

**ILLUMINATING PHYLOGENY AND FUNCTION OF MICROBIAL DARK
MATTER IN SAKINAW LAKE**

by

Esther Anna Gies

B.Sc., Radboud University Nijmegen, Netherlands, 2007

M.Sc., Radboud University Nijmegen, Netherlands, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Microbiology and Immunology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

June 2015

© Esther Anna Gies, 2015

Abstract

Microorganisms are the most abundant and diverse forms of life on Earth. Interconnected microbial communities drive matter and energy transformations integral to ecosystem functions and services through distributed metabolic networks innovated over 3.5 billion years of evolution. To effectively harness this metabolic potential it is necessary to chart uncultivated microbial community structure and function. Cultivation-independent studies indicate that over half of the microbial diversity on Earth belongs to uncultivated candidate divisions, also known as microbial dark matter (MDM). Here, I illuminate MDM structure and function in meromictic Sakinaw Lake on the Sunshine Coast of British Columbia Canada. Sakinaw Lake water column conditions quantified over 8 field campaigns were intimately associated with unprecedented abundance and diversity of MDM. Using network analysis and single cell genomics, co-occurrence patterns between MDM including OP9/JS-1, OP8 and methanogenic Archaea were linked with potential to perform syntrophic acetate oxidation, an important process in anaerobic digestion of organic matter in natural and engineered ecosystems. Single-cell and metagenome analysis revealed previously unrecognized nitrate reduction potential in candidate division OP3 and uncovered a novel archaeal lineage. In addition, numerous Fe-S oxidoreductases associated with MDM in Sakinaw Lake indicate the potential to couple sulfur oxidation to iron reduction. Taken together, my work establishes Sakinaw Lake as a natural laboratory in which to explore MDM structure and function, shines a spotlight on known and novel interactions and metabolic capabilities among these most enigmatic microorganisms, and points to potential biotechnological innovations based on cooperative interactions between MDM populations.

Preface

The work presented in this thesis was made possible by support of collaborators, former and current members of the Hallam laboratory, and contractors as described in the following paragraphs. As research advisor, Dr. Steven Hallam was involved in the design and conduction of field trips and experiments, as well as in the interpretation, presentation, and publication of the collected data.

In chapter 2, I used metadata from samples collected in Sakinaw Lake between June 2007 and August 2013 to describe physicochemical characteristics of the ecosystem. On all field trips, operation of the research vessel, and analysis of samples for oxygen, and nutrients (nitrite, nitrate and phosphate), as well as the operation of the conductivity, temperature, and depth device (CTD) was done by sea going technicians from the Department for Earth, Ocean and Atmospheric Sciences including Larysa Pakhomova and Chris Payne. Samples for transition metals (iron, manganese, and arsenic), sulfate, dissolved organic carbon, alkalinity and phosphorus for October 2007 and May 2011 were collected by the Hallam laboratory, and analyzed by the contractor Maxxam analytics in Burnaby, Canada. Current and past members of the Hallam laboratory planned, prepared, and conducted field trips prior to May 2010, and analyzed samples for sulfide, ammonium, and total microbial cell counts. I planned, prepared, and conducted two field trips in 2011 and one field trip in 2013. On all three field trips I had support from staff and students of the Hallam laboratory for sample collection in the field. I analyzed samples for sulfide and ammonium. Total microbial cell counts using flow cytometry was done by technical staff in the Hallam laboratory (Sam Keihrandish and Craig Mewis).

Data presented in chapter 2 were also published in main articles and supplementary material of:

Gies, EA, KM Konwar, J Thomas Beatty, Steven J Hallam (2014). "Illuminating Microbial Dark Matter in Meromictic Sakinaw Lake." *Appl Environ Microbiol*.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF, Darling A, Malfatti S, Swan BK, **Gies EA**, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu WT, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. (2013) "Insights into the Phylogeny and Coding Potential of Microbial Dark Matter", *Nature*.

And were used for supplementary material of manuscripts in preparation:

Dodsworth JA, Masaru N, Murugapiran S, Rinke C, Schwientek P, **Gies EA**, Hallam SJ, Kille P, Liu W, Tsiamis G, Webster G, Weightman A, Woyke T & Hedlund BP et al, (2015) "Phylogeny and physiological potential of the candidate phylum "Atribacteria"", *ISME (accepted)*.

In chapter 3, I used 454 sequencing data of the small subunit ribosomal RNA (SSU or 16S rRNA) gene generated from 66 DNA samples collected in Sakinaw Lake between 2007 and 2011 for a survey of the Sakinaw Lake microbial community with a focus on uncultivated organisms. Past members of the Hallam laboratory extracted DNA from samples collected prior to May 2010. I extracted DNA for samples collected in January 2011 and May 2011. With assistance of Sam Keihrandish, I performed PCR reactions for 42 samples that were submitted for 454 sequencing of the 16S rRNA gene at the McGill University and Génome Québec Innovation Center. PCR amplification and 454 sequencing of the remaining 24 samples was performed at the Department of Energy's Joint Genome Institute (JGI) in Walnut Creek, CA, USA. Dr. Kishori Konwar performed OTU clustering and chimera removal analyses of the 454 sequencing

data, and wrote a perl script for the construction of co-occurrence networks. For the published article I focused on sequencing data from May 2011. I assigned taxonomy, identified candidate divisions with the assistance of Dr. Christian Rinke from the JGI, performed rarefaction, statistical and co-occurrence analyses, generated all Figures, and wrote the manuscript with the assistance of Dr. Steven Hallam. Dr. Thomas Beatty provided constructive feedback for data presentation in the manuscript.

A version of chapter 3 is published as:

Gies, EA, KM Konwar, J Thomas Beatty, Steven J Hallam (2014). "Illuminating Microbial Dark Matter in Meromictic Sakinaw Lake." *Applied and Environmental Microbiology*.

In chapter 4, I used previously described 454 sequencing data of the 16S rRNA gene, full-length bacterial 16S rRNA gene sequences generated from 6 samples collected in June 2007, and 454 genomic shotgun sequencing data generated from 16 samples collected between June 2007, and May 2011, to identify candidate divisions that were associated with the redox transition zone (RTZ) of Sakinaw Lake. Moreover, I compared the population profile of RTZ associated candidate division RF3 to an active population of green sulfur Bacteria, that was identified in an unpublished study designed and conducted by Dr. Steven Hallam, Dr. Thomas Beatty, Dr. Jörg Overmann, and their students Susanne Meyer and Ovidu Rucker. Generation of full-length bacterial 16S rRNA sequences, as well as quality control of sequences and chimera removal was done by undergraduate student Claudia Biskupski as part of her UBC summer session MICB 448C directed studies project. Dr. Kishori Konwar and PhD candidate Niels Hanson provided assistance for the use of Metapathways, and generated a database of "Microbial Dark Matter"

proteins. I prepared DNA for genomic shotgun sequencing performed at the McGill University and Génome Québec Innovation Center. I performed taxonomic and statistical analysis, generated phylogenetic trees, performed metabolic pathway prediction analysis, and wrote the manuscript with assistance of Dr. Steven Hallam. Dr. Sean Crowe, Dr. Richard Pawlowicz and Dr. Thomas Beatty provided constructive criticism for data presentation, terminology and data analysis.

A version of chapter 4 is in preparation for submission to a peer reviewed journal.

In chapter 5, I used previously described 454 sequencing data and full-length archaeal 16S rRNA gene sequences generated from 6 samples collected in June 2007, and one previously published archaeal Sakinaw Lake single cell genome, to identify phylogeny and function of previously unassigned Archaea in the microbial population of Sakinaw Lake. Additionally, I performed CARD-FISH analysis with Archaea specific probes to identify relative abundance. Generation of full-length 16S rRNA sequences, as well as quality control of sequences and chimera removal was done by undergraduate student Yun-Wen (Wayne) Hsu as part of his UBC summer session MICB 448C directed studies project. I did CARD-FISH analysis, performed taxonomic and statistical data analysis of the sequencing data, generated phylogenetic trees, performed analysis for metabolic pathway prediction, and wrote the manuscript with assistance of Dr. Steven Hallam.

A version of chapter 5 is in preparation for submission to a peer reviewed journal.

Table of Contents

Abstract.....	ii
Preface.....	iii
Table of Contents	vii
List of Tables	xii
List of Figures.....	xiii
List of Abbreviations	xv
Acknowledgements	xvii
Chapter 1: Microbial dark matter and its natural enrichment in meromictic lakes	1
1.1 Synopsis	1
1.2 The “uncultivated microbial majority”	1
1.2.1 Methods to illuminate “Microbial Dark Matter”	5
1.2.2 Metagenomic sequencing.....	6
1.2.3 Entering the single cell genomics era	8
1.3 Meromictic lakes, a natural enrichment site for MDM.....	10
1.4 Meromixis	12
1.4.1 Formation of meromictic lakes	15
1.4.2 Geography of meromictic lakes	16
1.4.3 Meromictic lakes: natural laboratories with persistent redox gradients	19
1.4.4 Methanogenesis: a communal effort to trick thermodynamics.....	24
1.4.5 The microbial community of meromictic lakes	26
1.4.5.1.1 The mixolimnion: an equivalent to freshwater lakes.....	27
1.4.5.1.2 The redox transition zone: passage from known to unknown	28
1.4.5.1.3 The monimolimnion: a methanogenic community enriched in MDM.....	28
1.4.6 Shining a spotlight on MDM	29
1.4.6.1 Candidate division OP3: “Omnitrophica”	29
1.4.6.2 Candidate division OP9: “Atribacteria”	31
1.4.6.3 Candidate division OP11: “Microgenomates”	32

1.4.6.4	Insights into archaeal candidate divisions	34
1.5	Objectives.....	35
Chapter 2:	Meromictic Sakinaw Lake.....	37
2.1	Synopsis	37
2.2	Material and methods	38
2.2.1	Chemical profiling	38
2.3	Results and discussion.....	39
2.3.1	Permanent water column stratification	39
2.3.2	Sulfur chemistry.....	40
2.3.3	Iron and manganese	44
2.3.4	Arsenic	46
2.3.5	Nitrogen	48
2.3.6	Phosphate	49
2.3.7	Carbon.....	51
2.3.7.1	Dissolved organic carbon	51
2.3.7.2	Alkalinity	52
2.4	Conclusion.....	54
Chapter 3:	Taxonomic composition, population structure and diversity of MDM in	
Sakinaw Lake	56
3.1	Synopsis	56
3.2	Material and methods	57
3.2.1	Sampling	57
3.2.2	Enumeration of cells by flow cytometry.....	57
3.2.3	Environmental DNA	57
3.2.4	PCR amplification of 16S rRNA gene for pyrotag sequencing.....	58
3.2.5	Pyrotag sequence analysis.....	59
3.2.6	Statistical analyses	60
3.2.7	Co-occurrence network.....	61
3.2.8	Pathway reconstruction.....	62
3.3	Results	62

3.3.1	Site location and physicochemical properties.....	62
3.3.2	Relationship between community structure and physicochemical characteristics	66
3.3.3	Taxonomic composition.....	73
3.3.4	Co-occurrence analysis.....	77
3.4	Discussion	80
3.5	Conclusion.....	85

Chapter 4: Phylogeny, population structure and putative function of MDM in the redox transition zone of meromictic Sakinaw Lake 86

4.1	Synopsis	86
4.2	Material and methods	87
4.2.1	Water sampling and processing for tag sequencing of the 16S rRNA gene.	87
4.2.2	PCR amplification of 16S rRNA gene, clone library construction and sequencing.....	88
4.2.3	Genomic shotgun sequencing and metabolic pathway prediction.....	90
4.2.4	Pigment analysis	91
4.2.5	Environmental RNA	92
4.2.6	Physiochemical parameter	93
4.2.7	Statistical analyses	93
4.3	Results.....	93
4.3.1	Stability of water column stratification in Sakinaw Lake.....	93
4.3.2	Temporal stability in microbial community composition.....	94
4.3.3	Partitioning of the microbial community	95
4.3.4	OTU partitioning along the water column.....	98
4.3.5	Phylogeny of candidate divisions OP3, OP9/JS1, OP11, OD1, and RF3...	102
4.3.6	Population structure and dynamics of Chlorobi and RF3	107
4.3.7	Metabolic potential in the RTZ of Sakinaw Lake.....	110
4.4	Discussion	114
4.5	Conclusion.....	119

Chapter 5: Phylogeny and function of novel and known Archaea in Sakinaw Lake..... 120

5.1	Synopsis	120
5.2	Material and methods	121
5.2.1	Study site and sampling	121
5.2.2	Methods to infer taxonomic composition based on the 16S rRNA gene....	122
5.2.3	Quantification of Archaea with CARD-FISH.....	122
5.2.4	PCR amplification of 16S rRNA gene, clone library construction and sequencing.....	123
5.2.5	Phylogenetic analysis and matching of Arch915 probe.....	124
5.2.6	Metagenome assembly and analysis	124
5.2.7	<i>mcrA</i> clone library.....	125
5.2.8	SAG analysis.....	127
5.3	Results	128
5.3.1	Archaeal cell counts	128
5.3.2	Archaeal phylogeny	131
5.3.3	Assigning the unassigned.....	140
5.3.4	Inferring metabolic potential of Archaea in Sainkaw Lake	142
5.3.5	Single cell genome analysis	144
5.4	Discussion	150
5.5	Conclusion.....	155
Chapter 6: Concluding chapter		157
6.1	Meromictic Sakinaw Lake: remarkable stability of the salinity gradient and remaining questions about the sulfur chemistry	158
6.2	The Sakinaw Lake microbial community is dominated by bacterial and archaeal candidate divisions.....	159
6.3	Metabolic potential for dissimilatory nitrate reduction, iron and sulfur cycling in RTZ and monimolimnion	160
6.4	The archaeal community	161
6.5	Future directions.....	163
6.6	Closing remarks.....	166
Bibliography		167

Appendices.....	188
Appendix A Design and testing of CARD-FISH probe targeting candidate division OP9/JS-1	188
A.1 Motivation for the project	188
A.2 Methods.....	189
A.3 Results and discussion	190
Appendix B Phylogeny of Bacteria in Sakinaw Lake	192

List of Tables

Table 1.1 List of recognized bacterial candidate divisions.	4
Table 2.1 Sulfur compounds typically found in aquatic ecosystems.....	43
Table 3.1 Baseline physicochemical characteristics of geographically isolated meromictic lakes.	65
Table 3.2 Operational taxonomic units with highest abundance and their function as indicator species.	71
Table 4.1 Average abundance of MDM in individual water column compartments.....	99
Table 4.2 Abundance distribution of abundant MDM and RTZ associated OTUs.	101
Table 5.1 Methods and primers used to infer taxonomic composition based on the 16S rRNA gene.....	122
Table 5.2 General genomic features of the Sakinaw Lake SAG.	145

List of Figures

Figure 1.1 Phylogenetic distribution of sequenced genomes.....	6
Figure 1.2 Global distribution of meromictic lakes.....	12
Figure 1.3 Simplified model of water column stratification in a meromictic lake.....	14
Figure 2.1 Permanent water column stratification in Sakinaw Lake 2007 - 2013.....	40
Figure 2.2 Sulfide and sulfate profiles in Sakinaw Lake.....	44
Figure 2.3 Iron and manganese profiles in Sakinaw Lake.....	46
Figure 2.4 Arsenic profiles in Sakinaw Lake.....	47
Figure 2.5 Nitrogen profiles in Sakinaw Lake.....	49
Figure 2.6 Phosphorus profiles in Sakinaw Lake.....	50
Figure 2.7 Dissolved organic carbon profiles in Sakinaw Lake.....	52
Figure 2.8 Alkalinity profiles in Sakinaw Lake.....	54
Figure 3.1 Environmental parameters in Sakinaw Lake May 24 th 2011.....	64
Figure 3.2 Relationship between community structure and physicochemical characteristics.....	67
Figure 3.3 Partitioning of OTUs between water column compartments.....	68
Figure 3.4 Multilevel indicator species analysis.....	70
Figure 3.5 Fine scale partitioning of OTUs.....	72
Figure 3.6 Relative abundances of Bacteria, Archaea and Eukaryotes.....	76
Figure 3.7 Co-occurrence network analysis.....	78
Figure 3.8 Potential for syntrophic acetate oxidation.....	80
Figure 4.1 Variability in microbial population structure.....	95
Figure 4.2 The stratified microbial community.....	98
Figure 4.3 Niche partitioning along the water column.....	100

Figure 4.4 Phylogeny of candidate division OP3.	103
Figure 4.5 Phylogeny of candidate division OD1.....	104
Figure 4.6 Phylogeny of candidate division JS-1.	105
Figure 4.7 Phylogeny of candidate division RF3.....	106
Figure 4.8 Metabolic niche of Chlorobi in Sakinaw Lake.....	108
Figure 4.9 Population dynamics in RTZ associated phyla.	110
Figure 4.10 Partitioning of predicted metabolic pathways.	112
Figure 4.11 Phylogenetic anchoring of MDM reads.	114
Figure 5.1 Total cell counts and relative abundance of Archaea.....	130
Figure 5.2 Phylogenetic tree of Thaumarchaeota in Sakinaw Lake.	133
Figure 5.3 Phylogenetic tree of the DPANN.	135
Figure 5.4 Phylogenetic tree of Euryarchaeota in Sakinaw Lake.....	137
Figure 5.5 Phylogenetic analysis of Sakinaw Lake <i>mcrA</i> sequences.	139
Figure 5.6 Assigning the unassigned.	141
Figure 5.7 Metabolic potential of the archaeal community.....	144
Figure 5.8 PCA analysis of euryarchaeal and DPANN genomes.....	146
Figure 5.9 Metabolic model for archaeal Sakianw Lake SAG.....	150
Figure A.1 Microscopic evaluation of CARD-FISH probe JS1-640-a-A-23.	191
Figure B.1 Phylogeny of Bacteria in Sakinaw Lake.....	191

List of Abbreviations

APHA	American Public Health Association
AMD	acid mine drainage
AMP	adenosine monophosphate
ANME	anaerobic methane oxidizing
AOM	anaerobic oxidation of methane
As	arsenic
ATP	adenosine triphosphate
Bchl	bacteriochlorophyll
BLAST	basic local alignment search tool
BLASTP	protein BLAST
bp	base pair
BP	before present
CARD-FISH	catabolized reporter deposition fluorescent in situ hybridization
CH ₄	methane
CO ₂	carbon dioxide
CO ₃ ⁻	carbonate
COG	clusters of orthologous genes
contig	contiguous read
DIC	differential interference contrast
DNA	deoxyribonucleic acid
eDNA	environmental DNA
DNRA	dissimilatory nitrate reduction to ammonium
DPANN	Diapherotrites-Aenigmarchaeota-Parvarchaeota-Nanohaloarchaeota-Nanoarchaeota
DSMZ	German collection for cell cultures and microorganisms
Eh	redox potential
FACS	fluorescence activated cell sorting
Fe	iron
Fe (II)	ferrous iron
Fe (III)	ferric iron
FeS ₂	pyrite
Gb	giga bases
GS	glutamine synthetase
GOGAT	glutamine:2-oxoglutarate aminotransferase
H ₂	hydrogen
H ₂ CO ₃	carbonic acid
HCO ₃ ⁻	bicarbonate
H ₂ S	sulfide
HS ⁻	bisulfide
HNO ₃	nitric acid
HGT	horizontal gene transfer
JGI	Joint Genome Institute
kb	kilo base

Mb	mega bases
MDM	microbial dark matter
MGE	mobile genetic elements
MG1	marine group 1
MCG	miscellaneous crenarcheotic group
Mn	manganese
MTB	magnetotactic bacteria
NO ₃ ⁻	nitrate
NH ₄ ⁺	ammonium
NH ₃	ammonia
OAA	oxaloacetate
OD1	OP11 derived 1
OP	obsidian pool
ORF	open reading frame
OTU	operational taxonomic unit
PCA	principal component analysis
ppGpp	5'-diphosphate 3'-diphosphate guanosine
PVC	Planctomycetes – Verrucomicrobia – Chlamydiae
rRNA-SIP	16S rRNA stable isotope probing
RTZ	redox transition zone
Rubisco	ribulose-1,5-bisphosphate carboxylase
RuMP	ribulose monophosphate pathway
S _A	absolute salinity
SAG	single amplified genome
SAO	syntrophic acetate oxidation
SIMS	secondary ion mass spectrometry
SO ₄ ²⁻	sulfate
S ₂ O ₃ ²⁻	thiosulfate
S _n ⁻	polysulfide
SR1	sulfur river 1
SSU rRNA	small subunit of the 16S rRNA gene
Sud	sulfide dehydrogenases
TACK	Thaumarchaeota – Aigarchaeota – Crenarchaeota – Korarchaeota
TEM	transmission electron microscopy
TEA	terminal electron acceptor
WGA	whole genome amplification
YNP	Yellowstone National Park

Acknowledgements

I would like to express my sincere appreciation and thanks to Professor Dr. Steven Hallam, your guidance and mentorship always pushed me to the next level, allowing me to grow as a an independent researcher. I would also like to thank Professor Dr. Tom Beatty, Professor Dr. Richard Pawlowicz and Professor Dr. Sean Crowe, for serving as my committee members. Your constructive criticism improved the quality of my work and made me a better scientist. I would like to particularly thank Rich, my field trips with you were not only highly entertaining, I also learned a lot from you! I also would like to thank Dr. Jörg Bohlmann, who first served as committee member and later as a mentor on our occasional meetings on campus. I am grateful to Lora Pakhomova and Chris Payne who kept boat and crew safe; you were irreplaceable for my sampling. Big thanks to my collaborators at the JGI, PNNL and IOW, I have learned a lot from you! I offer many thanks to Darlene Birkenhead for your administrative support and the always-open door to your office! My special thanks goes to all members of the Hallam lab who accompanied me on this journey: thank you for your support and for letting me draft in your slipstream whenever I was running out of steam! Thank you for all the good laughs and your advice Farzin! Thank you Keith for introducing me to Georgia! Den aller größten Dank richte ich an meine Familie, die mich nach Kanada hat gehen lassen! Ich denke oft an euch und vermisse euch alle! Danke Mama, dass Du mir immer das Gefühl gibst alles richtig zu machen und Danke Papa dass Du immer alles wieder in die richtige Perspektive rückst. Lastly I would like to thank Devin, for always keeping me on belay, teaching me to watch the currents and tides, helping me to get up when I fall (after you had a good laugh about it), and for providing me with the love and fuel that I need to thrive and to keep on pedaling no matter what.

To my father!

Chapter 1: Microbial dark matter and its natural enrichment in meromictic lakes

1.1 Synopsis

Microorganisms represent the invisible majority of living things on Earth and interconnected microbial communities drive matter and energy transformations integral to ecosystem function. To effectively harness this metabolic potential it is necessary to chart uncultivated microbial community structure and metabolic potential. Indeed, cultivation-independent studies indicate that over half the microbial diversity on Earth belongs to uncultivated candidate divisions, also known as microbial dark matter (MDM) (Rappe and Giovannoni 2003, Rinke et al 2013). Despite ongoing advances in cultivation-independent approaches to study microbial communities in natural and engineered ecosystems, fundamental questions related to population structure, activity and metabolic interactions remain. In this introductory chapter I present an overview of cultivation-independent approaches and how recent technological advances have reinforced the realization that our current knowledge about genetic diversity and metabolic potential of the microcosmos is strongly biased towards a “cultivated minority”. I go on to emphasize the enormous potential of permanently stratified (meromictic) lakes as natural laboratories or model ecosystem in which to study uncultivated microbial structure and function. I specifically review the evolution of meromixis, provide an overview of the geographic distribution of meromictic lakes, and summarize their contemporary use as model ecosystems.

1.2 The “uncultivated microbial majority”

The fundamental role of microbes for the global biogeochemical cycling of elements was first discussed by Ferdinand J. Cohn in the 19th century, and his early contributions to the field were recently reviewed by Drews and colleagues (Drews 2000). Shortly after, Beijerinck and

Winogradsky pioneered enrichment and cultivation techniques (Ferris 1996). Traditional cultivation in laboratory settings has since helped elucidate the phylogeny and function of many microbial taxa spanning all three domains of life. Having a taxon of interest in culture allows one to specifically test hypotheses, measure physiological responses in a controlled environment, and thereby elucidate the microbe's potential roles in a given ecosystem. However, already in 1933 it was pointed out by Henrici that cultivation of microbes using artificial media often results in an incomplete representation of the microbial community of interest (Henrici 1933). Moreover, cultivation is not a selection for the most abundant or functionally most significant or active species. Rather, it is a selection for species that are able to grow rapidly on a given concoction of nutrients and thereby outcompete other community members.

With the advent of cultivation-independent molecular surveys, the true extent of our skewed perspective on microbial diversity became increasingly clear. In the early 1980's Norman Pace, Gary J Olsen, and their colleagues outlined first approaches to assess microbial diversity based on the cultivation independent amplification of the small subunit ribosomal (SSU or 16S rRNA) gene, providing a phylogenetic anchor for assessing uncultivated microbial diversity and abundance patterns (Olsen et al 1985, Pace et al 1986, Pace 1995). Subsequent adaption of this technique to samples from diverse natural and engineered environments has led to rapid increase in discovery of novel taxa including candidate divisions. Among the first ecosystems that revealed an outstanding diversity of microbes that resisted isolation in pure cultures were the geothermal springs of Yellowstone National Park (YNP). The first of many cultivation-independent studies in the obsidian pools (OP) of YNP uncovered 12 bacterial (OP1-OP12), and one archaeal (the Korarchaeota) candidate division (Hugenholtz et al 1998). Back then it was thought that these OP-phyla were indigenous extremophiles relegated to extreme

environments. This view changed with the development of next generation sequencing technologies, leading to higher throughput and lower cost for cultivation-independent surveys. The deep sequencing of numerous environments has since uncovered more than 30 candidate divisions (Table 1.1), and revealed the presence of OP phyla as ubiquitous and abundant members of various environments including flooded paddy soils and lakes (Rohini Kumar and Saravanan 2010).

Today it is generally accepted that an invisible “uncultured microbial majority” dominates the Earth (Rappe and Giovannoni 2003). This is directly reflected in the fact that public databases list more than 500,000 16S rRNA sequences (Silva database of rRNA sequences, version 115, <http://www.arb-silva.de/>) affiliated with more than 60 phyla, while 88% of the currently ~ 20,000 available cultivable isolates (German Centre for Microorganisms and Cell Cultures (DSMZ) November 2014, <http://www.dsmz.de/>) are affiliated with only 4 of these phyla, the *Proteobacteria*, *Firmicutes*, *Actinobacteria* and *Bacteroidetes*. Current estimates indicate that less than 1% of the microbes in nature have been cultured (Amann et al 1995), and that our knowledge about the phylogeny and function of the “uncultured majority” is in its infancy. To give credit to the vastness of the “known-unknown”, researchers recently compared the “uncultured microbial majority” to the “Dark Matter” problem in Astrophysics and coined the term “Microbial Dark Matter” (MDM) (Marcy et al 2007b, Rinke et al 2013). The analogy stems from the common keywords that describe both phenomena: (Microbial) Dark Matter is “ubiquitous”, “abundant” and only observable with “indirect techniques” (Hedlund et al 2014).

Table 1.1 List of recognized bacterial candidate divisions.

Acronym	Proposed Name	Superphylum	Genome Completeness	Reference*
ACD80			90%	1
BRC1			22%	2
BD1-5		Patescibacteria	93%	2,1
CD 12/BHI80-139	"Aerophobetes"		67%	2
CKC4				
EM3			22%	2
EM19	"Calescamantes"			2
GN02	"Gracilibacteria"		36%	2
GAL08				
GOUTA4				
Hyd24-12				
JL-ETNP-Z39				
KB1				
Kazan-3B-28				
LD1-PA38				
MVP-21				
NPL-UPA2				
Oct-Spa1-106			46%	2
OC31	"Fervidibacteria"			2
OD1	"Parcubacteria"	Patescibacteria	70%	2,1
OP1	"Acetothermia"		18%	2
OP3	"Omnitrophica"		80%	2,3,4
OP8	"Aminicenantes"	ANA	67%	2
OP9/JS1	"Atribacteria"		73%	2,5
OP11	"Microgenomates"	Patescibacteria	67%	2,1
PER			80%	
RsaHF231				
SAR 406/MG A				2
S2R-29				
SBYG-2791				
SHA-109				
SM2F11				
SR1				
TA06				
TM6				
TM7				
WCHB1-60/WS5				
WD272				
WS1			7%	2
WS3	"Latescibacteria"	FCB	73%	2
WS6				
WWE1	"Cloacimonetes"	FCB	99%	2

*** References:**

- 1 Wrighton KC, et al. 2012. Science 337:1661-1665.
- 2 Rinke et al 2013
- 3 Glöckner et al,
- 4 Kolinko et al
- 5 Dodsworth et al

It is currently thought that MDM escape cultivation in isolation because of their need of metabolic partners. Indeed, first genomic evidence indicated that many candidate divisions harbor genes encoding fermentation, hydrogen (H₂) or sulfur (S) metabolic pathways supporting co-metabolic or syntrophic growth modes under anaerobic conditions, and dependencies on exogenous vitamins sourced from surrounding microbial community members (Dodsworth et al 2013, Pelletier et al 2008, Wrighton et al 2012, Wrighton et al 2014). Such, public good dynamics could represent a common organizing principle in structuring microbial community interaction networks (Cordero et al 2012, Morris et al 2012) and help explain the resistance of most environmental microorganisms including candidate divisions to clonal isolation (Cordero et al 2012, Haruta et al 2009).

1.2.1 Methods to illuminate “Microbial Dark Matter”

The majority of sequenced genomes belong to representatives of the four cultivated phyla described above (Figure 1.1). This skewed representation is in the midst of dramatic change as the focus shifts from cultivation-dependent to cultivation-independent sequencing of MDM. In 1998, Handelsmann and colleagues first coined the term “metagenomics” to describe cultivation independent approaches to study microbes from the environment (Handelsman et al 1998). Sequencing technologies and experimental design has significantly advanced since, and provided deeper insight into phylogeny, function and biotechnological potential of microbes from the environment (Handelsman 2004). The following section provides an overview of methods used to illuminate MDM; from initial approaches that targeted single functional genes to the reconstruction of entire genomes from abundant as well as rare MDM using metagenomic and single-cell genomic sequencing methods.

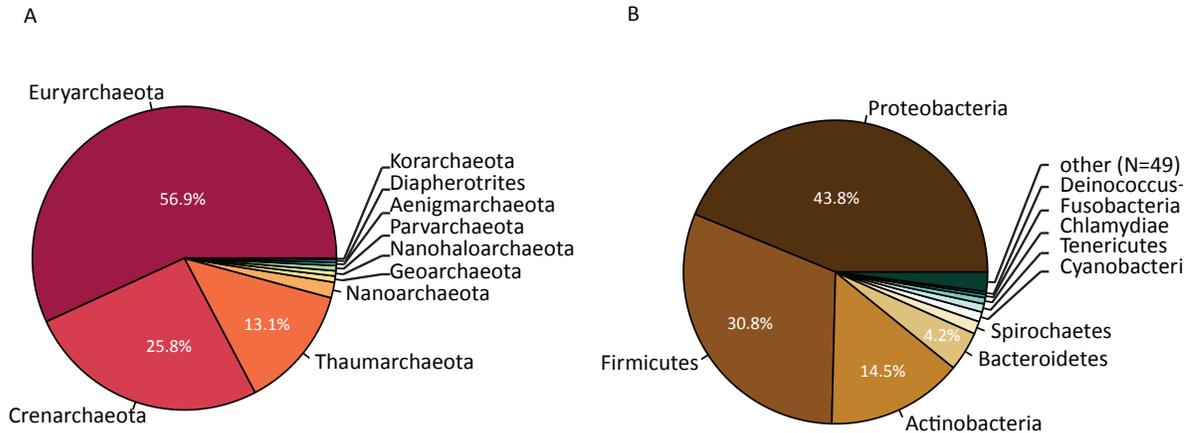


Figure 1.1 Phylogenetic distribution of sequenced genomes.

Phylogenetic representation of (A) archaeal genomes and (B) bacterial genomes. Data for this plot were downloaded from JGIs GOLD database in November 2014.

1.2.2 Metagenomic sequencing

Approximately 25 years ago DeLong and colleagues kickstarted cultivation independent genome analysis from environmental DNA (eDNA) (Stein et al 1996). Their work focused on the improvement of functional and phylogenetic understanding of planktonic Archaea in the eastern North Pacific. To gain insight into the genomic potential of these abundant community members, they adapted Norman Pace's earlier cultivation independent efforts to capture 16S rRNA genes (Pace et al 1986). DeLong and colleagues extracted high molecular weight eDNA from planktonic water samples for the construction of large insert (40 kb) clone libraries (Stein et al 1996). A screen using domain-specific primers identified clones encoding archaeal 16S rRNA genes. DNA from selected clones was then digested enzymatically, subcloned into smaller plasmids and shotgun sequenced. Bypassing the targeted amplification of individual genes allowed the identification of several functional properties that were not previously associated with Archaea.

