) showed a greater separation between SukB-L and SukS-L and the rest of the samples. In either case, the Siq samples from different depths clustered closely together, and in the case of Siq the aRNA sequences were also clustered more tightly with SukB-U and –M, and with SukS-U and –M. Based on a Mann-Whitney test, DNA sequences from Siq were more similar to SukB than to SukS ($p = 0.0001$, significant in 1000/1000 resamples) and DNA sequences from SukS were more similar to SukB than to Siq ($p = 0.0053$, significant in 1000/1000 resamples). The aRNA sequences from the three sites were not significantly different (0/1000 resamples).

3.2. **Thermokarst lake sediment bacterial assemblage composition**

In chapter 2, the three sampling sites presented different gradients of nitrate, sulfate, dissolved iron, and dissolved manganese concentrations (Matheus Carnevali et al., 2015). We were interested in how these environmental parameters influence the microbial community structure (or viceversa), and what microbial guilds inhabit
Table 3.2 Alpha diversity of samples subsampled to 28,000 sequences per group. Sediment intervals correspond to three sections in the first 30 cm of sediment: upper (U), middle (M), and lower (L). aRNA data is in bold.

<table>
<thead>
<tr>
<th>Lake</th>
<th>Depth</th>
<th>Coverage</th>
<th>$S_{obs}$</th>
<th>Simpson's Reciprocal Index</th>
<th>Simpson's Reciprocal Index low confidence interval</th>
<th>Simpson's Reciprocal Index high confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>Siq</td>
<td>U</td>
<td>DNA 0.93</td>
<td>3473.1</td>
<td>132.8</td>
<td>129.1</td>
<td>136.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aRNA 0.93</td>
<td>3874.5</td>
<td>130.8</td>
<td>125.8</td>
<td>136.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNA 0.92</td>
<td>3942.4</td>
<td>112.5</td>
<td>107.5</td>
<td>117.9</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>DNA 0.93</td>
<td>3372.8</td>
<td>112.5</td>
<td>108.9</td>
<td>116.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aRNA 0.95</td>
<td>2951.5</td>
<td>72.7</td>
<td>70.2</td>
<td>75.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNA 0.93</td>
<td>3359.7</td>
<td>148.6</td>
<td>143.8</td>
<td>153.6</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>DNA 0.94</td>
<td>2784.3</td>
<td>76.5</td>
<td>74.2</td>
<td>78.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aRNA 0.93</td>
<td>3586.7</td>
<td>129.2</td>
<td>125.1</td>
<td>133.6</td>
</tr>
<tr>
<td>SukB</td>
<td>U</td>
<td>DNA 0.94</td>
<td>3050.1</td>
<td>138.8</td>
<td>134.3</td>
<td>143.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aRNA 0.95</td>
<td>2722.8</td>
<td>68.9</td>
<td>66.5</td>
<td>71.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNA 0.94</td>
<td>3061.5</td>
<td>67.7</td>
<td>65.2</td>
<td>70.3</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>DNA 0.94</td>
<td>2863.2</td>
<td>104.1</td>
<td>100.7</td>
<td>107.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aRNA 0.95</td>
<td>2547.2</td>
<td>116.3</td>
<td>113.0</td>
<td>119.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNA 0.95</td>
<td>2989.0</td>
<td>117.3</td>
<td>113.8</td>
<td>121.1</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>DNA 0.96</td>
<td>2185.7</td>
<td>36.0</td>
<td>34.7</td>
<td>37.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aRNA 0.97</td>
<td>1714.7</td>
<td>53.0</td>
<td>51.2</td>
<td>54.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RNA 0.96</td>
<td>2666.4</td>
<td>86.4</td>
<td>83.1</td>
<td>90.0</td>
</tr>
<tr>
<td>SukS</td>
<td>U</td>
<td>DNA 0.96</td>
<td>2193.8</td>
<td>33.5</td>
<td>32.4</td>
<td>34.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aRNA 0.96</td>
<td>2722.1</td>
<td>86.3</td>
<td>83.4</td>
<td>89.3</td>
</tr>
<tr>
<td></td>
<td>M</td>
<td>DNA 0.94</td>
<td>3282.1</td>
<td>72.0</td>
<td>69.6</td>
<td>74.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aRNA 0.92</td>
<td>4205.3</td>
<td>223.0</td>
<td>215.0</td>
<td>231.6</td>
</tr>
<tr>
<td></td>
<td>L</td>
<td>DNA 0.97</td>
<td>1741.0</td>
<td>40.5</td>
<td>39.4</td>
<td>41.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>aRNA 0.96</td>
<td>2018.3</td>
<td>65.5</td>
<td>65.8</td>
<td>69.4</td>
</tr>
</tbody>
</table>
Figure 3.1a Non-metric multidimensional scaling (NMDS) analysis of DNA iTag sequences. Calculated stress (0.08) using Bray-Curtis similarity.

the sediments. We used the aRNA sequences as a proxy for the members of the community that may be potentially active in the sediments. The RNA sequences were also included in the analysis as a reference for the results obtained by RNA amplification, and we observed a relatively high degree of similarity (42-68 %) between these two sets of sequences for paired samples.
Figure 3.1b Non-metric multidimensional scaling (NMDS) analysis of aRNA iTag sequences. Calculated stress (0.138) using Bray-Curtis similarity.

In the database normalized by bootstrap resampling, aRNA and RNA OTUs comprising 1% of sequences or more accounted for 85 OTUs and represented at least 15 Bacteria phyla. These phyla included: Acidobacteria (1 OTU), Actinobacteria (11 OTUs), Proteobacteria (34) with the classes Alphaproteobacteria (4 OTUs), Betaproteobacteria (5 OTUs), Deltaproteobacteria (14 OTUs), and Gammaproteobacteria (11 OTUs) represented among the OTUs; Armatimonadetes
(2 OTUs), Bacteroidetes (13 OTUs), Chlorobi (2 OTUs), Chloroflexi (5 OTUs), Cyanobacteria (4 OTUs), Fibrobacteres (1 OTU), Firmicutes (2 OTUs), Nitrospirae (2 OTUs), Aminicenantes (OP8; 1 OTU), Artribacteria (OP9; 1 OTU), Planctomycetes (1 OTU), Verrucomicrobia (3 OTUs), and two unclassified OTUs (Fig. 3.2).

OTUs related to the order Cytophagales in the Bacteroidetes were the most abundant among the three sites, followed by OTUs related to the Deltaproteobacteria. OTUs related to the order Syntrophobacterales and the order Desulfobacterales were more abundant in SukB than in Siq and SukS. Similarly, the Cyanobacteria OTUs of the orders Gloebacterales, Synechococcales, and Chroococcales were more abundant in SukB, as well as OTUs related to the order Pseudomonadales in the Gammaproteobacteria. Nitrospira OTUs were more abundant in Siq than in SukB and SukS, and the SB-45 Artribacteria (OP9) were mostly present in Suk.

Significant differences ($p < 0.05$) in the abundance of phyla and Proteobacteria classes were observed. Log-transformed abundances of aRNA-OTUs across sites were compared using a Tukey’s test, where phyla and site were interacting factors and depth was the independent variable. In Siq, Actinobacteria, Bacteroidetes, Deltaproteobacteria, and Gammaproteobacteria were more abundant than Cyanobacteria; in SukB and SukS Deltaproteobacteria were more abundant than Nitrospira; and in SukS Bacteroidetes were also more abundant than Nitrospirae. The major difference between Siq and SukB and SukS was the
Figure 3.2 Relative abundance of bootstrap resampled OTUs observed with ≥1% relative abundance in either the aRNA or the RNA datasets, representing the potentially active microorganisms. Circle area is proportional to relative abundance ($r^2 \times$ relative abundance). There are three circles plotted on top of one another: the largest circle is the darkest, and represents the maximum resampled relative abundance (upper 95% confidence interval), the next largest circle represents median and is a moderate shade of color. The smallest circle is also the lightest and represents the minimum relative abundance over the resampling (lower 95% confidence interval). Confidence intervals were close to the median because these are the abundant OTUs. Names on the far left represent Bacteria phyla or class for the Proteobacteria, and are followed by Orders, and number of related OTUs in parentheses.
abundance of Cyanobacteria in SukB and SukS. Conversely, the abundance of Nitrospirae in SukB was lower than the abundance of Bacteroidetes in Siq; and the abundance of Nitrospirae in SukS was lower than the abundance of Deltaproteobacteria in Siq. Therefore, the abundance of Nitrospirae in Siq was a major difference between this lake and the two sites in Suk.

3.3. Thermokarst lake sediment-associated archaeal diversity

Exceedingly divergent between site patterns of archaeal OTUs were evident upon two-dimensional hierarchical clustering (average linkage) analysis based on non-parametric Spearman’s rank correlations (Fig. 3.3). Only DNA, aRNA, and RNA OTUs related to Archaea that had 95% or greater confidence of being observed in the bootstrap resampled community are shown in Figure 3.3.

OTUs related to the Miscellaneous Crenarchaeal Group (MCG) were present among the DNA sequences in all three sites, but only among the aRNA sequences in Siq-M, and the RNA sequences in Siq-M, SukB-M and –L. Methanogen-related OTUs dominated Siq OTUs, with OTUs related to Methanosarcinales present among the DNA, aRNA, and RNA sequences at all depths (Fig. 3.3). There were also OTUs related to the order Methanomicrobiales among DNA sequences from all three depths, and among the aRNA or RNA from each depth; and OTUs related to the order Methanobacteriales among DNA sequences from Siq-U and –L, and among Siq-U aRNA and Siq-M RNA. Methanogenic Archaea were absent from most SukB and SukS samples, which were populated by different kinds of Archaea. Notably, the only
methanogen-related OTUs that were present in SukB were found among the DNA sequences.

Other OTUs that could be possibly involved in CH\(_4\) cycling (i.e., potentially in anaerobic CH\(_4\) oxidation) were found among the DNA and aRNA sequences in SukB–L, and were related to ANME-2a (DNA) and ANME-2d (aRNA). Furthermore, among the SukS-U DNA and aRNA; SukB-M and SukS-M aRNA, and SukB-L RNA, some OTUs in the Class Thermoplasmata were related to the family Methanomassiliicoccaceae, which includes methanogenic members (Iino et al., 2013). OTUs related to the putative group, Parvarchaea YLA114 were found among SukS-U DNA and aRNA sequences; and SukS-M, SukB-L, and SukB-M aRNA sequences as well. Among SukS-M DNA and aRNA sequences there were OTUs related to the order Nitrososphaerales in the Class Thaumarchaeota.

Based on the aRNA sequences, Archaeal groups that could have been potentially active at the different sites include OTUs related to Methanoseta in Siq; OTUs related to ANME-2d (Candidatus Methanoperedens nitroreducens), and the putative Parvarchaea YLA114 in SukB; the Candidate division WCHD3-30 and the Methanomassillicoccaceae in SukB and SukS; and the miscellaneous crenarcheal group clone pGrfC26 at all three sites.

The diversity profile of Archaea obtained by DGGE of the hypervariable region 3 (V3) of the SSU rRNA gene (Fig. S3.1 and S3.2) from Siq10, SukS10, and Ikr10, supported the partition of Archaea among Siq and SukS (Fig. S3.3); and
indicated that there was a high number of archaeal phylotypes per site (18-43). Likewise, phylotypes from the other two lakes (Qaalluraq Lake and Ikroavik Lake) that were sampled as part of the diversity survey carried out in 2010, clustered with Siq (NW in the Fig. S3.1, S3.3) and separate from Suk.

Phylogenetic analysis of the archaeal SSU rRNA gene clone libraries showed that archaeal sequences were closely related to members of the Classes Methanomicrobia and Thermoplasmata in the phylum Euryarchaeota, and to members of the phylum Thaumarchaeota and the Miscellaneous Crenarchaeotic Group (MCG), which forms a separate clade from the Thaumarchaeota and the Crenarchaeota (Kubo et al., 2012). Interestingly, there were three clusters of sequences that appeared to be only related to candidate archaeal phyla or novel uncultured microorganisms (Fig. 3.3). One of these groups clustered with sequences related to the Order Themorplasmaetes, Class Thermoplasmata. Another one of these groups was related to sequences in the Class Methanomicrobia, and included matches to members of the so-called Rice Cluster II (Großkopf et al., 1998). The remaining group formed a cluster within the Euryarchaeota, which was separate from the rest. We refer to this cluster as an ‘Arctic Cluster’ because the closest matches to our sequences were other sequences retrieved from the Arctic (Galand et al., 2006; Galand et al., 2008), but it also included groups previously identified as the Rice Cluster V (Großkopf et al., 1998), the Lake Dagow Sediment cluster (Glissman et al., 2004) and one member of the Candidate division WCHD3-30.
Figure 3.3 Archaeal OTUs that were observed with >95% confidence over all the resamplings. The abundances were rank-transformed within each resampling and then average rank of OTUs within each sample over all resamplings was used to cluster samples and OTUs together using Spearman’s correlation. Distances were calculated as 1-Spearman and clustering is average neighbor. Colors on the right represent the following archaeal groups: genera of Methanogens (orange), Thermoplasmatales (fuchsia), Marine Benthic Group B (MBGB; gray), Parvarchaea clone YLA114 (teal), Anaerobic Methanotrophs (ANME; olive), Miscellaneous Crenarchaeal Group (MCG; purple), Candidate division WCHD3-30 (maroon), Candidatus genus of Thaumarchaeota (blue), other Archaea groups (black).
Figure 3.4a Maximum likelihood phylogenetic tree of SSU rRNA gene Sanger sequences related to Euryarchaeota. Hollow symbols indicate that the given clone represents other clones from the same lake. Filled symbols indicate that the given clone represents other clones from the same lake and from other lakes as well. Lakes are coded by symbol as follows: Siq – circles, SukB – triangles, SukS – squares. Only bootstrap values of 50 or more are shown. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jukes-Cantor model. The tree with the highest log likelihood (0.0000) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 88 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 574 positions in the final dataset. Evolutionary analyses were conducted in MEGA6. Clusters correspond to Großkopf et al. (1998), Glissman et al. (2004), and Galand et al. (2006; 2008).

Figure 3.4b Maximum likelihood phylogenetic tree of SSU rRNA gene Sanger sequences related to other Archaea phyla. The analysis was done as described in Figure 3.4a and it involved 30 nucleotide sequences. There were a total of 540 positions in the final dataset.
Sequences from the phylogenetic tree (Fig. 3.4) were also aligned with OTUs related to Archaea among the iTag sequences (Fig. S3.3). Except for those OTUs related to the Crenarchaeae Marine Benthic Group B, the phylogeny presented in Figure 3.4 was supported by clustering observed in Fig. S3.3, and the topology of the two trees was very similar to each other.