While the extraction of high molecular weight eDNA proved pivotal for marine water column sequencing, this approach can be more problematic when processing biofilms, soil, and sediments resulting in smaller molecular weight eDNA extracts. In 2004 Tyson and colleagues came up with an approach to bypass this limitation, by cloning eDNA from Acid Mine Drainage (AMD) samples into small insert plasmids (3.2 kb) followed by random shotgun sequencing and paired-end assembly (Tyson et al 2004). Using Sanger capillary technology the authors generated 76.2 Mb of AMD sequence information. In contrast to the large insert approach, the AMD shotgun assemblies provided a snapshot of the metabolic potential of the entire microbial community. The authors used a binning approach based on GC content (low versus high) and average depth of coverage to identify contigs associated with specific taxonomic groups. A total of 5 different bins were resolved, enabling reconstruction of two complete and 3 draft genomes (Tyson et al 2004). The success of this study was greatly influenced by the low biodiversity (six phyla), and the relative dominance (74%) of only one phylum in the AMD sample.

Tyson's study greatly benefitted from the low microbial diversity in the AMD samples (Tyson et al 2004). However, most environments are highly diverse and the functional significance of a community member is not equivalent to its abundance, underlining the need for an improved technology that allows genome reconstruction of rare community members. Only eight years later, the deep sequencing of water samples from acetate amended aquifers led to the recovery of 49 population genomes, all of which belonged to MDM with average abundance of <1% (Wrighton et al 2012). The success of this study was facilitated by the exponential increase in sequencing throughput using the Illumina platform, generating 20 Gb of shotgun sequences (~286-fold increase over the AMD study) for population genome assembly.

1.2.3 Entering the single cell genomics era

One of the biggest challenges for the reconstruction of genome equivalents from environmental sequence information using metagenomic sequencing data is accurate binning and taxonomic classification of assembled contigs (Rinke et al 2013). Annotation in this context can be confounded by the skewed representation of sequenced genomes in available reference databases especially when MDM are abundant community members. To address this challenge, the Department of Energy's Joint Genome Institute (JGI) initiated "The Microbial Dark Matter project" (Rinke et al 2013), with the aim to increase the availability of complete or near complete genomes from ubiquitous abundant as well as low abundant candidate divisions using single-cell genomics. Single-cell genomic approaches contrast metagenomic sequencing in that resulting sequence information originates - by definition - from an individual cell with a defined taxonomic origin, and thereby remove the need to separate assembled contigs into taxonomic bins. Moreover, successful genomic reconstruction can be uncoupled from the species abundance in the environment when using anonymous sorting and amplification methods as described below.

The process starts with the isolation of single cells using microfluidic devices or fluorescence activated cell sorting (FACS), followed by lysis of the cell envelope to allow efficient whole genome amplification (WGA) and subsequent library preparation, culminating in high throughput sequencing, assembly and annotation. To increase the success, quality and validity of SAG reconstruction, it is imperative to limit the introduction of contaminating DNA, which can come from the environmental sample itself, the laboratory environment or WGA reagents. Despite meticulous efforts to reduce contamination by working in clean rooms, by pre-treating WGA reagents with UV-irradiation (Woyke et al 2011), and by reducing the sample

volume (Marcy et al 2007a), contamination issues remain. However, most contaminating sequence can be extracted from the assembled data *in silico*. This can be accomplished by manually searching contigs with divergent taxonomic origins using BLAST (e.g. contigs with human origin) or kmer frequency analysis (e.g. (Albertsen et al 2013, Rinke et al 2013)). Horizontal gene transfer (HGT) and mobile genetic elements (MGE) can be confounding in the decontamination procedure, and one has to be cautious of identifying false positives. The identification of MGE and HGT events are by no means new to single cell genomics, there is an historical interest in these genome rearrangements to improve our understanding of eukaryotic and prokaryotic evolution (Gogarten et al 2002, Keeling and Palmer 2008), and for microbial engineering purposes (Strachan et al 2014). Carefully curated MGE databases (Leplae et al 2010), and improvements for algorithms to identify HGT, for example through gene signature analysis (e.g. GC content and codon usage), allow a relatively confident identification of HGT and MGE in genomic sequencing data, including SAGs.

The next challenge in the reconstruction of SAGs is the assembly of the sequencing data. WGA comes with its own set of biases, such as heterogeneous read depth along the genome and increased occurrence of chimeric reads (Binga et al 2008, Lasken and Stockwell 2007). Traditional assembly algorithms have functions implemented that assume chimeric sequences are rare, and read depth along the genome follows a Poisson-like distribution. Violation of these “rules” leads to poor assemblies with higher risks of erroneous assemblies. One approach to overcome these limitations is to increase sequencing depth, leading to better coverage of low read depth regions and to then remove redundant genomic information from high read depth regions *in silico* (Rinke et al 2013, Swan et al 2011). The co-assembly of taxonomically identical SAGs can also significantly improve the successful assembly of the genome (Blainey et

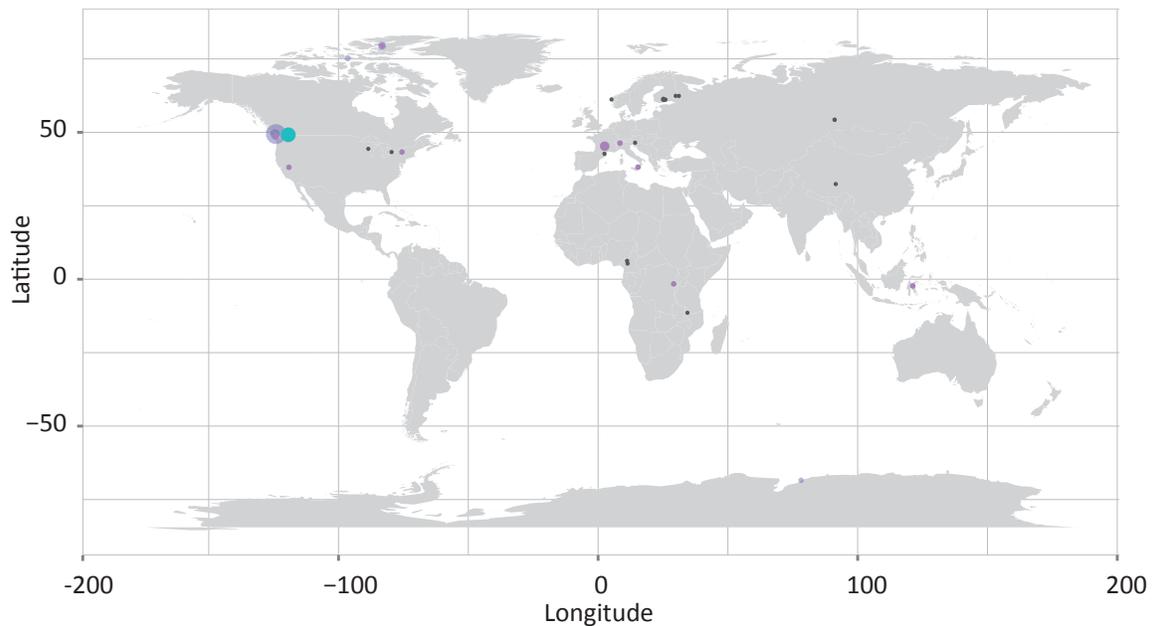
al 2011, Dodsworth et al 2013). This method makes use of the fact that WGA and sequencing biases are stochastic events, and are consequently not repeated in independently amplified SAGs.

Within the last five years, single cell genomics has gained more and more popularity and the increasing success is reflected in the number of published high impact studies (Dodsworth et al 2013, Dupont et al 2012, Field et al 2014, Swan et al 2011). Among these studies is also the result from the previously mentioned “Microbial Dark Matter Project” from the JGI (Rinke et al 2013). The work by Rinke and colleagues led to the sequencing, assembly and annotation of 201 single cell genomes from nine different environments. The selection of environments was based on the natural occurrence of MDM phyla. The majority, 81, of the SAGs originated from only one sample from permanently stratified, meromictic Sakinaw Lake, which was highly enriched with uncultivated microbes. To underline the value gained by the sequencing of these 201 MDM genomes, Rinke and colleagues compared the ability to phylogenetically anchor assembled contigs from metagenomic shotgun sequencing data before and after the addition of the novel genomes to non-redundant genomic databases. The ability to anchor Sakinaw Lake metagenome contigs increased by almost 20%, suggesting substantial improvements for the analysis of samples that are diverse and naturally enriched with MDM.

1.3 Meromictic lakes, a natural enrichment site for MDM

While the onset of the “omics” era provided first insights into the phylogeny and metabolic potential of many candidate divisions, fundamental questions about their biogeochemical roles, population structure and interactions remain largely unresolved. Recent cultivation independent surveys revealed an enrichment of MDM in natural and engineered environments with anoxic, sulfidic and /or methane-rich conditions (Borrel et al 2010, Lykidis et al 2011, Rybak and

Dickman 1988, Wrighton et al 2012). Permanently stratified meromictic lakes manifest these biogeochemical conditions, and molecular surveys in geographically isolated meromictic lakes confirmed abundance and diversity of bacterial as well as archaeal candidate divisions (Figure 1.2). Based on these observations meromictic lakes have the potential to become model ecosystems in which to study MDM population structure, function and dynamics.



Sequencing Technology	Candidate Divisions
• 454 sequencing	• 5
• Sanger sequencing	• 10
• Phylochip	• 15
• No sequencing	• 20

Figure 1.2 Global distribution of meromictic lakes.

This map shows the global distribution of meromictic lakes and their natural enrichment in uncultivated MDM. Enrichment of MDM is reflected in the size of the data point, and experimental method used for determining microbial community composition in the given ecosystem is indicated by the color of the data point.

1.4 Meromixis

Lakes in which the main water mass does not circulate are defined as meromictic and are in contrast to lakes in which the water undergoes complete circulation in regular intervals and that are therefore defined as holomictic (Boehrer and Schultze 2008). The lack of mixing is induced by a permanent density gradient, which causes the evolution of chemical gradients and allows partitioning of the water column into three compartments generally referred to as mixolimnion,

redox transition zone (RTZ) and monimolimnion (Figure 1.3). The naming of the surface waters as “mixolimnion” refers to the existence of periodic water circulation within this compartment. The circulation pattern within the mixolimnion compares to water circulation in holomictic lakes and is driven by temperature gradients and wind regiments. During times of temperature stratification, oxygen (O_2) and nutrients deplete within the lower depth of the mixolimnion and then redistribute when the water becomes homothermal. Not only the water mixing causes changing level of O_2 and nutrient distribution: higher levels of light penetration in the summer month allow the proliferation of photosynthetic organisms leading to elevated O_2 production and nutrient consumption. This dynamic nature of the mixolimnion has a direct impact on the food web, which through the combined presence of O_2 , micro- and macronutrients transfers energy over multiple trophic levels and allows higher predators to thrive. In the RTZ O_2 concentrations deplete and cause a shift in energy transfer from higher predators to microorganisms. Depending on the lake’s geological setting, the RTZ provides various dissolved ions and elements, such as nitrate (NO_3^-), sulfate (SO_4^{2-}), iron (Fe), manganese (Mn) and arsenic (As), that function as alternative terminal electron acceptors (TEA) for the energy metabolism of diverse microbial communities. The monimolimnion is completely anoxic and provides an environment for microbial communities that reduce acetate and carbon dioxide (CO_2) to methane (CH_4). Depending on the geological settings, the monimolimnion can be highly sulfidic or ferruginous, and ammonium (NH_4^+) concentrations can reach the mM range. Moreover, the lack of mixing and the hydrostatic pressure of the overlying water allow gases, such as CH_4 , H_2S , and CO_2 to accumulate well above atmospheric saturation levels (Boehrer and Schultze 2008).

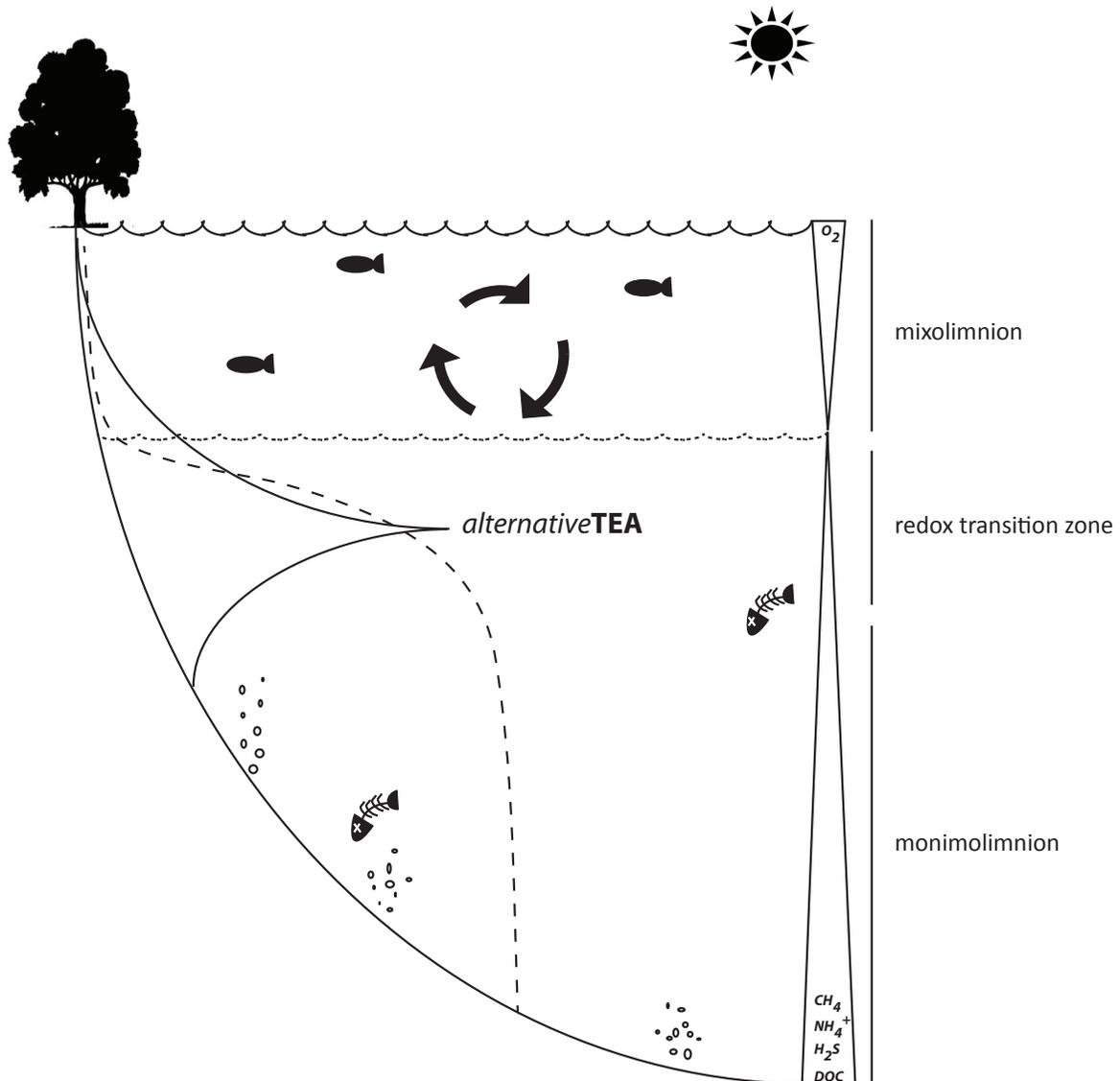


Figure 1.3 Simplified model of water column stratification in a meromictic lake.

The mixolimnion in meromictic lakes is oxygen (O_2) rich and undergoes seasonal change influenced by light availability, temperature, and wind regiments in different seasons. Along the redox transition zone (RTZ) watermasses become more stable as water density increase. Within the RTZ, O_2 depletes and alternative terminal electron acceptors (TEA) become more favourable for the microbial energy metabolisms. Peak concentrations of alternative TEA are typically measured just below the O_2 interface. The monimolimnion never intermixes, is completely devoid of O_2 , and often oversaturated with reduced compounds such as methane (CH_4), ammonium (NH_4^+), and sulfide (H_2S). Moreover, the monimolimnion is often marked by high concentrations of dissolved organic matter (DOC).

1.4.1 Formation of meromictic lakes

Many lakes undergo periods of water column stratification for extended periods of time, however permanent stratification is considered as rare. Hakala studied the origin of meromixis in Finnish lakes and estimated that only 1 in 800 lakes can be considered as meromictic (Hakala 2004). Altogether the genesis of meromixis typically underlies a combination of geomorphological, geochemical and climatic causes. Prominent factors include great depth of the lake while the surface area is small, a lack of seasonality in regions that are close to the equator, permanent ice cover in Arctic and Antarctic regions, submersed springs that feed mineral rich water into the lake or isolation of former fjord openings from the ocean through isostatic rebound caused by deglaciation at the end of the last ice age (Wetzel 2001).

Since Findenegg first described meromixis in 1935 (Findenegg 1935), limnologists have tried to classify different types of meromixis. The most recognized classification was introduced by Hutchinson in 1957 (Hutchinson 1957), who identified three categories of meromixis: crenogenic meromixis originating from internal biogeochemical processes, such as mineral supply from submerged springs, ectogenic meromixis originating from external causes such as intrusion of saline water and biogenic meromixis originating from decomposition of organic matter in the deep waters of the lake and resulting increase in mineral content. However, despite its popularity Hutchinson's category biogenic meromixis was criticized as being vague, and in 1975 Walker and Liekens modified the original classification into the two main categories ectogenic and endogenic meromixis (Walker and Likens 1975).

1.4.2 Geography of meromictic lakes

Meromictic lakes occur all over the world and their geographic location can vary from remote coastal, over industrial mining sites, volcanic craters and mountain ranges, to permanently ice covered sites in Antarctica and the high Arctic (e.g. (Anderson 1958, Lauro et al 2011, LI Wan-chun 2005, Macintyre and Melack 1982, Natacha Pasche 2009, Ramirez-Moreno et al 2005)). Forces that drive research in these ecosystems are often directly linked to their geographic location and environmental parameter that maintain meromixis. Permanently ice covered meromictic lakes, which are typically saturated with greenhouse gases CO₂ and CH₄, are used as model systems to assess effects of potential melt out due to global warming and its consequences on the functioning on the ecosystem (Comeau et al 2012, Coolen 2006, Vincent et al 2008). A well-studied example is meromictic Lake A, located at the northern coast of Ellesmere Island, high Arctic, Canada. Lake A evolved in times of deglaciation between 2,500 BP- 4,000 BP when it got separated from the ocean through isostatic rebound. It was recently established that the lake has been completely ice covered over the last few thousand years and perennially ice covered in most years since the 1930s (Vincent et al 2008). Due to an unusual warm summer in 2008, the ice cover of Lake A completely melted and allowed a comparison of the microbial community composition in ice-covered (May 2008) and ice free (August 2008) conditions (Comeau et al 2012).

Meromictic lake research in the boreal landscape of northern Europe is not necessarily driven by dramatic changes in the environment and their contribution to the global carbon cycling, but there is an historic curiosity on their evolution (Hakala 2004, Hongve 2002, Magadza 1979). Hakala recently reviewed the evolution and origin of meromixis in Finland and grouped meromictic lakes into three categories (Hakala 2004): (1) coastal and inland lakes that

establish meromixis due to a salinity gradient, (2) bog lakes, forest lakes and kettlehole lakes that sustain meromixis due to oxygen deficiency by load from run-off water or anthropogenic nutrient or salt load (most prominent cause of meromixis in Finland), (3) meromixis that is sustained solely based on the lake's morphogenesis. However, the contribution of boreal meromictic lakes to the global carbon cycling and greenhouse gas emission has also been the topic of a recent study of Swedish meromictic lakes, in which Peura and colleagues proposed that candidate divisions OD1 might have an important role in the anaerobic oxidation of methane (Peura et al 2012, Richter et al 2012).

Two alpine meromictic lakes in central and western Europe are - from a microbial ecology perspective - probably among the most extensively studied meromictic lakes: Lake Cadagno located in the Swiss Alps (Bosshard et al 2000, Camacho et al 2001, Gregersen et al 2009, Habicht et al 2010, Habicht et al 2011, Halm et al 2009, Musat et al 2008, Peduzzi et al 2003, Tonolla et al 2004, Tonolla et al 2005), and Lake Pavin located in the Central Massif, France (Biderre-Petit et al 2011a, Bonhomme et al 2011, Borrel et al 2010, Lehours et al 2005, Lehours et al 2007). In Lake Cadagno, meromixis is sustained by high mineral content through a sulfate-rich subsurface spring that feeds into the monimolimnion of the lake. The H₂S-O₂ transition zone in Lake Cadagno is shallow (20 m), and allows penetration of high levels of sunlight in the summer month. The combination of these biogeochemical parameters allows primary production to be dominated by highly-active anoxygenic phototrophic green sulfur bacteria affiliated with *Chlorobi* (Habicht et al 2010). The lake consequently gained popularity as a model system to study *Chlorobi* (e.g. (Gregersen et al 2009, Musat et al 2008, Tonolla et al 2004)). In contrast, SO₄²⁻ and H₂S concentrations in Lake Pavin are low, and the population of anoxygenic phototrophic sulfur bacteria is therefore rather small (Biderre-Petit et al 2011a). Lake

Pavin is a volcanic crater lake, estimated to be 6000 years old and meromixis is sustained through its morphology as well as through the high content of mineral salts originating from a sublacustrine spring (Bonhomme et al 2011). The partitioning of microbial populations between water column compartments was addressed in several cultivation-independent surveys (Biderre-Petit et al 2011b, Lehours et al 2005, Lehours et al 2007).

Another well-studied meromictic lake that evolved under the influences of subsurface springs and volcanic activity is Lake Kivu, located in the East African Rift valley on the border between Congo and Rwanda. The deep water maintains a higher density due to inflow of hydrothermal (warm) saline ground water, whereas the shallower water receives less saline (cold) water through subsurface springs that feed into the lake (Pasche et al 2009). The monimolimnion is marked by exceptionally high CH₄ and CO₂ concentrations, and while CO₂ was identified to be mainly of geogenic origin, CH₄ was proposed to originate mainly from microbial activity in the lakes sediments and the anoxic water column (Pasche et al 2011). The close proximity to active volcanoes pose a risk for a limnic eruption as has been seen 1986 in Lake Nyos (Lanigan 1989), and has driven fundamental research on the lakes biogeochemistry and physical properties (Findenegg 1935, Natacha Pasche 2009). However, the research interest is not only fostered by the need to estimate the risk of a catastrophic event, but also by the potential to harvest the CH₄ as an energy source for the lake's surrounding communities (<http://www.lake-kivu.org/>). At 485 m depth, Lake Kivu is ranked among the deepest lakes in the world, however meromictic Lake Matano in Indonesia is significantly deeper with a maximum depth of 600 m. Lake Matano differs greatly from any of the previously described lakes: high concentrations of ferrous iron (Fe (II)) make it the largest ferruginous basin in the world (Crowe et al 2011). Lake Matano is therefore a unique study site from an Earth history perspective, as it

provides the opportunity to study microbial ecology in the biogeochemical surroundings of the ferruginous Archean Eon (~2,500 million years ago).

1.4.3 Meromictic lakes: natural laboratories with persistent redox gradients

Meromictic lakes have been subjects of scientific research across many disciplines, ranging from geology and limnology, physics and microbial ecology (Hamilton et al 2014, Moreira et al 2011, Parkin 1980, Prokopkin et al 2014, Vagle et al 2010, Valero-Garcés and Kelts 1995, William M. Last a 2002). From a microbiological perspective, meromictic lakes provide natural laboratories for the study of phylogeny, community structure, and function of microbes along persistent redox gradients. Microbial degradation of organic matter in the stratified water column proceeds from high energy delivering redox reactions in the oxygenated mixolimnion, over intermediate to low energy delivering reactions using alternative TEA in the microaerophilic RTZ, to reactions that proceed close to the thermodynamic limit in the anoxic monimolimnion (Boehrer and Schultze 2008, Humayoun et al 2003). The redox conditions can be compared to the conditions prevalent in pore waters of anoxic marine sediments, with the advantage that metabolic niches span over meters in the water column compared to centi- or millimeters in sediments (Fenchel T. 1996, Sorensen et al 1979).

The net energy gain for microbial respiration is determined by the thermodynamic favorability of the corresponding oxidant, with O₂ being the most favorable followed by NO₃⁻, Mn, As, Fe, SO₄²⁻ and lastly CO₂. Organic matter in the oxygen-rich mixolimnion is therefore decomposed through aerobic respiration by microorganisms that use O⁰ in O₂ as TEA according to following equation (CH₂O represents organic matter):



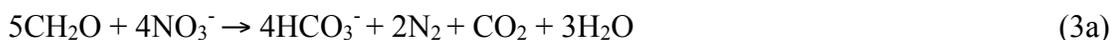
However, next to O_2 availability, solar radiation also shapes the metabolic niche in the mixolimnion by allowing phototrophic organisms to thrive. Among the phototrophic organisms are often also photosynthetic *Cyanobacteria*, which produce O_2 and degrade organic matter according to equation 2.



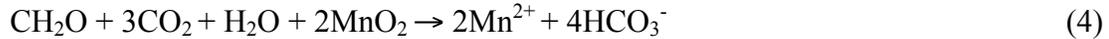
Next to O_2 , *Cynaobacteria* seem to indirectly influence the lake's CH_4 production as well: Gossart and colleagues have recently established that oxygen tolerant methanogenic Archaea produce CH_4 in the surface waters (the epilimnion) of Lake Stechlin, Germany using H_2 produced by *Cyanobacteria* as substrate (Grossart et al 2011). Methanogenesis in oxygenated waters is not restricted to lakes, and has recently been described for the ocean as well (Karl et al 2008, Metcalf et al 2012). However, while Karl and colleagues suggested that methane production in the O_2 rich waters of the ocean is a by-product of methylphosphonate decomposition (Karl et al 2008), Gossart speculated that methanogenesis through methylphosphonate decomposition might not be relevant in the mixolimnion/epilimnion of lake ecosystems (Grossart et al 2011).

With decreasing light availability and decreasing oxygen concentrations (<1mg/L) in the RTZ, the redox potential (E_h) decreases as well, and alternative TEAs become favorable for the microbial energy metabolism. Heterotrophic NO_3^- reduction and dissimilatory nitrate reduction to ammonium (DNRA) through denitrifying bacteria is the first alternative to aerobic respiration

and proceeds according to reactions (3a) and (3b) respectively. Alternative pathways of NO_3^- reduction are driven by chemolithotrophs using H_2S or Fe as electron donors. To date little is known about microbial drivers of heterotrophic and chemolithotrophic NO_3^- reduction in meromictic lakes, however rate measurements in the RTZ of several meromictic lakes provide evidence for its existence (Halm et al 2009, Mallet et al 1998). Upon NO_3^- depletion, transition metals and metalloids such as Mn, As and Fe become favorable as oxidants of organic matter.



Manganese (II) oxidation in the mixolimnion or upper part of the RTZ leads to the formation of sinking particles. Just below the depth of oxygen depletion, these particles are reduced according to equation (4), resulting in a peak of dissolved Mn. Below the RTZ, dissolved Mn concentrations decrease rapidly, which is thought to be a result of precipitation (Newton et al 2011b). Microbial Mn reduction can be driven by diverse microbes within the bacterial and archaeal domains, the most prominent examples are the *Shewanella* and *Geobacter* which are used as model systems to study Mn as well as Fe reduction (Lovley et al 2004). Metal reduction can be coupled to the oxidation of a wide array of carbon compounds, which can be completely oxidized to CO_2 . Moreover, in marine methane seep sediments it has been demonstrated that Mn reduction can be coupled to the anaerobic oxidation of methane (AOM) (Beal et al 2009), a process speculated as pathway for methanotrophy in meromictic lakes as well (Crowe et al 2011, Newton et al 2011b).

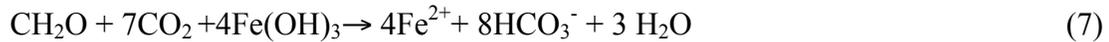
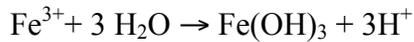


Arsenic exists in four different oxidation states, the predominant form in the oxygen rich mixolimnion is arsenate [As(V), H_2AsO_4^- , HAsO_4^{2-}] while the anoxic parts of the water column are dominated by arsenite [As(III), H_3AsO_3 , H_2AsO_3^-] (Oremland and Stolz 2003). The structural similarity of H_2AsO_4^- to phosphate (PO_4^{3-}) is the reason for its toxicity to most forms of life: once it has entered the cell it undermines PO_4^{3-} based energy generation, such as oxidative phosphorylation, resulting in cell death (Oremland 2006). Microbes have evolved mechanisms to tolerate moderate to high concentrations of As, and use it as an energy delivering substrate during chemolithotrophic energy generation and anaerobic respiration according to equation (5) (Laverman et al 1995).



After Mn and As follows Fe as next favorable substrate for microbial chemolithotrophic energy generation. The distribution of soluble Fe within the water column of meromictic lakes is often similar to the distribution of Mn. At circumneutral pH, Fe rapidly oxidizes into insoluble ferric iron (Fe (III)) according to reaction (6), and sinks down towards into the microaerophilic redox transition zone and monimolimnion of the lake. Once oxygen is depleted, microbial Fe reduction regenerates the pool of soluble Fe (II) according to equation (7).





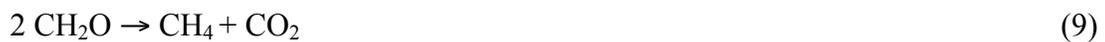
Microbial Fe reduction typically results into a peak of Fe (II) just below the oxygen interface, the soluble Fe (II) then diffuses back to the microaerophilic zone where it is re-oxidized. In meromictic lakes with high sulfide concentrations in RTZ and monimolimnion, Fe is removed from the water column through precipitation into iron-sulfide minerals, such as pyrite (FeS_2). The H_2S content of the monimolimnion is directly correlated to the lake's SO_4^{2-} pool, which is biologically reduced to H_2S according to equation (8). Sulfate concentrations in lake ecosystems vary largely, depending on the mineral composition of rock in the basin, as well as the SO_4^{2-} content of inflowing water and upward diffusion from the sediments (Wetzel 2001). Sulfate reducing microbes are diverse and belong to at least five bacterial and two archaeal phyla (Muyzer and Stams 2008). The electron donor used for the reduction of SO_4^{2-} can vary from organic carbon compounds, such as ethanol, formate, lactate, pyruvate, malate and succinate, over short chain fatty acids, such as propionate, butyrate and acetate to aromatic compounds such as benzoate and phenol (Muyzer and Stams 2008). Moreover, sulfate-reducing microbes are often not restricted to the reduction of SO_4^{2-} , but can also use alternative S compounds such as polysulfide (S_n^{2-}) or thiosulfate ($\text{S}_2\text{O}_3^{2-}$).



Biotic H₂S oxidation in the RTZ of meromictic lakes plays an equally important role as SO₄²⁻ reduction, and is driven by aerobic chemolithotrophic sulfur bacteria as well as by anaerobic anoxygenic phototrophic H₂S oxidizing bacteria. Green and purple sulfur bacteria often thrive in the O₂-H₂S transition zone, and have been studied extensively in meromictic lakes all over the world (Gregersen et al 2009, Meyer et al 2011, Rogozin et al 2009).

Methanogenesis is typically regarded as the last step in the decomposition of organic matter, using substrates such as H₂ and CO₂, acetate or formate, and proceeds according to equation (9). Energy gains during methanogenesis are low, and biological methane production is therefore often dependent on co-metabolic/syntrophic interactions to drive the process forward. The monimolimnion of meromictic lakes as well as its sediments are well known for high methane contents (Biderre-Petit et al 2011b, Crowe et al 2011, Pasche et al 2011, Vagle et al 2010). To date all known methanogens belong to the Archaea, which are discussed in more detail in the following sections.

C⁰ in CH₂O



1.4.4 Methanogenesis: a communal effort to trick thermodynamics

Methanogenic microbial communities can be regarded as prime example for an efficient coordination of metabolic processes to overcome energy barriers and to transform endergonic reactions into exergonic (Stams and Plugge 2009). Traditionally methanogenic Archaea comprise the six lineages *Methanobacteriales*, *Methanococcales*, *Methanosarcinales*, *Methanomicrobiales*, *Methanocellales* and *Methanopyrales*. However, evidence from cultivation

independent surveys in guts of termites and cockroaches suggests that a clade of uncultured *Thermoplasmatales* might form a seventh lineage of methanogenic Archaea, referred to as *Methanoplasmatales* (Paul et al 2012). The physiology of archaeal methanogens is highly specialized and allows for a subdivision into three groups: hydrogenotrophs using H₂/CO₂, acetoclasts using acetate, and methylotrophs using methylated compounds. Their metabolic partners are primary and secondary fermenting bacteria, which oxidize polymeric substrates and break them down into suitable substrates, such as acetate, methylamine, and CO₂/H₂. Acetate and CO₂/H₂ have been described as the most important intermediate substrates for methanogenesis (Conrad 1999).

In environments in which TEA acceptors are readily available, acetoclastic methanogens are often outcompeted by sulfate reducing Bacteria for acetate. The absence of TEA in the monimolimnion of meromictic lakes eliminates this competition and acetate oxidation is energetically constrained by the low midpoint redox potential for the H⁺/H₂ couple ($E^{0'}$ = -414 mV). This energetic problem can however be overcome in a process called syntrophic acetate oxidation (SAO). In this process homoacetogens run the Wood Ljungdahl pathway in reverse and transfer four H₂ molecules to hydrogenotrophic methanogens through interspecies electron transfer (Muller et al 2013, Nusslein et al 2001, Stams and Plugge 2009). Interspecies electron transfer between H₂ producing bacteria and hydrogenotrophic methanogens has been studied in great detail and revealed to be a highly specific syntrophic interaction (Stams and Plugge 2009). The interaction between the partners is tightly regulated such that they adapt their activity to each other's metabolic needs. Notably, interspecies hydrogen transfer with methanogens is not restricted to homoacetogens/acetate oxidizers. In fact, the composition of the syntrophic assemblages is mainly dependent on the substrate availability. For example, fatty acids can be

syntrophically degraded by *Deltaproteobacteria* affiliated with the *Syntrophomonadaceae* and *Syntrophaceae* (Cavaleiro et al 2010), and amino acids can be syntrophically degraded by *Firmicutes* affiliated with *Clostridium sporogenes* (Wildenauer and Winter 1986), or *Synergistes* affiliated with *Aminobacterium colombiense* (Worm et al 2010).

1.4.5 The microbial community of meromictic lakes

Biotic interactions (e.g. predator-prey relationships), and abiotic environmental parameters (e.g. geography, landscape position, trophic status) are known to shape microbial communities in the environment, including lakes (Kent et al 2004, Kent et al 2006, Shade et al 2008). A recent study by Kent and colleagues compared community structure and dynamics of five bog lakes in Wisconsin, USA underlining distinct differences in microbial populations between lakes (Kent et al 2007). One factor that was suggested to greatly influence the microbial community was the quality of dissolved organic matter (DOM), e.g allochthonous organic matter derived from terrestrial plants and soils versus autochthonous organic matter produced by phytoplankton, aquatic plants, and benthic algae (Kent et al 2007). However, despite undisputable differences in the microbial community composition of geographically isolated lakes, it has been shown that many bacterial groups are shared. Newton and colleagues recently pinpointed this observation in a global survey of freshwater bacteria (Newton et al 2011b), but it can certainly be extrapolated to the microbial community of meromictic lakes. The most prevalent characteristic of meromictic lakes is the division into subpopulations of the microbial community specifically associated with the distinct ecological niches provided in the mixolimnion, RTZ and monimolimnion, as described in the following sections.

1.4.5.1.1 The mixolimnion: an equivalent to freshwater lakes

The mixolimnion is an ecological niche in which resources in form of nutrients, light and O₂ are spatially relatively homogeneously distributed but vary according to the time of the day and season. Most of the bacteria in the mixolimnion show dynamically changing abundance patterns as a response to either grazing pressure, seasonal differences in water temperature, or light availability (Salcher 2013, Shade et al 2011). The ability to adapt to change, and the resilience of the microbial community after mixing events is one of the major characteristics that distinguish the surface community from the communities in RTZ and the monimolimnion (Shade et al 2011). The contrast of the temporal variability in the community structure of the surface waters and the stability in RTZ and monimolimnion was recently confirmed using 454 sequencing data from 5 different meromictic bog lakes in Sweden collected over a period of four years (Peura et al 2012).

The taxonomic composition of the microbial community in the mixolimnion compares to the typical community of freshwater lakes recently described by Salcher, and by Newton and colleagues (Newton et al 2011b, Salcher 2013). The community is marked by abundant freshwater bacteria affiliate with the *Alphaproteobacteria*, *Betaproteobacteria*, *Burkholderiales*, *Bacteroidetes*, *Actinobacteria*, *Cyanobacteria*, and *Verrucomicrobia*. This was consistently observed in cultivation-independent surveys of globally distributed meromictic lakes such as Ace Lake, Antarctica (Lauro et al 2011), Lake Faro, Italy (Gugliandolo et al 2011), several Swedish bog lakes (Peura et al 2012), and Lake A, Nunavut Canada (Comeau et al 2012).

1.4.5.1.2 The redox transition zone: passage from known to unknown

The RTZ in meromictic lakes is the part of the water column with the highest density and activity of microbes (Camacho et al 2001, Jorgensen et al 1979, Overmann 1997). It provides a microaerophilic environment that often hosts an abundant population of aerobic methane oxidizers affiliated with the *Gammaproteobacteria* (Biderre-Petit et al 2011b, Pasche et al 2011). Once O₂ is depleted and H₂S concentrations increase, conditions are ideal for H₂S oxidizing anoxygenic phototrophs affiliated with the *Chlorobi* (Gregersen et al 2009, Lauro et al 2011). Indeed, primary production can be dominated by *Chlorobi* in meromictic lakes where the O₂-H₂S transition zone is exposed to high levels of sunlight (Habicht et al 2011).

The availability of alternative TEA in the RTZ further supports the growth of diverse anaerobic respiring organisms, often affiliated with the *Deltaproteobacteria* (Humayoun et al 2003, Junier et al 2010). Moreover, bacterial candidate divisions can make up large proportions of the microbial community. In Lake Pavin, for example, members of the SR1, OD1, OP11 make up 15% of the microbial community in the oxygen depleted parts of the RTZ (Borrel et al 2010), transitioning into the monimolimnion, which is often dominated by MDM.

1.4.5.1.3 The monimolimnion: a methanogenic community enriched in MDM

Physicochemical conditions in the monimolimnion of meromictic lakes provide an environment for methanogenic Archaea (Lauro et al 2011, Lehours et al 2005), as well as for abundant and diverse candidate divisions. Indeed, cultivation-independent surveys in geographically isolated meromictic lakes confirmed a natural enrichment of these enigmatic microbial groups in the H₂S and CH₄ saturated waters of the monimolimnion. Small subunit ribosomal RNA (SSU or 16S rRNA) gene clone library sequencing in Lake Pavin identified at least five bacterial candidate

divisions including OP1, OP3, OP10, OP11 and WS5 (Lehours et al 2007). A more recent study using a combination of 16S rRNA gene clone library sequencing and PhyloChip hybridization in Mahoney Lake identified at least eight bacterial candidate divisions including OP1, OP8, OP9/JS1, OP11, TM6, WS1, WS3 and ZB2 (Klepac-Ceraj et al 2012). A higher-throughput study in Arctic Lake A identified 16S rRNA gene pyrotag sequences affiliated with at least four candidate divisions including OP3, OP8, OP9 and OP11 that increased in abundance within the monimolimnion (Comeau et al 2012). Moreover, a single-cell genomic study of MDM that elucidated the coding potential of 201 uncultivated archaeal and bacterial cells from nine different environments sourced 81 of these SAGs from a single depth interval in monimolimnion of meromictic Sakinaw Lake, British Columbia Canada. The Sakinaw Lake SAGs belonged to 11 different candidate divisions affiliated with CD 12, OD1, OP1, OP3, OP8, OP9/JS1, OP11, WS3, WWE1 and one novel Euryarchaeote. We are just beginning to understand the functional role of various MDM groups in natural and engineered ecosystems, and cultivation-independent technologies including single cell genomics provided first insights into their metabolic potential.

1.4.6 Shining a spotlight on MDM

1.4.6.1 Candidate division OP3: “Omnitrophica”

Members of the Candidate phylum OP3 thrive in a variety of anoxic environments, such as rice paddy soils (Derakshani et al 2001), methanogenic bioreactor (Tang et al 2007), and in the redox transition zone (RTZ) of stratified lakes (Comeau et al 2012). A phylogenetic study by Glöckner and colleagues revealed that OP3 can be subdivided into four different lineages with correspondence between phylogenetic placement and sampling source (Glöckner et al 2010). Results from this study indicated that metabolic niches inhabited by OP3 are often defined by

redox cycling of Fe/Mn or S. To further gain insight into the metabolic potential of OP3, Glöckner and colleagues generated a fosmid library from rice paddy soil samples (Glöckner et al 2010). They were able to identify three clones that belonged to the OP3, and BLASTP analysis of 90 predicted open reading frames (ORFs) revealed closest relatives among the *Deltaproteobacteria*. Moreover, the identification of 9 ORFs that encoded for NADH dehydrogenases I, and the described niche selection led the authors speculate on a metabolic potential for anaerobic respiration.

Kolinko and colleagues provided insight into the function of OP3 by combining targeted sampling techniques, genomic sequencing and microscopic imaging (Kolinko et al 2012). They isolated individual cells from magnetically collected multi-species assemblages of magnetotactic bacteria (MTB) from lake Chiemsee sediments, and the subsequent generation of SAGs provided genomic evidence that the proposed novel magnetotactic species belonged to candidate division OP3. Magnetotaxis is driven by specific organelles, called the magnetosomes that support orientation along Earth geomagnetic fields. It has been suggested that Bacteria thriving in microaerophilic, chemically stratified environments use magnetotaxis for navigation along redox gradients (Flies et al 2005). A comprehensive morphological study of cell shape and structure using CARD-FISH, differential interference contrast (DIC) light microscopy, and transmission electron microscopy (TEM) confirmed the presence of magnetosomes, and hence the ability for magnetotaxis in the isolated OP3 cell (Kolinko et al 2012). Moreover, energy dispersive X-ray spectroscopy revealed intracellular S inclusions and iron content within the magnetosome. This observation led the authors to speculate about a potential for S oxidation coupled to the reduction of Fe (Kolinko et al 2012), which would be in agreement with previous findings from Glöckner and colleagues (Glöckner et al 2010). Kolinko's phylogenetic analysis of the 16S and 23S rRNA

genes further suggested the placement of OP3 within the Planctomycetes –Verrucomicrobia – Chlamydiae (PVC) superphylum (Kolinko et al 2012), which was later confirmed by Rinke and colleagues (Rinke et al 2013).

1.4.6.2 Candidate division OP9: “Atribacteria”

Members of the candidate division OP9 thrive, similar to the OP3, in anoxic environments such as geothermal springs (Costa et al 2009), petroleum reservoirs (Gittel et al 2009), thermal bioreactors (Tang et al 2011), wastewater sludge treatment plants (Riviere et al 2009) and in the anoxic waters of stratified lakes (Klepac-Ceraj et al 2012). First insights into the genomic potential of OP9 were gained by Dodsworth and colleagues (Dodsworth et al 2013), who were able to generate a near complete OP9 genome with a combined assembly from 15 OP9 SAGs, as well as one near complete OP9 genome generated from metagenomic shotgun sequencing data. Samples for single cell genomics originated from hot spring sediments at Little Hot Creek, California and samples for the metagenomic analysis were taken from in situ cellulose enrichments in Great Boiling Springs Nevada (Dodsworth et al 2013). Results from Dodsworths genome analyses are described in the following section. Phylogenetic 16S rRNA gene analysis and the analysis of 31 conserved protein-coding genes supported the placement of the OP9 as an independent phylum that was comprised of four distinct clades. Based on 16S rRNA phylogeny, candidate division JS-1 was identified as a member of the OP9, and an in-depth phylogenomic study confirmed the affiliation of the JS-1 with the OP9 (Dodsworth et al 2015). Genome wide comparison of the OP9 revealed a close relationship to the phylum *Firmicutes*. Several marker genes, including genes that encode for membrane protein assembly (BamA/YaeT, OmpH), suggested that OP9 have a diderm cell wall structures. The ability to enrich for OP9 in cellulosic

biomass incubations suggested that OP9 are capable of cellulose degradation, further supported by the presence of several genes that encode for glycohydrolases. Complete coverage of the Embedden-Meyerhoff and the pentose phosphate pathways as well as the presence of oxygen sensitive genes such as pyruvate-ferredoxin oxidoreductases, suggested that the OP9 are strictly anaerobic, fermenting organisms. Genes for carbon fixation pathways were not detected, suggesting a heterotrophic lifestyle. Anabolic pathways reconstruction revealed the ability to synthesize all amino acids, except from methionine, but the lack of pathways for the synthesis of several essential vitamins. The authors further suggested that a lacking genomic coverage for the synthesis of essential vitamins suggests a dependency on exogenous vitamins provided by other members of the community, which might be the underlying reasons why OP9 resist the growth in pure cultures.

1.4.6.3 Candidate division OP11: “Microgenomates”

Members of the candidate divisions OP11 inhabit similar ecological niches as the previously discussed OP-phyla, such as marine sediments (Li et al 1999), contaminated aquifers (Dojka et al 1998), wastewater digesters (Hatamoto et al 2007), and the anoxic waters of permanently stratified lakes (Borrel et al 2010). When the OP11 were first described, the phylogenetic divergence within its own clade (33%) exceeded observations for all previously described phyla (Harris et al 2004). Harris and colleagues doubted that the observation represented the true divergence of OP11, and to gain better insight into their phylogeny, they used 240 16S rRNA sequences that were generated with OP11 specific as well as universal bacterial primer from samples that were collected in more than 10 different environments to construct robust phylogenetic trees (Harris et al 2004). With their analysis they were able to show that the

original OP11 phylum could be divided into three separate phyla. The proposed names for the novel OP11 derived candidate divisions were OD1 (OP11 derived-1), SR1 (sulfur river 1) and OP11 (three remaining OP11 clades). Despite the division into three separate phyla, OP11 remains to be one of the most divergent phyla (29.8%), exceeding even the divergence observed in the *Proteobacteria* (23%) (Harris et al 2004), potentially suggesting that OP11 might comprise members with diverse metabolic potential.

First insight into the genomic potential for OP11 was gained from one metagenomic (Wrighton et al 2012), and two single cell genomic studies that were able to reconstruct collectively more than 20 near complete OP11 genomes (Rinke et al 2013, Youssef et al 2011). Phylogenomic analysis of 38 conserved protein-coding regions placed the OP11, together with the candidate division BD-1 and candidate division GN02, in the novel superphylum “Patescibacteria” (Rinke et al 2013). None of the three above-mentioned studies detected genes that encode for the electron transport chain, and genomic evidence pointed congruently towards a non-respiring, fermentative lifestyle. Based on the detection of genes that encode celluloseA, Youssef and colleagues speculated the potential to degrade starch and cellulose into fermentable end products (Youssef et al 2011). Wrighton and colleagues reported the potential for glycolysis via the Embdedden-Meyerhoff pathway as well as the potential for ATP generation by a proton motif force through a membrane bound pyrophosphatase (Wrighton et al 2012). Moreover, the detection of a membrane bound Ni-Fe hydrogenase, closely related to homologues in Archaea, suggested the potential to either produce H during fermentation, or sulfide (H_2S) if polysulfide (S_n^{2-}) is available. Also Rinke and colleagues detected genes closely related to the Archaea, specifically genes that encode for purine biosynthesis (Rinke et al 2013).

1.4.6.4 Insights into archaeal candidate divisions

When Woese first introduced Archaea as third domain of life, they were represented by only two phyla, the Euryarchaeota and the Crenarchaeota, (Woese et al 1990). In the first years after their discovery, it was generally thought that Archaea exclusively thrive in extreme environments such as haline brines (Andrei et al 2012), anoxic waters and sediments (Liu et al 2012) and in environments with geothermal activity (Jannasch et al 1992). This hampered view of the Archaea was due to limitations in the early days of cultivation independent approaches that overlooked Archaea with low abundance and/or small genome size (Baker et al 2010), or missed abundant community members because of biases in presumably universal archaeal primer and FISH-probes (Teske and Sorensen 2008, Youssef et al 2014). However, higher sequencing throughput, and technologies that bypass the use of primers led to a vastly expanded view of their geographic distribution. Indeed, recent environmental studies have identified Archaea as abundant members in temperate environments, such as the Ocean (Karner et al 2001), lakes (Vila-Costa et al 2013) and soil (Tripathi et al 2013).

The acquisition of novel archaeal 16S rRNA gene sequences led to the description of several new phyla, including Nanoarchaeota (Huber et al 2002), Korarchaeota (Barns et al 1996), Thaumarchaeota (Torna et al 2011), Aigarchaeota (Nunoura et al 2011), and Geoarchaeota (Kozubal et al 2013). Moreover, in 2011 Guy and colleagues proposed the first archaeal superphylum comprising the Thaumarchaeota, Aigarchaeota, Crenarchaeota and Korarchaeota (TACK) (Guy and Ettema 2011). The previously mentioned MDM study from Rinke and colleagues followed with the preposition of a second archaeal superphylum, which included four novel phyla denominated as Diapherotrites, Parvarchaeota, Aenigmarchaeota, and Nanohaloarchaeota as well as the previously described Nanoarchaeota (DPANN). In addition to

the insights into the phylogenetic relationships of DPANN, Rinke and colleagues uncovered several genomic novelties reversing the traditional view on Archaea. Representatives of the Nanoarchaea and Diapherotrites were described as the first Archaea with complete coverage for bacterial sigma factors as well as the first Archaea with genomic potential for the bacterial stringent response, which is an alarmone-based stress response initiated by the deployment of 5'-diphosphate 3'-diphosphate guanosine (ppGpp) signaling molecules.

1.5 Objectives

The overall goal of this dissertation was to study the uncultivated microbial majority that thrives in meromictic Sakinaw Lake, to gain insight into population structure and phylogeny, as well as genomic and economic potential of microbial dark matter.

1. **Chapter 2: Evaluate the temporal evolution and physicochemical characteristics of Sakinaw Lake.** I discussed physicochemical measurements from samples collected between 2007 and 2013 in the context of previous geochemical studies in Sakinaw Lake.

2. **Chapter 3: Determine the population structure, ecological interactions and potential metabolic roles of MDM groups.** I used 454 sequencing of the 16S rRNA gene sequencing from 12 samples collected along the water column of Sakinaw Lake in May 2011, to gain insight into the microbial population structure, diversity and co-occurrence patterns with emphasize on MDM.

3. **Chapter 4: Assess population structure and functions of MDM that thrive in the RTZ of Sakinaw Lake and identify potential temporal perturbation.** I used 454 sequencing of the 16S rRNA gene from 66 samples collected along the Sakinaw Lake water column between June 2007 and May 2011, 16S rRNA clone libraries generated with Bacteria specific primers from 6 samples collected in June 2007, and shotgun metagenome sequencing of 16 samples collected between August 2009 and May 2011, to describe phylogeny, function and population structure of MDM in the RTZ of Sakinaw Lake.

4. **Chapter 5: Establish diversity, abundance, population structure and function of Archaea in Sakinaw Lake.** I used a combination of catabolized reporter deposition fluorescence in situ hybridisation (CARD-FISH) with a previously published Archaea specific probe, 16S rRNA and *mcrA* gene clone libraries, previously described 454 genomic and 16S rRNA gene sequences, and one previously published Sakinaw Lake single-cell genome to describe diversity, abundance, population structure and function of Archaea in Sakinaw Lake.

Chapter 2: Meromictic Sakinaw Lake

2.1 Synopsis

Sakinaw Lake is situated at an elevation of ~ 5 m above sea level, near the northern end of the Sechelt Peninsula on the Sunshine coast of British Columbia, Canada (49°40.8'N, 124°00.39'W). Originally a fjord opening to the Strait of Georgia, Sakinaw Lake was almost completely isolated from the ocean due to coastal uplift ~11,000 BP, following the last ice age (Clague and James 2002). It consists of two basins: a relatively shallow freshwater basin 49 m deep and a more extensive salt stratified basin 140 m deep (Vagle et al 2010). The larger basin has a width of ~ 0.7 km and is ~ 9 km long. For thousands of years a small stream named Sakinaw Creek was the only conduit between Sakinaw Lake and the Strait of Georgia. In 1952, Sakinaw Creek was dammed to better manage water levels for development projects in the surrounding watershed preventing seawater ingress. In the mid 1960's, Northcote and colleagues first described meromictic Sakinaw Lake as a permanently stratified ecosystem (Northcote 1964). In the early 1990's, Perry and colleagues explored S chemistry with focus on the potential for pyrite formation (Perry 1990), and the rather unusual biogeochemistry of Sakinaw Lake became evident for the first time. Then, in 2010 Vagle and colleagues identified one of the most intriguing features of the deep waters in Sakinaw Lake: a methane bubble curtain caused by the high CH₄ concentrations in sediments and monimolimnion (Vagle et al 2010). Next to the interest in Sakinaw Lake from a scientific perspective, it has always been a point of interest from the economic side as well: it is not only a well-situated recreation site with many vacation homes along the shoreline, but also the habitat for an endangered species of Sockeye Salmon (http://www.dfo-mpo.gc.ca/species-especies/species-especies/sakinaw_sockeyesalmon-saumonsockeye-eng.htm), and an important watershed for the surrounding coastal communities.

Here, I describe the biogeochemical characteristics of Sakinaw Lake as measured in eight surveys between 2007 and 2013, and relate my observations to previous studies. Overall, I could show that the water column stratification can be regarded as stable, and that more contemporaneous measurements generally agreed with previous observations made by others (Perry 1990, Vagle et al 2010). However, despite the remarkably stable salinity gradient, I observed variability in H_2S , NH_4^+ and PO_4^{2-} concentrations in the monimolimnion of the lake, which due to the lack of time series sampling efforts in previous studies, have not been described.

2.2 Material and methods

2.2.1 Chemical profiling

All salinities are on the TEOS-10 Reference Composition Salinity Scale, with the salinity anomaly assumed to be zero (IOS 2010). Dissolved oxygen concentrations were determined by Winkler titrations (Winkler 1888b). Hydrogen sulfide concentrations were measured from water samples fixed with 2% final concentration of Zinc Acetate and analyzed in the lab using the methylene blue method as described by Cline and colleagues (Cline 1968). Transition metals including iron (Fe), manganese (Mn), and arsenic (As), and SO_4^{2-} concentrations were determined at Maxxam Analytics (Burnaby, BC Canada) using standard methods for the examination of water and wastewater published by the American Public Health Association (American Public Health Association (APHA) 1999). Sulfate was determined with automated colorimetry according to the standard protocol SM 4500 SO_4^{2-} within 24h after sample collection. Samples for dissolved fractions of Fe, Mn and As were filtered through a 0.45 μm membrane filter and preserved with nitric acid (HNO_3) prior to analysis with inductively coupled

plasma mass spectrometry based on standard protocol EPA 200.8 within 14 days of sampling. Dissolved Methane (CH_4) was measured from 13 samples (5 m, 10 m, 25 m, 30 m, 33 m, 36 m, 40 m, 45 m, 50 m, 55 m, 60 m, 80 m, 120 m) collected in June 2007 by GC-MS using a static headspace-equilibrium technique as previously described by Zaikova et al (Zaikova et al 2010). As CH_4 concentrations in the monimolimnion of Sakinaw Lake exceed detection limits of the applied method, our data are estimates and indicate that CH_4 concentrations in the deep Sakinaw Lake waters exceed atmospheric saturation values.

2.3 Results and discussion

2.3.1 Permanent water column stratification

Absolute Salinity (S_A), temperature and O_2 profiles in Sakinaw Lake were determined between 2007 and 2013 and revealed persistent water column stratification. Based on the salinity profiles the water column could be divided into the mixolimnion spanning from 5 m to 30 m, the halocline spanning from 30 m – to 55 m, and the monimolimnion spanning from 55 m to 120 m (Figure 2.1 A). Water temperatures in the mixolimnion were indicative of the sampling time point, with maximum temperatures around 19°C in June and minimum temperatures around 6°C in January. Below 30 m temperature changes between sampling dates were less than 0.3°C, and below 40 m less than 0.05°C (Figure 2.1 B). Oxygen concentrations at 30 m changed up to 5 ml/L between sampling time points (Figure 2.1 C). The variation of O_2 in the upper part of the mixolimnion can be attributed to higher O_2 production during the summer month, where light levels and daylength are longer and promote the activity of photosynthetic organisms.

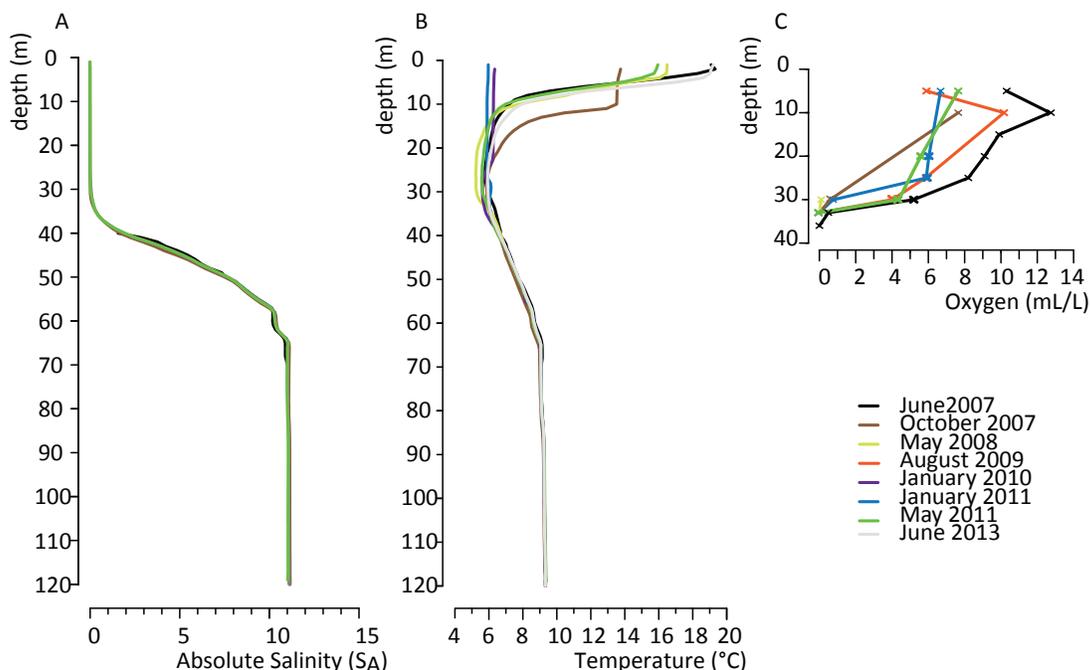


Figure 2.1 Permanent water column stratification in Sakinaw Lake 2007 - 2013.

(A) Measurements of Absolute salinity (S_A) in the Sakinaw Lake water column revealed a stable gradient with fresh water in the mixolimnion (0 m – 30 m) increasing salinity along the halo-/pycnocline (30 m - 55 m), a prominent step-wise salinity increase around 65 m, and constant salinity in the monimolimnion (65 m and 120 m). (B) Temperature profiles revealed season dependent change in the mixolimnion, a gradual increase in temperature along the water masses of the halocline, and constant temperatures in the monimolimnion. (C) Oxygen (O_2) concentrations in the mixolimnion revealed seasonal changes. The depth of O_2 depletion was inconsistent and was measured between 30 m and 33 m.

2.3.2 Sulfur chemistry

Sources of S in inland waters originate mainly from weathering of rocks, fertilizers, atmospheric precipitation and dry deposition (Wetzel 2001). Additionally, S is an essential component of all living organisms and can range from 0.05 – 5% dry weight of biomass. Moreover, it is well established that S respiring Bacteria store inorganic S compounds intra and extracellularly

reaching mM concentrations (Berg et al 2014). Indeed, it has been shown that species of the Gammaproteobacteria can store up to 1,000 times higher SO_4^{2-} in their cells than found in the surrounding freshwater (Kojima et al 2007). Furthermore, it was shown that depending on environmental conditions S can be stored as a variety of species such as cyclooctasulfur (S_8), inorganic polysulfides (S_n^{2-}), polythionates ($\text{S}_n\text{O}_6^{2-}$), and long-chain organosulfanes ($\text{R}-\text{S}_n-\text{R}$) (Berg et al 2014). Organic S compounds comprise mainly sulfur-containing amino acids (cysteine and methionine), sulphonates, organic esters and sulphuric acid. However, in lakes the biological pool of S is considered small compared to the inorganic forms. The most stable form of inorganic S in aerobic environments is SO_4^{2-} , whereas anaerobic waters are dominated by H_2S and bisulfide (HS^-). Other S compounds typically found at lower concentrations are partly oxidized species, such as thiosulfate ($\text{S}_2\text{O}_3^{2-}$), sulfite (SO_3^-) and polythionates ($\text{S}_n\text{O}_6^{2-}$). Karen Perry assessed the Sakinaw Lake S chemistry in great detail in a doctoral study in the early 1990's, and results from her measurements are summarized in Table 2.1. Her analysis revealed high levels of H_2S , and additionally unexpectedly high levels of S_n^{2-} . Perry speculated in detail about potential artifacts in her measurements for S_n^{2-} detection, but came to the conclusion that her numbers were accurate. Perry never discussed a potential biological origin of S_n^{2-} , however the extraordinary S_n^{2-} profile in Sakinaw Lake next to the natural enrichment of MDM with potential contribution on S cycling (e.g. OP11), provide an interesting opportunity to study biological components of S speciation in more detail.

For this study SO_4^{2-} concentrations were measured at two time points and H_2S concentrations at six time points between 2007 and 2011 (Figure 2.2). The mixolimnion revealed low SO_4^{2-} concentrations reaching $\sim 25 \mu\text{M}$ in October 2007 and $\sim 45 \mu\text{M}$ in May 2011. Maximum SO_4^{2-} concentrations of $\sim 57 \mu\text{M}$ were measured at 36 m in both years. The SO_4^{2-}

profile in the lower part of the RTZ (40 m – 55 m) varied, and revealed SO_4^{2-} depletion in October 2007, and SO_4^{2-} rich water in May 2011. The monimolimnion (60 m- 120 m) was devoid of SO_4^{2-} in both years. Sulfide was absent from the oxygenated surface waters, but concentrations rapidly increased below the oxygen interface (~ 33 m see also O_2 profiles above), and varied up to four fold between 2007 and 2011. Maximum H_2S concentrations in the monimolimnion varied as well, and were measured between 2.5 mM and 7 mM. The underlying reasons for the drastic changes in H_2S concentrations remain unknown. Given the relatively low SO_4^{2-} concentrations in the surface water, it seemed unlikely that the inflow of stream water or run-off supplies enough SO_4^{2-} to explain H_2S changes in the mM range. Karen Perry was also staggered by the high H_2S concentrations in Sakinaw Lake. She compared measurements with meromictic Powell Lake, which is located a few kilometers further north on the Sunshine Coast, and has evolved in the same geological timeframe as Sakinaw Lake. However, while Sakinaw Lake is only 5 m above sea level, Powell Lake is located more than 50 m above sea level. Powell Lake had only half as much H_2S in its monimolimnion, compared to Sakinaw Lake, and Perry suggested that Sakinaw Lake might have received additional SO_4^{2-} through intrusion of seawater during high tides and strong onshore winds. Visual inspection of the stream that connects Sakinaw Lake with the ocean, and the dam that controls water levels in the lake are not consistent with recent ingress of SO_4^{2-} rich seawater suggesting an alternative subsurface inflow of S-rich water.

Table 2.1 Sulfur compounds typically found in aquatic ecosystems.*

Inorganic Molecules

Compound	Name	Oxidation State	Detected in Sakinaw Lake
SO ₄ ²⁻	sulphate	(+6)	yes
SO ₃ ²⁻	sulphite	(+4)	No
SO ₂	sulphurdioxide	(+4)	ND
S _n O ₆ ²⁻	polythionate ion	(+5)	No
S ₂ O ₃ ²⁻	thiosulphate	terminal S -2, central S +6	No
S ₈	elemental sulfur	0	yes
S _n ²⁻	polysulphide	terminal S (-1)	yes
H ₂ S	hydrogen sulphide	(-2)	yes
FeS	iron monosulphide	(-2)	yes
Fe ₃ S ₄	greigite	(-2)	yes
FeS ₂	pyrite	(-1)	yes

Organic Molecules

Name	Acronym	Oxidation State	
dimethyl sulfide	DMS	(-2)	ND
dimethylsulfonio- propionate	DMSP	(-2)	ND
methanethiol	MT	(-2)	ND
methyonine	met	(-2)	ND
cysteine	cys	(-2)	ND

* The table summarizes results from Karen Perrys study (Perry 1990), see also main text.