3.4. Functional potential of the microbial community

To investigate the functional potential of the microbial community and to look into differences in the potential for organic matter degradation at the different sites we employed a new tool (PICRUSt; Langille et al., 2013), which is based on function inference for related, completely sequenced genomes. Energy metabolism, membrane transport, and signal transduction KO pathways were targeted as a means to identify processes that may be related to the different geochemical environments previously described for each site (Chapter 2; Matheus Carnevali et al., 2015).

Consistently, predicted gene counts related to transporters, two-component systems, oxidative phosphorylation, and ABC transporters were significantly different between sites in the upper sediment layer (Fig. 3.5). ABC transporter and oxidative phosphorylation predicted gene counts were found in higher proportion (~ 1.5 and ~ 4 % respectively) in Siq, whereas other transporters were more abundant in SukS. A similar trend was observed in the middle sediment layers, except that Siq-M estimated gene counts related to bacterial secretion systems,
oxidative phosphorylation, methane metabolism, and carbon fixation pathways in prokaryotes were also significantly higher in Siq than in SukS-M. At the bottom sediment intervals there were significant differences between proportions in predicted gene counts for secretion systems, CH₄ metabolism, carbon fixation pathways in prokaryotes, nitrogen metabolism, and transporters, all of which were higher in proportion in Siq-B. Oxidative phosphorylation and ABC transporters predicted gene counts were more abundant in SukS-L.

Differences in proportions between Siq and SukB were reflected in estimated gene counts related to transporters (~6-8% higher in SukB-U and -M), two-component systems (~2.5% higher in Siq-U), and oxidative phosphorylation (~1-2.5% higher in Siq-U and -M). In the bottom sediment interval, secretion systems, CH₄ metabolism, carbon fixation pathways, and other transporters were ~1-3% higher in Siq-L than in SukB-L, whereas oxidative phosphorylation, ABC transporters, and two-component systems were ~1-3% higher in SukB-L than in Siq-B (Fig. S3.5). Methane, nitrogen, and sulfur metabolism estimated gene counts were generally higher in proportion (~0.2-2%) in Siq than in SukB, except that in SukB-L sulfur metabolism was higher in proportion (~0.2%) than in Siq. Photosynthesis-related gene counts were also higher in proportion in SukB-L.

The most notable differences in proportion between sites within Sukok Lake (Fig. S3.6) were in estimated gene counts related to ABC transporters and other transporters, which were higher in proportion (~2-3%) in SukB-U, and higher in proportion (~1-2.5%) in SukS-M. Two-component system estimated gene counts
Figure 3.5 Extended error bar plots for the proportion of Siq and SukS sequences related to energy metabolism and environmental information processing KEGG categories filtered by effect size (> 0.2 %). KEGG Orthologs were ordered according to the difference between proportions (note the different scales). The Benjamini-Hocherg's FDR approach was used as a multiple test correction at a false discovery rate of 0.05 % (q-value > 0.05 %).
were higher in proportion (~ 2.5 %) in SukS-U; and photosynthesis proteins, bacterial secretion systems, methane metabolism, and carbon fixation pathways in prokaryotes were ~ 1 % higher in proportion in SukB-M. Nitrogen metabolism estimated gene counts were slightly higher in proportion in SukB-L.

4. Discussion

To fully understand the functioning of ecosystems at risk of undergoing dramatic changes due to climate warming, it is necessary to know the diversity and structure of the microbial communities that inhabit them. Yet, in the case of arctic thermokarst lakes, which are sensitive to the warming trend in the Arctic (Osterkamp, 2005; McGuire et al., 2009; Thienpont et al., 2013) very few studies have been conducted. Our work aimed (i) to understand how different geochemical environments in two arctic thermokarst lake sediments on the Coastal Plain of Alaska influence the diversity and composition of the microbial community (and vice versa) and, (ii) to identify what members of the community were potentially involved in organic matter degradation processes that ultimately lead to methanogenesis

4.1. Bacterial assemblage diversity, composition, and function

Based on an NMDS analysis of the Bray-Curtis dissimilarity between DNA sequences determined by iTag sequencing (DNA iTags) (Fig. 3.1a), the OTU abundance of the subsampled microbial community was different between sites, particularly between
Siq and SukS. Furthermore, analysis of aRNA iTags (Fig. 3.1b) suggested that the portion of the community that was potentially active in the sediments was more similar between depths within each site than between sites. These results were also supported by hierarchical clustering of archaeal SSU rRNA gene bands analyzed by DGGE in 2010 (Fig. S3.3). The differences in geochemical profiles (Chapter 2) may drive the differences observed here between sites, while at the same time the interaction between the microbial community and the environment may be altering these differences.

Microbial diversity as estimated by Simpson's reciprocal index was high in comparison to other environments (e.g., snowpacks in Chapter 3, where the highest value for this index estimated for DNA was 19), particularly in Siq and SukB. Diversity among the DNA iTag sequences decreased with depth, following the trend of decrease in TOC with depth in Siq and the low amounts of TOC present at higher depths in all three sites (Chapter 2). The high levels of diversity observed among the aRNA iTags in SukS-M were surprising, and could perhaps be explained by a more efficient amplification process of the total RNA extracted from this sediment interval. The lower amounts of TOC present at this site suggest a lower presence of humic and fulvic acids that may interfere with the amplification reaction.

Given that we were interested in bacterial OTUs that may be actively involved in organic matter degradation processes, we used the aRNA sequences to indicate potentially active bacteria and to study the functional potential of the community. The ribosome to cell ratio is proportional to growth rates, and RNA
molecules are less stable than DNA when exposed to the environment, therefore rRNA is considered a reflection of active populations (Putkinen et al., 2009). Good correlation ($r > 0.8$) has also been found between distributions of SSU rRNA sequences, metatranscriptomes, and metagenomes from permafrost soils (Hultman et al., 2015). In view of the similarity between the aRNA and RNA sequences (~ 42-68 % Bray-Curtis similarity), especially in the upper sediment intervals, we looked at the most frequently occurring OTUs in the aRNA and RNA data sets (Fig. 3.2). In only a few cases the aRNA may have overrepresented a particular OTU as a result of the linear amplification process. The aRNA and RNA OTUs representing potentially active microorganisms in the sediments were related to three main bacterial phyla: Actinobacteria, Bacteroidetes, and Proteobacteria of the classes Deltaproteobacteria and Gammaproteobacteria. At least two of these phyla have been associated to organic matter degradation processes (Madigan et al., 2014).

Though in different proportions and based on different experimental approaches, a similar composition of bacteria (including all the taxa related to OTUs comprising 1 % of sequences or more) has been observed in other arctic environments: permafrost, active layer and a thermokarst bog around Fairbanks, AK (Hultman et al., 2015); permafrost-affected polygonal soil in the Canadian Arctic (Frank-Fahle et al., 2014); permafrost-affected soils of Northeast Greenland (Ganzert et al., 2014); peat layers of a wetland in European North Russia (Serkebaeva et al., 2013); High-Arctic peat soils in Svalbard (Tveit et al., 2013); active layer and permafrost from a wetland in the Canadian High Arctic (Wilhelm et
al., 2011); permafrost from Nunavut, Canada (Steven et al., 2007); and most importantly in Toolik Lake on the North Slope of Alaska (Crump et al., 2012). Similar Archaea composition has also been found in the freshwater lakes of the Yunnan Plateau (Zhang et al., 2015).

Among the OTUs that comprised 1 % of the aRNA or RNA sequences or more, the Bacteroidetes were dominant members in the Siq bacterial assemblage, and may include obligate anaerobic, fermentative, and saccharolytic bacteria that can ferment sugars and proteins, and make acetate and succinate (Madigan et al., 2014). Other members of this group specialize in the degradation of complex polysaccharides like cellulose (e.g., Cytophagas sp.), though others are obligate aerobes with limited fermentative capabilities (Madigan et al., 2014). Members of this phylum have also been found to reduce NO$_3^-$, degrade starch, and to be psychrophilic or psychrotolerant (e.g., many members of the order Flavobacteriales) (Madigan et al., 2014).

Another phylum that was common to all sites and that could potentially participate in organic matter degradation processes is the Actinobacteria, which includes members that are facultative anaerobes and are able to degrade cellulose, lignin and chitin (e.g., Actinomyces sp.) (Madigan et al., 2014). Likewise, the phylum Verrucomicrobia and the Class Betaproteobacteria include chitinolytic species (e.g., Chitinomonas sp.). Together, members of these bacterial groups may lead the first steps of organic matter decomposition in artic lake sediments.
Other abundant OTUs that were commonly found in all three sites were related to the Deltaproteobacteria (Fig.3.2). Members of this class often have defined biogeochemical characteristics that are phylogenetically constrained, and mostly involved in sulfate reduction. This is an important finding, because there was at least 50 times more sulfate in upper SukB and SukS sediments than in Siq sediments (Chapter 2), and the highest aRNA Deltaproteobacteria signal that were potentially active were found in SukB. In SukB-L, members of the order Syntrophobacterales (Deltaproteobacteria) comprised ~10% of the aRNA and RNA sequences. In Siq, the Syntrophobacterales comprised 2.5–5% of the sequences, and OTUs related to the genus Desulfobacca, which is an acetate-degrading sulfate-reducing bacterium (Oude Elferink et al., 1999) were most common. The presence of this group in Siq, may explain the trend of decrease in SO\textsubscript{4}\textsuperscript{2-} concentration with depth in the upper sediment intervals in Siq.

Among the Syntrophobacterales there are species capable of fermenting pyruvate, fumarate, and malate (Madigan et al., 2014), \textit{i.e.} they are able to carry out the final fermentative steps that precede methanogenesis, and could help explain the findings from Siq regarding CH\textsubscript{4} as well. There are also species that are commonly found in syntrophic association with methanogenic Archaea in the deep biosphere (\textit{e.g.}, Smithella sp.), and participate in methanogenic degradation of petroleum hydrocarbons (Kallmeyer & Wagner, 2014). Other species have been found to oxidize propionate to make acetate, CO\textsubscript{2}, and H\textsubscript{2}, and can also live in syntrophy with H\textsubscript{2}-consuming bacteria (Madigan et al., 2014). Yet, other species
(e.g., *Syntrophomonas wolfei*) can oxidize butyrate to produce either H\textsubscript{2} or formate, both substrates for hydrogenotrophic methanogenesis. In our previous work with these sediments, we detected a small petroleum contribution in the organic matter of Suk (Matheus Carnevali et al., 2015). Petroleum is usually composed of various kinds of hydrocarbons (e.g., alkanes such as ethane, propane, and butane) with different degrees of susceptibility to bacterial degradation, a process that can be affected by low temperature and binding of hydrocarbons to other soil components (Barathi & Vasudevan, 2001; Olajire & Essien, 2014). These factors may affect the ability of microorganisms like the Syntrophobacterales-related OTUs to degrade organic matter. This may help explain the lag phase observed prior to biological CH\textsubscript{4} production in SukB sediments under experimental conditions, as well as the lower amount of CH\textsubscript{4} produced, and what looks like the lack of utilization of the high amounts of SO\textsubscript{4}\textsuperscript{2-} available in SukB sediments.

Another group that was abundant in SukB and SukS (2.5-10 % of the sequences) was Desulfobacterales-related OTUs. Sulfate-reducing bacteria (SRB) in this group were related to members of the families Desulfobacteraceae, Desulfobulbaceae, and Desulfarculaceae. Some species (e.g., *Desulfococcus oleovorans*) are capable of utilizing C\textsubscript{12}-C\textsubscript{20} alkanes as growth substrates, producing fatty acids (Aeeckersberg et al., 1998; So et al., 2003), and others like members of the genus *Desulfobulbus* are considered “power cords” (Pfeffer et al., 2012). Most importantly, this last genus has been found in association with anaerobic methanotrophs (ANME) (Orphan et al., 2001; Orphan et al., 2002; Holmes et al.,
These bacteria can use alternative electron acceptors like iron (Holmes et al., 2004; Beal et al., 2009), which was found in high concentrations in the sediments of SukB and Siq (Chapter 2), suggesting that in these sediments these microorganisms may be involved in iron reduction, and could potentially support CH$_4$ oxidation this way. The trend of SO$_4^{2-}$ consumption in SukS remains to be explained, as well as the possible mechanisms for its formation by other members of the community (e.g., green sulfur bacteria in the phylum Chlorobi) in both SukB and SukS.

Other notable groups of bacteria that were abundant in specific samples like SukS-U include the Sphingomonadales (2.5% of sequences), some of which can metabolize a broad range of organic and aromatic compounds; and the Rhodospiralles (2.5% of sequences), which are a metabolically versatile group of anoxygenic phototrophs, N$_2$-fixing bacteria, denitrifiers, methylotrophs and acetic acid bacteria (Madigan et al., 2014). In addition, Pseudomonadales-related OTUs in SukB-L (10% of sequences) are known chemoorganotrophs that can use NO$_3^-$ for anaerobic respiration (Madigan et al., 2014), and several can perform complete denitrification.

Among the OTUs with the highest number of sequences in Siq (~5%), those related to the phylum Nitrospirae and to the order Methylococcales (2.5-5% of the sequences) stood out. Some members of the phylum Nitrospirae are known for the autotrophic oxidation of NO$_2^-$ to NO$_3^-$. This group of bacteria, together with Rhodospiralles- and Pseudomonadales-related OTUs, may be involved in the cycling of N detected in all three sites. The Methylococcales are a well-known group of type
I methanotrophs (Madigan et al., 2014) and may be indicative of mixing of lake water and sediments, because they typically perform aerobic CH₄ oxidation and were also found in abundance in the water column (Priscu, unpublished). In SukB, the presence of unicellular Cyanobacteria of the orders Gloebacterales, Synechococcales, and Chroococcales was also surprising. These bacteria are oxygenic photoautotrophs and require sunlight as their main source of energy.

4.2. Arctic thermokarst lake Archaea in the North Slope of Alaska

Given the essential role that methanogenic Archaea play in biological CH₄ production, which was detected in the sediments of Siq and SukB (Chapter 2), we investigated the previously unknown structure, composition, and phylogeny of Archaea in the sediments of these thermokarst lakes. Based on iTag sequencing and DGGE, different groups of Archaea were distributed between these arctic lakes. OTUs related to methanogens were mostly found in Siq, and OTUs related to other groups of Archaea were mostly found in Suk. This partition of methanogens is highly consistent with the differences in biological CH₄ production observed in these lakes.

Among the groups found in SukB there were aRNA OTUs related to Candidatus Methanoperedens nitroreducens (ANME-2d), and Parvarchaea YLA114. Microorganisms in the ANME-2d group have been found to use NO₃⁻ as the terminal electron acceptor in the anaerobic oxidation of CH₄ in consortia with anaerobic ammonium oxidizing bacteria (Haroon et al., 2013). Sequences that were related to the Parvarchaea YLA114, an environmental sequence retrieved from Yellowstone
Lake (Kan et al., 2011) were part of a small group of sequences that does not have cultured representatives in the databases, thus supporting the novelty of this group. Parvarchaeota are part of an archaeal superphylum that comprises Archaea of small cell sizes and genomes (Rinke et al., 2013; Castelle et al., 2015; Baker et al., 2010). The only group that was consistently found in all three sites despite their biogeochemical differences was composed of sequences related to the MCG, perhaps supporting the idea that these Archaea are metabolically diverse (Kubo et al., 2012).