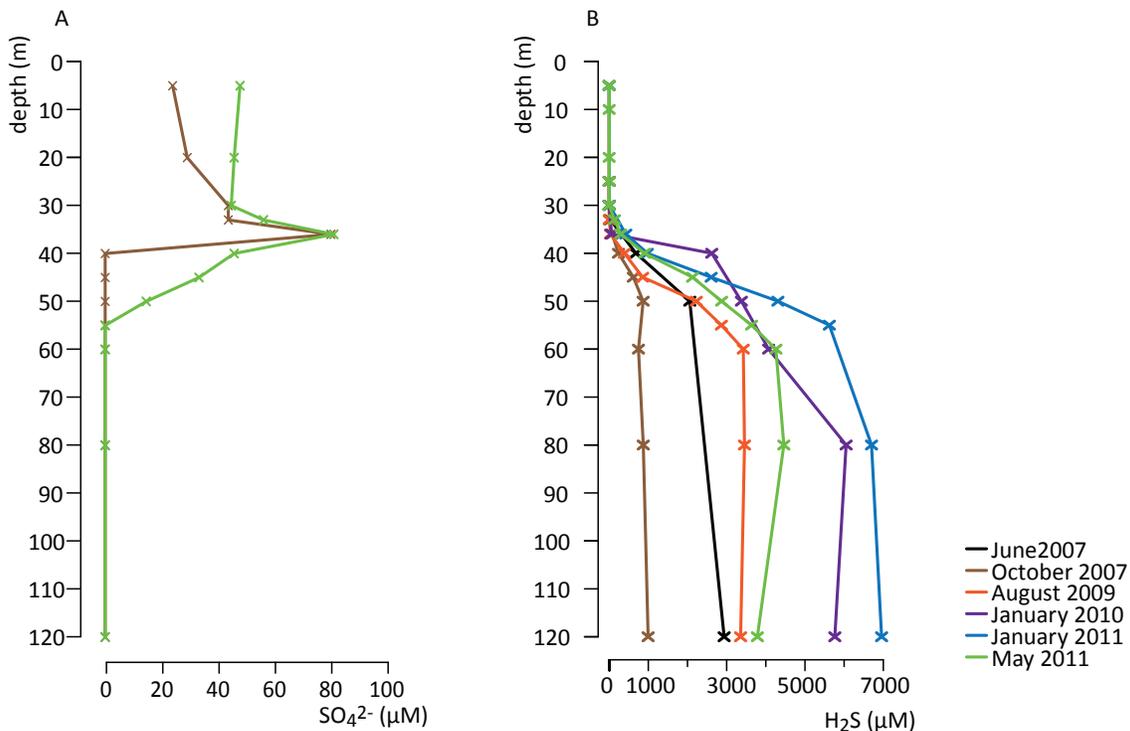


Figure 2.2 Sulfide and sulfate profiles in Sakinaw Lake.

(A) Sulfate (SO₄²⁻) was measured in October 2007 and May 2011 and revealed a consistent depth for peak concentration (~80 μM) at 36 m. However, while SO₄²⁻ was depleted at 40 m in October 2007, SO₄²⁻-rich water masses intruded the lower part of the redox transition zone (40 m – 55 m) in May 2011. (B) Sulfide (H₂S) was measured on 6 field campaigns and revealed up to four-fold changes in redox transition zone and monimolimnion of Sakinaw Lake.

2.3.3 Iron and manganese

The availability of transition metals iron (Fe) and manganese (Mn) plays a key role for anaerobic respiration in the RTZ of stratified lakes. The distribution of soluble Fe and Mn species in the water column are affected by biotic and abiotic factors as discussed in chapter 1 (1.3.4). In this study soluble fractions of Fe and Mn were measured in October 2007 and May 2011, and both transition metals revealed to be almost completely absent from oxygenated surface waters,

displayed peak concentrations in the RTZ, and had a distinctly lower concentration in the anoxic-sulfidic part of the water column (Figure 2.3). This profile was consistent with previous measurements (Perry 1990), and is typical for H₂S-rich meromictic lakes. Comparison of Fe profiles revealed close overlap in October 2007 and May 2011, with peak concentrations of 3.2 μM and 4.0 μM at 36 m respectively. However, in May 2011 a second Fe peak of 1.2 μM was measured at 55 m, which was not observed in October 2007. Manganese profiles were similar in the two years, but maximum concentrations measured in October 2007 were 86 μM at 33 m and 57 μM at 36 m in May 2011.

As already mentioned, availability of soluble Fe and Mn in the water column is dependent on abiotic and biotic processes. Abiotic processes could be increasing oxygen concentrations in the microaerophilic zone of the RTZ due to a mixing event in the mixolimnion, and changing concentrations of H₂S, which affect Fe and Mn removal through the formation of insoluble metal-oxides and metal-sulfide complexes. Biotic factors that influence peak concentrations of soluble Mn and Fe fractions could be microbial driven metal reduction. Given the consistency in the depth for the Fe peak at 36 m, it is unlikely that the changes in the depth of the Mn peak were caused by abiotic factors, as one would expect to see a similar change in the Fe profile. Therefore I suspect that the observed differences might be the result of different activity levels in microbial activity associated with Mn reduction.

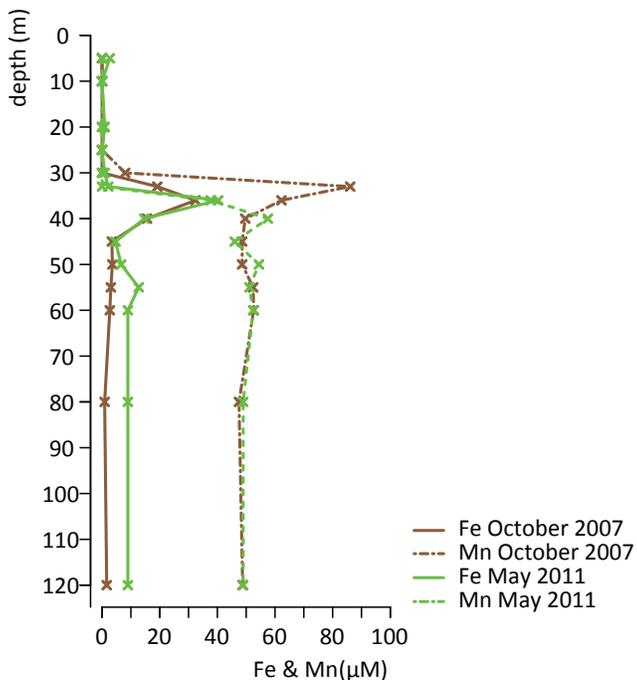


Figure 2.3 Iron and manganese profiles in Sakinaw Lake.

Iron (Fe) and manganese (Mn) were measured in October 2007 and May 2011 and revealed similar profiles for Fe with peak concentrations of 3.2 μM and 4.0 μM at 36 m respectively. Profiles for Mn revealed a distinct shift of 8.6 μM at 33 m in October 2007 to 5.7 μM at 36 m in May 2011.

2.3.4 Arsenic

Arsenic is not only an anthropogenic pollutant, it also occurs naturally in terrestrial and aquatic ecosystems, including lakes (Oremland and Stolz 2003). Its natural forms are mobilized by weathering of rocks, biologic/microbial activity, and volcanic emission. As discussed previously, microbes have not only evolved As detoxification pathways, but they can also use this metalloid as TEA for energy generation. To infer whether the RTZ and monimolimnion of Sakinaw Lake provide a metabolic niche for As respiring organisms, we measured concentrations in October 2007 and May 2011. In both years As was detected in the low nM range (<10 nM) in the

mixolimnion, and concentrations increased along the RTZ to maximum concentrations of ~ 90 nM at 60 m and below. The concentration profile of As in Sakinaw Lake is not unexpected, and has been observed in other meromictic lakes (Seyler and Martin 1989). It is widely accepted that As adsorbs to Mn oxide or co-precipitates with ferric oxyhydroxide oxygen rich waters, leading to its removal from the mixolimnion, while As (III) is mobilized - most likely through microbial activity - along RTZ and monimolimnion causing this particular distribution curve (O'Day et al 2004).

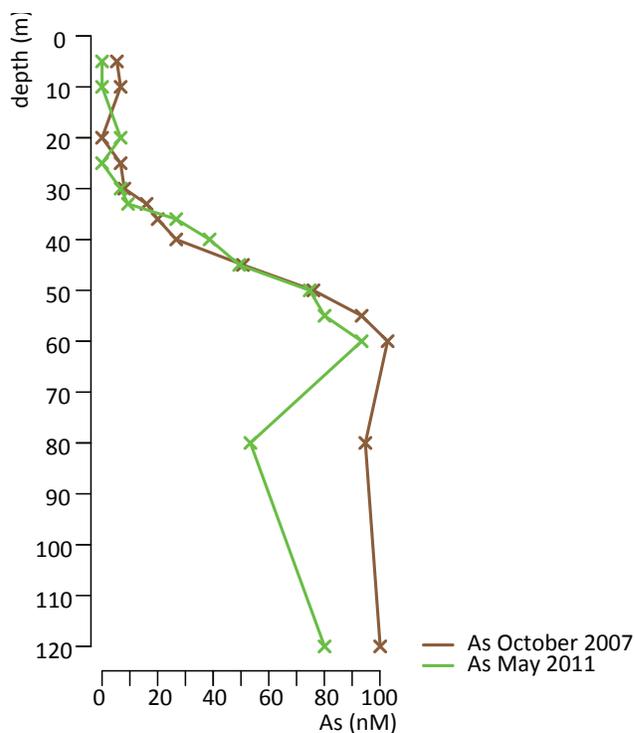


Figure 2.4 Arsenic profiles in Sakinaw Lake.

Arsenic (As) concentrations in Sakinaw Lake measured in October 2007 (brown) and May 2011 (green) were similar with increasing As concentrations along the RTZ (33 m - 60 m) and maximum concentrations of 102 nM and 93 nM at 60 m respectively.

2.3.5 Nitrogen

Nitrogen (N) is a major constituent of all cellular life and is one of the key nutrients in aquatic ecosystems. Nitrogen inputs in lakes can be through inflowing water, N rich fall out, as well as through microbial N fixation (Wetzel 2001). Nitrogen losses occur through sedimentation of N-rich organic molecules, outflow of N rich water and microbial denitrification. The cycling of N through the ecosystem underlies a complex interplay of biochemical processes, which in the case of lakes, is mainly driven by microbial participants (Wetzel 2001). Bacterial mediated redox reactions result in the oxidation and reduction of N compounds, which directly affect assimilation and utilization by algae and larger plants. The major N compounds in lake ecosystems are nitrogen gas (N_2), ammonia (NH_4^+), nitrite (NO_2^-), nitrate (NO_3^-), and organic compounds, such as amino acids and proteins and nucleotides. In meromictic lakes NO_3^- , NO_2^- and NH_4^+ concentrations are typically low in the oxygen rich mixolimnion, where plants and photosynthetic bacteria assimilate N quickly. Peak concentrations of NO_3^- and NO_2^- are generally measured just above the oxygen interface where nitrifying bacteria oxidize NH_4^+ , which diffuses upwards from the anoxic RTZ and monimolimnion. This is also the part of the water column where microbial denitrification occurs, leading to the reduction of NO_3^- and NO_2^- to N_2 . The monimolimnion is typically marked by high concentration of NH_4^+ , caused by deamination of organic compounds through heterotrophic Bacteria, promoting DNRA as second denitrifying pathway in the RTZ.

Karen Perry has determined NO_3^- at maximum concentrations of 7 μM and NO_2^- at maximum concentrations of 0.24 μM both measured at 30 m (Perry 1990). I measured NO_3^- and NO_2^- combined with maximum values between 4.5 μM to 10 μM at 30 m depth. Only in October 2007 maximum concentrations were measured at 25 m. Ammonium concentrations in the

mixolimnion were below the detection limit of the method, and varied between 4.5 mM and 10 mM in the monimolimnion. The cause variations in NH_4^+ and NO_3^- remain unknown.

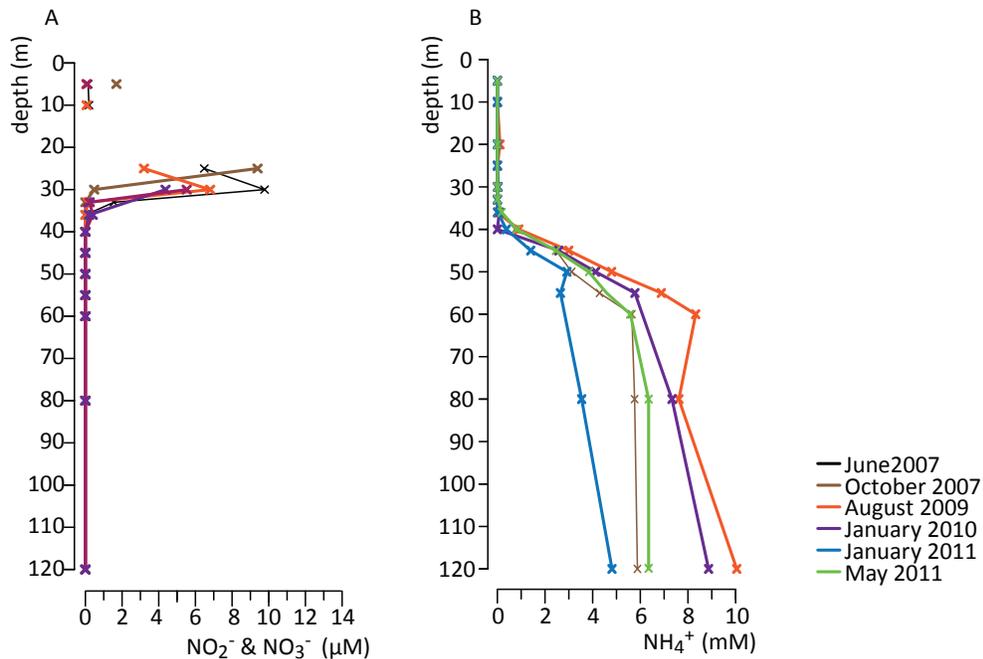


Figure 2.5 Nitrogen profiles in Sakinaw Lake.

(A) Nitrate (NO_3^-) and nitrite (NO_2^-) measured at five time points between 2007 and 2011 in Sakinaw Lake. Sample collection in the mixolimnion (0 m - 30 m) was not sufficient to determine shape of the curve in the surface waters. Maximum concentrations were measured 4.5 μM and 10 μM at 30 m. (B) Ammonium (NH_4^+) was absent from the oxygen rich surface waters, concentrations increased distinctly along the RTZ (33 m – 55 m), and reached maximum concentrations of 10 mM in the monimolimnion (60 m - 120 m).

2.3.6 Phosphate

In comparison to other nutrients, phosphorus (P) is typically the least abundant in lake ecosystem that are unaffected by anthropogenic eutrophication, and often limits biological activity. Phosphorus is overall very reactive and forms insoluble compounds with cations (e.g. Fe), and absorbs to colloids and particulate compounds, which leads to its removal from the water

column. Inorganic phosphorus is only available for biological processes if it occurs in the form of PO_4^{3-} , organic forms of phosphorus are phosphate esters, sugar phosphates, phosphorylated hydroxyamines, amino acids, nucleotides and phospholipids (Wetzel 2001). The majority, 90%, of P in natural waters is of organic origin and is present as part of the living biota or dead cells.

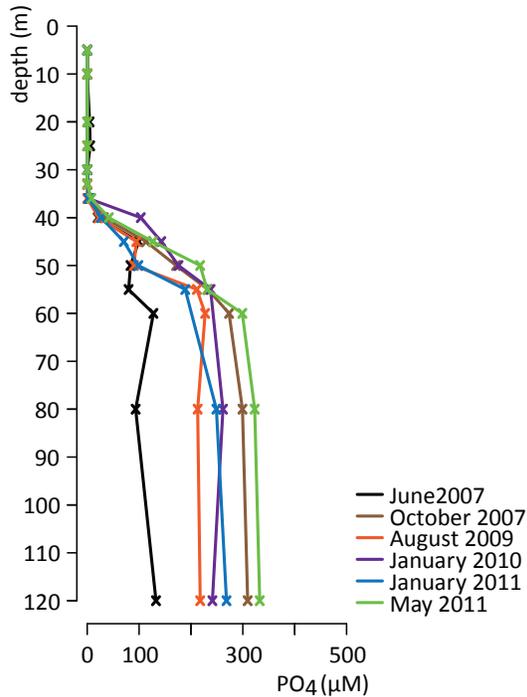


Figure 2.6 Phosphorus profiles in Sakinaw Lake.

Phosphorus (PO_4^{3-}) was almost completely absent from the oxygenated mixolimnion, and concentrations increased along the RTZ reaching up to $\sim 300 \mu\text{M}$ in the monimolimnion (60 m - 120 m).

In Sakinaw Lake dissolved PO_4^{3-} was almost completely absent from the oxygenated mixolimnion, and concentrations increased along the RTZ reaching up to $\sim 300 \mu\text{M}$ (Figure 2.5), Phosphate concentrations as high as $\sim 300 \mu\text{M}$ are quite unusual, and Karen Perry has underlined in her study that nearby meromictic Powell Lake had maximum PO_4^{3-} concentrations of only $0.7 \mu\text{M}$ (Perry 1990). The origin of the high PO_4^{3-} concentrations in Sakinaw Lake remains

unknown, it is however well documented that Bacteria excrete PO_4^{3-} if P-rich organic matter is available in excess (Jansson 1993).

2.3.7 Carbon

2.3.7.1 Dissolved organic carbon

Substantial part of autochthonous dissolved organic carbon (DOC) in aquatic ecosystems are the products of photosynthetic organisms, such as carbohydrates, amino acids peptides and organic acids (Sundh 1992). However, in lakes where a large proportion of the surface area is part of the littoral zone and with substantial inflow of fresh water, much of the DOC is of allochthonous terrestrial origin (Wetzel 2001). Dissolved organic carbon is an important carbon source for the heterotrophic part of the microbial community, and quantity as well as quality of DOC can greatly influence the composition of a microbial community in lakes (Fujii et al 2012, Kent et al 2007). Moreover, DOC impacts productivity of a lake on many other levels, for example through attenuation of ultraviolet radiation, which reduces harmful UV damage to phytoplankton and zooplankton, and through the binding of metals and nutrients, which reduces their bioavailability (Zhang et al 2010).

In Sakinaw Lake the shorelines are the remainders of a steep walled fjord opening and the freshwater input is small. Karen Perry speculated that most of the DOC in Sakinaw Lake was derived from autochthonous sources. Active logging along the shoreline gives however rise to a previously unrecognized source of carbon (C): obvious in the shallow waters and hidden in the deep waters of Sakinaw Lake are remainders of sunken, partially degraded wooden logs providing C in form of cellulose, hemicellulose and lignin. The DOC concentration in the mixolimnion of Sakinaw Lake was around ~ 2 mg/L and increased steadily along the pycnocline,

culminating at ~ 10 mg/L in the monimolimnion (Figure 2.6). The general DOC profile seems to be typical for stratified lakes e.g. Lake Pavin, France (Lehours et al 2005). Also arctic meromictic lake A was shown to have a similar DOC profile albeit with lower DOC concentrations (3.6 mg/L in the monimolimnion) than in Sakinaw Lake. While similar in distribution, Karen Perry measured higher DOC concentrations with a maximum of 16 mg/L in the monimolimnion of the lake.

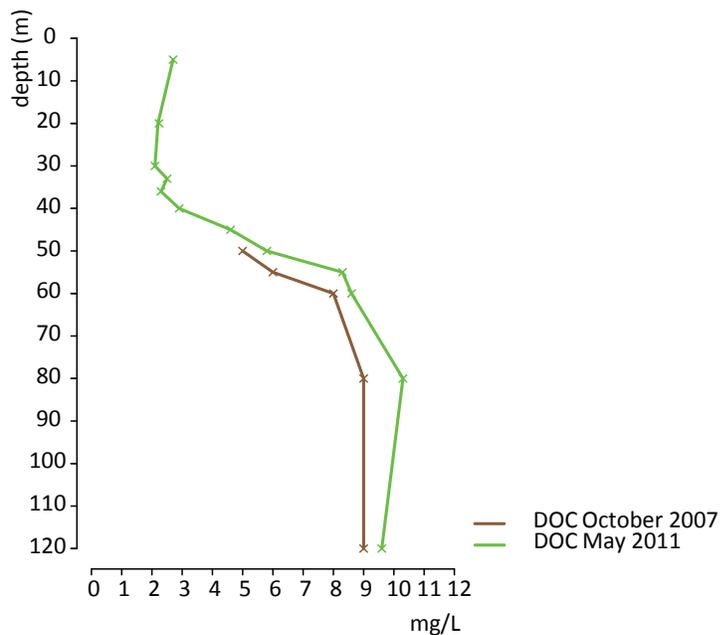


Figure 2.7 Dissolved organic carbon profiles in Sakinaw Lake.

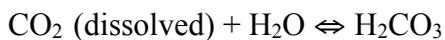
Dissolved organic carbon (DOC) concentration in the mixolimnion of Sakinaw Lake was around ~ 2 mg/L and increased steadily along the pycnocline, culminating at ~ 10 mg/L in the monimolimnion.

2.3.7.2 Alkalinity

The majority of C in fresh waters occurs as equilibrium products of dissolved CO₂ and carbonic acid (H₂CO₃) following reactions (1) (from Wetzel). Dissolved CO₂ slowly hydrates into H₂CO₃, which then dissociates into bicarbonate (HCO₃⁻) and carbonate (CO₃⁻). The amount of

free CO₂ and H₂CO₃ is directly linked the pH of the ecosystem, resulting in a dominance of CO₂ species in acidic, a dominance of HCO₃⁻ species in neutral and a dominance of CO₃²⁻ in alkaline waters.

Carbonate alkalinity is a measure of the buffering capacity, and in freshwaters it is commonly defined as the total concentration of dissolved HCO₃⁻ species, plus the total concentration of dissolved CO₃²⁻ species. Redox reactions driven by anaerobic microbial respiration as described in chapter 1.3.4 are important processes that produce HCO₃⁻, and cause a distinct increase of alkalinity in the RTZ and monimolimnion of meromictic lakes. For example, the reduction of one SO₄²⁻ molecule results in the production of 2 HCO₃⁻ molecules. Alkalinity in Sakinaw Lake was measured in June 2007 and May 2011, where CO₃²⁻ was measured as CaCO₃²⁻ (Figure 2.7). The HCO₃⁻ fraction of the total alkalinity was almost identical in the two years, however CaCO₃²⁻ was higher in June 2007 compared to May 2011, indicating differences in pH that resulted into a shift of the carbonate equilibrium and a decrease of total alkalinity in May 2011.



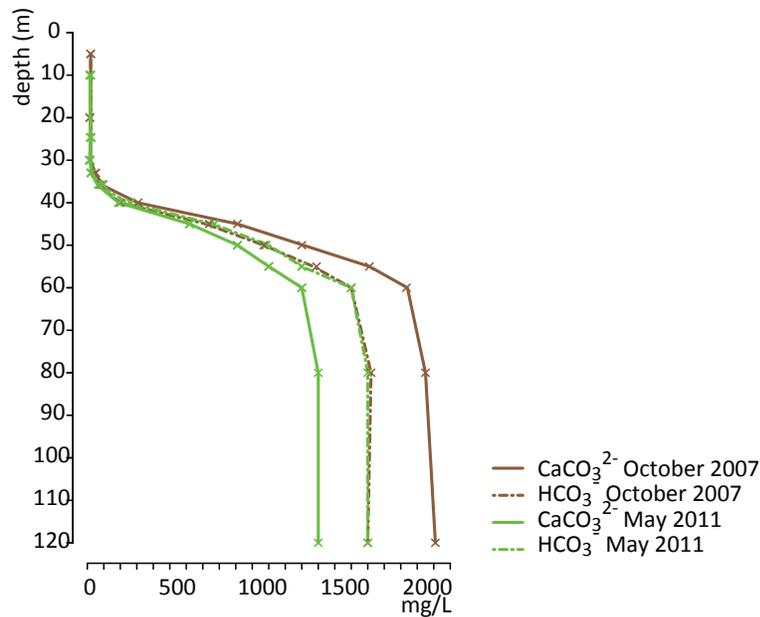


Figure 2.8 Alkalinity profiles in Sakinaw Lake.

Alkalinity was measured in June 2007 and May 2011, where carbonate (CO_3^{2-}) was measured as calcium carbonate (CaCO_3^{2-}). The bicarbonate (HCO_3^-) fraction of the total alkalinity was almost identical in the two years, however CaCO_3^{2-} was higher in June 2007 compared to May 2011.

2.4 Conclusion

This survey of the Sakinaw Lake water column gave first insights into the temporal stability and variability of the water composition. My measurements were overall consistent with results from previous studies in Sakinaw Lake, however H_2S , NH_4^+ and PO_4^{2-} concentrations in the monimolimnion revealed previously unrecognized variations. The cause for the variability remains unknown, however I speculate that there might be either an intrusion of mineral rich water from a subsurface spring, an upward diffusion of nutrients from sediment pore water, or intrusion of seawater during strong onshore winds. Given the irregular sampling intervals, it is currently not possible to link changes in nutrient composition to a season, or a particular weather

condition (e.g. heavy rainfall, or strong onshore winds). The results from this study underline the need for a thorough hydrological study of the lake, and provide a framework to design process-oriented experiments to evaluate function and activity of the microbial community in Sakinaw Lake.

Chapter 3: Taxonomic composition, population structure and diversity of MDM in Sakinaw Lake¹

3.1 Synopsis

This chapter presents the first comprehensive survey of the microbial community in Sakinaw Lake. To infer microbial population composition and structure, samples were collected along the water column and processed for 454 sequencing of the 16S rRNA gene, targeting all three domains of life. Standard multivariate statistics were applied to identify operational taxonomic units (OTUs) differentially associated with the mixolimnion, RTZ, and monimolimnion. The taxonomic composition of the Sakinaw Lake community revealed an unprecedented diversity and abundance of MDM: 25 out of the currently 30 recognized bacterial candidate divisions were recovered from the Sakinaw Lake water column accounting for up to 40% of the microbial community at some depths. To uncover potential metabolic interactions between candidate divisions and other members of the community, a co-occurrence network was generated and putative co-metabolic interactions were linked to genomic data from previously published Sakinaw Lake SAGs. Combined results support the hypothesis that candidate divisions OP8 and OP9/JS1 have the potential for syntrophic acetate oxidation (SAO) with hydrogenotrophic Archaea affiliated with *Methanomicrobiales*, an integral syntrophic growth mode in methanogenic bioreactors.

¹ A version of the chapter is published as: Gies, E.A., et al., *Illuminating microbial dark matter in meromictic sakinaw lake*. Applied and environmental microbiology, 2014. **80**(21): p. 6807-18.

3.2 Material and methods

3.2.1 Sampling

Sakinaw Lake water samples were taken at deep basin station S1 (49 40.968 N 124 00.119 W) using a combination of 12L Niskin and 8L GO-FLO bottles on June 6th 2007, October 23rd 2007, May 21st 2008, August 5th 2009, January 5th 2010, January 27th 2011 and May 24th 2011. A total of 66 samples were collected covering the mixolimnion (5 m - 30 m) the transition zone (33 m - 55 m) and the monimolimnion (60 m - 120 m). Conductivity temperature and depth were measured with a Seabird SBE19 CTD-device (Sea-Bird Electronics Inc., Bellevue USA). Samples were kept at 4°C in the dark and subsequently processed for environmental DNA extraction. Additionally, transition metals, H₂S and sulfate SO₄²⁻ concentrations were determined for samples collected on May 24th 2011.

3.2.2 Enumeration of cells by flow cytometry

Enumeration of total cells was performed as described by Zaikova and colleagues (Zaikova et al 2010). Briefly, water samples were fixed in 4% (w/v) formaldehyde and stored at 4°C in the dark for approximately 18 hours prior to processing for flow cytometry. Nucleic acids were stained with SYBR® Green (Invitrogen) and cells were counted with a FACS LSRII flow cytometer equipped with an air cooled argon laser (Becton Dickinson). Cell counts were estimated using a known concentration of 6 µm fluorescent beads (Invitrogen).

3.2.3 Environmental DNA

Approximately 6 hours after sample collection 2L water was filtered through a 0.22 µm Sterivex-GV-filter (Millipore) without pre-filter using a Masterflex L/S 7553–70 peristaltic pump (Cole-

Parmer) for DNA extraction. DNA was extracted from Sterivex filters as described by Zaikova and colleagues (Zaikova et al 2010) and DeLong and colleagues (DeLong et al 2006). The DNA extraction protocol can be viewed as a visualized experiment at <http://www.jove.com/video/1352/> (Wright et al 2009).

3.2.4 PCR amplification of 16S rRNA gene for pyrotag sequencing

Environmental DNA extracts described above were amplified using previously published three domain primers targeting the V6-V8 region of the 16S rRNA gene (Engelbrektson et al 2010): 926F (5'-cct atc ccc tgt gtg cct tgg cag tct cag AAA CTY AAA KGA ATT GRC GG-3') and 1392R (5'-cca tct cat ccc tgc gtg tct ccg act cag-<XXXXXX>-ACG GGC GGT GTG TRC-3'). Primer sequences were modified by the addition of 454 A or B adapter sequences (lower case). In addition, the reverse primer included a 5 bp barcode designated <XXXXXX> for multiplexing of samples during sequencing. Twenty-five microliter PCR reactions were performed in triplicate and pooled to minimize PCR bias. Each reaction contained between 1 and 10 ng of target DNA, 0.5 µl Taq DNA polymerase (Bioshop inc. Canada), 2.5 µL Bioshop 10 x buffer, 1.5 uL 25 mM Bioshop MgCl₂, 2.5 µL 10 mM dNTPs (Agilent Technologies) and 0.5 µL 10 mM of each primer. The thermal cycler protocol started with an initial denaturation at 95°C for 3 min and then 25 cycles of 30s at 95°C, 45s at 55°C, 90s at 72°C and 45s at 55°C. Final extension at 72°C for 10 min. PCR products were purified using the QiaQuick PCR purification kit (Qiagen), eluted in 20 mM Tris pH 8, quantified using the Quant-it Picogreen dsDNA reagent. SSU rRNA amplicons were pooled at 30 ng DNA for each sample. Emulsion PCR and sequencing of the PCR amplicons was performed at the McGill University and Génome Québec Innovation Center on a Roche 454 GS FLX Titanium platform according to the manufacturer's instructions. The

sequences reported in this study have been deposited in the NCBI BioProject database, www.ncbi.nlm.nih.gov/bioproject (BioProject Accession:PRJNA247822, ID: 247822).

3.2.5 Pyrotag sequence analysis

Pyrotag sequences were analyzed using the Quantitative Insights Into Microbial Ecology (QIIME) software package (Caporaso et al 2010). To minimize the removal of false positive singleton OTUs, 901,664 pyrotag sequences generated from 66 samples collected in Sakinaw Lake between 2007 and 2011 were clustered together. Reads with length shorter than 200 bases, ambiguous bases, and homopolymer sequences were removed prior to chimera detection. Chimeras were detected and removed using the chimera slayer provided in the QIIME software package. Sequences were then clustered at 97% identity using uclust with furthest linkage algorithm. Prior to taxonomic assignment singleton OTUs (OTUs represented by one read) were omitted, leaving 23,231 OTUs. To generate an OTU table specific for the May 2011 dataset the `filter_otus_by_sample.py` script was used leaving 181,464 sequences and 12,908 OTUs for downstream analysis. Average number of reads per sample in the Sakinaw Lake May 2011 dataset was 16,323, with exception of the 33 m sample, which had ~50% fewer reads. Representative sequences from each non-singleton OTU were queried against the SILVA database release 111 (Quast et al 2013) and the Greengenes database (DeSantis et al 2006) using BLAST (Altschul et al 1990).