To study the phylogeny of Archaea from these unexplored arctic lake sediments, we cloned ~800-900 bp fragments of the SSU rRNA genes using primers that targeted Archaea and that spanned overlapping regions of the gene (either 109F-915R or 20F-958R). Most of the sequences in the clone libraries were related to the MCG described by Kubo et al. (2012) as a separate clade from the Thaumarchaeota and the Crenarchaeota. The MCG is a very diverse (76% - 87.7% intragroup similarity of the SSU rRNA gene) and ubiquitous group of Archaea that includes clones isolated from freshwater environments (e.g., pRfc26), deep sea hydrothermal vents, mud volcanoes, methane hydrates in marine sediments, among several others. They were thought to be involved in CH₄ cycling, but new evidence suggests that does not seem to be the case (Kubo et al., 2012), and that they thrive in low energy subsurface environments. Further, the archaeal assemblage in sediments with abundant MCG has been shown to use fossil organic matter as a carbon source (Biddle et al., 2006). The high number of sequences that were related to this group may be a result of primer bias (20F-958R) when compared to Archaea.
detected with the other primer set (109F-915R). However, their presence was supported by the iTaq sequences (Fig. S3.4), and some of the closest relatives in the GenBank and the Silva databases were obtained in geographical regions that are also cold environments (e.g., other arctic environments, the Tibetan Plateau, wetlands and other freshwater bodies, and permafrost environments).

Our findings also corresponded with the phylogeny of Archaea described by Großkopf et al. (1998) and Glissman et al. (2004), whose classification system has been used as a reference by other authors studying permafrost and water bodies in the Arctic. Sequences related to the Euryarchaeota in our clone libraries were related to methanogens in the Class Methanomicrobia, mostly with members of the family Methanoregulaceae in the Methanomicrobiales, and the family Methanosoaetacea in the Methanosarcinales. The remaining sequences formed clusters with uncultured members of the Euryarchaeota that have been previously identified as Rice Clusters I-V (Großkopf et al., 1998). Rice Cluster II (RC-II) also included sequences from sediments in the Tibetan Plateau, biodegraded soil from a Canadian reservoir, and submarine permafrost from the Laptev Sea. Sequences retrieved from polygonal ponds of northeastern Canada (Negandhi et al., 2013) and Finnish fen-type peatlands (Putkinen et al., 2009) also fall in this cluster. Other sequences were related to the Rice Cluster III (RC-III) (Großkopf et al., 1998) and belonged to a different group of Euryarchaeota, the family Methanomasilicoccaceae in the Thermoplasmatales, which includes a new group of methanogens that reduces methanol with H₂ (Dridi et al., 2012).
A cluster composed of uncultivated sequences only, which we refer to here as an ‘Arctic Cluster’, comprised two groups of sequences: i) sequences mostly related to the Rice Cluster V (RC-V), many of which have been reported from Arctic habitats (Großkopf et al., 1998; Galand et al., 2006; Galand et al., 2008; Høj et al., 2008), and ii) sequences retrieved from an arctic river, a stamukhi lake, and the Arctic Ocean (Galand et al., 2006; Galand et al., 2008), Lake Dagow Sediment (LDS) in Germany (Glissman et al., 2004), and the Candidate division WCHD3-30, which fall into the LDS cluster (Galand et al., 2006; Høj et al., 2008). Sequences in this cluster were also related to sequences from deep sea hydrothermal vents (DHVEG, Deep Sea Hydrothermal Vent Euryarchaeal Group-6) (Takai & Horikoshi, 1999), in correspondence with sequences related to Parvarchaea YLA114 (Kan et al., 2011) and findings by Galand et al. (2006). In parallel to the last study, the diversity of Archaea in these arctic lakes was unusually high (Fig. 3.3, Fig. S3.1 and S3.2), in which many phylotypes were apparent upon DGGE analysis of amplified DNA, archaeal SSU rRNA gene sequences and iTags. Sequences in this cluster, though common to a variety of arctic environments, are cosmopolitan as well, yet their functional significance remains unknown.

4.3. **Functional potential of the microbial community**

We also studied the functional potential of the microbial community in a separate analysis that involved the use of predicted metagenomes that may be associated to the SSU rRNA sequences we obtained, and the abundance of these sequences was a proxy for the abundance of the KEGG orthologs that may be expressed by the
community. A high proportion of predicted transporter genes and oxidative phosphorylation genes were observed at all sites, followed by two-component signal transduction systems, secretion system, methane metabolism, and carbon fixation genes. The prominent proportion of predicted transporter genes found in this study correlated with recent findings of abundant protein transporters in the proteome of a thermokarst bog (Hultman et al., 2015). Predicted nitrogen metabolism (ammonia oxidation, nitrite reduction, nitrate reduction, nitric oxide reduction, and nitrogen fixation), photosynthesis, and sulfur metabolism (sulfite reduction and sulfate transport) genes were observed in a minor proportion.

Significant differences in proportion (> 0.2 %) between Siq-M/B and SukS-M/B (Fig. 3.5), and SukB-M and SukS-M (Fig. S3.5) were observed for CH$_4$ metabolism, supporting previous observations of major differences in biological CH$_4$ production potential between the three sites. Furthermore, the proportion of predicted genes related to CH$_4$ metabolism (~ 9 %) was usually greater than the proportion of predicted genes associated to nitrogen and sulfur metabolisms, except for the upper sediment layers of both Suk sites. There was a significant difference ($q$-value < 0.0001) between proportions of predicted nitrogen metabolism genes between Siq-L (> 1 %) and SukS-L, and between SukB-L and SukS-L (Fig. S3.6), suggesting that this kind of energy metabolism may be more active in Siq and SukB. Predicted sulfur metabolism genes were consistently present at all three sites.

These differences in predicted gene proportion also reflect the observed distribution of Archaea in the lakes, where the majority of methanogen-related
OTUs were found in Siq (Fig. 3.3), and the majority of OTUs related to archaeal ammonia oxidizers of the phylum Thaumarchaeota were found in Suk. Conversely, the abundance of bacterial OTUs (Fig. 3.3) related to the phylum Nitrospirae in Siq are indicative of the occurrence of nitrite oxidation to $\text{NO}_3^-$ uncoupled to ammonia oxidizers. The higher abundance of SRB in SukB may be explained by the higher concentrations of $\text{SO}_4^{2-}$ available there.

Even though the abundance of Cyanobacteria detected in Suk was significantly higher than the abundance of Cyanobacteria in Siq, there was only a minor difference in proportion ($< 0.5 \%$) in predicted genes related to photosynthesis and carbon fixation in photosynthetic organisms. The presence of a substantial Cyanobacteria signal is a puzzling finding that warrants further interrogation, and their presence in the sediments may be explained by mixing with the water column. Alternatively, there may be potentially long turnover times resulting in preservation of cyanobacterial RNA molecules.

Together, our findings indicate that the composition of the microbial community inhabiting these arctic thermokarst lakes may be influenced by the geochemical environment that characterizes the sediments of each lake. Likewise, the microbial community may be altering the cycling of nutrients in these lakes, as shown by the abundance of methanogens in Siq sediments and the higher potential for biological $\text{CH}_4$ production observed in this lake. However, further studies are necessary to establish the degree of correlation with the environment. Overall, the bacterial assemblage has the potential to efficiently degrade organic matter of mixed
composition, and together with the Archaea, to contribute to the cycling of nitrogen, sulfur, and CH$_4$. Finally, the phylogeny of Archaea indicates the presence of novel groups of Archaea that merit further investigation.

**Acknowledgements**

Special thanks to Megan Rorhssen, Pamela Santibañez, Dan Berisford, Heather Adams, John Priscu, Andy Klesh, John Leichty, and the Barrow Arctic Science Consortium for assistance in the field. We are very grateful to Olivia Rassuchine, Vivian Peng, Ema Kuhn, and Gareth Trubl for technical assistance in the lab; and to John Mejia for writing scripts for iTag sequence processing. We specially thank the Joint Genomic Institute for sequencing effort. iTag sequencing was supported by an award to AE Murray, JGI-634 by the DOE Joint Genome Institute. PMC was supported in part by the Division of Earth and Ecosystem Sciences, DRI. KPH acknowledges support through the Jet Propulsion Laboratory (JPL), California Institute of Technology, under contract with the National Aeronautics and Space Administration (NASA). Financial support for this research was provided by the NASA Astrobiology Institute, Astrobiology of Icy Worlds program at JPL.
References


**Supplementary Tables and Figures**

Table S3.1: Tukey’s test comparing the log transformed abundances of phyla (based on aRNA sequences comprising ≥ 1% of the sequences). Phyla and sites were interacting factors and depths act as an independent variable.

<table>
<thead>
<tr>
<th>X-Y variables</th>
<th>Difference</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cyanobacteria:SIQ-Actinobacteria:SIQ</td>
<td>2.901</td>
<td>0.041</td>
</tr>
<tr>
<td>Cyanobacteria:SIQ-Bacteroidetes:SIQ</td>
<td>3.807</td>
<td>0.002</td>
</tr>
<tr>
<td>Nitrospirae:SUFB-Bacteroidetes:SIQ</td>
<td>-2.981</td>
<td>0.032</td>
</tr>
<tr>
<td>Nitrospirae:SUFS-Bacteroidetes:SIQ</td>
<td>-3.399</td>
<td>0.008</td>
</tr>
<tr>
<td>Deltaproteobacteria:SIQ-Cyanobacteria:SIQ</td>
<td>3.601</td>
<td>0.004</td>
</tr>
<tr>
<td>Gammaproteobacteria:SIQ-Cyanobacteria:SIQ</td>
<td>3.092</td>
<td>0.022</td>
</tr>
<tr>
<td>Actino:SUFB-Cyanobacteria:SIQ</td>
<td>3.458</td>
<td>0.006</td>
</tr>
<tr>
<td>Bacteroidetes:SUFB-Cyanobacteria:SIQ</td>
<td>3.267</td>
<td>0.012</td>
</tr>
<tr>
<td>Delta:SUFB-Cyanobacteria:SIQ</td>
<td>3.962</td>
<td>0.001</td>
</tr>
<tr>
<td>Actino:SUFS-Cyanobacteria:SIQ</td>
<td>3.114</td>
<td>0.020</td>
</tr>
<tr>
<td>Bacteroidetes:SUFS-Cyanobacteria:SIQ</td>
<td>3.558</td>
<td>0.004</td>
</tr>
<tr>
<td>Deltaproteobacteria:SUFS-Cyanobacteria:SIQ</td>
<td>3.657</td>
<td>0.003</td>
</tr>
<tr>
<td>Nitrospirae:SUFS-Deltaproteobacteria:SIQ</td>
<td>-3.191</td>
<td>0.016</td>
</tr>
<tr>
<td>Nitrospirae:SUFS-Actinobacteria:SUFB</td>
<td>-3.050</td>
<td>0.026</td>
</tr>
<tr>
<td>Nitrospirae:SUFS-Bacteroidetes:SUFB</td>
<td>-2.856</td>
<td>0.048</td>
</tr>
<tr>
<td>Nitrospirae:SUFB-Deltaproteobacteria:SUFB</td>
<td>-3.137</td>
<td>0.019</td>
</tr>
<tr>
<td>Nitrospirae:SUFS-Deltaproteobacteria:SUFB</td>
<td>-3.554</td>
<td>0.004</td>
</tr>
<tr>
<td>Nitrospirae:SUFS-Bacteroidetes:SUFS</td>
<td>-3.148</td>
<td>0.018</td>
</tr>
<tr>
<td>Nitrospirae:SUFS-Deltaproteobacteria:SUFS</td>
<td>-3.247</td>
<td>0.013</td>
</tr>
</tbody>
</table>
Figure S3.1 Denaturing gradient gel electrophoresis (DGGE) of the variable region 3 (V3) in the SSU rRNA gene of Archaea from different sediment intervals in Ikrl10 (A, April 2010 samples, O, October 2010 samples) and Siq10. Genomic DNA was amplified with the Archaea-specific primer GC_358F and the UNIV517R. PCR reactions were pooled to obtain 600 ng of DNA. The denaturing gradient on the gel was between 30 % and 65% (top to bottom). The electrophoresis was performed at 1000V/hr for 16 hr at 60 °C. Most phylotypes were grouped by depth and at least one phylotype was common to all lakes.
Figure S3.2 Denaturing gradient gel electrophoresis (DGGE) of the variable region 3 (V3) in the SSU rRNA gene of Archaea from different sediment intervals in SukS10. Genomic DNA was amplified with the Archaea-specific primer GC_358F and the UNIV517R. PCR reactions were pooled to obtain 600 ng of DNA. The denaturing gradient on the gel was between 30% and 65% (top to bottom). The electrophoresis was performed at 1000V/hr for 16 hr at 60 °C. Most phylotypes were grouped by depth and at least one phylotype was common to all lakes.
Figure S3.3 Unweighted pair group method with Arithmetic Mean (UPGMA) hierarchical clustering analysis of DGGE phylotypes based on Euclidean distances. Samples for this analysis were collected in 2010. Numbers next to the lake initials are sediment intervals in cm. *Ik 7-9 corresponds to April 2010.
Figure S3.4 Maximum likelihood phylogenetic tree of archaeal SSU rRNA gene Sanger sequences and iTag sequences. Hollow symbols indicate that the given clone represents other clones from the same lake. Filled symbols indicate that the given clone represents other clones from the same lake and from other lakes as well. Lakes are coded by symbol as follows: Siq – circles, SukB – triangles, SukS – squares. Only bootstrap values of 50 or more are shown. The evolutionary history was inferred by using the Maximum Likelihood method based on the Jukes-Cantor model. The tree with the highest log likelihood (0.0000) is shown. The percentage of trees in which the associated taxa clustered together is shown next to the branches. Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 81 nucleotide sequences. All positions containing gaps and missing data were eliminated. There were a total of 250 positions in the final dataset. Evolutionary analyses were conducted in MEGA6.
Figure S3.5 Extended error bar plots for the proportion of Siq and SukB sequences related to energy metabolism and environmental information processing KEGG categories filtered by effect size (> 0.2%). KEGG Orthologs were ordered according to the difference between proportions (note the different scales). The Benjamini-Hocherg’s FDR approach was used as a multiple test correction at a false discovery rate of 0.05 % (q-value > 0.05 %).
Figure S3.6 Extended error bar plots for the proportion of SukB and SukS sequences related to energy metabolism and environmental information processing KEGG categories filtered by effect size (> 0.2 %). KEGG Orthologs were ordered according to the difference between proportions (note the different scales). The Benjamini-Hocherg’s FDR approach was used as a multiple test correction at a false discovery rate of 0.05 % (q-value > 0.05 %)
Chapter 4

Unique structure of the microbial community associated with snow algae in the Western U.S.

Co-authorship statement

P. B. Matheus Carnevali, T. H. Painter, K. P. Hand, and A. E. Murray

PM, TP, KH, and AM designed the sampling strategy and collected samples. AM designed the experiments and performed statistical analyses. PM and AM executed the experiments. PM analyzed the data, and wrote this chapter under the supervision of AM. TP, KH, AM secured funding for the research at CNS, and AM was part of a field course in which samples at WP and PCT were collected. AM secured support for the NGS.
Abstract

The snowpack in high mountain ranges across the world is an ephemeral environment that allows the exchange of nutrients and microorganisms between the atmosphere, the ground, and water bodies that are fed by meltwater. A variety of microorganisms including unicellular algae, bacteria, fungi, and ciliates inhabit the snowpack; and during spring and summer, when liquid water is available algal blooms (e.g., *Chlamydomonas nivalis*) are observed as red snow. Studies about the diversity, composition and function of the microorganisms inhabiting seasonal snow cover are sparse. The present study aimed to uncover the diversity and structure of the microbial community in snowpacks with visible snow algae, and to understand the mechanisms that drive community assembly. Snow samples were collected from Mt. Conness (CNS) in the Sierra Nevada of California, which was considered our primary study site. Chemical analyses of these samples to determine concentrations of dissolved inorganic nutrients, dissolved organic carbon (DOC), particulate carbon (POC), particulate nitrogen (PN) and Chlorophyll *a* (*Chla*) were conducted. The snow microbiota was studied by quantifying algal and bacterial cells, and by MiSeq (Illumina) next-generation sequencing (NGS) of the SSU rRNA gene (iTag sequencing) of the three domains of life from some of these samples. An additional site in the Sierra Nevada (the Pacific Crest Trail, PCT), and another site in the Snake Range in eastern Nevada (Wheeler Peak, WP1) were sampled for comparison of the microbial communities. Analysis of a resampled dataset of bacterial iTag sequences...
indicated that the bacterial assemblage had a unique structure. In addition to the traditional ‘singletons’ (OTUs with just 1 sequence) that dominate NGS datasets, OTUs that comprised a high number of sequences were found at only one site (‘singleton OTUs’). Interestingly, these OTUs were mostly affiliated to two different families of Betaproteobacteria (the Comamonadaceae and the Oxalobacteraceae) in the genera *Polaromonas* and *Herminiimonas*. Even though at the OTU level the bacterial assemblage was only 5% similar between some sites, there was a high resemblance in taxonomic composition among all three sites, and 7 ‘core’ OTUs were affiliated to the Phyla Bacteroidetes and Actinobacteria. Variations in structure of the abundant OTUs between sites (and depths) were reflected in the proportion of sequences related to the same families of Betaproteobacteria and groups of Bacteroidetes. To elucidate possible mechanisms for microbial community assembly, and to explain this unique bacterial assemblage structure we tested two hypotheses: that the presence of snow algae determined bacterial assemblage structure, and that environmental factors affected the structure. An inverse relationship between Chla and bacterial diversity was found, and this finding was supported by significant, positive, and negative Spearman’s rank correlations between algal OTUs (from a parallel dataset) and bacterial OTUs, particularly between core bacterial OTUs and OTUs affiliated to the Chlorophyceae. Only two significant correlations were found between core bacterial OTUs and environmental factors (DOC and PN). Further testing using multivariate statistical approaches may allow us to formulate possible explanations to similar microbial community
composition geographically distant regions, and determine what mechanisms drive community assembly in these snowpacks.

1. Introduction

The so-called “red snow”, commonly found in snowpacks in high mountain ranges across the world, owes its coloration to the UV-resistant carotenoid pigments produced by green algae such as *Chlamydomonas nivalis* (Duval & Hoham, 2000; Thomas & Duval, 1995; Thomas, 1972). These algae inhabit an environment that is around freezing temperatures (Kawecka & Drake, 1978) most of the year, but during the spring and the summer, when liquid water is available, nutrients entrapped in the snow become available and promote algal growth (Komárek & Nedbalová, 2007). These conditions are also suitable for the presence of prokaryotes and a variety of uni- and multi-cellular eukaryotic organisms (*e.g.*, fungi, ciliates, insects, rotifers); and some bacteria have even been found to be attached to the algal cells (Weiss, 1983; Segawa et al., 2005; Takeuchi & Kohshima, 2004).

Studies about the diversity, composition and function of the microorganisms inhabiting seasonal snow cover are sparse (Larose et al., 2013b) and most studies have focused on the snow algae (Takeuchi & Kohshima, 2004; Duval & Hoham, 2000; Thomas & Duval, 1995; Muller et al., 1998; Kviderova, 2010; Lutz et al., 2014) which are composed mostly of different species of green alga (*e.g.*, *C. nivalis*,...
Chloromonas spp., Mesotaenium berggrenii, and Raphidonema nivale) (Takeuchi & Kohshima, 2004; Hoham, 1989). Studies about the bacterial assemblage inhabiting the snow in cold environments have been conducted, but not necessarily in connection with snow algae, and these studies encompass a range of alpine and polar environments such as glaciers in the Tibetan plateau (Liu et al., 2009; Zhang et al., 2010), the Tateyama Mountains in Japan (Segawa et al., 2005), Antarctica (Fujii et al., 2010), and the Arctic (Larose et al., 2010; Harding et al., 2011; Moller et al., 2013). Only recently, a study looked at microbial communities in red snow as a whole in Antarctica, and detected the presence of green algae, fungi, and bacteria by analysis of the SSU rRNA gene, with a predominant signal of bacteria in the genus Hymenobacter (Fujii et al., 2010). Also, next-generation sequencing was used to study the microbial community in High-Arctic snow, where the top and middle layers of the snowpack were dominated by bacteria in the phyla Proteobacteria, Bacteroidetes and Cyanobacteria (Moller et al., 2013).

The ephemeral nature of the snowpack adds a layer of complexity to the study of snow as an ecosystem, because the snowpack is an interphase that lasts for varying amounts of time (< 1 to 6 months in the Sierra Nevada depending on elevation and aspect) and allows the exchange of nutrients, microorganisms, and contaminants between the atmosphere, the ground, and water bodies that are fed by meltwater (Larose et al., 2013a). Little is known about how environmental factors affect the microbial community overall. However, light, temperature, snowmelt, and nutrients affect the life cycle of green algae flagellated stages, and other factors like
pH can affect algal growth in the lab (Hoham, 1989). Conversely, the presence of snow algae can have an effect on the snow albedo (Thomas & Duval, 1995) and on CO₂ fluxes from snowfields (Sommerfeld et al., 1991). Therefore, studying the microbial community inhabiting the snowpack would allow: i) to understand how the presence of microorganisms in the snow may contribute to cooling or warming of the climate, and how the drought conditions that the Western U.S. is currently experiencing may ultimately affect this ecosystem; ii) to aid future studies using remote sensing technologies to estimate microbial biomass in snowfields, because they may benefit from taking into consideration the microbial community composition and the variety of cell wall properties of the different microorganisms that make up the community. This knowledge would also contribute to expanding the database of spectral biosignature that can be detected remotely.

The present study aimed to uncover the diversity and structure of the microbial community in snowpacks with visible snow algae (red snow), and to understand the mechanisms that drive community assembly. For the purposes of this chapter, we will refer to all members of the snow bacteria occupying a given sample as an assemblage, given that we do not a priori know about the interactions between these organisms, though the concept of community assembly will be discussed. Using the conventional terminology, community is a group of potentially interacting organisms that co-occur in space and time (Magurran, 2003). To address these issues, we looked at assemblage diversity and structure and tested two hypotheses: (1) that an algal bloom (visible red snow) would be usually associated
with the same groups of microorganisms, even in geographically distant locations, and (2) that environmental selection could determine the structure and composition of the entire community. In order to test these hypotheses we sampled two sites in the Sierra Nevada of California and another site in the Snake Range in eastern Nevada. Chemical analyses of snow samples were conducted to determine dissolved inorganic nutrients, dissolved organic carbon (DOC), particulate carbon (POC), particulate nitrogen (PN) and Chlorophyll \( a \) (Chla) concentrations. The snow microbiota were studied by quantifying algal and bacterial cells, and by next-generation sequencing of the SSU rRNA gene of the three domains of life from some of these samples.

2. Methods

2.1. Area of study and samples collection

Mount Conness (CNS), located north of Tioga Pass (37° 58′01″ N, 119° 19′17″ W), on the border between Inyo National Forest and Yosemite National Park, is one of the peaks (12,649 ft) of the Sierra Nevada range in California. Samples were collected on 7 July 2009 in the snow fields of CNS at a location of a long term study site where previous surveys tracked the presence of Chlamydomonadaceae-related algae, \textit{e.g.} \textit{Chlamydomonas nivalis}, on ground-based and airborne imaging spectrometer data (Painter et al., 2001). At the base of the mountain, a \( \sim \) 12 m long horizontal transect (Fig. S4.6a) was established along the snow field. Between 0.05 cm\(^3\) and 0.1 cm\(^3\) of snow were collected at surface using sterile instruments, at 1, 2, 3, 4, 5, 7, 9 and 11
m from one end of the transect, and then stored in sterile Whirl Pak bags. Between meter 4 and meter 5, a snow pit was excavated and samples were collected along a vertical profile at the surface (0 m), and 0.2, 0.3, 0.6 and 0.9 m below the surface, in the same way samples were collected from the horizontal transect (Fig. S4.6b). At the time of collection three whirlpak bags (∼ 1 L ea.) of snow were obtained from each of the sites for subsequent analysis (e.g., chemical analyses, cell counts, nucleic acids extractions). All samples were stored under cold conditions during transportation to the lab either on the same or subsequent day. Other samples were collected in a similar fashion in July 2010 from the Pacific Crest Trail (PCT; 39° 17′ 06″N, 120° 19′ 10″ W; 18 July), which is another sampling site in the Sierra Nevada of California, and from Wheeler Peak (WP1; 38° 59′ 34″ N, 114° 18′ 26″ W; 14 July) in Great Basin National Park, which is the second tallest peak in Nevada. Samples from the PCT were collected at 2,406 m.a.s.l. and samples from WP (WP1) were collected at 3,292 m.a.s.l. CNS was our primary sampling site, and the other two sites were mainly used for microbial assemblage comparisons.

2.2. Confocal microscopy

In preparation for microscopy, snow samples from all three sites were thawed at room temperature, mixed well, and aliquoted in 50 mL conical tubes. Ten percent of the volume of the sample was then added as formalin 37 % to fix the cells. Staining was performed using 1 mL of 4′,6′-diamidino-2-phenylindole (DAPI, 0.1 mg mL⁻¹) and 10 mL of each sample. Samples were kept in the dark for at least 10 min with
rotation, and subsequently passed through a 0.2 µm GTBP filter and a 0.45 µm HAWP filter using a vacuum pump. Finally, the filters were placed on clean glass slides with immersion oil and cover slips. The slides were examined by confocal microscopy (Olympus, FluoView™ FV1000). Visualization of the algae under the microscope was done at a magnification of 200X, and bacteria were visualized at a magnification of 600X. Only CNS samples counts are presented here.

2.3. Chemical analyses

Snow samples from CNS were thawed and 20-60 mL were passed through combusted GF/F filters and collected into clean bottles. The filtrates were used for analysis of dissolved organic carbon (DOC), nitrite (N-NO₂), nitrate (N-NO₃), ammonia (N-NH₃), and phosphate (O-PO₄) concentrations, according to procedures outlined in Standard Methods for the Examination of Water and Wastewater (Greenberg et al., 1992). Filters were kept for particulate organic carbon (POC) and particulate nitrogen (PN) analysis, following Karl et al. (1991). Another 6-50 mL portion of each sample (from CNS and from WP1) was filtered through GF/F filters and kept at -20 °C until processing. Filters were then incubated in 10 mL of 90 % acetone in the dark at -20 °C, for less than 24 hr in order to extract chlorophyll. Chlorophyll a was measured following the method by Welschmeyer (1994) using a fluorometer (Tuner Designs Model 10-AU).
2.4. Nucleic acids extraction and next generation sequencing

The snow samples that had been kept at -80 °C were thawed for two days at 4 °C and filtered through a 0.2 μm pore size Sterivex filter (Millipore) using a peristaltic pump. Sucrose lysis buffer (1.8 mL; 40 mM EDTA, 50 mM Tris-HCl, 0.75 M sucrose) was added to each filter, and the filters were preserved at -20 °C until further processing. Cell lysis was completed by adding lysozyme (1 mg mL⁻¹) and the filters were incubated at 37 °C for 30 minutes, combined with an addition of proteinase K (0.5 mg mL⁻¹) and 20 % sodium dodecyl sulfate (SDS), and incubation at 55 °C for 2
hr. A standard phenol:chloroform nucleic acids extraction was performed on the lysates following methods in (Massana et al., 1997), with the exception that two rinses with chloroform-isoamylalcohol (24:1) were done in place of just one. The aqueous phase was finally passed through a Microcon filter to clean the DNA using sterile DI water and concentrate it to a volume of ~ 80 µL. DNA quantification was done using a NanoDrop 100. DNA extracts were preserved at -80 °C.

DNA extracts from selected samples and all three sampling sites was used to amplify and sequence the variable region 4 (V4) of the SSU rRNA gene of Bacteria and Archaea, with primers 515F-806R (Caporaso et al., 2010) and the V4 region of the SSU rRNA gene of Eukarya with primers V4Forward – 5’-CCAGCASCYGCGGTATTCC-3’ (Saccharomyces cerevisiae positions 565-584) and V4Reverse – 5’-ACTTTCGTTCTTGATYRA-3’ (S. cerevisiae positions 964-981) (Hadziavdic et al., 2014). Extracts were processed in the Joint Genomic Institute (JGI) for paired-end sequencing (2 x 300 bp) using the MiSeq Illumina platform (iTaq sequencing), and contaminants, barcodes, and primers were removed from the raw fastq files by JGI.

2.5. iTag sequences analyses and statistics

Sequences were analyzed following the MiSeq SOP outlined for Mothur (Kozich et al., 2013). In brief, sequences were separated into individual sample files containing forward and reverse reads using deinterleave_fastq.sh

(https://gist.github.com/nathanhaigh/3521724). These sequences were joined into
contigs, and quality filtered by removing sequences with any ambiguous bases, or that were longer than 325 bp. Unique sequences were identified and aligned against the newest release of the Greengenes database (August, 2013), preclustered, classified, and clustered into operational taxonomic units (OTUs) at a distance of 0.03 following the MiSeq SOP (Kozich et al., 2013). The remaining sequences were used to create a database (using the mothur script, create.database) containing representative sequences for each OTU formed at a distance of 0.03 and the OTUs were classified as well. A perl script (Daisychopper.pl) was used to subsample the sequences to the lowest number of sequences in a given sample (103,385 sequences). The subsampled data set was used for alpha-diversity and beta-diversity statistical analyses. Spearman’s rank order correlations were calculated with the statistical package Statistica Data Miner.