Overall, the comparison between Silva and Greengenes databases revealed similar taxonomic assignments for bacterial OTUs. However, significant differences for archaeal OTU assignments were observed between the two databases. Approximately twice as many reads were assigned to *Methanomicrobia* using Greengenes when compared to Silva. This difference could

be mapped back to a single abundant OTU, approaching 10% of total sequences at 45 m that was assigned to *Methanomicrobiales* using Greengenes and to Halobacteriales using Silva. BLAST-based comparisons indicated that sequences comprising this OTU shared ~90% identity with *Halobacteriales* reference sequences in Silva and 90% identity with *Methanomicrobiales* reference sequences in Greengenes. Moreover, a number of putative archaeal OTUs associated with the redox transition zone and monimolimnion (33 m – 120 m) were identified with “no blast hits” using Greengenes but assigned to *Halobacteriales* using Silva. Because both databases are not well annotated for Halobacteria (Youssef et al 2012), the Sakinaw Lake dataset was mapped onto the full-length Halobacteria SSU rRNA gene database generated by Youssef and colleagues (Youssef et al 2012) using the program CD-HIT (“cd-hit-est-2d”) (Li and Godzik 2006). The dataset could not be clustered with the *Halobacteria* reference sequences at 99%, 97% or 95% identity. Even for 90% sequence identity, only 981 reads (0.5%) clustered with the *Halobacteria* reference sequences. Given these uncertainties in putative archaeal OTU affiliation, we designated them “unassigned Archaea” until full-length sequences or reference genomes become available to support more in-depth phylogenetic analysis.

3.2.6 Statistical analyses

Microbial community richness was determined using native scripts implemented in the QIIME package. OTU tables were rarefied starting with 10 sequences to a maximum of 6,000 sequences. Ten iterations per sample were calculated with 100 sequences between each step. Hierarchical cluster analysis of microbial community compositional profiles and environmental parameters was done using the software package R (Development Core Team, 2011; <http://www.R-project.org/>) using Manhattan Distance measures for clustering microbial community

compositional profiles (Mc Cune 2002) and Euclidean distance for environmental parameters (Guler et al 2002). Prior to analysis, pyrotag datasets were normalized to the total number of reads per sample, and environmental parameter data were transformed to the same order of magnitude so that each variable had equal weight.

Multilevel indicator species analysis (ISA) was performed to identify OTUs specifically associated with different water column compartments (mixolimnion = 5 m, 20 m, 30 m; upper part of the transition zone = 33 m, 36 m, 40 m, 45 m; lower part of the transition zone = 50 m, 55 m and monimolimnion = 60 m, 80 m, 120 m), Legendre and colleagues established a method to determine an indicator species by its relatedness to a user-defined environment, e.g. group of samples, where each species is treated individually and its indicator value is determined according to its abundance value (Dufrene and Legendre 1997). De Caceres and colleagues recently extended the algorithm for ISA to include a multilevel pattern analysis (De Caceres et al 2010). Arguing that species with higher adaptability for different environmental factors are indicative for specific environment combinations, this analysis additionally determines indicator species for combinations of environments. The ISA/multilevel pattern analysis calculates p values with Monte Carlo simulations and returns indicator values (IV) and p values with $\alpha \leq 0.05$. The IVs lay between zero and one, where one is considered a true indicator.

3.2.7 Co-occurrence network

To generate a robust network emphasizing co-occurrences between prevalent OTUs in water column compartments rather than individual depth intervals, the Spearman's rank correlation was used. Spearman correlation coefficients were calculated using a custom perl script, "correlation_network.pl" (https://github.com/hallamlab/utilities/tree/master/correlation_network).

The initial dataset consisted of 12,900 OTUs. To simplify the network, we retained OTUs with at least 10 reads appearing in at least three samples leaving 1,528 OTUs with Spearman correlations equal or greater than 0.99. The resulting co-occurrence network contained 130,101 edges, each with positive correlation. The network was visualized with a force directed layout, using Cytoscape 2.8.3 (Shannon, Markiel et al. 2003). Network properties were calculated with the “Network Analysis” plug-in. Nodes in the co-occurrence network corresponded to individual OTUs and edges were defined by computed correlations between corresponding OTU pairs. The layout revealed four distinct modules, which persisted after lowering the correlation coefficient cut-off for edge creation to 0.90 reinforcing the robustness of the network.

3.2.8 Pathway reconstruction

To evaluate the metabolic potential for syntrophic acetate oxidation (SAO) among highly connected candidate divisions, publicly available single-cell genomes from Sakinaw Lake affiliated with OP9/JS1, OP8 and WWE1 were searched for coverage of the Wood-Ljungdahl pathway using the Joint Genome Institute integrated microbial genomes expert review portal (IMG-ER) (<http://img.jgi.doe.gov/>).

3.3 Results

3.3.1 Site location and physicochemical properties

To determine the availability of TEA for microbial energy metabolism in different water column compartments O_2 , Fe, Mn, As and SO_4^{2-} measurements were plotted as a function of depth and compared to measurements of S_A , H_2S , and total bacterial cell counts (Figure 3.1). The mixolimnion in Sakinaw Lake (5 m - 30 m) was constituted of entirely fresh, oxygen-rich water.

In the transition zone (30 m – 55 m), salinity and the concentration of alternative TEAs including, Fe, Mn and SO_4^{2-} gradually increased whereas O_2 concentrations decreased. The oxic-anoxic interface was located at 33 m. Methane concentrations, which were determined for samples collected in June 2007 increased below 33 m and reached saturation at 45 m (data not shown) forming a sulfate methane transition zone (SMTZ). Peak concentrations of Mn, Fe and SO_4^{2-} were observed in the μM range (5.7 μM , 4.2 μM and 81 μM respectively) at 36 m and corresponded with increased microbial cell abundance ($\sim 2.8 \times 10^5$ cells/ mL). Sulfide concentrations in the lower part of the redox transition zone and the monimolimnion (50 m - 120 m) were high (4.5 mM) and resulted in the removal of soluble Fe from the water column by precipitation into iron sulfide (FeS) and FeS_2 (Perry 1990). The peak concentration of As was in the nM range (93.4 nM) at 60 m. The peak concentration of As was in the nM range (93.4 nM) at 60 m. Sakinaw lake revealed unique features with regards to the physicochemistry compared to other geographically isolated meromictic lakes (Table 3.1). The combination of high H_2S and high CH_4 has to our knowledge not been documented before.

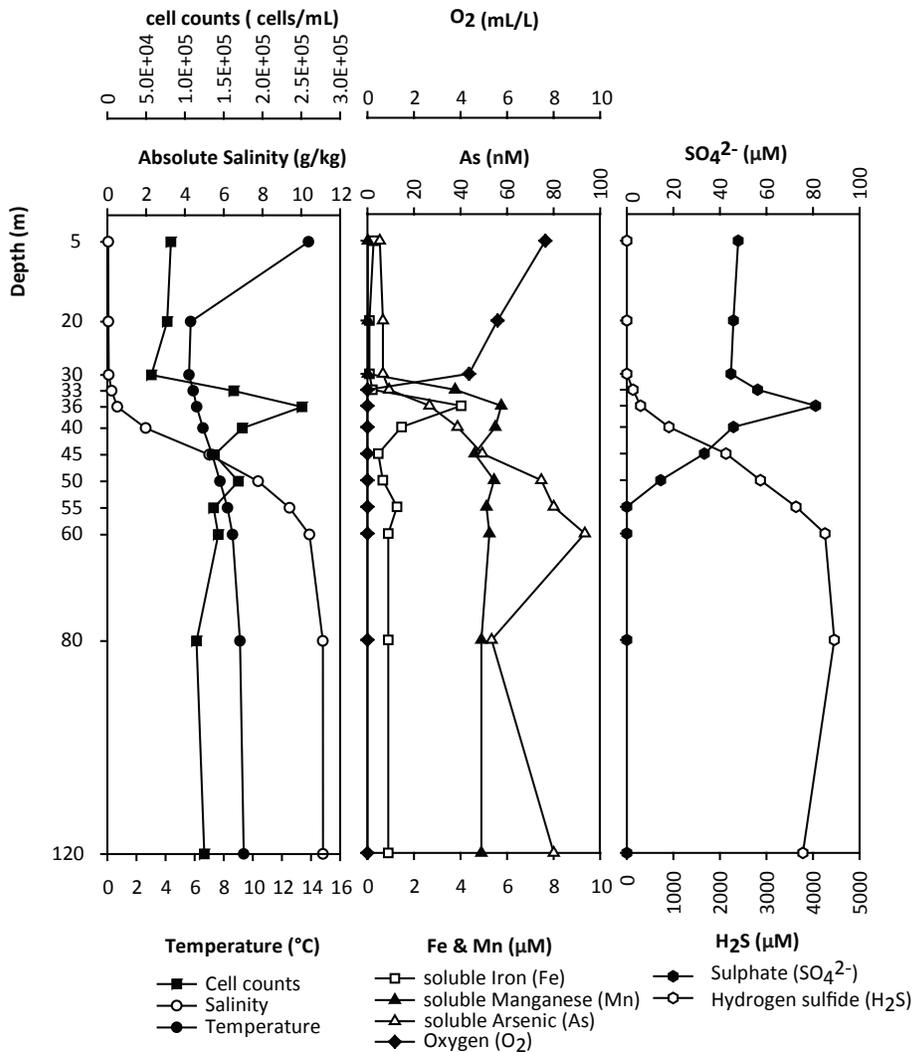


Figure 3.1 Environmental parameters in Sakinaw Lake May 24th 2011.

The salinity gradient revealed water column compartmentalization into mixolimnion (5 m - 30 m), redox transition zone (33 m – 55 m) and monimolimnion (60 m - 120 m). Seasonal temperature influences resulted in a temperature gradient within the mixolimnion whereas temperatures in transition zone and monimolimnion remained unchanged. The peak in microbial cell counts appeared at 36 m where oxygen was depleted and maximum concentrations of alternative terminal electron acceptors Iron (Fe), Manganese (Mn) and sulphate (SO₄²⁻) were measured. Maximum concentrations of Arsenic (As) were located at 60 m. Sulfide was absent in the mixolimnion and increased to mM concentrations in transition zone and monimolimnion.

Table 3.1 Baseline physicochemical characteristics of geographically isolated meromictic lakes.

Lake	Lake type	Geographical location	Depth of chemocline (total depth)	Max H ₂ S μM/L (depth)	Max. SO ₄ ²⁻ μM/L (depth)	Max soluble Fe μM/L (depth)	Max soluble Mn μM/L (depth)	Max CH ₄ μM/L (depth)	Max cell counts cells mL ⁻¹	Reference
Sakinaw Lake	ex-fjord	Sunshine Coast British Columbia, Canada	33 m (140 m)	4500 (120 m)	81.81 (36 m)	4.02 (36 m)	5.75 (36 m)	3000 (below 40 m)	2.8 * 10 ⁵	Vagle, S., et al., 2010.
Lake Cadagno	alpine lake	Canton of Ticino, Switzerland	12 m (20 m)	205 (18 m)	1560 (16 m)	1.6 (20 m)	NA	44 (7.6 cm above sediment)	10 ⁷	Peduzzi, S., et al. 2003. Wagener, S., et al, 1990.
Lake Pavin	crater lake	Auvergne, Central France	60 m (90 m)	3.7 (85 m)	16.9 (60 m)	1211 (90 m)	25.7 (85 m)	6200 (90 m)	1.2 * 10 ⁷	Lehours, et al. 2005 Bura-Nakic, E., et al. 2009
Lake Mahoney	saline lake	Okanagan Valley British Columbia, Canada	5 m (15 m)	30000–35000 (15 m)	400000–500000 (15 m)	NA	NA	NA	10 ⁸	Klepac-Ceraj, V., et al., 2012.
Lake A	ex-fjord	High Arctic, Canada	13 m (60 m)	NA	NA	35.8 (30 m)	176.5 (10-29 m)	NA	2.52 * 10 ⁷	Van Hove, P., et al., 2006.

3.3.2 Relationship between community structure and physicochemical characteristics

To better understand how the physicochemical properties of the Sakinaw Lake water column influence microbial community diversity we performed 454 pyrosequencing of the SSU rRNA gene with three-domain resolution. Although OTUs were generated from a time series we focused on datasets obtained from the May 24th 2011 sampling campaign given the availability of extensive physicochemical information spanning defined water column redox gradients. More detailed analysis of this dataset can be found in Chapters 5 and 6 focused on MDM in the RTZ and archaeal MDM respectively. With this approach, the number of non-singleton OTUs available for downstream analysis increased when compared to clustering of the May 2011 dataset alone. Indeed a number of singletons in May 2011 were found in clusters containing >200 reads in the time series consistent with the presence of conditionally rare taxa in the Sakinaw Lake water column (Shade et al 2014).

Approximately 75% from the May 2011 dataset were assigned to Bacteria and 25% of these were assigned to candidate divisions. Collectively, bacterial OTUs contained the majority of pyrotag reads in all but the 45 m sample where almost 60% of the reads were assigned to Archaea (Figure 3.2 A). Eukaryotes, although prevalent in oxygenated surface waters vanished almost completely below the mixolimnion. Richness estimates based on count data indicated that the mixolimnion had fewer OTUs than samples from the transition zone and monimolimnion (Figure 3.2 B). This pattern was consistent for all samples collected between 2007 and 2013.

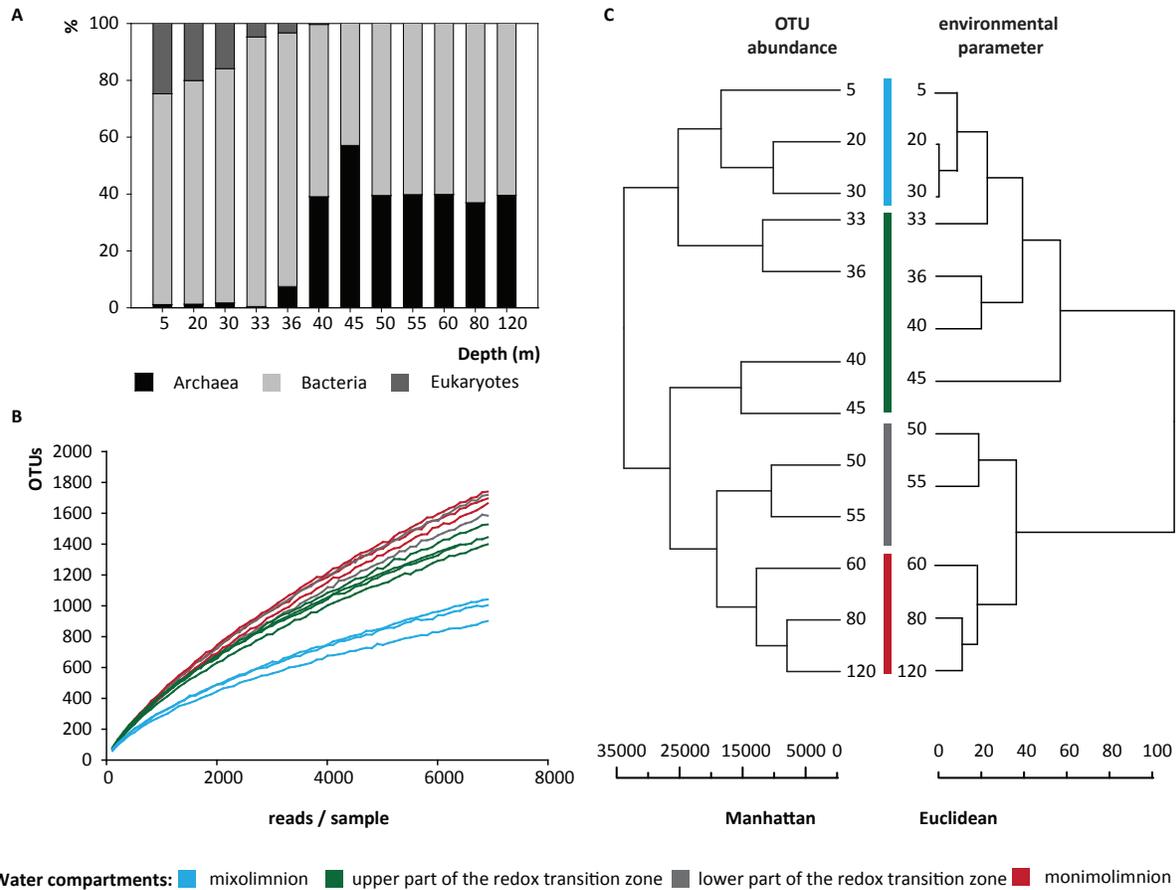


Figure 3.2 Relationship between community structure and physicochemical characteristics.

(A) Bacterial OTUs contained the majority of pyrotag reads in all but the 45 m sample where almost 60% of the reads were assigned to Archaea. Eukaryotes vanished almost completely below the mixolimnion (B) Richness estimates based on count data indicated that the mixolimnion had fewer OTUs than samples from the transition zone and monimolimnion. (C) Hierarchical clustering of the OTU abundance and environmental parameter data ($^{\circ}\text{C}$, S_{A} , TEAs, and H_2S).

Hierarchical clustering of the microbial community composition profiles and selected environmental parameter data ($^{\circ}\text{C}$, S_{A} , TEAs, and H_2S) reflected the stratified nature of the lake ecosystem (Figure 3.2 C). Resulting dendrograms revealed identical clustering patterns between OTU abundance and environmental parameters for almost all depth intervals. Both dendrograms

suggest that water column partitioning in Sakinaw Lake did not easily conform to a three-compartment meromixis model i.e., mixolimnion, redox transition zone and monimolimnion. This was not unexpected, as the division of the water column into three compartments is based solely on hydrodynamic conditions and does not reflect partitioning of the water column into distinct environmental niches. Based on these observations we operationally divided the transition zone into an upper and lower part to more accurately reflect potential niche partitioning.

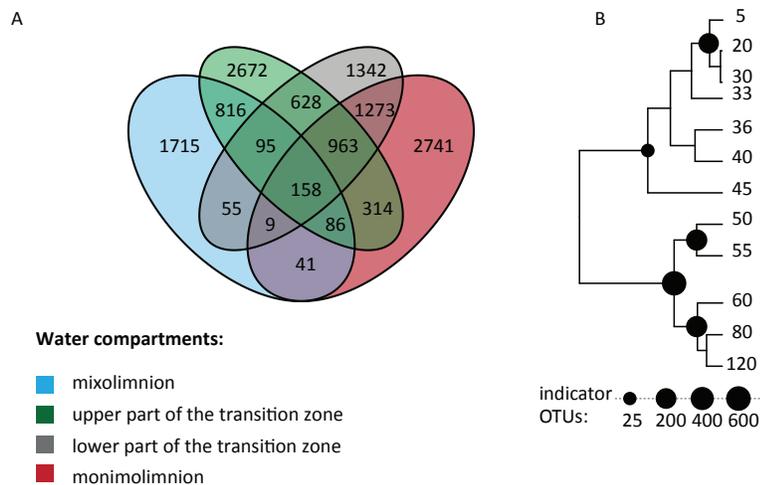


Figure 3.3 Partitioning of OTUs between water column compartments.

(A) The Venn Diagram revealed mutually exclusive distribution patterns for the majority of operational taxonomic units (OTUs). (B) Number of indicator OTUs in different water column compartments.

To identify how OTUs partitioned between the mixolimnion, upper part of the transition zone, lower part of the transition zone and monimolimnion, we conducted a four-way set difference analysis of OTU presence (Figure 3.3 A). Approximately 1% of the OTUs were shared between all four water column compartments, with the majority of OTUs exhibiting mutually exclusive distribution patterns (13% unique to mixolimnion, 21% unique to upper

transition zone 10% unique to lower transition zone and 21% unique to monimolimnion). Operational taxonomic units associated with a given water column compartment that are absent or rare in other compartments are potential ecological indicators. To identify potential indicator OTUs we performed a multilevel indicator species analysis. With this analysis we were able to determine indicator species for individual water column compartments and compartment combinations. Only 5-10% of the compartment specific OTUs were identified as indicator species (Figure 3.3 B & 3.4).

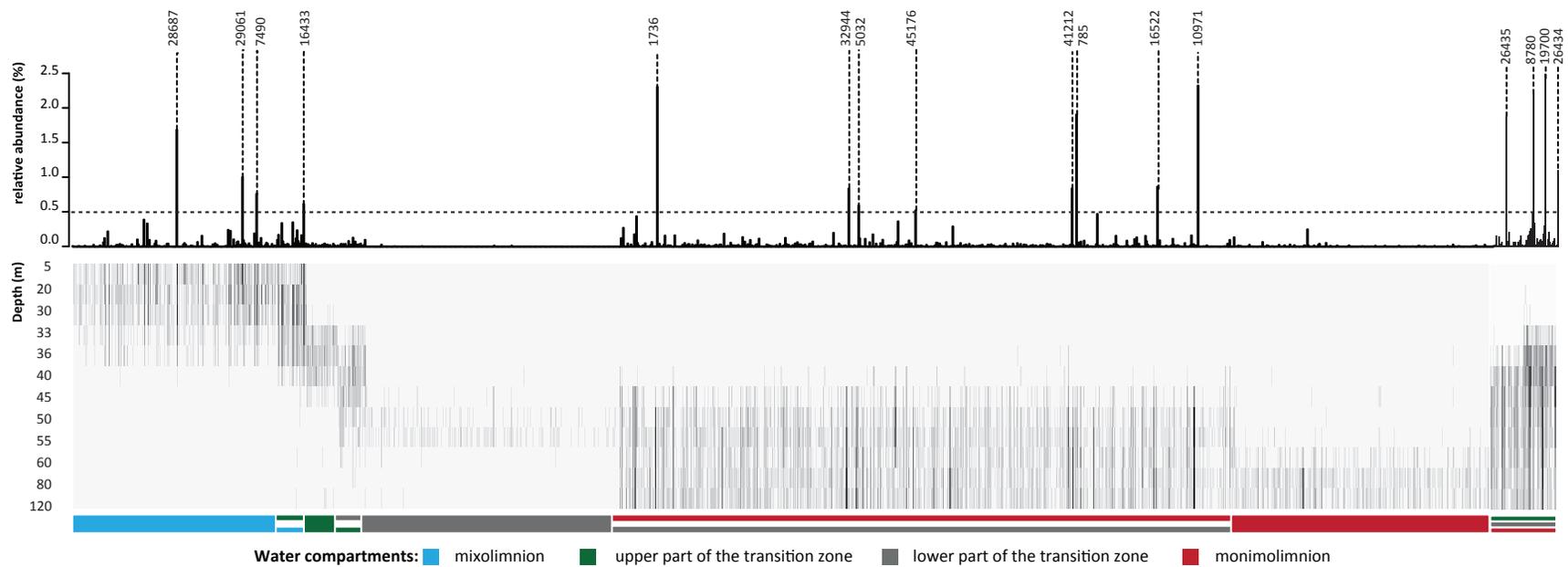


Figure 3.4 Multilevel indicator species analysis.

Abundant and low abundant indicator species (IS) mark the partitioning of the microbial community between water column compartments. Taxonomic affiliation of abundant IS is listed in Table 3.2.

Table 3.2 Operational taxonomic units with highest abundance and their function as indicator species.

Taxonomy	Indicator/ Multilevel Indicator	SAG available	Water column compartment of highest abundance
Bacteria;Proteobacteria;Alphaproteobacteria;SAR11 clade;LD12 freshwater group	MI		MI
Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacteriales;Syntrophaceae;Smithella	MI & UTZ		MI
Eukaryota;Opisthokonta;Metazoa;Arthropoda;Crustacea;Maxillopoda	MI		MI
Eukaryota;SAR;Alveolata;Ciliophora;Intramacronucleata;Spirotrichea;Hypotrichia;Halteria	MI		MI
Bacteria;Actinobacteria;Actinobacteria;Frankiales;Sporichthyaceae;hgcl clade	-		MI
Bacteria;Actinobacteria;Actinobacteria;Frankiales;Sporichthyaceae;hgcl clade	-		MI
Archaea;Euryarchaeota;Methanomicrobia;Methanomicrobiales;Methanoregulaceae;Methanoregula	-		UTZ
Archaea;unassigned	UTZ & LTZ & MO		UTZ
Archaea;unassigned	UTZ & LTZ & MO		UTZ
Archaea;unassigned	-		LTZ
Bacteria;Chloroflexi;GIF9	UTZ & LTZ & MO		LTZ
Bacteria;Chloroflexi;GIF9	LTZ & MO		LTZ
Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosaetaceae;Methanosaeta	-		LTZ
Bacteria;Bacteroidetes;Sphingobacteriia;Sphingobacteriales;WCHB1-69	-		LTZ
Archaea;Euryarchaeota;Methanomicrobia;Methanomicrobiales;Methanospirillaceae;Methanospirillum;uncultured Methanospirillaceae archaeon	LTZ & MO		MO
Archaea;unassigned	LTZ & MO		MO
Archaea;Euryarchaeota;Methanomicrobia;Methanosarcinales;Methanosaetaceae;Methanosaeta	LTZ & MO		MO
Bacteria;Candidate division OD1	LTZ & MO	YES	MO
Bacteria;Candidate division OP9	UTZ & LTZ & MO		MO
Bacteria;Chloroflexi;GIF9	LTZ & MO		MO
Bacteria;Candidate division OP8	LTZ & MO	YES	MO
Bacteria;Candidate division WWE1	LTZ & MO	YES	MO

MI = Mixolimnion

UTZ = Upper part of the transition zone

LTZ = Lower part of the Transition Zone

MO = Monimolimnion

Indeed, the majority of OTUs that were identified as compartment specific in the four-way set difference analysis exhibited fine scale depth partitioning (Figure 3.5). Indicators shared between compartments were also uncommon with the exception of the lower part of the transition zone and monimolimnion. In these two compartments ~44% of the OTUs were identified as indicators. Overall, the majority of OTU indicators were affiliated with unassigned Archaea (25.63%) and candidate divisions (21.3%), followed by *Chloroflexi* (16.2%).

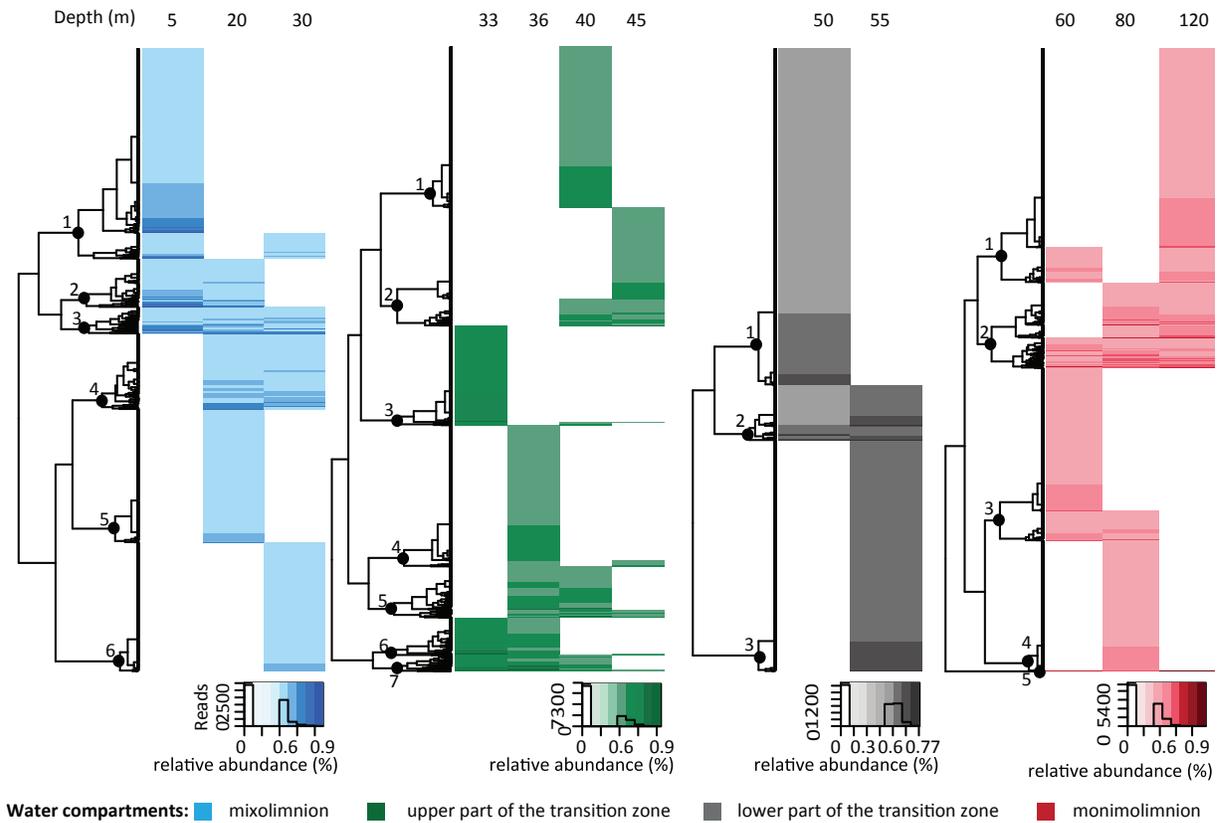


Figure 3.5 Fine scale partitioning of OTUs.

Operational taxonomic units that were identified as water compartment specific using the four-way set difference analysis in 3 (A) exhibited a fine scale depth partitioning as revealed by a cluster analysis of the associated OTUs.

3.3.3 Taxonomic composition

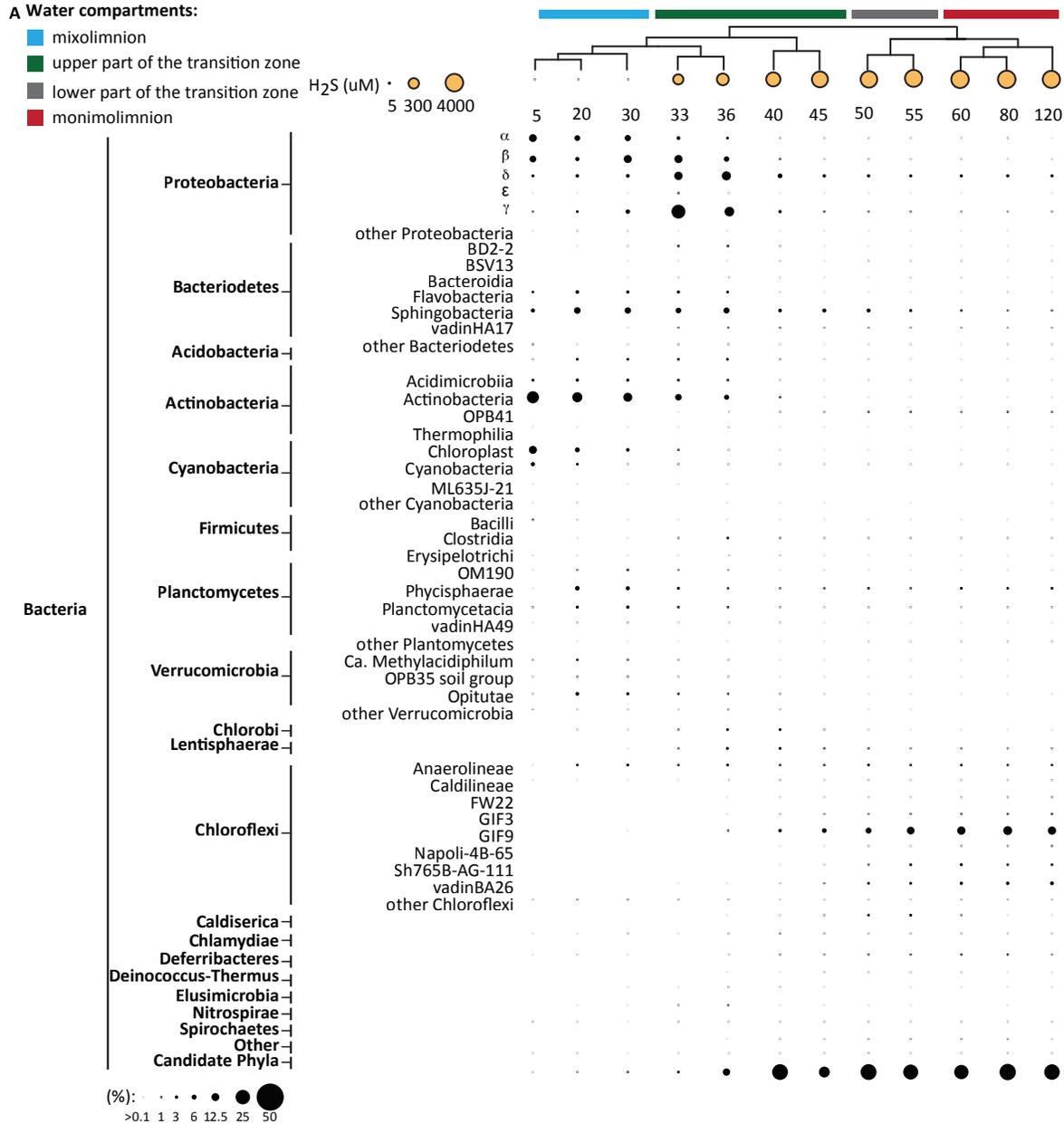
To determine how the taxonomic composition of the microbial community was distributed within and between water column compartments we plotted bacterial and archaeal OTUs based on relative abundance (Figure 3.6). All taxa exhibiting intermediate abundance ($> 0.1\%$) or above are represented in the figures. However, for visualization purposes only taxa $> 1\%$ are shown to scale.