3. Results

3.1. Cell counts

Confocal microscopy was used to observe and quantify algal and bacterial cells in samples collected from CNS. The number of both kinds of cells was on the same order of magnitude ($10^3$ cells mL$^{-1}$), but the number of bacterial cells was at least 2.5-fold higher. The highest numbers of algae were observed at meter 4 on the horizontal transect, and at the surface (0.0 m) of the vertical profile (between meters 4 and 5), which correlated with field observations of the highest amounts of “red snow” in the area sampled. The number of algal cells was higher between
meters 7 and 11 (~ 1.4 x 10^3 cells mL^-1) than at meters 1, 2, 3 and 5 (< 1.3 x 10^3 cells mL^-1; Fig. 1A). The opposite pattern was observed for bacteria, whose highest numbers were detected between meters 1 and 3, particularly at meter 2, where the highest number of bacteria of the whole horizontal transect was found. The number of both, algal and bacterial cells decreased with depth, but it was evident that the algae showed similar numbers of cells between the horizontal transect and the vertical profile (1.1-1.8 x 10^3 cells mL^-1).

3.2. Chemical analyses

Snow samples from all sites were collected in July, a time when the snowmelt was well underway and the snowpack was changing. At CNS, a suite of chemical parameters that could potentially affect or be affected by microbial life were determined. The concentration of chlorophyll a (Chl a) was measured as a proxy for the presence of photosynthetic microbial life, particularly green algae, in addition to performing cell counts, mentioned in the previous section. The Chl a concentration was relatively constant along the horizontal transect (0.011-0.164 µg L^-1), except at meter 9, where it was the highest (1.000 µg L^-1) (Fig. 1B). A similar trend was observed for NO_2-N (1-2 µg L^-1), as well as the concentration of PO_4-P, although at meters 1 and 9 there were slightly higher concentrations (5 µg L^-1) of the later (Fig. 1B). The highest DOC concentration (18 µg L^-1) was also observed at meter 9 (Fig. 1A), as well as the highest concentration of Chl a. At this same location, an increase of NO_3-N and NH_3-N concentration was observed, which was only surpassed by the
concentrations of these nutrients observed at the end (11 m) of the horizontal transect (33 and 24 µg L\(^{-1}\) respectively). Particulate organic carbon (POC) was found in low amounts (< 11 mg L\(^{-1}\)) along the horizontal transect (Fig. 4.2A), and the highest concentration was detected at meter 3, where the concentration of Chl\(a\) also showed a small peak, and the number of bacterial cells was higher than in other locations (Fig. 4.2A). The concentration of PN was consistently lower than that of POC (≤ 0.16 mg L\(^{-1}\)) along the horizontal transect. Only samples collected at meters 2 and 11 on the horizontal transect, and at 0.2 m below surface, had N to P ratios that were close to the Redfield ratio 16:1 (Redfield, 1934) with values of 14.00, 14.74, and 15.50 respectively. The C to N ratios that were closest to the Redfield ratio (6:625) were found at meters 5 (5.845) and 11 (5.96) on the horizontal transect and at 0.9 m below surface (7.17). Only at these sites where the Redfield ratios were close to the expected values nutrients may not have been limiting.

Nitrogen in the form of nitrite (NO\(_2\)-N) was the only nutrient whose concentrations (1 µg L\(^{-1}\)) were low and consistent throughout the different depths. Across the depth profile, the highest amount of POC (37.88 mg L\(^{-1}\)) was detected at the surface, with the concentration being ~ 4 times the concentration of POC detected at meter 3 of the horizontal transect. This result is not surprising, because at this site, abundant red snow was observed at the time of collection, and high numbers of algae were observed under the microscope, explaining the high content of particulate matter. However, at a nearby location (meter 4 of the horizontal transect) the concentration of POC was much lower (1.87 mg L\(^{-1}\)), as well as the
Figure 4.2 Cell counts and snow chemistry along the horizontal transect and the vertical profile in CNS on July 2009. Algal and bacteria counts, particulate carbon (POC), particulate nitrogen (PN) and dissolved organic carbon (DOC) concentrations along the transect (A) and the vertical profile (C). Concentration of chlorophyll a (Chl a), nitrogen in the form of nitrate (NO$_3$-N), ammonia (NH$_3$-N), and nitrite (NO$_2$-N), and orthophosphate (OPO$_4$-P) along the transect (B) and the vertical profile (D).
concentration of DOC and inorganic nutrients, exemplifying a relatively patchy distribution of the chemical constituents measured. The concentration of PN across the vertical profile was consistently low (≤ 1.59 mg L⁻¹).

### 3.3. Microbial community structure and diversity based on iTag sequences

To gain deeper insights into the structure of the microbial community inhabiting snow, iTag next-generation sequencing (MiSeq, Illumina) of the bacterial and archaeal SSU rRNA gene (V4 region) was applied to samples collected during summer from three locations CNS, PCT and WP1. Surface or near-surface snow algae were present at each location at varying levels, enabling us to compare the structure of the community between areas dominated by snow algae and areas with less visible snow algae. After sequence processing following MiSeq SOP (Kozich et al., 2013) there was between 103,385 and 147,860 sequences per sample. A subset of the CNS samples (10) were sequenced, and only samples from the first 40 cm in the snowpack were compared between sites at WP1 and PCT. These sequences initially clustered into 70,326 OTUs at a distance of 0.03, then following subsampling to the lowest number of sequences (103,385), resulted in a total of 45,992 OTUs.

Given that the majority of OTUs were singletons (35,679 OTUs were only represented by a single sequence), doubletons and tripletons, all OTUS with 3 or few sequences were removed prior to comparative analyses. The remaining OTUs (4,416) encompassed between 99,048 and 101,406 sequences, and fell into three groups: those OTUs that comprised 1.00 % of the sequences or more (127 OTUs),
OTUs that comprised between 0.01 and 0.99 % of the sequences (1,947 meso-abundant OTUs), and those with less than 0.0099 % of the sequences (2,411 OTUs). This last group corresponds to our definition for rare OTUs, which was within the range used by Hugoni et al. (2013). To simplify the analysis of the meso-abundant and the rare OTUs, only OTUs that were present in 3 samples or more were retained for analysis.

The structure of the OTU data set was distinctive such that 71 % of all OTUs were present in only one sample (e.g. 55 OTUs out of the 127 OTUs comprising ≥ 1 % of the sequences were singleton OTUs; Fig. 4.3). These are referred to as ‘singleton OTUs’ and could even dominate a given snow sample, i.e., in 5 of the 15 samples singleton OTUs represented > 10% of the sequences per sample. Similar assemblage structures were observed among the meso-abundant and rare ‘singleton OTUs’ (Fig. S4.1). Among the abundant, meso-abundant and rare groups there usually was 1 abundant OTU that comprised a higher percentage of sequences (0.60 - 0.95 % in the last two groups), particularly at the surface snow layers of all three sites, and the outer sites on the CNS horizontal transect (Fig. S4.1). Strikingly, even though the most abundant OTUs were unique, the taxonomic composition of ‘singleton OTUs’ was similar among the abundant, mesoabundant and rare OTU groups (data not shown). In the CNS samples the majority of ‘singleton OTUs’ were related to *Herminiimonas* (Class Betaproteobacteria, Order Burkholderiales, family Oxalobacteraceae). In the PCT and the WP1 samples there was a mixture of *Herminiimonas*, and *Polaromonas*, another genus of Burkholderiales in the family
Figure 4.3 OTUs found in only one sample (singleton OTUs) that comprised ≥ 1% of sequences (subsampled sequence dataset including OTUs with 3 sequences or more). Only 10 out of the 15 samples were represented on this figure. Missing samples did not have ‘singleton OTUs’ comprising 1% of the sequences or more, but they have representative ‘singleton OTUs’ that comprised < 1% of the sequences (Fig. S4.1). Coloring scheme matches Fig. S4.1 for comparative purposes.
Comamonadaceae. Other well-represented genera observed in these sites were *Rhodoferax* (Comamonadaceae) and *Massilia* (Oxalobacteraceae).

When considering the assemblage as a whole, OTUs related to the families Comamonadaceae and Oxalobacteraceae were among the abundant OTUs, and comprised the largest number of sequences in all samples (Fig. 4.4a). In the family Comamonadaceae, OTUs related to *Polaromonas* were more abundant in the depth profiles of PCT and WP1, with increasing numbers of sequences with depth. In CNS, *Collimonas*, another genus within this family was more prevalent among samples. However, except for the sample collected at meter 7 on the CNS transect, most samples from CNS were dominated by OTUs related to members of the Chitinophagaceae, Sphingomonadaceae and other unclassified OTUs in the phylum Bacteroidetes.

There was a common “core” set of OTUs related to Bacteroidetes (OTUs 1-4 and OTU8) and Actinobacteria (OTUs 6 and 7) present in all samples, though the proportions varied widely amongst the 15 samples (*e.g.* ranging from 0.01 - 32 % for a single OTU; Fig. S4.2). Notably, there was a sharp difference in abundance between the core OTUs present in the Sierra Nevada and those present in Wheeler Peak. WP1 was dominated by Bacteroidetes-related OTUs, OTU2 (average 24.5 ± 13.0 %; Class Cytophagia) and mostly OTU1 (average 4.7 ± 3.2 %; Class Saprospirae), as opposed to a more even distribution of the other OTUs observed in the Sierra sites. Across the Sierra sites Actinobacteria, classes Saprospirae (OTU1) and Sphingobacteria (OTU4), and other unclassified Bacteroidetes (OTU3) averaged
Figure 4.4a Relative abundance of the abundant OTUs (≥ 1% of the sequences) grouped by the highest taxon possible. Taxa were grouped by color, each color representing a different phylum or class where appropriate. From top to bottom: Phylum Actinobacteria (Family Microbacteriaceae, Order Actinomycetales); Phylum Bacteroidetes (Family Chitinophagaceae, Genus Hymenobacter, Family Sphingobacteriaceae, Unclassified Bacteroidetes); Phylum Cyanobacteria, Class Alphaproteobacteria (Family Acetobacteraceae, Genus Novosphingobium, Family Sphingomonadaceae); Class Betaproteobacteria (Genera Polaromonas and Rhodoferax, Family Comamonadaceae, Genera Collimonas and Herminiimonas, Family Oxalobacteraceae); Class Gammaproteobacteria (Genus Pseudomonas). Sample IDs include the location in capital letters (CNS: Mt. Conness, PCT: Pacific Crest Trail, WP1: Wheeler Peak) followed by distance from one end of the horizontal transect (1 m, 3 m, 4 m, 7 m, 11 m) or by snow pit depth. Depths represent the bottom of the depth range, except for CNS-0.0 m that represents the upper end.
Figure 4.4b Relative abundance of the meso-abundant OTUs (0.01 – 0.99 % of the sequences) grouped by Phylum or Class (same color scheme as Fig. 4.4a) to show degree of resemblance to abundant OTUs at the indicated taxonomic level. Sample IDs include the location in capital letters (CNS: Mt. Conness, PCT: Pacific Crest Trail, WP1: Wheeler Peak) followed by distance from the end of one horizontal transect (1 m, 3 m, 4 m, 7 m, 11 m) or by snow pit depth. Depths represent the bottom of the depth range, except for CNS-0.0 m that represents the upper end.
Figure 4.4c Relative abundance of the rare OTUs (< 0.0099% of the sequences) grouped by Phylum or Class (same color scheme as Fig. 4.4a, b) to show degree of resemblance to abundant OTUs at the indicated taxonomic level. Sample IDs include the location in capital letters (CNS: Mt. Conness, PCT: Pacific Crest Trail, WP1: Wheeler Peak) followed by distance from one end of the horizontal transect (1 m, 3 m, 4 m, 7 m, 11 m) or by snow pit depth. Depths represent the bottom of the depth range, except for CNS-0.0 m that represents the upper end. Note: the % of Gemmatimonadetes was negligible, and color of the bars cannot be distinguished.
11.7 ± 8.4 %. The main different between the two Sierra sites was the abundance of OTU1, which in the PCT site was ~2.4-fold higher than at CNS.

Most OTUs in the meso-abundant group were ‘singleton OTUs’, and the taxonomic composition of meso-abundant OTUs was different from the other two groups of OTUs (Fig. 4.4b). OTUs in this group were consistently related to the phylum Bacteroidetes, the same phylum containing the core OTUs, in which they comprised 0.24-0.66 % of the sequences. Additionally, there were OTUs related to Gammaproteobacteria present in all samples (0.13-0.82 % sequences). Also, in WP1, OTUs related to the Betaproteobacteria were most abundant in comparison to the other OTUs (1-3 % sequences).

In the group of rare OTUs (Fig. 4.4c), there was a much lower percentage of sequences related to the Betaproteobacteria (0.001-0.02 % sequences), but OTUs related to the Bacteroidetes were consistently present across samples, in addition to Acidobacteria, Actinobacteria, Alphaproteobacteria and Gammaproteobacteria. Some other groups that were not detected in the other two groups appeared in this group, the Deltaproteobacteria, SAR406 and OTUs related to phyla of the domain Archaea.

Sequence coverage was > 96 % for all the samples, as determined using Good’s coverage (Good). The bacterial community presented low alpha-diversity, and the inverse Simpson’s index (Table S4.1) indicated that the diversity within samples was low (CNS: 7.6-15.7; PCT: 11.4-12.3; WP1: 5.6-18.8; CNS-Transact: 7.4-
The sample collected at meter 4 was the most diverse (Inverse Simpson’s Index of 16.16) on the CNS transect, while the 0.2 m horizon in the CNS profile had highest diversity (Inverse Simpson’s Index of 15.66) in the CNS vertical profile. However, the deepest (0.4 m) samples in the PCT and WP1 were the most diverse in those profiles (Inverse Simpson’s Index of 12.31 and 18.80 respectively). Intra-sample rarefaction curves for the observed number of OTUs reflected similar results (Fig. S4.3). Furthermore, there were low levels of similarity at the OTU level (Fig. 4.5) among samples. Similarity between CNS and PCT was 23 to 29 %, between CNS and WP1 it was 8 to 13 %, and between PCT and WP1 it was 5 to 27%, in all cases increasing with depth. The Unifrac weighted approach was used to test the probability that the dissimilarity among samples was not greater than expected by chance. Results from this test indicated that this probability was low, and that dissimilarity among vertical profile samples from the different sites was significant for all three sites ($p < 0.001$).

3.4. **Factors that drive community assembly**

The most important questions we intended to answer regarding microbial community diversity and structure in summer snow with varying levels of snow algae was whether there were specific relationships between algal signals and certain groups of bacteria, or if there was a relationship that leads to higher diversity levels (Connell & Orias, 1964). Alternatively, the prevalence of certain bacterial groups over others may be the result of the effect of the environment
(selection by environmental factors) on the entire microbial community (Bacteria and Eukaryotes).

![Dendrogram of Bray-Curtis dissimilarities clustered with the unweighted pair group method using the arithmetic averages (UPGMA) algorithm.]

The relationship between algae and bacteria was first tested by studying the Spearman's rank correlation between Chla (as a proxy for green algae in the Class Chlorophyceae) and the inverse Simpson's diversity index (Fig. S4.4). A negative relationship between snow samples with higher algal signals and bacterial diversity was observed in the vertical profiles from CNS and WP1, where Chla data was obtained. There was a significant negative linear relationship (p < 0.05) in
which diversity decreased, while at the same time Chla increased. Furthermore, in WP1 diversity clearly increased with depth (Table S1) while Chla decreased in concentration (Fig. S4.7). A similar pattern was observed for CNS, but not as strong as in WP1.

To further test the hypothesis that there was a relationship between specific algal and bacterial OTUs we used the iTag sequencing data of the SSU rRNA of Eukaryotes that were obtained for some of the samples and analyzed by JGI (CNS-4m, CNS-7m, CNS-0.2m, CNS-0.3m, and all the samples from PCT and WP1) in parallel with the bacterial data set described here. Spearman’s rank correlations were calculated with the most abundant bacterial OTUs (≥ 1 % of sequences or more) and 8 OTUs affiliated with the Chlorophyceae class of green algae in the Eukaryota iTag data set. A positive correlation (r > 0.693, p < 0.05) was found between core OTUs (OTU 1, 3, 6 and 8) and some of the algae-related OTUs (Fig. S4.5). Additional positive correlations were found with other abundant bacterial OTUs, most of which were found in higher proportion in WP1 than in the other two sites. Relationships observed (Fig. S4.5): (i) positive relationships, e.g. OTU65397 increased in abundance with the increase in abundance of a Chlamydomonadaceae-related iTag (OTU0000033; r=0.937, p< 0.000); (ii) Bacterial OTUs co-occurred with chlorophyte OTUs following a “threshold” of bacterial OTU occurrence. For example, OTU00001 (affiliated with Saprospirae) was present at levels between 5000 to 10,000 tags with little corresponding chlorophyte signal, then increased in samples where the abundance of chlorophyte OTUs related to the families
Chlamydomonadaceae (OTU000067; $r=0.706$, $p=0.015$) and Mamiellophyceae (OUT000112; $r=0.814$, $p=0.002$) were detected at high levels ~ 20,000 tags. OTU3 had a similar pattern, in which the iTag abundance increased once a *Chloromonas nivalis*-related OTU (OTU000037) was at a level > 15,000 tags. (iii) OTU4638 (affiliated to Comamonadaceae) appeared to be inversely related ($r=-0.956$, $p=0.044$) to the abundance of *Chloromonas hohamii* (OTU000039).

To test the second hypothesis, the relative abundance of the same ‘core OTUs’ was used for Spearman’s rank order correlations with inorganic nutrients, Chla, DOC, POC, and PN from the CNS transect. Only two significant correlations were found, a correlation between OTU8 (affiliated with Saprospirae) and PN ($r=0.9; p = 0.037$), and between OTU2 (affiliated with Cytophagia) and DOC ($r=0.9; p = 0.037$).

4. Discussion

4.1. Cell counts from Mt. Conness

Mean algal concentration on the top 10 cm of the snowpack in the east drainages of Mt. Conness was similar to those estimated by Painter et al. (2001) using imaging spectrometry (e.g., $2.5 \times 10^3$ cells mL$^{-1}$). Thomas & Duval (1995) studied standing crops of algae and bacteria at locations nearby during the late spring of 1994, and the numbers of bacterial cells they encountered were greater than the numbers of algae, particularly in snow samples where algal blooms were observed. Similar
findings were reported in our study, even though the numbers of algal cells obtained were one order of magnitude smaller than the numbers obtained here. In comparison to another site in the Southern Hemisphere, algal cell counts were an order of magnitude lower than the cell concentration observed for snow algae in a Patagonian glacier, where cell numbers also decreased with altitude. The authors attributed this trend to the occurrence of an earlier snowmelt at lower altitudes, leading to higher exposure to the harsh environmental conditions during the reproductive season, and to lower nutrient content in their sites (Takeuchi & Kohshima, 2004).

The snow in CNS contained bacterial cell counts that were well within the range of bacterial abundance reported for other snow samples from the Antarctic, the Arctic, and other high mountain regions including the Tibetan Plateau (Amato et al., 2007; Liu et al., 2009; Segawa et al., 2005; Zhang et al., 2010). Bacterial abundance was similar throughout the vertical profile, while numbers of algal cells showed a sharp drop from the surface to the greater depths, as Thomas (1972) previously reported. Additionally, bacterial counts were higher than the number of algal cells at all the depths at CNS, suggesting that there may have been ample nutrients to support growth in all the snow layers. Note that growth rates were not directly determined. However, this trend of higher abundance of bacteria than snow algae in CNS is inverted relative to what other authors found in nearby locations, where algae in red snow were tens to thousands of times more abundant than bacteria in red snow, and 6 to 32 times more abundant than bacteria in white snow;
and both algae and bacteria were more abundant in red snow than in white snow (Thomas & Duval, 1995). Studying succession through the stages of snow algal bloom development and demise is likely required to best address these dynamics.

4.2. Snowpack dynamics and abiotic parameters related to the microbial community

The snowpack is a heterogeneous environment, composed of grains of snow in a range of sizes, ice layers, and different water saturation levels, all of which reduce the permeability to air and water (Larose et al., 2013a). Major mobilization of nutrients and particles only happens during snowmelt, late in the spring and early in the summer, when the temperature of the snowpack pack rises above 0 °C, or when there is intense solar radiation (Larose et al., 2013a; Kuhn, 1987). The first 30 % of meltwater removes about 80 % of solutes from the snowpack via percolation of meltwater, and repeated melting and freezing increases solute concentration, especially in the deeper layers of the snowpack (Larose et al., 2013a).

Snow algae are known to thrive only in wet snow that lasts throughout the summer months, as well as in old summer snow that has undergone cycles of freeze-thaw. This old summer snow contains snow crystals surrounded by a continuous water film, thick enough to create a microenvironment around the algae (Remias, 2012). The thin water layer between the algae and the snow crystals is the main source of inorganic nutrients, and inorganic elements that accumulate on cell surfaces may be taken and stored inside the cells (Lütz-Meindl & Lütz, 2006). The red-pigmented snow algae represent the climax stage of the life cycle of green algae
(Remias, 2012), in which the accumulation of carotenoids has been attributed not only to high levels of radiation in the surface snow, but also to N-, P-, S-, K- and Fe-deficiency (Bidigare et al., 1993; Muller et al., 1998). However, Kviderova (2010) found that phosphorous input from surrounding vegetation led to algal blooms in snowfields of the Giant Mountains in the Czech Republic. Algae can also deplete the nutrient content in the snow, particularly the pool of nitrogen, which has been correlated with lower conductivity in the snow (Hoham, 1989; Hoham et al., 1993; Jones, 1991).

Our samples were collected during the summers (early July) of 2009 and 2010 in areas where visible red snow algae were present (e.g. vertical profile from all three sites) in addition to surrounding areas with low concentrations of snow algae (some samples in the CNS transect). Considering the low the levels of nutrients in CNS samples (most samples were far from the Redfield ratio) in comparison to other snow samples adjacent to green snow algae (e.g. 0.13 mg L\textsuperscript{-1} O-PO\textsubscript{4}, 31.24 mg L\textsuperscript{-1} NO\textsubscript{3}-N, 0.049 mg L\textsuperscript{-1} NH\textsubscript{3}-N; Hoham & Mullet, 1977), it is likely that the CNS samples had undergone several rounds of snow melt, perhaps also explaining the spikes of DOC, NO\textsubscript{3}-N and NH\textsubscript{3}-N in the deepest samples. The concentration of nutrients in the transect samples was also, in general, much lower than the concentration of the same nutrients measured by Kviderova (2010). For instance, this study of snow algae in the Giant Mountains, Czech Republic (June 2008) detected 1,000 to 3,000 times more Chla than we did in areas covered with snow algae, although algal cell counts also were two orders of magnitude higher in
that study. Additionally, Thomas & Duval (1995), detected much higher levels of Chl\(a\) (200 µg L\(^{-1}\)) in patches of snow algae in the area around Mt. Conness.

Lack of correlation between the environmental factors measured in CNS and the cell abundance, suggested that other factors (e.g., pH, altitude, slope, time of sampling) may have a significant effect on the abundance of microorganisms in the snow. However, abundance is perhaps not the most indicative proxy for use in determining snow ecological relationships; functional ecology and species-specific interactions stand more informative. Therefore, we tested the relationship between Chl\(a\) concentrations and diversity in the vertical profiles from CNS, and a site adjacent to WP1.

4.3. **Diversity of the microbial community in comparison to other places**

MiSeq next-generation sequencing of the SSU rRNA gene indicated that the microbial community inhabiting the snow presented low levels of diversity, and it was composed of only 4 abundant phyla of organisms. These phyla corresponded to the same groups reported by other studies of snow samples from all over the world, and include the Proteobacteria (Classes: Betaproteobacteria, Alphaproteobacteria, and Gammaproteobacteria), Bacteroidetes, Actinobacteria, and Cyanobacteria (Harding et al., 2011; Liu et al., 2009; Fujii et al., 2010; Larose et al., 2010; Moller et al., 2013). Archaea were only present in the rare OTUs data set and were never more than 0.007 % of the sequences in a given sample. Even at higher taxonomic levels there was correspondence with groups reported by other studies, for
instance, *Hymenobacter*-related phylotypes (Phylum Bacteroidetes) were the dominant group detected in Antarctic snow (Fujii et al., 2010), and were among the most abundant OTUs detected in WP1 surface samples. This group is also commonly found in other kinds of environments, of which air samples (Buczolits et al., 2002) are of interest to this study. Some of the genera that were observed in glaciers on the Tibetan Plateau (Liu et al., 2009; Zhang et al., 2010), like *Sphingomonas* (Phylum Alphaproteobacteria), *Polaromonas*, and *Rhodoferax* (these last two in the family Comamonadaceae, Order Burkholderiales, Class Betaproteobacteria), were also observed in this study, and the structure of the microbial community described in those glaciers also showed variation in the relative abundance of these groups among different glaciers. Similar to this study, Harding et al. (2011) also reported high abundance of OTUs related to the order Burkholderiales, constituting 64% of the bacterial clone sequences detected in High Arctic snow, in addition to OTUs related to the Bacteroidetes. Interestingly, the genus *Herminiimonas* in the family Oxalobacteraceae (Order Burkholderiales) contains bacteria that have been isolated from spring and mineral waters, but have not been reported in other snow ecosystems to the best of our knowledge.

4.4. **Theory of community assemblage and sources of microorganisms**

According to Vellend’s framework for community assembly (Nemergut et al., 2013) there are two forces that may bring new organisms into communities: speciation and dispersal. These can be contrasted with processes such as ecological drift and
selection by environmental factors that may affect presence/absence, as well as the
relative abundance of microorganisms over time.

In this study, we hypothesized that in the snow ecosystem, the presence of
snow algae may be an important factor driving bacterial community structure. Snow
algae may provide a protective layer against the high irradiation at the snow surface
(Remias, 2012). Furthermore, they are considered essential primary producers in
the ecosystem, and CO₂ sinks (Painter et al., 2001; Williams et al., 2009), which may
favor the presence of groups of heterotrophic microorganisms. Thomas & Duval
(1995) detected more organic matter in red snow than in white snow, which was
favorable for bacterial growth in the late summer, supporting the hypothesis that
snow algae provide autochthonous sources of carbon for bacteria. Also, Fujii et al.
(2010) suggested that bacteria associated with snow algae in Antarctica are
secondary inhabitants of the snow. Our study showed that there was an inverse
relationship between Chlα concentration in the CNS and WP1 vertical profiles and
bacterial diversity, which decreased with higher Chlα concentrations, indicating that
only certain groups of bacteria were favored by the presence of the algae.
Furthermore, we found that certain bacterial OTUs were correlated with certain
algal OTUs, supporting the idea that snow algae may create a microenvironment
that is favorable for other microorganisms to grow.

We also tested the hypothesis that environmental factors could explain the
relative abundance of the core bacterial OTUs, however we only found two
significant correlations between an OTU affiliated to Saprospirae with PN, and
between an OTU affiliated to Cytophagia with DOC. Saprospirae and Cytophagia are unicellular gliders that may be able to degrade complex organic matter (Reichenbach, 2006). One aspect of microbial communities in snow ecosystems is the patchiness of its distribution, which may explain the overall lack of correlations observed here with environmental parameters. This kind of distribution could be due to eolian transport (Harding et al., 2011), because snowpacks receive inorganic nutrients (e.g., CO₂, N₂O, salt particles, mineral dust, and sulfates), organic matter (e.g. pollen and dead foliage), and contaminants by wet and dry deposition, i.e., either incorporated into falling precipitation or transferred directly to the snow surface (Larose et al., 2010).

Wet and dry deposition may drive community assembly by dispersal processes. Part of this deposition may be spore-forming bacteria, gram-positive bacteria, and fungi that are adapted to mobilization via aerial transport (Harding et al., 2011; Brinkmeyer et al., 2003). Segawa et al. (2005) also observed that a great proportion of bacteria detected in surface snow samples from the Tateyama Mountains in Japan, was delivered by mineral particles; and similar results were reported by Liu et al. (2009) for the Guoqu Glacier on the Tibetan Plateau. Harding et al. (2011) sampled air and snow in the High Arctic, and reported 71 SSU rRNA gene clones from air samples related to Bacteroidetes, Acidobacteria, Firmicutes, Alpha-, Beta- and Gamma-proteobacteria. Among these clones, the genera *Rhodoferax* and *Polaromonas* were detected in both snow and air samples, suggesting adaptation of these groups to aerial distribution, which makes them very
ubiquitous. Both were found in our study – in particular, microdiverse *Polaromoas*-affiliated iTags were among the dominant OTUs at all sites (up to 32% of the sequences in WP1). Darcy et al. (2011) have suggested a mechanism by which bacteria in the genus *Polaromonas* may withstand the stresses of atmospheric transport.

Different processes may be operating in conjunction and at different scales to determine community structure, for instance, a combination of dispersal and selection by environmental factors may explain the observed distribution of ‘singleton OTUs’. The singleton OTUs could also be a result of stochastic processes influencing the relative abundance of these groups; features such as weak selection, low alpha diversity, in addition to low numbers of community members have been described as characteristic features leading to ecological drift in macro-organisms (Chase & Myers, 2011). There is a possibility that the bacterial assemblage may be enabling algal growth as well, and an alternative mechanism by which bacteria might create a favorable microenvironment for algae should be investigated.