The most abundant Bacteria in the mixolimnion were *Alphaproteobacteria* (11.8%) the majority of which were affiliating *SAR11* (9.7%), *Betaproteobacteria* (10%) mostly affiliated with *Burkholderiales* (8.1%), *Bacteroidetes* (13.2%), *Actinobacteria* (22.3%), *Cyanobacteria* (16.9%), *Planctomycetes* (8.3%) and *Verrucomicrobia* (1.1%) (Figure 4A). The most abundant Eukaryotes in the mixolimnion were *Metazoa* (11.8%) and *Alveolata* (11.4%). Archaea were almost absent from the mixolimnion, although *Methanomicrobia* were identified among members of the rare biosphere ($< 0.1\%$) (Figure 4B). Depth-specific trends within the mixolimnion were observed for several taxa, including *Cyanobacteria*, decreasing rapidly below 5 m, and *Planctomycetes* and *Verrucomicrobia* increasing below 5 m. Overall, the bacterial composition in the mixolimnion of Sakinaw Lake was similar to other freshwater and meromictic lake ecosystems (Comeau et al 2012, Gregersen et al 2009, Pasche et al 2011).

Proteobacteria dominated the upper part of the transition zone, including *Betaproteobacteria* (12.5%) mostly affiliated with *Burkholderiales* (7.7%), *Deltaproteobacteria* (12.9%) mostly affiliated with *Desulfurobacterales* (4%) and *Syntrophobacterales* (9.6%), and *Gammaproteobacteria* (23.8%) mostly affiliated with the *Methylococcales* (22.4%). In addition to proteobacterial groups, *Bacteroidetes* (13.2%), *Actinobacteria* (22.3%), *Cyanobacteria* (16.9%), *Plantomycetes* (8.3%), *Verrucomicrobia* (1.1%), *Chlorobiales* (1.8%), and *Chloroflexi*

(8.5%) were also abundant in the upper part of the transition zone. Interestingly, between 33 m and 40 m, a ten-fold increase in candidate divisions including OP3 (4%), OP8 (4%), OP11 (4%), OP9/JS1 (8%), RF3 (5.6%) and WWE1 (7%) was observed (Figure 4A & 4B). Eukaryotes were virtually absent in all depth below the mixolimnion. The upper part of the transition zone was further marked by a distinct increase in archaeal OTU abundance. Several of these OTUs were affiliated with the methanogenic *Methanomicrobiales* (7%) and *Methanosarcinales* (3%), ammonia-oxidizing *Thaumarchaeota* (1.4%), and anaerobic methane-oxidizing (ANME) archaea belonged to the rare biosphere (<1%). The remaining archaeal OTUs (44.5%) shared similar identity to *Halobacteriales* and *Methanomicrobiales* and could not be assigned to a specific phylum. Overall, the microbial community structure in the upper part of the transition zone was similar to other methane-rich meromictic lake ecosystems (Comeau et al 2012, Lehours et al 2005). However, the diversity and abundance of bacterial candidate divisions and the relative proportion of unassigned Archaea is unprecedented. Indeed, 25 of the currently estimated 30 bacterial candidate divisions were recovered from the upper part of the transition zone. The taxonomic composition of the lower part of the transition zone and the monimolimnion was similar to the 40 m and the 45 m sample of the upper part of the transition zone. The most abundant Bacteria were affiliated with candidate divisions OP3 (1.5%), OP8 (3.6%), OP9/JS1 (6.1%), OP11 (4.1%), WWE1 (6.1%) and *Chloroflexi* (22.3%). Depth-specific trends were observed for the *Chloroflexi*, which increased in abundance between 45 m and 120 m. This trend has also been reported for *Chloroflexi* in meromictic Lake A and Lake Pavin (Comeau et al 2012, Lehours et al 2007). Similar to the upper part of the transition zone, OTUs affiliated with methanogenic *Methanomicrobiales* (5.2%) and *Methanosarcinales* (4.5%), and unassigned

Archaea (30.5%) were also recovered. Moreover, ANME-1 abundance approached 1% at 50 m and remained present throughout the monimolimnion.



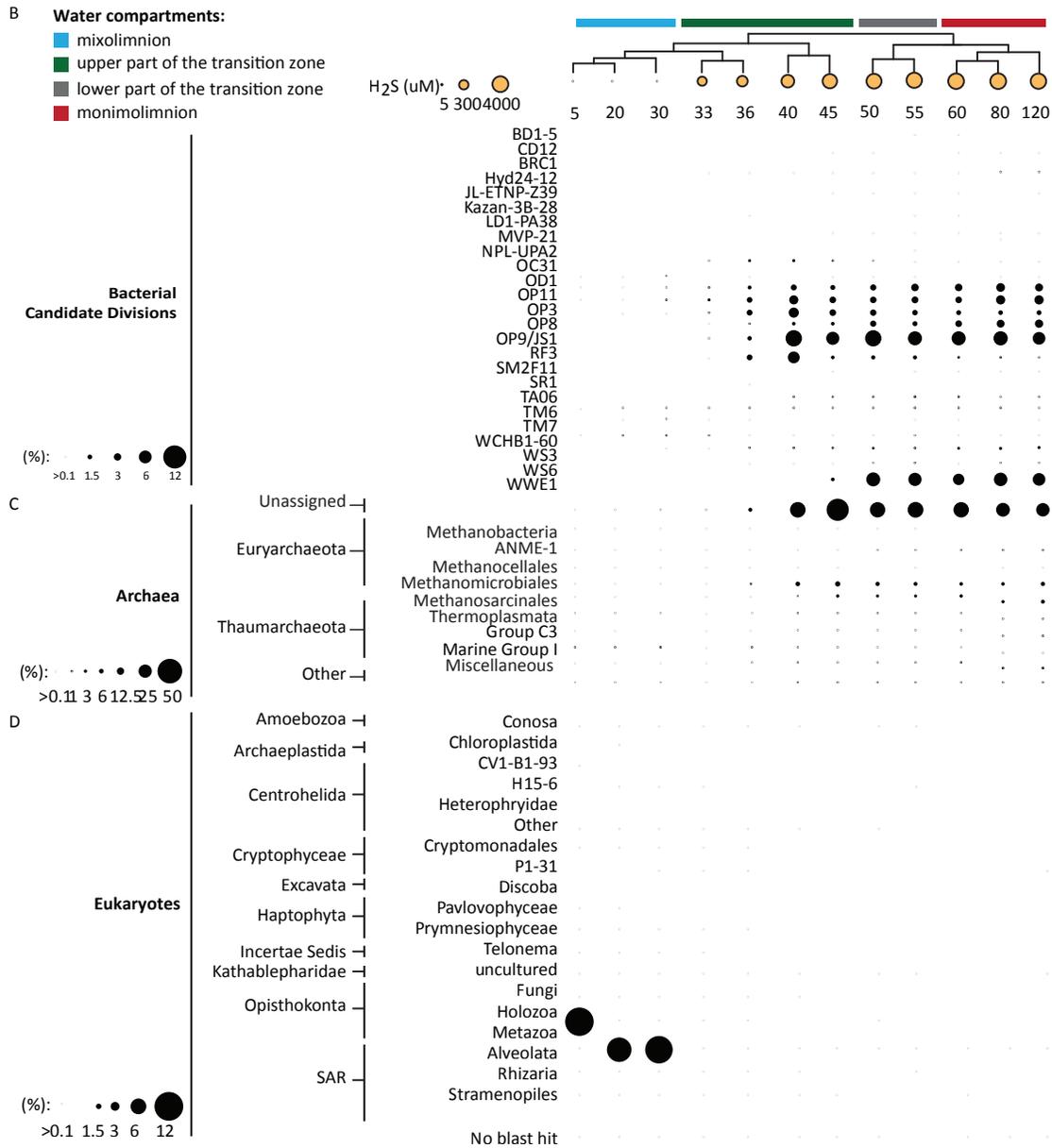


Figure 3.6 Relative abundances of Bacteria, Archaea and Eukaryotes.

All taxa exhibiting intermediate abundance (> 0.1%) are represented but only taxa > 1% are shown to scale. Sulfide concentrations are represented as dots where the radius correlates to the sulfide concentration. (A) Bacteria exhibited water compartment specific distribution patterns. Microbial Dark Matter was predominately recovered from the sulfide rich part of the water column. (B) A total of 25 bacterial candidate divisions were identified. (C) Archaea increased in abundances in the sulfidic part of the water column. 80% of the Archaea could not be assigned to a taxonomy (D) Eukaryotes were only present in the mixolimnion.

3.3.4 Co-occurrence analysis

To identify putative interactions between microbial community members in the Sakinaw Lake water column, we constructed a co-occurrence network. The resulting network contained 130,101 positively correlated co-occurrences (edges) between 1,528 OTUs (nodes) and was composed of four modules corresponding to previously defined water column compartments. Twenty-two OTUs in the network contained >1000 reads and collectively these OTUs represented 40% of total reads in the network (Figure 3.7). Three of these OTUs were indicators for the mixolimnion, one was a multilevel indicator for the mixolimnion and upper part of the transition zone, 8 were multilevel indicators for the lower part of the transition zone and monimolimnion and 4 were multilevel indicators for upper part of the transition zone, lower part of the transition zone and the monimolimnion (Figure 3.4 & Table 3.2). Eight of the multilevel indicators revealed highest abundance in the monimolimnion and displayed a unique correlation pattern that was not observed for other abundant OTUs in the network.

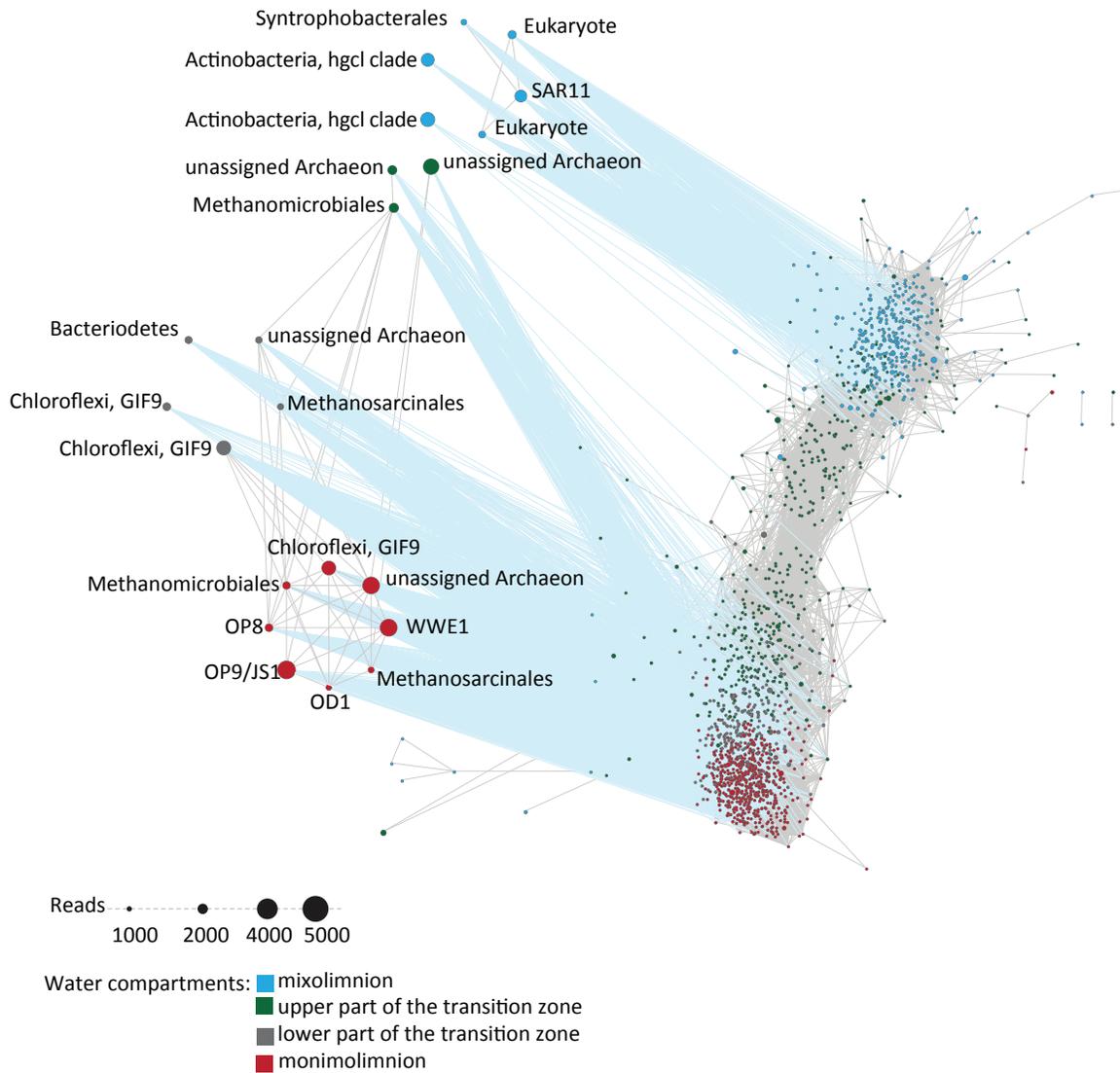


Figure 3.7 Co-occurrence network analysis.

Co-occurrence network depicting OTUs present in ≥ 3 samples with a minimum of 10 reads using Spearman's rank correlations at a correlation co-efficient of 0.99. To better resolve potential interdependencies between abundant candidate division indicator OTUs and other abundant OTUs (≥ 1000 reads), associated nodes were presented outside of the main network, without altering edge sets. Edges that connected these 22 nodes to nodes in the main network were colored blue for visual purposes.

Closer inspection of these OTUs revealed linkages between OP8, OP9/JS1, OD1, WWE1, *Chloroflexi*, *Methanomicrobiales*, *Methanosarcinales* and one unassigned archaeal OTU (Figure 3.8A). Previous studies have posited a role for OP8, OP9/JS1, OD1, WWE1 and *Chloroflexi* in providing methanogenic substrates, specifically acetate and H₂ (Dodsworth et al 2013, Hug et al 2013, Lykidis et al 2011, Wrighton et al 2012, Wrighton et al 2014).

Linkages between OP8, OP9/JS1 and hydrogenotrophic methanogenic *Methanomicrobiales* in the monimolimnion are consistent with interspecies hydrogen transfer and competition for acetate during syntrophic acetate oxidation (SAO). In support of this observation, no linkages were observed between OP8 and OP9/JS1 or between OP8, OP9/JS1 and acetoclastic methanogens affiliated with *Methanosarcinales*. Conversely, linkages between WWE1, *Methanomicrobiales* and *Methanosarcinales* suggest that this candidate division is unlikely to perform SAO. To strengthen hypotheses drawn from network results we evaluated whether publicly available near complete Sakinaw Lake OP8, OP9/JS1 and WWE1 SAGs encode the Wood-Ljungdahl pathway, which is proposed to run in reverse during SAO (Rinke et al 2013). Candidate divisions OP8 and OP9/JS1 harbored a complete set of genes encoding the carbonyl and methyl-branches of the Wood-Ljungdahl pathway while WWE1 harbored only a subset, including genes encoding 5,10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate cyclohydrolase and CO-dehydrogenase/acetyl-CoA synthase (Figure 3.8B).

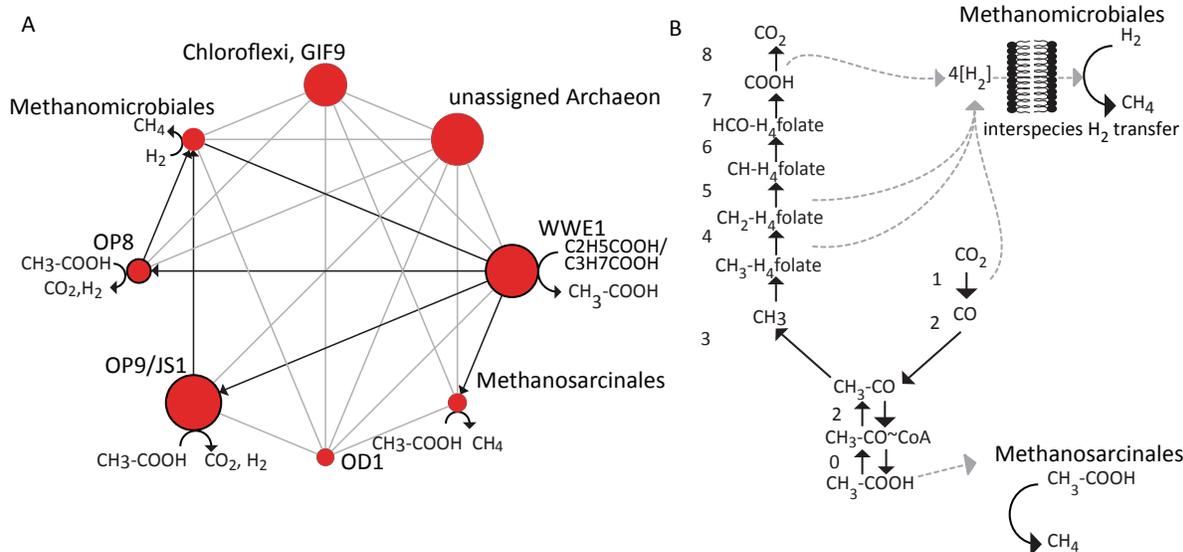


Figure 3.8 Potential for syntrophic acetate oxidation.

(A) Eight of the most prevalent multilevel indicators in the monimolimnion displayed a unique correlation pattern that was not observed for other abundant OTUs in the network. (B) Single amplified genomes affiliated with OP9/JS1 and OP8 encode all components of the Wood-Ljungdahl pathway, graphic representation is adapted from (Muller et al 2013) (1/2: CO-dehydrogenase/acetyl-CoA synthase, 3: trimethylamine:corrinoid methyltransferase, 4: ATP:corrinoid adenosyltransferase, 5: methylenetetrahydrofolate reductase, 6: 5,10-methylene-tetrahydrofolate dehydrogenase/methenyl tetrahydrofolate cyclohydrolase, 7: formyltetrahydrofolate synthetase, 8: formate dehydrogenase). Contrarily WWE1 genomes revealed only genes for the carbonyl branch of the Wood-Ljungdahl pathway. (1/2: CO-dehydrogenase/acetyl-CoA synthase and 3: 5,10-methylene-tetrahydrofolate dehydrogenase/Methenyl tetrahydrofolate cyclohydrolase. Additionally WWE1 encodes 0: ADP-forming acetyl-co~A synthase).

3.4 Discussion

The Sakinaw Lake water column is a highly stratified ecosystem, in which microbial community members partition into distinct sub-populations on the basis of water column redox gradients. This stratification is stabilized by a steep salinity gradient that persists below the oxygen

interface. Richness estimates revealed higher diversity in the redox transition zone and monimolimnion than in the mixolimnion, which is likely supported by a continuous supply of nutrients and substrates as previously discussed for other stratified lakes (Shade et al 2011, Shade et al 2012). Microbial community cluster analysis mirrored patterns observed for physicochemical parameters. This distribution pattern promotes hypotheses related to redox-driven niche partitioning and metabolic coupling in the Sakinaw Lake water column.

The microbial community structure in the oxygen rich, sunlit and entirely fresh mixolimnion (5 m – 30 m), is dominated by *Actinobacteria*, *Cyanobacteria* and *Alphaproteobacteria* affiliated with the *SAR11* consistent with other fresh-water ecosystems recently described by Newton and colleagues (Newton et al 2011b). Moreover, eukaryotic OTUs were also abundant in surface waters and strong correlations (Spearman's correlations ≥ 0.99) between *SAR11* and eukaryotic *Opisthokonta* as well as SAR OTUs were identified, indicating potential grazing relationships.

The microaerophilic upper part of the redox transition zone (between 33 and 45 m) provides a habitat for abundant aerobic methane oxidizers affiliated with the *Methylococcales* (22.4% at 33 m). As O_2 concentrations decrease below 33 m Proteobacteria abundance increases, consistent with patterns observed in other meromictic lakes ecosystems (Biderre-Petit et al 2011b, Comeau et al 2012, Pasche et al 2011, Peduzzi et al 2003, Tonolla et al 2004). Between 33 and 36 m, where concentrations of SO_4^{2-} , Fe and Mn are at their highest, OTUs affiliated with putative sulfate-reducing *Deltaproteobacteria* are abundant. Interestingly, candidate division RF3 OTUs increased between 36 and 40 m within the SMTZ. In marine sediments the AOM is associated with similar gradients of H_2S and CH_4 (Knittel and Boetius 2009). AOM is driven by syntrophic interactions between sulfate-reducing bacteria (SRB) and ANME (Hallam et al 2003,

Hallam et al 2004). While no genomic sequence information is currently available for RF3, its niche in the Sakinaw Lake water column suggests a potential role in sulfur or methane cycling. Moreover, OTUs affiliated with ANME-1 were recovered from the rare biosphere in the upper part of the redox transition zone and increased in abundance within the monimolimnion consistent with water column AOM potential.

As the water becomes more sulfidic in the lower part of the redox transition zone between 50 and 55 m and the monimolimnion between 60 and 120 m, the diversity and abundance of MDM, with the potential to mediate co-metabolic or syntrophic interactions, increases (Dodsworth et al 2013, Pelletier et al 2008, Wrighton et al 2012, Wrighton et al 2014). With more than 25 candidate divisions accounting for 40% of SSU rRNA gene sequences and high numbers of unassigned Archaea, MDM enrichment in Sakinaw Lake is unprecedented. Abundant bacterial candidate divisions were affiliated with OD1, OP3, OP8, OP9/JS1 OP11 and WWE1. With the exception of WWE1, all of these candidate divisions have been previously recovered from other meromictic lake ecosystems (Borrel et al 2010, Comeau et al 2012, Klepac-Ceraj et al 2012). Furthermore, genomic potential for OD1, OP3, OP9/JS1, and OP11 has been recently inferred from metagenomic or single cell genomic approaches (Dodsworth et al 2013, Glöckner et al 2010, Kolinko et al 2012, Wrighton et al 2012).

Metagenomic reconstruction of OD1 and OP11 genomes revealed the potential for fermentative metabolism as well as S_n^{2-} reduction to H_2S (Wrighton et al 2012). Karen Perry inspected sulfur speciation in Sakinaw Lake in the early 90's and reported a remarkably high concentration of S_n^{2-} below the oxygen-sulfide interface (Perry 1990). Based on this evidence, OD1 and OP11 could contribute to the high H_2S concentrations in Sakinaw Lake through S_n^{2-} reduction. Candidate divisions OP3 has been proposed to be magnetotactic with the ability for

anaerobic respiration (Glöckner et al 2010, Kolinko et al 2012). The availability of alternative TEAs in Sakinaw Lake is restricted to the RTZ, suggesting that OP3 in the monimolimnion could encode an alternative energy metabolism to previously studied OP3 genomes. Genomic potential for OP9/JS1 recovered from hot spring sediments revealed a saccharolytic, fermentative lifestyle with the potential for cellulose degradation and hydrogen production (Dodsworth et al 2013). In contrast, enrichment cultures from acetate amended sulfate-rich, anoxic marine sediments revealed ^{13}C -acetate uptake by OP9/JS1 (Webster et al 2006). Based on these observations Webster and colleagues suggested that OP9/JS1 could be acetate-oxidizers using SO_4^{2-} as a terminal electron acceptor. In Sakinaw Lake, this would only be possible in the upper part of the redox transition zone where SO_4^{2-} is available suggesting a fermentative lifestyle below the transition zone. In support of this conclusion sulfate-reducing genes such as *dsr* could not be recovered from recently published OP9/JS1 SAGs from Sakinaw Lake (Rinke et al 2013)

In addition to bacterial candidate divisions, *Chloroflexi*, and methanogenic Archaea affiliated with *Methanomicrobiales* and *Methanosarcinales* are abundant in the lower part of the redox transition zone and monimolimnion. In many aquatic ecosystems, sediments are considered the main source of CH_4 production. However, the potential role of water column methanogenesis in Sakinaw Lake should not be underestimated and has been previously reported for other meromictic lakes (Kallistova et al 2006, Winfrey and Zeikus 1979). Under methanogenic conditions, microbial communities commonly consist of primary and secondary fermenting bacteria, which degrade polymeric substrates into hydrogen (H_2), carbon dioxide (CO_2) and organic acids including acetate, formate, propionate and butyrate. These substrates in turn are used by hydrogenotrophic and acetoclastic methanogens to convert CO_2 and acetate into methane respectively (Sieber et al 2012). Energy yields during the degradation of organic matter

into CH₄ are low and syntrophic interactions between community members are needed to make the process energetically more favorable (Schink 1997). A common strategy used by well-studied syntrophs affiliated with the *Syntrophobacterales* and hydrogenotrophic methanogens is to overcome energy constraints by interspecies hydrogen transfer (Plugge et al 2011, Stams and Plugge 2009). A similar interaction is established during syntrophic acetate oxidation (SAO), where acetogenic bacteria are proposed to run the Wood-Ljungdal cycle in reverse while transferring four H₂ molecules to methanogens (Muller et al 2013).

Consistent with syntrophic growth modes associated with water column methanogenesis, *Syntrophobacterales* and *Methanomicrobiales* were abundant in the lower part of the transition zone and monimolimnion of Sakinaw Lake. Moreover, several MDM including WWE1, OP9/JS1, OP8 and OD1 manifested statistically significant co-occurrence (Spearman's correlations ≥ 0.99) patterns among themselves and between known methanogens and putative fermentative *Chloroflexi* (Hug et al 2013, Wrighton et al 2012, Wrighton et al 2014). These co-occurrence patterns support hypotheses for syntrophic interactions that drive interspecies hydrogen transfer between OP8, OP9/JS1 and *Methanomicrobiales* as well as competition for acetate between OP8, OP9/JS1 and *Methanosarcinales*. Metabolic reconstruction focused on the Wood-Ljungdahl pathway using recently published Sakinaw Lake OP8 and OP9/JS1 SAG sequences revealed complete pathway coverage in both candidate divisions with the potential to mediate SAO. In contrast, 5,10-methylene-tetrahydrofolate dehydrogenase/ Methenyl tetrahydrofolate cyclohydrolase and CO-dehydrogenase/acetyl-CoA synthase were the only Wood-Ljungdahl pathway components identified in WWE1. Given that WWE1 but not OP8, OP9/JS1 were linked to acetoclastic methanogens in the network we suggest that this candidate divisions has the potential to produce acetate as a metabolic end-product from butyrate oxidation

as recently proposed for WWE1 in a terephthalate degrading methanogenic bioreactor (Lykidis et al 2011). Syntrophic acetate oxidation has been speculated to be an important syntrophic pathway in methanogenic bioreactors (Karakashev et al 2006, Muller et al 2013). The Sakinaw Lake water column thereby provides an interesting convergence of natural and engineered ecosystems with potential application to design and operation of anaerobic bioreactors built for bioremediation and energy generation.

3.5 Conclusion

In conclusion, Sakinaw Lake is a natural laboratory in which to study population structure and metabolic potential of MDM. By combining co-occurrence networks with SAG-enabled metabolic reconstruction we posit that SAO in the monimolimnion is linked to active methane cycling. Network analysis has previously been combined with biogeochemical measurements to predict physiology and function of candidate divisions (Peura et al 2012). However to our knowledge this study is the first to reconstruct syntrophic relationships based on convergent co-occurrence patterns and single-cell genomic evidence. Process oriented studies combined with phylogenetic staining and secondary ion mass spectrometry (NanoSIMS) are now needed to validate predicted syntrophic interactions and provide quantitative insights into methane cycling within the Sakinaw Lake water column.

Chapter 4: Phylogeny, population structure and putative function of MDM in the redox transition zone of meromictic Sakinaw Lake

4.1 Synopsis

Recent metagenomic and single-cell genomic sequencing studies have identified putative fermentative and syntrophic growth modes in several candidate divisions (Dodsworth et al 2013, Pelletier et al 2008, Wrighton et al 2012, Wrighton et al 2014). Here I combine phylogenetic and environmental sequence analysis of 88 samples, traversing water column compartments in meromictic Sakinaw Lake over a period of five years, to explore energy metabolism of multiple candidate divisions at the population and community levels. The specific focus was on the metabolic potential of MDM within the RTZ, to identify MDM with the potential for anaerobic respiration. Candidate division RF3 and subpopulations of OP3, OP9, OP11 and OD1 were identified as distinctly associated with the RTZ, and phylogenetic analysis of full-length 16S rRNA genes confirmed the existence to RTZ specific clades for the OD1 and OP3. Inspection of environmental sequence information suggested the potential for respiratory nitrate reduction in the RTZ associated OP3 population. Moreover, while not detected in 16S rRNA gene surveys, environmental sequencing revealed high abundance of open reading frames (ORFs) affiliated with candidate division EM19, including the potential for polysulfide reduction. In addition, numerous ORFs encoding Fe-S oxidoreductases were recovered. Many of these proteins were affiliated with candidate divisions manifesting differential abundance patterns, including OP9/JS1, OP8 and WS3. The prevalence of Fe-S oxidoreductases, together with the physicochemical properties of the Sakinaw Lake water column indicate a potential for sulfur oxidation coupled to ferric iron reduction, a metabolic pathway that might have been important in early Earth history.

4.2 Material and methods

4.2.1 Water sampling and processing for tag sequencing of the 16S rRNA gene

Sample collection and processing for 16S rRNA gene sequencing has been described in detail in chapter 3. Briefly, Sakinaw Lake samples were taken at deep basin station S1 (49° 40.968' N 124° 00.119' W) with a combination of 12L Niskin and 8L go-flow bottles on June 6th 2007, October 23rd 2007, May 21st 2008, August 5th 2009, January 5th 2010, January 27th 2011 and May 24th 2011. Samples were taken from water compartments mixolimnion (5 m - 30 m), upper part of the transition zone (33 m- 45 m), lower part of the transition zone (50 m and 55 m) and monimolimnion (60 m – 120 m). For environmental DNA extraction, 2L water was filtered through a 0.22 µm sterivex-GV-filter (Millipore) without pre-filter using a Masterflex L/S 7553–70 peristaltic pump (Cole-Parmer). Environmental DNA was extracted as described Zaikova and colleagues (Zaikova et al 2010).

Amplification of environmental DNA was done using the following previously published three Domain primers targeting the V6-V8 region of the 16S rRNA gene: 926F (5'-cct atc ccc tgt gtg cct tgg cag tct cag AAA CTY AAA KGA ATT GRC GG-3') and 1392R (5'-cca tct cat ccc tgc gtg tct ccg act cag-<XXXXXX>-ACG GGC GGT GTG TRC-3') (Engelbrekton et al 2010). Lower case letters indicate 454 A or B adapter sequences. The reverse primer included a 5 bp barcode designated <XXXXXX> for multiplexing of samples during sequencing. Samples were submitted to the McGill University and Génome Québec Innovation Center and to the Joint Genome Institute, California for emulsion PCR and sequencing on a Roche 454 GS FLX Titanium platform according to manufactures instructions.

A total of 901,664 pyrotag sequences were analyzed using the Quantitative Insights Into Microbial Ecology (QIIME) software package (Caporaso *et al.*, 2010) and sequences were

clustered at 97% followed by the removal of 24,241 singleton operational taxonomic units (OTU) (Gihring et al 2012, Kunin et al 2010, Tedersoo et al 2010). Representative OTUs for each cluster were queried against the SILVA database release 111 (Quast et al 2013) using BLAST (Altschul et al 1990).