### 4.5. Future directions and closing remarks

To properly test hypotheses regarding mechanisms that may lead to snow community assembly and to analyze the relationship between environmental parameters and indices of alpha and beta diversity of the microbial community, more extensive spatial and temporal data sets should be examined with multivariate statistical approaches. This may lead to formulation of possible explanations of how
geographically distant regions share the same taxonomic groups, such as we found in this study. What is more, the uniqueness of the microbial community structure observed in this study, demands further investigation, to explain both the resemblance in taxonomic composition between the abundant and the rare OTUs, and the differences in structure between these two groups of OTUs. These differences included the high levels of dissimilarity at the OTU level between the snow samples.

Moreover, other groups of microorganisms in the kingdom Eukarya co-inhabit the snow with green algae, including other unicellular flagellates and ciliates, fungi, and rotifers (Segawa et al., 2005; Harding et al., 2011). To obtain a more complete picture of the structure of the microbial community associated to snow algae, and to identify the specific groups of algae that constitute the red snow, analysis of the sequenced V4 region of the SSU rRNA gene of the snow eukaryal assemblage in parallel with the bacterial and archaenal dataset will offer a significant layer of information. The next stage of this work involves analyses of these data using similar procedures, and the microbial assemblage described here will be re-analyzed in the context of those microorganisms.

Future studies of the microbial community inhabiting snow fields flourishing with snow algae should encompass the temporal aspect of these ecosystems, and sampling of possible sources of microorganisms, like air and soil. We also recommend to sample i) the snowpack at its highest point of accumulation, just before snowmelt starts, and several times during the snowmelt season; ii) to collect
meltwater as well as snow because it should incorporate most taxa that were present in the snow that season, as well as the soil in contact with the snow (Larose et al., 2010); and iii) to analyze multiple environmental parameters of the snow for the different kinds of samples.

**Acknowledgements**

We appreciate the guidance of KP Hand and TH Painter, Jet Propulsion Laboratory (JPL), California Institute of Technology, and the field teams at CNS in 2009, and members of the Nevada Geobiology Summer Short Course in 2010 that was led by CH Fritsen and D Crowther. We also appreciate field assistance and technical support provided by Vivian Peng. Financial Support for this research was provided by the NASA Astrobiology Institute, Astrobiology of Icy Worlds program at JPL, NASA Cooperative Agreement/Grant # NNG05G587H to D Crowther and CH Fritsen, and iTag sequencing was supported by an award to AE Murray, JGI-634 by the DOE Joint Genome Institute.
References


Hoham RW (1989) Snow as a habitat for microorganisms.


### Supplementary Tables and Figures

#### Table S4.1 Alpha diversity statistics of the snow samples.

<table>
<thead>
<tr>
<th>Sample</th>
<th>N(^o) Sequences</th>
<th>Good’s coverage</th>
<th>OTU richness ((S_{obs}))</th>
<th>Simpson’s reciprocal Index ((SRI))</th>
<th>SRI low confidence interval</th>
<th>SRI high confidence interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNS-1 m</td>
<td>103,385</td>
<td>0.97</td>
<td>4,293</td>
<td>13.44</td>
<td>13.29</td>
<td>13.60</td>
</tr>
<tr>
<td>CNS-3 m</td>
<td>141,637</td>
<td>0.98</td>
<td>3,615</td>
<td>7.47</td>
<td>7.40</td>
<td>7.55</td>
</tr>
<tr>
<td>CNS-4 m</td>
<td>119,606</td>
<td>0.96</td>
<td>5,753</td>
<td>16.16</td>
<td>16.00</td>
<td>16.32</td>
</tr>
<tr>
<td>CNS-7 m</td>
<td>110,130</td>
<td>0.96</td>
<td>5,309</td>
<td>13.88</td>
<td>13.75</td>
<td>14.01</td>
</tr>
<tr>
<td>CNS-11 m</td>
<td>132,940</td>
<td>0.97</td>
<td>5,559</td>
<td>13.22</td>
<td>13.12</td>
<td>13.33</td>
</tr>
<tr>
<td>CNS-0.0 m</td>
<td>127,275</td>
<td>0.97</td>
<td>4,295</td>
<td>7.57</td>
<td>7.48</td>
<td>7.66</td>
</tr>
<tr>
<td>CNS-0.2 m</td>
<td>121,526</td>
<td>0.96</td>
<td>5,892</td>
<td>15.66</td>
<td>15.50</td>
<td>15.82</td>
</tr>
<tr>
<td>CNS-0.3 m</td>
<td>121,133</td>
<td>0.97</td>
<td>5,053</td>
<td>12.77</td>
<td>12.64</td>
<td>12.90</td>
</tr>
<tr>
<td>PCT-0.1 m</td>
<td>124,141</td>
<td>0.98</td>
<td>3,483</td>
<td>11.43</td>
<td>11.33</td>
<td>11.53</td>
</tr>
<tr>
<td>PCT-0.3 m</td>
<td>115,174</td>
<td>0.97</td>
<td>4,603</td>
<td>11.95</td>
<td>11.80</td>
<td>12.11</td>
</tr>
<tr>
<td>PCT-0.4 m</td>
<td>119,547</td>
<td>0.97</td>
<td>4,943</td>
<td>12.31</td>
<td>12.18</td>
<td>12.44</td>
</tr>
<tr>
<td>WP1-0.1 m</td>
<td>130,802</td>
<td>0.97</td>
<td>4,567</td>
<td>5.55</td>
<td>5.49</td>
<td>5.61</td>
</tr>
<tr>
<td>WP1-0.2 m</td>
<td>139,437</td>
<td>0.97</td>
<td>6,021</td>
<td>10.77</td>
<td>10.66</td>
<td>10.89</td>
</tr>
<tr>
<td>WP1-0.3 m</td>
<td>128,686</td>
<td>0.96</td>
<td>6,006</td>
<td>13.86</td>
<td>13.72</td>
<td>14.01</td>
</tr>
<tr>
<td>WP1-0.4 m</td>
<td>147,860</td>
<td>0.96</td>
<td>7,223</td>
<td>18.80</td>
<td>18.65</td>
<td>18.94</td>
</tr>
</tbody>
</table>
Figure S4.1 Abundance distribution of the ‘singleton OTUs’ comprising < 1 % of the sequences. Each bar represents a different OTU.
Figure S4.2 Relative abundance of the ‘core’ OTUs that comprised ≥ 1% of the sequences and that were present in all the samples. Coloring scheme according to Figure 4.4a.
Figure S4.3 Intra-sample rarefaction curves for the observed number of OTUs.
Figure S4.4 Spearman’s rank correlation between chlorophyll a and the inverse Simpson’s diversity index.
Figure S4.5 Spearman's rank correlations between algal OTUs and bacterial OTUs. a) Correlation ($r=0.937$, $p<0.000$) between algal OTU33 (Chlamydomonadaceae-related iTag) and bacterial OTU65397 (Polaromonas-related iTag - Black circles, from WP1). b) Correlation between algal OTU67 (Chlamydomonadaceae-related iTag) and bacterial OTU1 (Saprospirae-related iTag - Black circles) ($r=0.706$, $p=0.015$), algal OUT112 (Mamiellophyceae-related iTag) and bacterial OTU1 (Red circles) ($r=0.814$, $p=0.002$), and correlation between algal OTU37 (Chloromonas nivalis-related iTag) and bacterial OTU3 (Bacteroidetes-related iTag - Green triangles) ($r=0.864$, $p=0.003$). c) Correlation ($r=-0.956$, $p=0.044$) between algal OTU39 (Chloromonas hohamii-related iTag) and bacterial OTU4638 (Comamonadaceae-related iTag - Blue circles).
Figure S4.6 Snow sampling sites. a) Mt. Conness (CNS) snow field; transect sampled is indicated with the black dashed line. b) Closer up view of the transect sampled with meter flags indicating sampling sites. c) Pacific Crest Trail (PCT) snow field. d) Close up look at visible snow algae patch showing melted zones near densely colored areas. e) Wheeler Peak (WP1) snow pit top of snow was 77 cm above melt layer interface.
Figure S4.7 Wheeler Peak (WP) chlorophyll a concentration in the vertical profile.
Supplementary Information

S1. Supplementary Methods

S1.1 Amplification of the SSU rRNA gene of Bacteria, Archaea and Eukarya

To obtain a diversity profile of the microbial assemblage in the CNS samples, primers BactGC358f and Univ 517r were used to amplify the variable region 3 (V3) of the SSU rRNA gene, a ~220-240 bp fragment (Murray et al., 1996), including a 40 bp GC clamp added on to the forward primer. The PCR reaction mixture consisted of 1X PCR buffer with 1.5 mM MgCl$_2$, 0.2 mM of each dNTP, 0.5 µM of forward and reverse primers, 1 µg µl$^{-1}$ of BSA, 10 ng of template DNA, and 2.5 U of TaqDNA polymerase, in a 50 µl final reaction volume. Primers Euk GC960bf and Univ1200r with (Gast et al., 2004; Muyzer et al., 1993) were used to amplify the variable 6 (V6) region of the SSU rRNA gene, a ~ 300 bp fragment. PCR conditions were the same as above. The thermal-cycling program utilized to amplify the bacterial and eukaryal sequences was as follows: initial denaturation at 94 °C for 3 min, 10 cycles of 94 °C for 30 sec, touch-down primer annealing with the temperature decreasing from 65 °C to 55 °C (1 °C per cycle) for 30 sec, and 72 °C for 30 sec. Next, 18 cycles of 94 °C for 30 sec, 55 °C for 30 sec and 72 °C for 30 sec, and finally primer extension at 72 °C for 7 min. To screen for Archaea sequences, primers Arch20f and Univ958r were used (Delong, 1992) The expected size of the fragment was 950 bp. PCR conditions were the same as above, except that each reaction contained a 2 mM concentration of MgCl$_2$. Thermal cycling was done as follows: initial denaturation at 94 °C for 3
min, 38 cycles of 94 °C for 1 min, primers annealing at 55 °C for 1 min, 72 °C for 1 min 30 sec, and final extension at 72 °C for 7 min. Another set of primers for the variable region 3 of the SSU rRNA was used to screen for Archaea too, AR-GC_347F and UNIV519R. The PCR reaction mixture consisted of 1X PCR buffer, 2 mM MgCl\(_2\), 0.2 mM of each dNTP, 0.5 µM of forward and reverse primers, 10 ng of template DNA, and 2.5 U of TaqDNA polymerase, in a 50 µl final reaction volume. The same touch-down protocol utilized to amplify bacterial and eukaryal DNA was applied here.

S1.2. Diversity profiling of CNS samples using DGGE

Bacteria and Eukarya PCR amplicons from 5 x 50 µl PCR reactions were pooled and precipitated over night at -20 °C with two volumes of absolute ethanol and 1/10 of the volume of 3 M sodium acetate pH 5. After centrifugation in cold for 30 min, the DNA was washed with 70% ethanol, centrifuged for 5 to 8 min, air-dried, or dried in a DNA speed vac (Savant), and resuspended in 16 µl of sterile DI water. DNA was quantified using fluorometry and the PicoGreen dye. DNA samples (600 ng of Eukarya and 800 ng of Bacteria) were loaded onto two different 8 % polyacrylamide-bisacrylamide (37.5:1) gels. A 25 to 75 % gradient of denaturants (7 M urea and 40 % deionized formamide in 100 % concentration of denaturants) was used for Eukarya DNA, and a 25 % to 60 % gradient was used for Bacteria DNA. Electrophoresis was run at 60 °C for 15 h and 30 min at 63 V (Eukarya) or 16 h at 63 V (Bacteria). To prepare these gels and run the electrophoresis methods from
Murray (1996) were followed. Band identification and analyses of the banding patterns resulting from DGGE were done with the Quantity One software (BioRad).

S2. Supplementary Results

Figure S4.8a DGGE of PCR amplified fragments using general primers for Bacteria. (T) Horizontal transect. (V) Vertical profile. (T1) 1 m; (T2) 2 m; (T3) 3 m; (T4) 4 m; (T5) 5 m; (T7) 7 m; (T9) 9 m; and (T11) 11 m from one end of the transect. (V.0) Surface; (V.2) 0.2 m; (V.3) 0.3 m; (V.6) 0.6 m and (V.9) 0.9 m below surface. (S) Surface water sample from Alaska. (D) Deep water sample from Alaska.
Figure S4.8b Dendrogram of phylotypes observed by DGGE analysis of the SSU rRNA gene of Bacteria. Generated by UPGAMA (unweighted pair group method using arithmetic averages) cluster analysis using the Quantity One software (BioRad). Similarity scale: 0.0 being no similarity, and 1.0 being identical. (T) Horizontal transect. (V) Vertical profile. (T1) 1 m; (T2), 2 m; (T3) 3 m; (T4) 4 m; (T5) 5 m; (T7) 7 m; (T9) 9 m; and (T11) 11 m from one end of the transect. (V.0) Surface; (V.2) 0.2 m; (V.3) 0.3 m; (V.6) 0.6 m and (V.9) 0.9 m below surface. (A) Water samples collected in Alaska for comparison. (Surf) Surface. (Deep) Deep.
Figure S4.9a DGGE of PCR-amplified fragments using general primers for Eukarya. (T) Horizontal transect. (V) Vertical profile. (T1) 1 m; (T2), 2 m; (T3) 3 m; (T4) 4 m; (T5) 5 m; (T7) 7 m; (T9) 9 m; and (T11) 11 m from one end of the transect. (V.0) Surface; (V.2) 0.2 m; (V.3) 0.3 m; (V.6) 0.6 m and (V.9) 0.9 m below surface. (S) Surface water sample from Alaska. (D) Deep water sample from Alaska.
Figure S4.9b. Dendrogram of phylotypes observed by DGGE analysis of the SSU rRNA gene of Eukarya. Generated by UPGAMA (unweighted pair group method using arithmetic averages) cluster analysis using the Quantity One software (BioRad). Similarity scale: 0.0 being no similarity, and 1.0 being identical. (T) Horizontal transect. (V) Vertical profile. (T1) 1 m; (T2), 2 m; (T3) 3 m; (T4) 4 m; (T5) 5 m; (T7) 7 m; (T9) 9 m; and (T11) 11 m from one end of the transect. (V.0) Surface; (V.2) 0.2 m; (V.3) 0.3 m; (V.6) 0.6 m and (V.9) 0.9 m below surface. (A) Sea water samples collected in Alaska for comparison. (Surf) Surface. (Deep) Deep.
Supplementary References


Chapter 5

General Conclusions
What is life? Where did it come from? Where can we find it? We will continue the search for answers to these questions, not only as a quest for truth, but also as a quest for knowledge, because we—humans—yearn “to know”, and because we want to be surprised by the answers to these questions.