4.2.2 PCR amplification of 16S rRNA gene, clone library construction and sequencing

For the generation of bacterial full-length 16S rRNA gene sequences, Sakinaw Lake DNA extracts collected in October 2007 from 25 m, 30 m, 33 m, 40 m, 50 m, and 120 m were PCR amplified for clone library construction. Universal bacterial primers B27F (5'-AGAGTTTGATCCTGGCTCAG-3') and U1492R (5'-GGTTACCTTATGTACGACTT-3') were used under following PCR conditions: 3 min at 94°C followed by 35 cycles of 94°C for 40 s, 55°C for 1.5 min, 72°C for 2 min and a final extension of 10 min at 72°C. Each 50 µl reaction contained 1 µl of DNA, 1 µl each 10 mM forward and reverse primer, 2.5 U Taq (Qiagen, Germantown, MD, USA), 5 µl 10mM deoxynucleotides and 41.5 µl 1X Qiagen PCR Buffer. For quality control, amplicons were visualized on 1% agarose gels in 1X Tris-acetate-EDTA (TAE), and purified using a PCR Purification Kit (MinElute, Qiagen, CA) according manufacturer's instructions. Approximately 4 µl purified product was cloned into pCR4-TOPO vector using TOPO TA cloning kit (Invitrogen, Carlsbad CA) and transformed by chemical transformation into Mach-1-T1R cells, following manufacturer's instructions. Transformed cells were transferred into 96-well plates, which contained 180 µl LB, 50 µg/ml kanamycin and 10% glycerol to grow cells at 37°C overnight prior to storage at -80°C. DNA fingerprint screening was done using Rsa I (Invitrogen, CA) 4-base cutter. For fingerprint analysis inserts were amplified directly from glycerol stocks with M13F (5'-GTAAAACGACGGCCAG) and M13R

(5' - CAGGAAACAGCTATGAC) primers using above-mentioned PCR profile. To ensure adequate cell lysis samples were incubated for 10 min at 94°C prior to the start of the protocol. The reaction mixture consisted of 2 µl 10x React 1 buffer, 5 µl of M13 amplified PCR product, 16 µl sterile water and 10U Rsa I. Digestion reactions incubated for 2 hours at 37°C followed by Rsa I inactivation at 65°C for 10 min. The digested fragments were visualized by running 5 µl of each RSA I reaction for 90 min on a 2% agarose gel at 120 V (20 cm in width) using 1X TAE buffer. Restriction patterns were visually inspected and unique patterns selected for sequencing at the McGill University and Genome Quebec Innovation Centre (Montreal, Quebec, Canada)

Plasmids were sequenced bidirectional using M13F (5'-GTAAAACGACGGCCAG-3') and M13R (5'-CAGGAAACAGCTATGAC-3') primers. Sequence data was collected on an ABI Prism 3100 DNA sequencer (Applied Biosystems Inc, Foster, CA) using Big DyeTM chemistry (PE Biosystems, Foster, CA) according to manufacturer's instructions. Bidirectional sequence reads were assembled using Sequencher v4.8 (Gene Codes Corporation, Ann Arbor, MI, USA) and manually edited for base-calling errors. The resulting data sets were checked for chimeras with the open source application Mallard using default settings (<http://www.bioinformatics-toolkit.org/Mallard/index.html>).

A total of 176 non-chimeric Sakinaw Lake 16S rRNA gene sequences were used for the generation of a phylogenetic tree. Cluster representatives were determined at 99% identity, using the open source software mothur v.1.19.0 (Schloss et al 2009), resulting in the selection of 112 OTU representatives for downstream analysis. Sequences were aligned using online aligning software "Sina" (<http://www.arb-silva.de/aligner/>), and subsequently imported into the ARB software package using the ARB parsimony tool (default settings) ((Ludwig et al 2004), database 115). To allow best possible recovery of bacterial sequences that belong to "Microbial Dark

Matter” recently described by Rinke et al (Rinke et al 2013), 201 16S sequences recovered in Rinke’s study were clustered, aligned and imported using identical settings as described for Sakinaw Lake sequences, leading to the addition of 63 bacterial MDM reference sequences. To allow global comparison of the Sakinaw Lake population, additional 287 representative sequences from freshwater, marine and engineered ecosystems were included as well. A maximum likelihood tree was then generated using PHYML (Guindon et al., 2005). Positioning of the sequences was inferred with GTR+4G+I model of nucleotide evolution. The parameter for gamma distribution, and the transition/transversion ratio were estimated. Assembly of a consensus tree was done from 100 bootstrap replicates.

4.2.3 Genomic shotgun sequencing and metabolic pathway prediction

A total of 16 samples were taken for genomic shotgun sequencing on October 23rd 2007 (36 m, 50 m, 55 m, 60 m, 120 m), August 5th 2009 (36 m, 55 m, and 120 m), January 5th 2010 (36 m, 80 m) and May 24th 2011 (33 m, 36 m, 50 m, 55 m, 60 m, and 120 m). Environmental DNA was extracted as described in chapter 3, and approximately 700 ng DNA per sample were submitted to the McGill University and G enome Qu ebec Innovation Center for sequencing on a Roche 454 GS FLX Titanium platform according to manufactures instructions.

Samples were analyzed using the Metapathways pipeline (v.2.4.1) with standard settings (Konwar et al 2013). Unassembled sequences were loaded into the program and sequences shorter than 200 bp were removed. On average 283,179 sequences per sample passed this quality control step. Open reading frames were predicted using the Prokaryotic Dynamic Programming Gene-finding Algorithm (Prodigal) (Hyatt et al 2010), resulting in an average of 300,707 predicted ORFs per sample. A quality control step removed ORFs shorter than 180 bp (60 amino

acids) leading to an average of 207,747 ORFs for downstream analysis. For annotation, translated ORFs were queried against protein databases using the protein BLAST or optimized LAST algorithm (Altschul et al 1990, Kielbasa et al 2011). Annotation of ORFs was done using the following seven publicly available databases (release date in parentheses): metacyc-v4 (2011-07-16), InnateDB mus (2010-7-16), InnateDB human (2010-07-16), refseq-nr (2014-01-18), COG (2013-12-27), kegg-pep (2011-06-18), seed (2014-01-30). Moreover, a custom database with proteins predicted in 201 recently published single cell genomes (Rinke et al 2013) was also queried. MetaPathways BLAST/LAST searches were run on the WestGrid compute grid grex on the Compute/Comsul Canada Consortium (<https://www.westgrid.ca/>).

4.2.4 Pigment analysis

Cell material from 250 mL water was concentrated on 0.7 µm GF/F Whatman filter and stored at -20°C until further processing. For pigment extraction 4 ml methanol/acetone (2:7, v/v) were added to each filter. 15 min sonication in an ultrasonic bath was followed by over night pigment extraction at 4°C in the dark. Extracts were then purified through filtration (0.22 µm). To dry solvents, samples were exposed to a nitrogen flow. Pigment extracts from *Chl. phaeobacteroides* BS1 were used as standards for *bacteriochlorophyll e BChle* concentrations. For separation of pigment homologs, pigments were dissolved in 135 µl methanol/acetonitrile (1:5, v/v). 15 µl ammonium acetate was added as ion pairing agent, followed by reversed-phase HPLC on a Dionex system equipped with a P580 pump, STH585 column oven, PDA-100 photo diode array detector and a RF2000 online fluorescence detector (Dionex Softron, Sunnyvale, CA). For pigment separation a Spherisorb ODS2 column (3µm, 250mm by 4.6mm) in-line with a precolumn of the same material (CS Chromatographie Service, Langerwehe, Germany) was

used. A linear gradient of a acetonitrile, methanol, 0.01 M ammonium acetate and ethyl acetate mixture built the mobile phase at a flow rate of 0.7 ml min⁻¹ as described previously (Airs et al 2001). Fluorimetric detection was carried out at an excitation wavelength of 476 nm and an emission at 676 nm.

4.2.5 Environmental RNA

2 L water per sample taken in May 2008 and August 2009 were filtered directly after sample collections on 0.22 µm polycarbonate filter and extracted as described previously (Chomczynski and Sacchi 1987). Briefly, cells were lysed by addition of 600 µl lysis buffer (50 mM Na-acetate, 1% SDS, 10 mM EDTA, pH 4.2), 600 µl phenol (pH 4.5-5.5) and 1 g of 0.1-mm-diameter siliconized zirconia beads (BioSpec Products, Bartlesville, USA) followed by disruption in a beadbeater (BioSpec Products, Bartlesville, USA) for 5 min. RNA later solution was concentrated using Amicon Ultra – 0.5 ml 50K centricons (Millipore, Bedford, MA) and 0.7 ml lysis buffer (20 mM Na-acetate pH 5.5, 0.5% SDS, 1 mM EDTA, pH 8.0) were added. Total RNA from filter and liquid fractions was isolated using phenol-chloroform and subsequently pooled and purified with the RNeasy MiniElute Cleanup Kit (Qiagen, Hilden, Germany) according to manufactures protocol. To remove all remaining DNA, RNA was treated with Turbo DNA free (Applied Biosystems, Foster City, CA). RNA concentrations were determined with nanodrop ND-1000 (peqlab, Erlangen, Germany). Quality was assessed on a formaldehyde gel (3.1%, w/v). Reverse transcription was performed with the ImProm II Kit (Promega, Madison, WI) using 4,5 µl of RNA samples (total 16 µl) according to the manufactures protocol. Omitting reverse transcriptase from sample served as negative control. Quantitative real-time PCR (RT-qPCR) were performed using 25 µl reactions with iQ SYBR Green Supermix and 4 µl

of cDNA solution and at optimized PCR conditions in a iCycler (Bio-Rad, Hercules, CA). For standardization, the copy number of 16S rRNA gene transcripts was quantified in parallel for all samples using a plasmid containing the cloned fragment from the 16S rRNA and ITS region of the investigated bacterium. Each quantification was performed in triplicates.

4.2.6 Physiochemical parameter

Conductivity Temperature and Depth (CTD) were measured with a Seabird SBE19 (Sea-Bird Electronics Inc., Bellevue USA). Water samples for H₂S measurements were fixed with 2% final concentration Zink Acetate and analyzed in the lab using the methylene blue method as described by Cline et al (Cline 1968). Oxygen concentrations were determined using the Winkler method (Winkler 1888a). On the May 21st 2008 trip, light levels were measured with a LI-COR LI-192SA Underwater Quantum Sensor with a LI-1000 data logger.

4.2.7 Statistical analyses

Hierarchical cluster analysis, simplex plots, and MJ Andersons test for betadispersion were done with scripts implemented in the software package R (Development Core Team, 2011; <http://www.R-project.org/>) (Mc Cune 2002).

4.3 Results

4.3.1 Stability of water column stratification in Sakinaw Lake

The stability of the water column stratification based on salinity gradient and temperature was discussed in detail in chapter 2. Further it was highlighted that variations in the depth of the H₂S transition zone exist, and that H₂S concentrations in the RTZ and monimolimnion varied up to

four-fold between years. Based on these differences, I looked more closely at microbial community structure and function in the RTZ to better infer potential metabolic interactions among and between MDM.

4.3.2 Temporal stability in microbial community composition

To assess whether the microbial population structure displayed a temporal stable partitioning between mixolimnion, RTZ, and monimolimnion, beta-dispersion was estimated for 66 previously published Sakinaw Lake 454 16S rRNA gene tag sequences using MJ Anderson's permutation analysis (chapter 3 and (Gies et al 2014)) (Figure 4.1). Beta-dispersion is the average dissimilarity from individual observations to their group centroid, and can be used to estimate temporal variability in community composition (Anderson et al 2006). With regards to the physicochemical water column stratification in Sakinaw Lake, it was expected that variations were restricted to the mixolimnion and that the microbial community composition for RTZ and monimolimnion would be relatively homogenous. However, these expectations were not met: the estimated distance to the centroid for the RTZ (0.49) was larger than estimates for mixolimnion (0.39) and monimolimnion (< 0.2) (Figure 4.1 A). Previous observations from the May 2011 dataset revealed a high degree of microbial stratification along the RTZ (chapter 3 and (Gies et al 2014)), indicating that the depth intervals sampled from the RTZ cannot easily be summarized as one ecological niche. Grouping samples from the RTZ could therefore lead to an inflated estimation of temporal variability if the microbial community drastically differs between individual depths. To more accurately reflect temporal variability in the RTZ, estimates of beta-dispersion were determined for individual depths (Figure 4.1 B), revealing that the distance to the centroid for 33 m and 36 m (0.3) was indeed higher than estimated for samples from the

monimolimnion, while samples from 40 m, 45 m, 50 m and 55 m had a more homogenous population structure (distance to the centroid < 0.2).

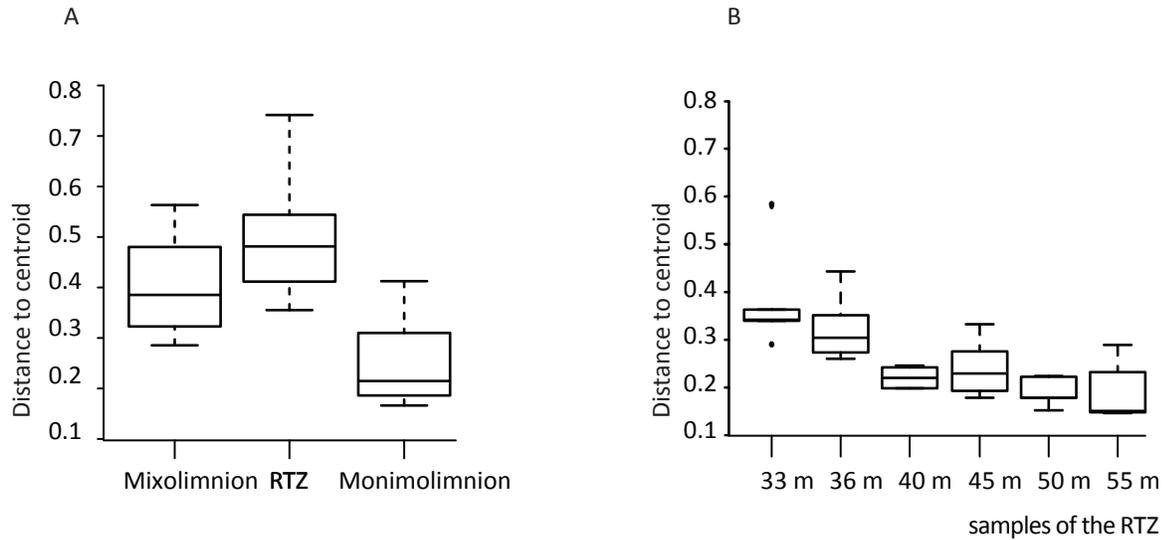


Figure 4.1 Variability in microbial population structure.

Distribution of the distance from the centroid from samples of the mixolimnion, RTZ, and monimolimnion collected between 2007 and 2011. (A) The distance from the centroid was determined using MJ Anderson’s permuted analysis of beta-dispersion representing a statistical measure for temporal variability in communities. (B) Distribution of the distance from the centroid determined from samples of the RTZ only revealed variability at 33 m and 36 m and more homogenous community composition in samples below 36 m.

4.3.3 Partitioning of the microbial community

To outline microbial community partitioning between water column compartments, and to identify MDM distinctly associated with the RTZ, a simplex diagram of average microbial class abundance in mixolimnion, RTZ, and monimolinion was produced (Figure 4.2). The population structure revealed partitioning between water column compartments, consistent with previous observations for the May 2011 dataset (Gies et al 2014). The mixolimnion (0 m – 30 m) was

represented by 12 abundant microbial classes of the bacterial and eukaryotic domains (> 0.5% relative abundance, which corresponds to ~ 1,000 reads). However, none of them belonged to MDM. The community composition was similar to other freshwater lakes recently described by Newton and colleagues (Newton et al 2011a). The most abundant classes were affiliated with *Actinobacteria* (6.9% *Actinobacteria* and 1.2% *Acidimicrobilia*), *Proteobacteria* (*Alpha-* and *Betaproteobacteria*, both 4.5%), *Bacteroidetes* (4.5% *Sphingobacteria* and 0.9% *Flavobacteria*), *Plantomycetes* (1.4% *Plantomycetacia*), *Acidobacteria* (0.9% *Holophagae*), *Cyanobacteria* (2%), SAR (1.5% *Alveolata*) and *Opisthokonta* (0.8% *Metazoa*). Abundant members of the *Gamma-* and *Deltaproteobacteria* (4.0% and 4.9% respectively) were detected at the interface between the RTZ and the mixolimnion. The RTZ (30 m – 60 m) was represented by four abundant bacterial classes affiliated with the *Lentisphaera* (1.2% *Lentisphaeria*), *Chlorobi* (0.7% *Chlorobiales*), candidate division OP3 (3.0%) and candidate division RF3 (1.2%).

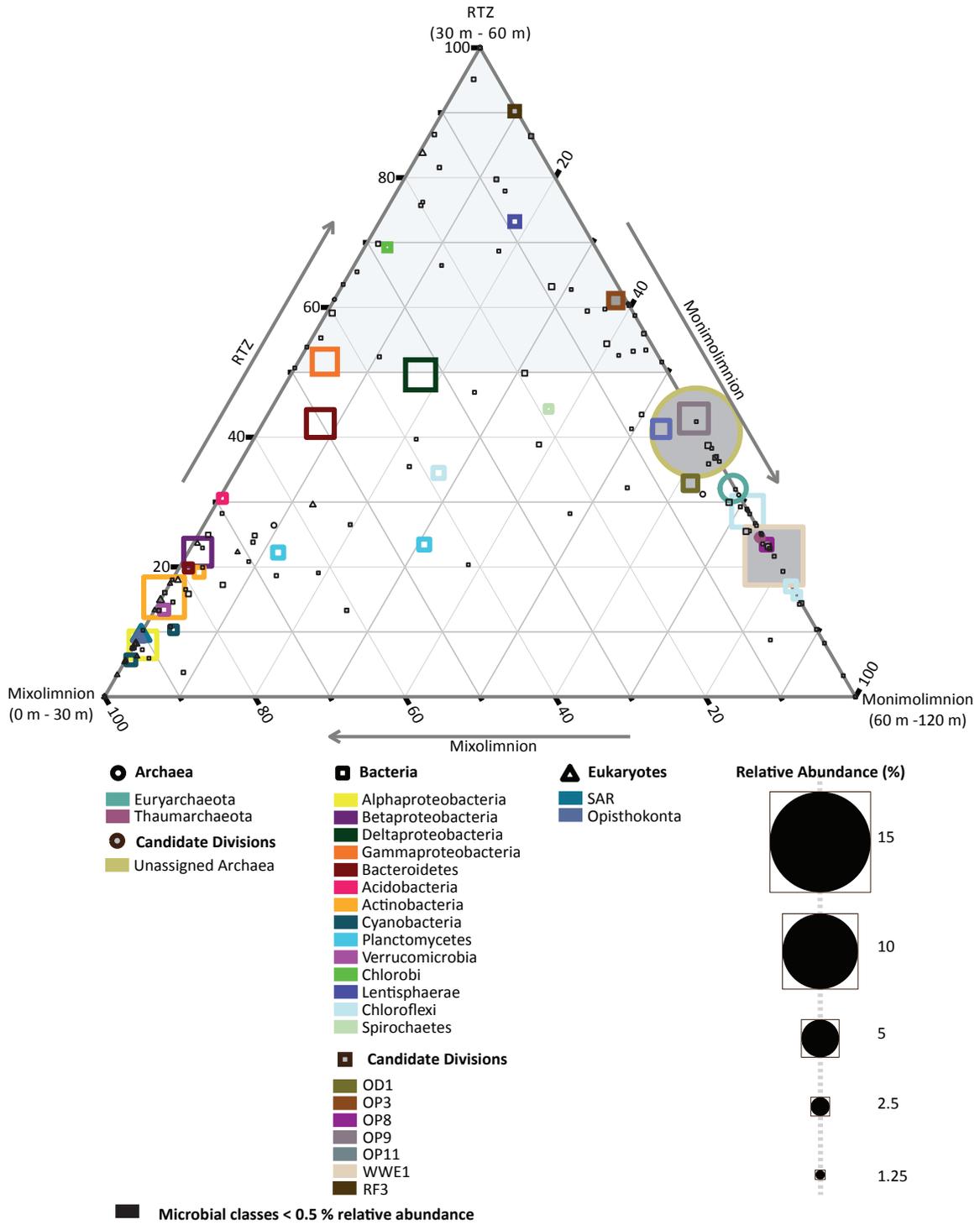


Figure 4.2 The stratified microbial community.

(legend on following page)

Figure 4.2 The stratified microbial community.

Simplex plot showing the distribution of microbial classes in the time-series data set from Sakinaw Lake (2007-2011). However, for visual purposes taxonomic affiliation was color coded on the phylum level. Microbial classes with relative abundance smaller than 0.5% were depicted using dark blue lineout. Shapes of the data points represent the three domains Eukaryotes (triangle), Bacteria (circle) and Archaea (square), while sizes represent relative abundance of the corresponding microbial class in the dataset. Candidate divisions are highlighted by grey shading of the datapoint. Axes represent the three water column compartments (mixolimnion, redox transition zone (RTZ) and monimolimnion) and the proportion of the microbial class associated with the compartment. Grey shading highlights the area of the simplex plot that represents microbial classes distinctly associated with the RTZ.

The monimolimnion (60 m – 120 m) was represented by 12 abundant bacterial and archaeal classes affiliated with the *Chloroflexi* (5.2% GIF9, 1.4% vadin BA26 and 0.8% Sh765B-AG-111), *Spirochaetes* (0.7% *Spirochaetes*), candidate divisions WWE1 (10.1%), OP9 (4.7%), OP8 (1.3%), OP11 (2.5%), and OD1 (2.1%), *Euryarchaeota* (3.7% *Methanobacteria*), *Thaumarchaeota* (0.5% *Miscellaneous Crenarchaeotic Group*), and unassigned Archaea (13.8%). Two abundant classes affiliated with the *Planctomycetes* (1.4% *Phicisphaerae*) and the *Chloroflexi* (1.5% *Anaerolineae*) were present throughout the water column.

4.3.4 OTU partitioning along the water column

Previous observations in Sakinaw Lake indicated that the majority of candidate divisions started to appear in the upper part of the RTZ, and increased in abundance in the lower part of the RTZ and monimolimnion (Gies et al 2014). In the simplex diagram this pattern is represented by the location of corresponding data points towards the middle of the axis shared between RTZ and monimolimnion. Hence, while generally associated with the monimolimnion, candidate divisions

WWE1, OD1, JS1/OP9 and OP11 manifested higher relative abundances in the RTZ than the majority of phyla previously described as RTZ specific (Table 4.1).

Table 4.1 Average abundance of MDM in individual water column compartment

Taxon	mixolimnion	rtz	monimolimnion
BD1-5	0.00	0.02	0.01
BHI80-139	0.00	0.01	0.11
BRC1	0.00	0.06	0.14
Hyd24-12	0.00	0.01	0.01
JL-ETNP-Z39	0.00	0.00	0.00
Kazan-3B-28	0.00	0.02	0.05
LD1-PA38	0.00	0.00	0.00
NPL-UPA2	0.05	0.50	0.08
OC31	0.02	0.01	0.01
OD1	0.35	2.08	3.88
OP3	0.09	3.75	2.30
OP8	0.00	0.96	3.14
OP9	0.01	6.08	8.09
OP11	0.41	3.20	4.14
RF3	0.01	3.18	0.33
SHA-26	0.00	0.01	0.01
SM2F11	0.00	0.02	0.10
SR1	0.00	0.01	0.00
TA06	0.00	0.27	0.21
TM6	0.17	0.29	0.29
TM7	0.08	0.04	0.03
WS3	0.02	0.33	0.96
WS6	0.00	0.09	0.15
WWE1	0.00	2.19	7.91

To identify partitioning of OTUs along the water column that could indicate RTZ specific subpopulations of WWE1, OD1, JS1/OP9 and OP11, affiliated OTUs were extracted from the dataset, hierarchically clustered, and compared to OTU representatives of the RTZ specific phyla OP3, RF3, *Lentisphaera* and *Chlorobi* (Figures 4.3 A and B). Four abundance classes were established to gain insight into relative abundances of depth specific OTU populations, and to establish the presence of dominating – abundant – OTUs (Table 4.2). Because of highest resolution with sample number (n=14), “August 2009” was chosen as a representative dataset.

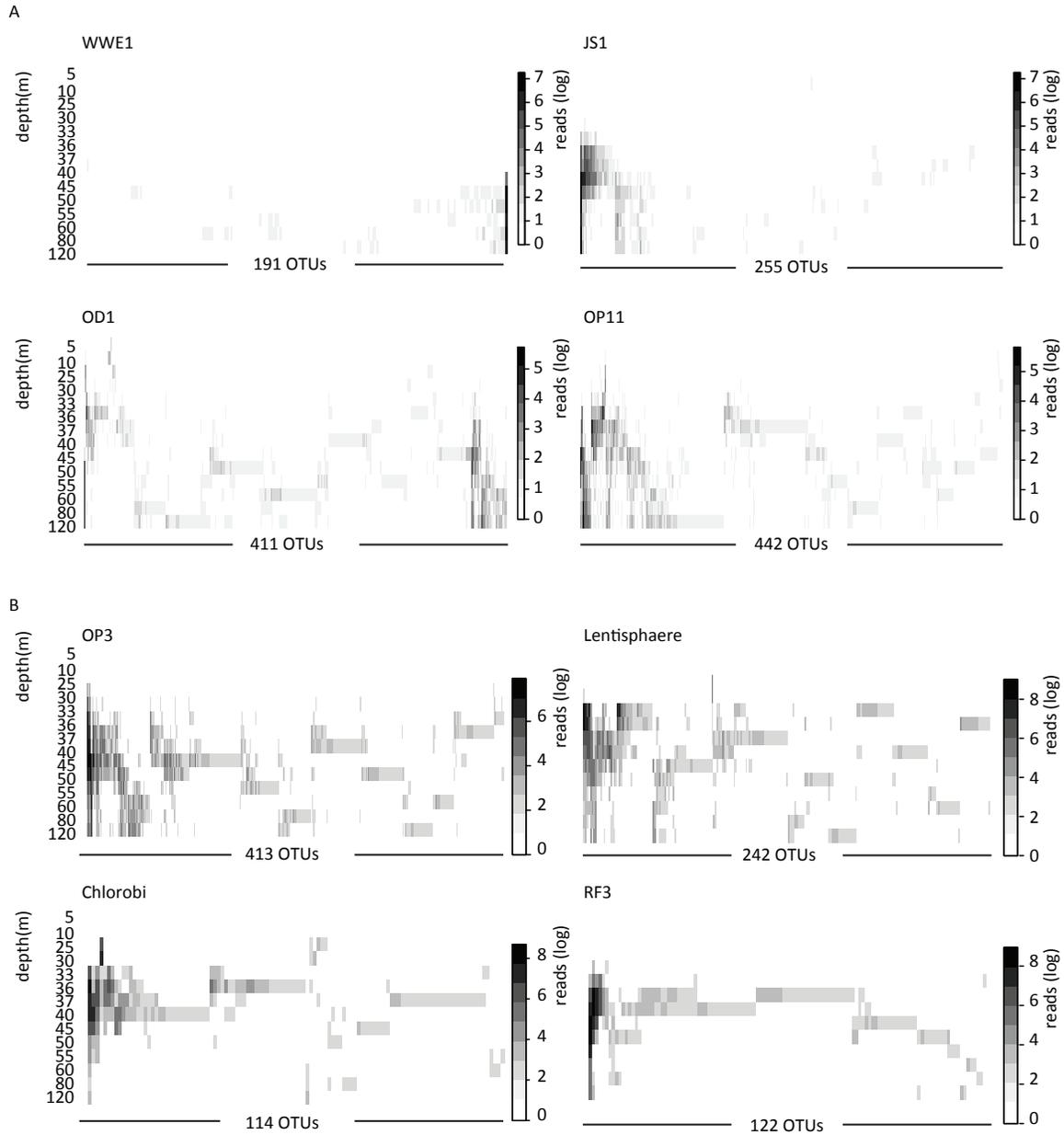


Figure 4.3 Niche partitioning along the water column.

Hierarchical clustering of OTUs affiliated with (A) abundant candidate divisions and (B) phyla that were distinctly associated with the RTZ, revealed niche partitioning within the RTZ for abundant OTUs affiliated with OP11, OD1 and OP9. Analysis was done with OTUs extracted from the August 2009 time point.

Table 4.2 Abundance distribution of abundant MDM and RTZ associated OTUs.

	total OTUs	abundant >1000	intermediate 100 - 1000	low 10 - 100	rare <10	total reads
WWE1	191	1	0	6	184	5,024
JS1	256	1	9	26	220	10,421
OD1	411	1	5	45	358	4,063
OP11	443	0	12	76	355	6,104
OP3	431	1	7	64	358	6,079
Lentisphaere	242	1	7	28	206	4,659
Chlorobi	114	1	3	8	102	2,177
RF3	122	2	2	2	116	4,054

total reads Auguts 2009: 205,121

Inspection of OTU abundance and distribution profiles revealed that candidate division WWE1 did not have a RTZ specific population (Figure 4.3 A and Table 4.2). The entire WWE1 population thrived in water masses of 45 m and below and was dominated by one abundant OTU, while the remaining OTUs were classified as low and rare. Operational taxonomic units affiliated with OP9/JS1 partitioned into two distinct populations, one was present throughout the anoxic-sulfidic part of the water column (33 m and below) and was dominated by one abundant OTU, and the other population was composed of intermediate, low and rare OTUs with peak abundance in the RTZ (at 45 m). Inspection of OTUs affiliated with OD1 revealed three distinct populations, one population thrived in water masses of 45 m and below and was dominated by one abundant OTU, the second population thrived in the anoxic-sulfidic part of the water column and was composed of intermediate, low and rare OTUs, and the third population thrived predominantly in the RTZ with peak abundances at 33 m and was composed of intermediate, low and rare OTUs. Inspection of the OTU distribution pattern of OP11 revealed absence of an abundant OTU, but revealed instead the highest number (12) of intermediate abundant OTUs. Two populations of intermediate, low and rare OTUs partitioned in the anoxic-sulfidic part of the water column; one population was restricted to water masses of 45 m and below, and one

population was restricted to the RTZ with peak abundances at 33 m. The OTU distribution profile for candidate division OP3, which was previously identified as RTZ specific phylum, was with regards to the formation of depth specific OTU populations similar to candidate divisions OP9/JS1, OD1, and OP11 (Figure 4.3 B). One abundant and six intermediate OTUs represented a population of candidate division OP3 with distinct dominance in the RTZ (33 m - 60 m), a second population of intermediate, low and rare OTUs exhibited distinct restriction to the RTZ at 45 m and 50 m, and a third population of low and rare OTUs was restricted to the monimolimnion. The *Lentisphaera* was represented by one population of abundant, intermediate, low and rare OTUs that were restricted to the RTZ, with one abundant OTU dominating at 33 m and 36 m. Only low and rare OTUs prevailed below the RTZ. The OTU partitioning of candidate division RF3 did not compare to any of the previously described candidate divisions. Operational taxonomic units affiliated with the RF3 were almost entirely restricted to the RTZ where they paralleled *Chlorobi* distribution patterns, with only one low and otherwise rare OTUs appearing in the monimolimnion. However, contrasting *Chlorobi*, candidate division RF3 was represented by two abundant OTUs.