1. **Positive and negative feedback loops as a property of self-regulating systems**

I started my dissertation, thinking that there was something very counterintuitive about finding methanogens in not cold, but extremely cold environments. These microorganisms of the domain Archaea excrete CH$_4$ as a way to reach chemical equilibrium, and can operate with amounts of energy that are barely enough to make 1 ATP molecule ($G^o$' = -31.8 KJ mol$^{-1}$) (Madigan et al., 2014). CH$_4$ turns out to be a potent greenhouse gas (IPCC, 2007), though estimating total CH$_4$ emissions to the atmosphere is no easy task, considering the different sources and mechanisms for CH$_4$ formation. Water logged sediments of wetlands, rice paddies, and lakes provide anoxic environments with low redox potentials ideal for methanogenesis, and if in addition to that, there is high availability of labile organic matter, as in permafrost, all of these habitats are among the best for biological CH$_4$ production on the planet.

Thermokarst lakes on the North Slope of Alaska are frozen for months at a time, preserving for thousands of years organic matter that had only been partially degraded; and when permafrost thaws, this organic matter can be decomposed and utilized by microorganisms inhabiting the sediments. These lakes are also exposed
to 24 hours of sunlight during the summer, when temperatures are maintained well above 0 °C, which promotes primary production in the water column contributing fresh organic matter to the sediments as well. Moreover, thermokarst erosion at the bottom and edges of the lakes, releases more of the organic matter trapped in permafrost; and all of these features make of thermokarst lakes hotspots for CH$_4$ production (Walter et al., 2007).

In chapter 2 my research demonstrated that there is one lake (Siqlukaq) on the North Slope of Alaska where biological CH$_4$ production takes place at both, in situ sediment temperatures and at higher temperatures. Another lake (Sukok) is a conduit for thermogenic gas formed deep under the permafrost, and its sediments are composed of one-fourth of the organic matter detected in the surface sediments of Siqlukaq Lake. Surprisingly, we detected biological CH$_4$ production from another area (SukB) in Sukok Lake, when temperatures were increased 8-10 degrees above the in situ surface sediment temperatures. This result suggested that even though there were low amounts of organic matter in Sukok, microorganisms that are present in this lake can mineralize it under warmer conditions. Together, these findings led us to contend that under scenarios of climate change and warmer temperatures (also experienced during summer), there is a potential for increased biological CH$_4$ production in sediments of arctic thermokarst lakes on the North Slope of Alaska (Matheus Carnevali et al., 2015).

However, the same way there are processes that lead to CH$_4$ production, there are processes that lead to CH$_4$ consumption (He et al., 2012; Lofton et al.,
Other members of our team detected CH$_4$ oxidation in the water column of the lakes, together with microorganisms that are able to carry out this process; and we detected the presence of aerobic type I methanotrophs among the bacteria OTUs, and anaerobic methane oxidizing archaea among the Archaea OTUs (Chapter 3). Future work on CH$_4$ emissions from these lakes will benefit from including CH$_4$ flux calculations that account for CH$_4$ production and consumption in each compartment of the lakes (i.e., sediments, water column, and the ice layer). Also the understanding of CH$_4$ cycling should include sample collections across different seasons (e.g., summer) and different areas (e.g., shores) of the lakes. This information would enhance future scenario development and allow determining (i) if CH$_4$ emissions from arctic thermokarst lakes on the North Slope of Alaska have an impact on the warming trend that is currently observed in the Arctic, and (ii) what the consequences are for the greenhouse gas budget of our planet. It is important to remember that in addition to biological CH$_4$ production sensitivity to temperature, CH$_4$ consumption (oxidative processes), are also biological processes mediated by enzymes, and are responsive to temperature (Lofton et al., 2014). Therefore, it is possible that an increase in one process may be counteracted by an increase in the other. Chapter 2 of this dissertation constitutes a first approximation to the potential for biological CH$_4$ production from lakes on the Coastal Plain of Alaska, and I hope it will help constraining estimates of CH$_4$ emissions from the Arctic.
It is undeniable that the Arctic is changing; we can see it in pictures of icebergs breaking apart, permafrost slumping on the coastal areas, and changes in polar animals’ behavior. However, the effects of climate warming go beyond what we can see with our own eyes, and may have deeper consequences that we anticipate. Microorganisms are the key players of biogeochemical cycling on the planet, and microbial interactions with essential elements in the environment (e.g., C, N, S, Fe) determine their fate globally. Understanding the structure of microbial communities inhabiting climate-sensitive environments like permafrost, will allow us to predict how ecosystem functioning could be affected under different scenarios of climate change in the future.

2. Microbial communities in cold environments and their global distribution

2.1. Microbial communities in thermokarst lake sediments

Microorganisms that decompose organic matter in arctic thermokarst lakes set up the conditions for biological CH$_4$ production, a process that results from a chain of redox reactions that rely on a variety of electron donors and electron acceptors. These microorganisms are individual engines that act in concert, and as a whole, to regulate nutrient cycling on the planet. However, the diversity and structure of the microbial communities that inhabit different environments are determined in large part by the environment itself. Methanogens and their enzymes may be more efficient at warm temperatures, but the fact that there is active biological CH$_4$ production in the sediments of arctic thermokarst lakes indicates that these
enzymes must be adapted to function under the stresses imposed by cold temperatures as well. Additionally, these microorganisms may have developed mechanisms to remain inactive for long periods of time, as demonstrated by Rivkina et al. (1998) who found viable methanogens in thousands of years old permafrost samples. This may also explain why we were able to obtain methanogens from enrichment cultures inoculated with sediments from Sukok Lake (Appendix), and why the mcrA gene was detected by PCR amplification of DNA extracted from this lake, although it was found at levels that were too low for our qPCR assay (Chapter 2).

In Chapter 3 we uncovered the phylogegetic and functional diversity of microorganisms inhabiting the sediments of thermokarst lakes, and demonstrated that the availability of certain carbon and energy sources in the sediments (e.g., organic matter, sulfate, nitrate, and dissolved iron) was paralleled by the presence of microorganisms that can use these substrates and nutrients as electron donors or electron acceptors in a variety of metabolic reactions. Furthermore, we detected unexpected groups of microorganisms like cyanobacteria, which are obligate phototrophs and require sunlight as their primary energy source. This finding caused me to question the biological-geological-physical mechanisms that would allow these types of microorganisms to survive in darkness in the sediments of these shallow lakes. For example, can these organisms exist in dormant states? What is the life-span or turnover time of ribosomal RNA in cold and frozen sediments? What are the important scales (both physical and temporal) of thermokarst erosion?
And, what is the influence of wind on rearrangement of these shallow lake sediments?

Perhaps one of the main contributions from chapter 3 lies in the use of MiSeq next-generation sequencing to uncover the diversity and structure of the bacterial and archaeal assemblages in these lakes. This aspect of the biogeochemistry of the lakes had not been studied until now, and it allowed exposing striking differences in the structure of the archaeal assemblage in the two lakes. Clearly, methanogens play a bigger role in sediments of Siqlukaq, as determined by the abundance of ribosomal RNA, but what determines the almost absence of methanogenic SSU rRNA in the sediments of Sukok remains to be explained. Future studies could investigate if there is an effect of active gas seepage on the structure of the microbial assemblage, and search for mechanisms to explain that influence.

By combining the sensitivity of next-generation sequencing with the nearly full length archaeal SSU rRNA gene sequences, we were also able to study the phylogeny of Archaea inhabiting the sediments of these lakes. Remarkably, we found a high degree of resemblance with the phylogeny of Archaea identified in other kinds of sediments, and most importantly in other places in the Arctic. Furthermore, expanding the short list of cold-adapted methanogens in culture was a driving force that sparked my interest on this line of work. The culturing effort allowed a glimpse into the cultivable portion of the microbial assemblage, and reinforced findings from the use of molecular methods alone. Even though the culture work was not included in this dissertation as a chapter on its own, it
constituted a major endeavor during my tenure as a Ph. D. student, and I expect to continue developing this work in the future in collaboration with my Ph.D. advisor.

In comparison to the snow samples collected from two different mountain ranges at lower latitude, the diversity of the microbial community in thermokarst lake sediments was very high. Even though the surface of these lakes is frozen for about 9 months of the year, this difference in diversity was expected, given the broad range of nutrients that are available for microorganisms in the lakes and the burst of activity that is expected to take place in the lakes during the summer months.

2.2. Microbial communities in snow samples

The snowpack is an ephemeral environment that not only experiences extreme diurnal temperature variations, but also undergoes changes within seasons with variability in air mass movement and storm systems resulting in different levels of snow deposition. The snow ecosystem is perhaps more heterogeneous from the biogeochemical point of view than lake sediments are, because of the geochemical gradients that are established in sediments, mostly by depth. Conversely, snowpacks accumulate over a short season and present a range of depths that are determined by factors such as topography, the amount of snow accumulation during the season, the direction of exposure to sunlight, and the slope of the terrain. This contributes to the formation of areas where the deepest layers of the snowpack may be anaerobic, and shallower zones can be exposed to more air circulation. Additionally, the main
source of nutrients in this ecosystem are likely dominated by allochthonous sources, and may be carried by wet and dry deposition, or originate from the surface of the soil where the snowpack resides (Larose et al., 2013; Williams et al., 2009). However, late in spring when the snowmelt begins, the snow becomes a more dynamic ecosystem where nutrients are transferred from the surface layers to increasingly high depths of the snowpack, and autotrophic microorganisms produce higher amounts of autochthonous organic matter.

Chapter 4 revealed incredible parallels in composition between snow samples that were collected on opposite ends of the state of Nevada, even though at the OTU level there was high dissimilarity and differences in structure. Likewise, we exposed a unique structure of the microbial community composed of OTUs with different levels of abundance but composed of similar taxa, some of which were also dominant in all three sites. Even more interesting perhaps, was the finding of unique OTUs that were highly abundant in individual samples, and that were also common to all sites (e.g., OTUs related to Polaromonas and Herminiimonas). Explaining the driving forces behind this unique community structure and finding mechanisms that may lead to community assemblage in these ephemeral snow ecosystems, opens up the door to a number of questions about the interconnectivity of similar ecosystems around the world and the adaptations that allow specific groups of organisms to inhabit them.
3. **Astrobiological implications of the study of life in cold environments on Earth**

The two primary elements that microbial communities inhabiting arctic thermokarst lake sediments and snowpacks may have in common are the molecular adaptations that allow microbial life forms to thrive in the cold, and the global distribution patterns of certain phylogenetic groups. However, microorganisms inhabiting all kinds of environments are one engine, whose interaction with the environment shape the face of the planet and determine what the Earth looks like from space. In the search for life elsewhere, we have to think of what life looks like from afar, under conditions that are not necessarily amenable to our standards for life. Only the study of life in extreme environments in our planet will allow us to expand our definition of life, and to predict how life could exist under analog planetary conditions. The identification of signatures that could be used as indicators of life is also essential to constrain our search for life, and to identify the most likely candidates that could harbor life today. The more we learn about the common features that microbial life forms share in extreme environments, the better chances we have to identify life when we find it. If I could only chose one thing to take with me from this work, it would be that life is extraordinarily versatile; and that we cannot let our preconceived limits for what life is to determine what we look for in other worlds, because even cold environments that may seem lifeless harbor some of the most powerful life forms on the planet.
References


Appendix
Table A1. Summary of relevant enrichment cultures from arctic lake sediments

<table>
<thead>
<tr>
<th>Lake</th>
<th>Depth (cm)</th>
<th>Medium</th>
<th>T (°C)</th>
<th>Sequence ID</th>
<th>No. clones</th>
<th>Min. % similarity (Silva database)</th>
<th>Tentative order</th>
<th>Tentative family</th>
<th>Tentative Genus</th>
<th>Tentative species</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ikroøvik</td>
<td>02−18</td>
<td>Acetate + Antibiotics</td>
<td>~ 20</td>
<td>A1</td>
<td>3</td>
<td>99</td>
<td>Methanosarcinales</td>
<td>Methanosarcinaceae</td>
<td>Methanosarcina</td>
<td>lacustris</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B1</td>
<td>7</td>
<td>99</td>
<td>Enterobacteriales</td>
<td>Enterobacteriaceae</td>
<td>Yersinia</td>
<td>--</td>
</tr>
<tr>
<td>Ikroøvik</td>
<td>02−18</td>
<td>H₂/CO₂ + Antibiotics</td>
<td>~ 20</td>
<td>A2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siqlukaq</td>
<td>68−96</td>
<td>H₂/CO₂ + Antibiotics</td>
<td>~ 20</td>
<td>A3</td>
<td>8</td>
<td>97</td>
<td>Methanobacteriales</td>
<td>Methanobacteriaceae</td>
<td>Methanobacterium</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B3</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siqlukaq</td>
<td>02−28</td>
<td>Methanol + Antibiotics</td>
<td>~ 10</td>
<td>A6</td>
<td>9</td>
<td>97</td>
<td>Methanomicrobiales</td>
<td></td>
<td>Methanosphaera</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B6</td>
<td>1</td>
<td>93</td>
<td>Clostridiales</td>
<td>XII Incertae Sedis</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Siqlukaq</td>
<td>02−28</td>
<td>H₂/CO₂</td>
<td>~ 10</td>
<td>A4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siqlukaq</td>
<td>02−28</td>
<td>H₂/CO₂</td>
<td>~ 10</td>
<td>A7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Siqlukaq</td>
<td>02−28</td>
<td>H₂/CO₂</td>
<td>~ 10</td>
<td>A8</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B8</td>
<td>1</td>
<td>98</td>
<td>Clostridiales</td>
<td>Clostridiaceae</td>
<td>Clostridium</td>
<td>--</td>
</tr>
<tr>
<td>Siqlukaq</td>
<td>02−28</td>
<td>H₂/CO₂</td>
<td>~ 10</td>
<td>A9</td>
<td>--</td>
<td>--</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B9</td>
<td>7</td>
<td>96</td>
<td>Clostridiales</td>
<td>Eubacteriaceae</td>
<td>Acetobacterium</td>
<td>--</td>
</tr>
<tr>
<td>Sukok Seep</td>
<td>02−16</td>
<td>Acetate + Antibiotics</td>
<td>~ 10</td>
<td>A5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sukok Seep</td>
<td>02−16</td>
<td>Acetate</td>
<td>~ 20</td>
<td>A10</td>
<td>10</td>
<td>99</td>
<td>Methanosarcinales</td>
<td>Methanosarcinaceae</td>
<td>Methanosarcina</td>
<td>lacustris</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B10</td>
<td>3</td>
<td>95</td>
<td>Clostridiales</td>
<td>Peptostreptococcaceae</td>
<td>Proteocatella</td>
<td>--</td>
</tr>
<tr>
<td>Sukok Seep</td>
<td>02−16</td>
<td>Acetate</td>
<td>~ 20</td>
<td>A11</td>
<td>10</td>
<td>99</td>
<td>Methanosarcinales</td>
<td>Methanosarcinaceae</td>
<td>Methanosarcina</td>
<td>lacustris</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>B11</td>
<td>11</td>
<td>97</td>
<td>Lactobacillales</td>
<td>Carnobacteriaceae</td>
<td>Trichococcus</td>
<td>--</td>
</tr>
</tbody>
</table>