4.3.5 Phylogeny of candidate divisions OP3, OP9/JS1, OP11, OD1, and RF3

The division of OP3, OP9/JS1, OP11 and OD1 into compartment specific groups was further investigated based on phylogenetic analysis of full-length 16S rRNA gene sequences generated from six clone libraries covering samples from mixolimnion (25 m, 30 m, 33 m), RTZ (40 m 50 m), and monimolinion (120 m). Phylogenetic analysis revealed high bootstrap support for the division of the OP3 clones into two distinct depth specific clades representing the RTZ and the monimolimnion (Figure 4.4). Reference sequences sourced from the Silva database (v111) with

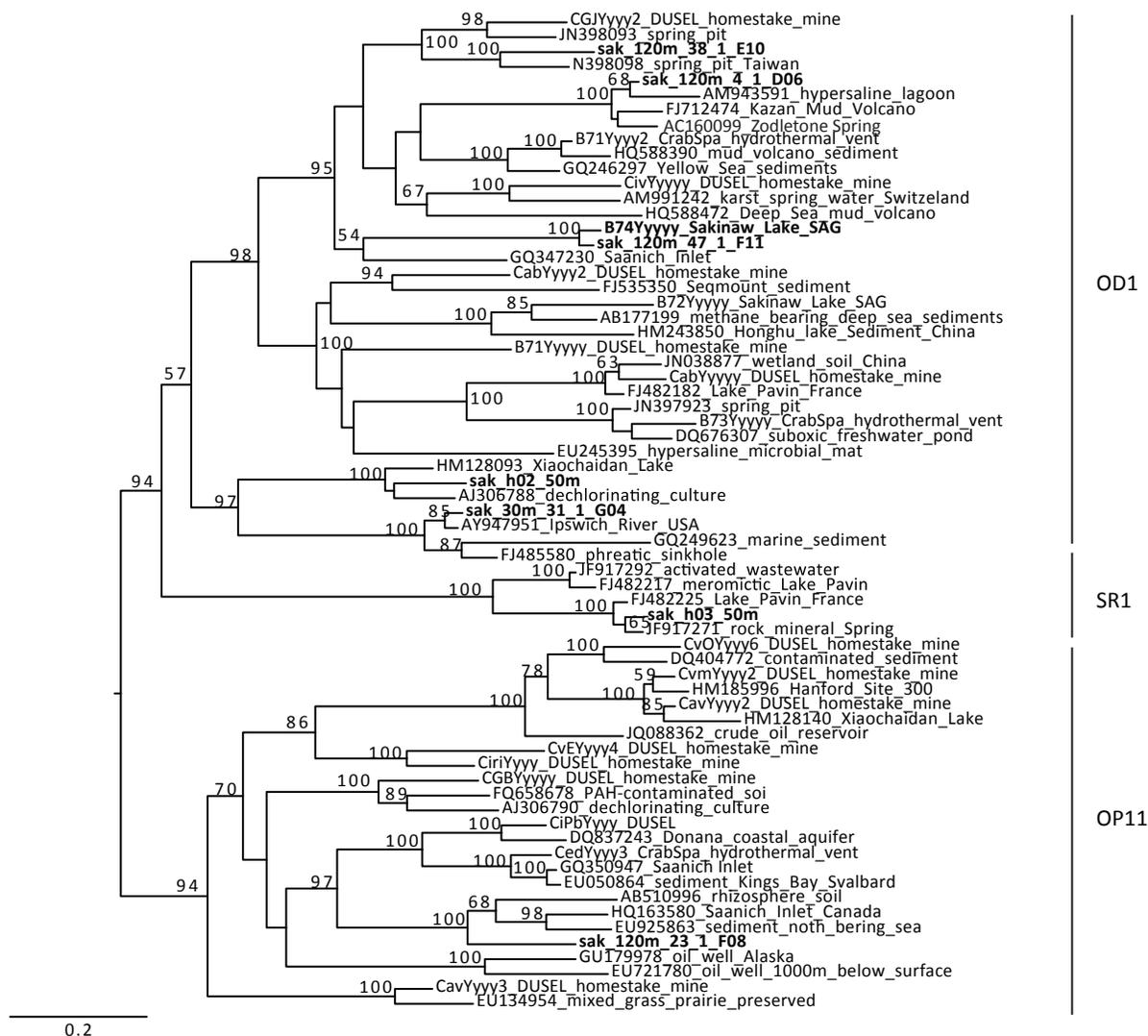


Figure 4.5 Phylogeny of candidate division OD1.

Phylogenetic tree generated with full-length 16S rRNA gene sequences of OD1 and its relatives SR1 and OP11. Sakinaw Lake cluster representatives (clustered at 99% identity) are indicated in bold face. The tree was calculated using maximum likelihood estimation and generalized time reversible (GTR) models implemented in open source software PhyML. The proportion of invariable sites was estimated and assembly of a consensus tree was done from 100 bootstrap replicates.

The partitioning of JS1 related Sakinaw Lake clones into depth specific clades had no bootstrap support. However, phylogenetic analysis revealed a clustering of Sakinaw Lake clones

with JS1 sequences, which distinctly diverged from original OP9 sequences of the Yellowstone hot springs (Figure 4.6).

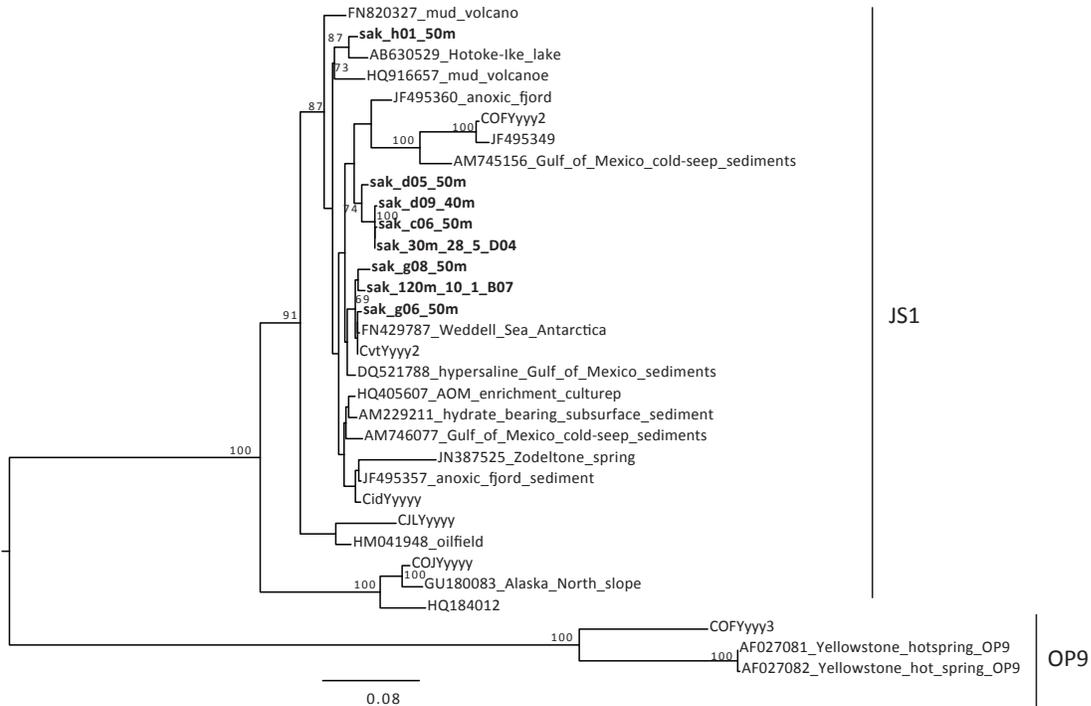


Figure 4.6 Phylogeny of candidate division JS-1.

Phylogenetic tree generated with full-length 16S rRNA gene sequences of OP9/JS-1. Sakinaw Lake cluster representatives (clustered at 99% identity) are indicated in bold face. The tree was calculated using maximum likelihood estimation and generalized time reversible (GTR) models implemented in open source software PhyML. The proportion of invariable sites was estimated and assembly of a consensus tree was done from 100 bootstrap replicates.

Clones affiliated with RF3 were only recovered from the RTZ (40 m), suggesting either absence or low abundance of RF3 in the monimolimnion, which is in agreement with previous observations using 454 sequencing of the 16S rRNA gene. Phylogenetic analysis of 125 bacterial 16S rRNA sequences from Sakinaw Lake and 385 reference sequences sourced from the Silva

database version 111, and MDM reference sequences from Rinke and colleagues resulted in a placement of RF3 within the *Firmicutes* (Appendix B). This affiliation was further confirmed using *Arb* with implementation of Silva database version 115, which eliminated candidate phylum RF3 entirely placing previous RF3 designated sequences into the *Firmicutes*. Within the *Firmicutes*, RF3 clustered distinctly away from phototrophic *Heliobacter* and revealed closest similarity to the *Erysipelotrichia*. Representative RF3 sequences from the Silva database formed two distinct cluster with sequences mainly originating from fecal samples, and a clade including the two Sakinaw Lake clones and environmental sequences originating from oil contaminated soils, dechlorinating cultures, and hypersaline mats (Figure 4.7).

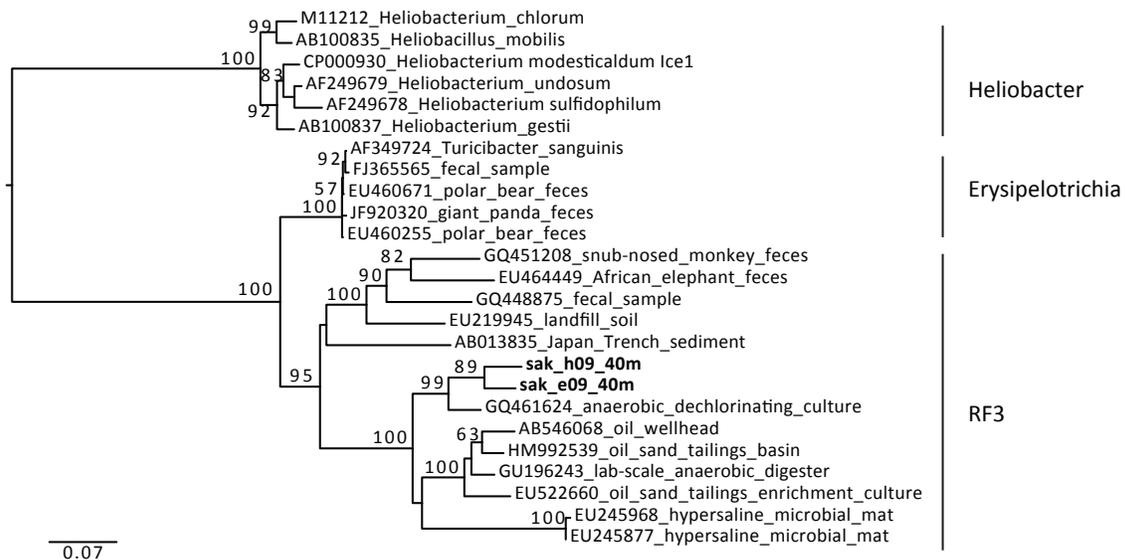


Figure 4.7 Phylogeny of candidate division RF3.

The phylogenetic tree generated with full-length 16S rRNA gene sequences of RF3 was calculated using maximum likelihood estimation and generalized time reversible (GTR) models implemented in open source software PhyML. The proportion of invariable sites was estimated and assembly of a consensus tree was done from 100 bootstrap replicates. Sakinaw Lake cluster representatives (clustered at 99% identity) are indicated in bold face.

4.3.6 Population structure and dynamics of *Chlorobi* and RF3

The previously described similarity between RF3 and *Chlorobi* OTU distribution patterns and the distant phylogenetic relationship between RF3 and phototrophic *Heliobacter*, suggested that RF3 might be a phototrophic organism as well. Phototrophic members of the *Chlorobi* (*Chlorobiales*) are known to be sensitive to seasonal changes of light availability and exposure to oxygen, and spatial shifts and fluctuations in relative abundance have been reported from other lakes (Mizoguchi et al 2011, Rogozin et al 2009, Tonolla et al 2005). To infer putative metabolic dependencies on resources, or sensitivities to environmental stressors between RF3 and *Chlorobiales*, we inspected temporal and spatial variability of the two populations in more detail. Samples were collected for biomass and activity measurements as well as for environmental parameter analysis on meter intervals in the upper part of the RTZ in May 2008 and August 2009 (Figure 4.8). Oxygen was depleted at 33 m in both years and H₂S concentrations were starting to increase below 30 m in August 2009, H₂S measurements for May 2008 failed. Light intensities in May 2008 were with 0.05 and 0.0002 $\mu\text{mol}\cdot\text{s}^{-1}\cdot\text{m}^{-2}$ between 33 m and 40 m sufficient for low light adapted species of the *Chlorobiales* (Manske et al, 2010). Bacteriochlorophyll *e* (Bchle) concentrations were measured in May 2008 and compared to abundance profiles from 454 tag sequences affiliated with *Chlorobi*. Both methods revealed a gradual increase in *Chlorobi* abundance between 33 m and 37 m, followed by a similarly gradual decrease below 37 m. At 37 m, maximum Bchle concentrations approached 140 ng/ml, which is below levels detected in many environmental studies where the sulfide-oxygen transition zone is exposed to higher levels of light (Bchl levels can reach up to 28 mg/L (Overmann et al 1994)), but well above Bchle concentrations in the chemocline of the Black Sea (8 ng/L) where light levels are in a similar range as in Sakinaw Lake (Manske et al 2005, van Gemerden and Mas 1995).

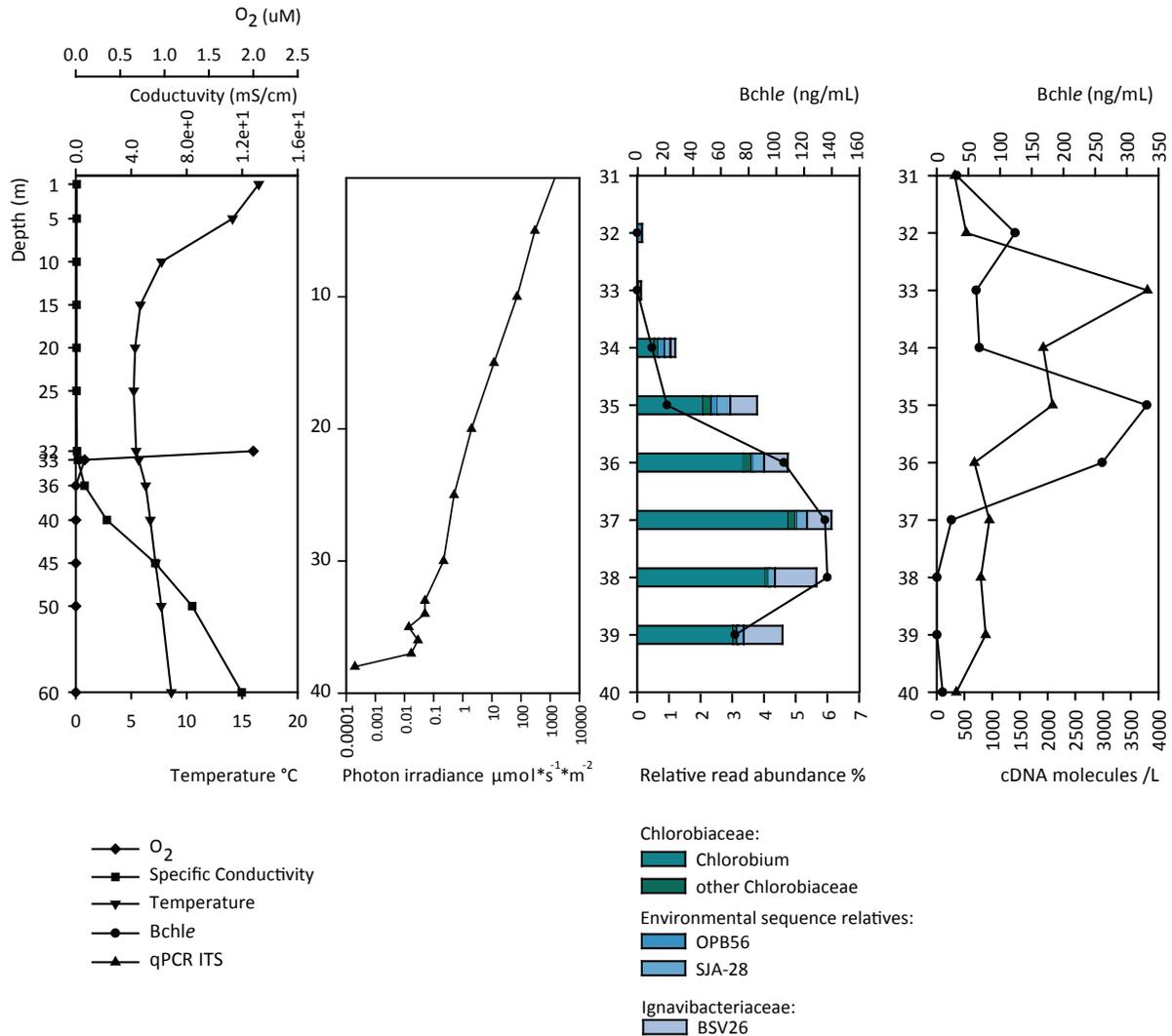


Figure 4.8 Metabolic niche of Chlorobi in Sakinaw Lake.

Oxygen, Conductivity, Temperature and light penetration revealed ideal metabolic conditions for *Chlorobiales* between 30 m and 40 m. Bchle measurements and 454 pyrosequencing confirmed presence of *Chlorobiales* with a maximum relative abundance of 6% at 37 m. Transcripts of the internal spacer region of the rRNA gene were used as estimate for the activity of *Chlorobiales* in August 2009 and detected higher activity in shallower depth.

Chlorobi tag sequence abundance peaked at 37 m reaching 6% of total sequences.

Taxonomic assignments revealed that the majority of these sequences were affiliated with

phototrophic *Chlorobiales* within the family *Chlorobiaceae* and the genus *Chlorobium*. Moreover, OTUs related to environmental sequences OBP56, SJA-28 and non-phototrophic *Ignavibacteriaceae* (BSV-26) were recovered as well.

To better constrain the activity of *Chlorobi* within the RTZ, specific internal transcribed spacer (ITS)-RNA transcripts were quantified at one-meter intervals between 30 m and 40 m in August 2009. Transcript abundance was compared to corresponding *Bchle* concentrations. Notably, the peak *Bchle* concentration in August 2009 was shifted to 35 m and approached 350 ng/ml compared to May 2008. The ITS transcripts were detected in all investigated samples with the highest *Chlorobiales* activity observed at 33 m (two meters above the peak in biomass) which can be ascribed to higher light availability at shallower depths. To infer changes in positioning along the water column and relative abundance of *Chlorobi* and RF3 over time, 454 tag sequences collected in October 2007, August 2009, January 2011 and May 2011 from the interface between mixolimnion and RTZ (30 m) and the RTZ (36 m, 40 m and 50 m) were compared. The proportion of non-phototrophic members of the *Chlorobi* was considered as small and therefore not subtracted from total *Chlorobi* reads. The population profile revealed distinct spatiotemporal changes at 30 m, 36 m and 40 m for *Chlorobi*, RF3 and other RTZ associated phyla (Figure 4.9). However, distribution and abundance patterns of *Chlorobi* and RF3 did not change in unison suggesting at least some degree of metabolic independence.

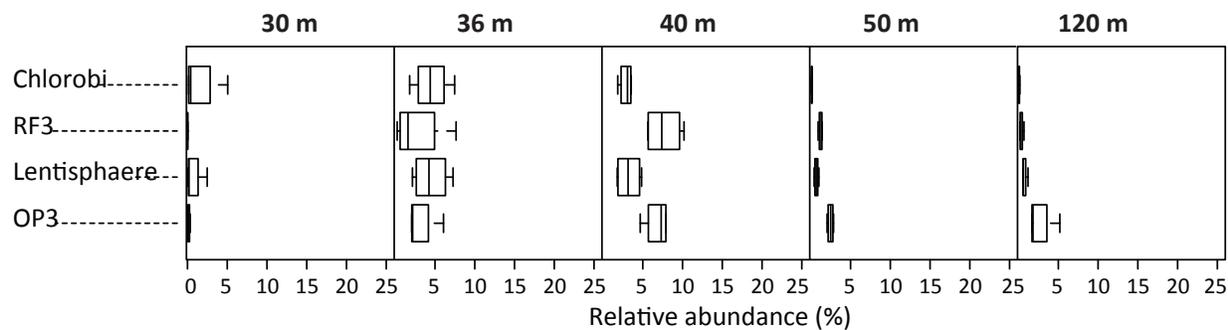


Figure 4.9 Population dynamics in RTZ associated phyla.

Box-whisker plots for relative abundances of *Chlorobi*, *Lentisphaera*, RF3, and OP3 at 30 m, 36 m, 40 m, 50 m and 120 m for samples collected in October 2007, August 2009, January 2011 and May 2011. Relative abundances for all phyla changed between time points. *Lentisphaera* and *Chlorobi* were occasionally also recovered from 30 m, the H₂S-O₂ interface, while candidate divisions RF3 and OP3 did not thrive that high in the water column. Candidate division OP3 was the only phylum that was abundantly recovered from 120 m monimolimnion samples.

4.3.7 Metabolic potential in the RTZ of Sakinaw Lake

To identify metabolic pathways for energy acquisition distinctly associated with the upper part of the RTZ (33 m and 36 m), the lower part of the RTZ (50 m and 55 m) and the monimolimnion (60 m, 80 m and 120 m), genomic shotgun sequences were generated from 16 samples collected between 2007 and 2011. Prediction of ORFs and annotation was done using the Metapathway pipeline (Konwar et al 2013), producing on average 207,747 predicted ORFs and 100,827 protein annotations per sample. Metabolic pathways were inferred using the Metacyc database (Caspi et al 2014), and were only included if coverage was either complete or only one step missing, resulting in 493 pathways used for downstream analysis. To separate predicted pathways into water column compartment-specific groups, and to specifically identify the metabolic potential unique to the upper part of the RTZ, a simplex diagram of biosynthetic, degradation, detoxification, carbon fixation and energy acquisition pathway families was

produced. The upper part of the RTZ revealed unique coverage within all pathway families (Figure 4.8). Abundant energy acquisition pathways were aerobic respiration (PWY-3781), dissimilatory nitrate reduction (DENITRIFICATION-PWY, PWY-6748), NAD(H) pathways related to respiration (PWY0-1334, PWY0-1335, NADPHOS-DEPHOS-PWY), and pathways that process acetyl-CoA to CO₂ and/or NADH (GLYOXYLATE-BYPASS, PWY-5046, PWY-5084). Less abundant energy acquisition pathways included arsenate reduction (PWY-4601), thiosulfate disproportionation (PWY-5277), succinate and proline to cytochrome transfer (PWY0-1353, PWY0-1544, PWY0-1329, dissimilatory nitrate reduction IV (PWY-5674), intra-aerobic nitrate reduction (PWY-6523). The majority of pathways were recovered from all three water column compartments (data points appearing near the center of the plot). The most abundant evenly distributed energy acquisition pathways were glycolysis and related pathways (PWY-5484, PWY66-400, PWY-1042, PWY-6118) followed by hydrogen production (PWY-6758, PWY-6759, PWY-6780), methanogenesis from CO₂ (METHANOGENESIS-PWY), and pyruvate fermentation (PWY-6587, PWY-5480, PWY-5482). Less abundant energy acquisition pathways that were evenly distributed throughout the anoxic-sulfidic part of the water column were hydrogen oxidation (P283-PWY), hydrogen production (PWY-6744), pyruvate fermentation to lactate (PWY-5481), sulfur reduction via polysulfide (PWY-5364), and methanogenesis from tetramethylammonium (PWY-5261). One pyruvate oxidation pathway (PYRUVOX-PWY) was predominantly recovered from the lower part of the RTZ, and two energy acquisition pathways associated with methanogenesis were predominantly recovered from the monimolimnion (methyl-coenzyme M reduction to methane (METHFORM-PWY) and methanogenesis from trimethylamine (PWY-5250)).

To identify potential of anaerobic respiration in candidate divisions, BLAST/LAST outputs from a query against proteins identified in 201 recently published MDM SAGs were searched for ORFs that encode genes of the electron transport chain and anaerobic respiration. The analysis was focused on the May 2011 data set, which included samples covering the upper (33 m and 36 m) and lower (50 m and 55 m) part of the RTZ, and the monimolimnion (60 m, and 120 m). Based on the ORF annotations, candidate divisions OP8, OP9, OP11, OD1, WS3, EM19 and WWE1 belonged to abundant representatives of the MDM population with distinct increase in abundance below 36 m, which was consistent with abundance profiles of previously published 454 16S tag sequencing data generated from the same samples (Gies et al 2014) (Figure 4.6 A).

Open reading frames with BLAST/LAST hits against proteins belonging to candidate division OP3 were recovered from all six samples. Inspection of the predicted proteins revealed respiratory nitrate reductases, and NADH-(ubi)quinone oxidoreductase in 33 m and 36 m samples (Figure 4.9 B). Below 36 m, denitrification pathway components disappeared, while ORFs encoding Fe-S oxidoreductases became more prevalent. Components of the electron transport chain and Fe-S oxidoreductase genes were associated with numerous candidate divisions including BRC1, CD 12, EM19, KSB1, NKB19, OD1, OP1, OP3, OP11, OP8, OP9, SAR406, WS3, WWE1 persisted at low coverage throughout the depth profile. The exception was candidate division EM19, which manifested high coverage of electron transport chain components in the 33 m and 36 m samples, as well as ORFs encoding polysulfide reductase. Additional polysulfide reductase ORFs were associated with candidate divisions BRC1 and SAR406 throughout the anoxic-sulfidic part of the water column albeit at lower coverage.

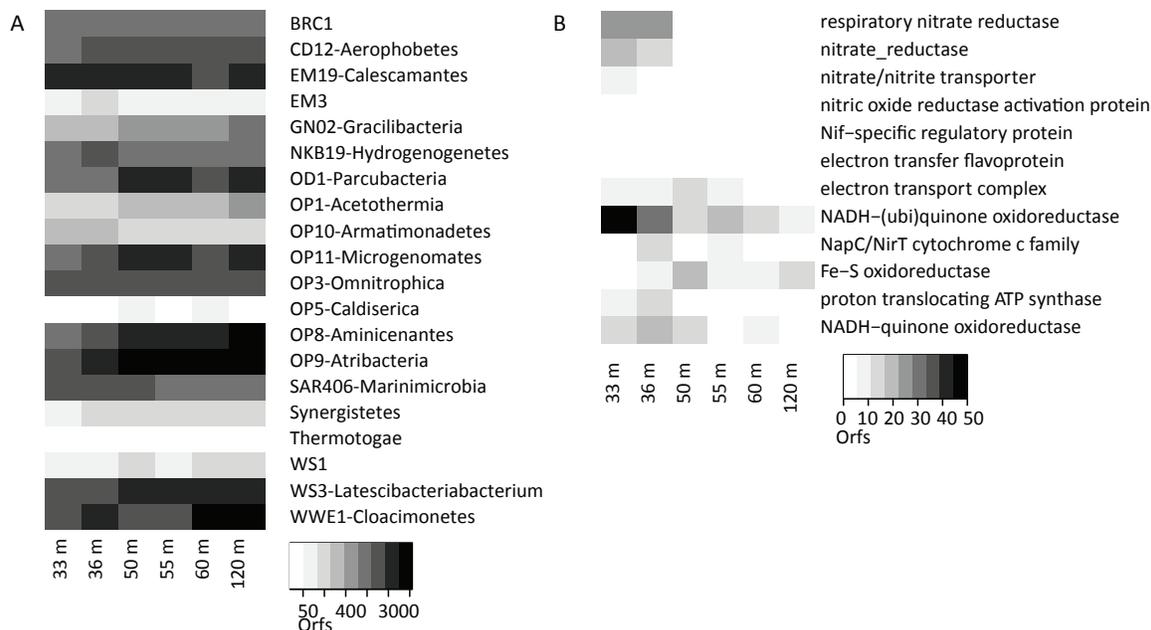


Figure 4.11 Phylogenetic anchoring of MDM reads.

Phylogenetic anchoring of predicted open reading frames (ORFs) based on BLAST/LAST outputs from Metapathways using a custom database of annotated protein sequences from 201 published MDM SAGs (Rinke et al 2013). (B) Genomic potential for respiratory nitrate reduction and presence of electron transfer chain genes in ORFs affiliated with OP3.

4.4 Discussion

The Sakinaw Lake water column revealed persistent water column stratification between June 2007 and June 2013 leading to a distinct partitioning into mixolimnion (0 m – 30 m), RTZ (30 m – 55 m) and monimolimnion (55 m- 120 m). Seasonal changes in temperature and O₂ concentrations were detectable in the mixolimnion, and at the interface of the mixolimnion and the RTZ (33 m), causing temporal changes in the depths of the H₂S transition zone (see also Chapter 2). Indeed, despite the stability in temperature and salinity profiles, H₂S concentrations were shown to change up to four fold below 33 m. The cause of the drastic changes in H₂S concentrations remains unknown.

Consistent with the physicochemical characteristics of the water column, MJ Anderson's analysis of beta-dispersion revealed temporal changes in microbial population structure in the mixolimnion, and upper part of the RTZ, and a stable composition of the microbial community in the lower part of the RTZ and the monimolimnion. The partitioning of the community between mixolimnion, RTZ, and monimolimnion was further resolved in the simplex plot of microbial classes. Moreover, hierarchical cluster analyses of OTU populations revealed the grouping of OP11, OD1, OP9 and OP3 into water compartment-specific OTU clusters. The division of OP3 and OD1 into water compartment-specific clades could further be confirmed with phylogenetic analysis. All candidate divisions were represented by abundant or intermediate abundant OTUs in each of the compartment-specific clusters, providing an ideal environment for comparative studies of different genotypes. Coverage of the metagenomic shotgun sequencing data was, however, too low to assemble population genomes, impeding a comparative analysis. High throughput sequencing and/or single cell genomics will be required to follow up on putative metabolic diversification along the Sakinaw Lake water column in candidate divisions OP11, OD1, OP9 and OP3.

Metabolic pathway reconstruction using unassembled 454 shotgun sequencing data revealed a less distinct partitioning into water compartment-specific pathways compared to the partitioning on the taxonomic and OTU level. It is to be noted that genome-sequencing data for samples of the mixolimnion was missing, which might be an underlying reason for the lower resolution. The majority of metabolic pathways were fermentation related and could be recovered from all water column compartments. The division into an upper and a lower part of the RTZ allowed the identification of unique metabolic pathways in the upper part of the RTZ including aerobic respiration, dissimilatory nitrate reduction, arsenate reduction and intra-aerobic

NO₃⁻ reduction. Genomic potential for respiratory nitrate reduction was the only pathway that could be linked to a candidate division, and is described in more detail below. The distinct separation of the upper part of the RTZ from the remaining depth and the coverage of aerobic as well as anaerobic respiration pathways is consistent with observed variations in physicochemical parameter and results from MJ Anderson's analysis of betadispersion, underlining that the upper part of the RTZ provides a unique metabolic niche for microbes that can tolerate variations in O₂ and H₂S concentrations.

Candidate divisions RF3 as well as *Chlorobi* and *Lentisphaere* have been identified as distinctly associated with the RTZ. *Lentisphaere*, described as putatively fermenting bacteria, have recently been isolated from the human gastrointestinal tract and from an upflow anaerobic sludge blanket reactor (Cho et al 2004, Qiu et al 2013, van Passel et al 2011). Additionally, *Lentisphaere* have been isolated from the ocean, where they are thought to exist as aerobic chemoheterotrophs (Cho, Vergin et al. 2004). The dominance of *Lentisphaere* in the upper part of the RTZ underlines specific metabolic needs that can only be supported within the RTZ, leading to the speculation that the Sakinaw Lake population might be chemo- or lithoheterotrophic and not fermentative. This hypothesis is supported by the above described partitioning of metabolic pathways with unique coverage for aerobic and anaerobic respiration to the upper part of the RTZ. The *Chlorobi* population in Sakinaw Lake was dominated by active and low light adapted *Chlorobiales*, which are a well-studied order of chemolithoautotrophs that use H₂S as an electron source for anoxygenic photosynthesis (Overmann 2008). The parallel population structure of candidate division RF3 suggests potential metabolic similarities or dependencies between RF3 and *Chlorobi/Chlorobiales*. A previous study by Lanzen and colleagues surveyed the microbial community of six Ethiopian Soda Lakes, and revealed co-

existence of RF3 and phototrophic purple sulfur bacteria affiliated with *Chromatiales*, supporting the hypothesis of shared metabolic traits (Lanzen et al 2013). However, inspection of temporal signals in abundance and positioning of RF3 and *Chlorobi* within the RTZ of Sakinaw Lake revealed differential population dynamics. Moreover, phylogenetic analysis of the 16S rRNA gene placed RF3 within the *Firmicutes*, but with distinct differentiation to the anoxygenic phototrophic class *Heliobacter*. Instead, RF3 were more closely related to *Erysipelotrichi*, which have been identified as common members of the gut microbiome (Turnbaugh et al 2009). The functional role of RF3 in the Sakinaw Lake water column remains elusive, but the strict niche partitioning within the RTZ points towards specific metabolic needs that can only be supported in the RTZ.

Phylogenetic anchoring of metagenomic shotgun sequences using an in-house generated database that consisted of proteins from 201 recently published MDM SAGs (Rinke et al 2013), recovered OP3 affiliated reads from all depths of the anoxic-sulfidic part of the water column. Genomic analysis suggests a potential metabolic differentiation of OP3 along the Sakinaw Lake water column. This observation was consistent with the above-mentioned phylogenetic differentiation into water compartment-specific clades. Samples of the upper part of the RTZ revealed unique coverage for genes encoding respiratory NO_3^- reductase, while ORFs recovered from samples of the lower part of the RTZ and the monimolimnion revealed gene coverage for Fe-S oxidoreductases. It has been shown that Fe-S oxidoreductase isolated from *Thiobacillus ferrooxidans* can couple sulfur-oxidation to ferric ion reduction under aerobic and anaerobic conditions (Pronk et al 1992, Suzuki et al 1990). Moreover, laboratory cultures of *Thiobacillus ferrooxidans* were shown to grow on pyrite as substrate as well (Suzuki et al 1990). The coverage of Fe-S oxidoreductase supports therefore hypotheses from previous studies that

suggested a potential role in Fe and S cycling for OP3 populations based on genomic and phylogenetic analyses (Glöckner et al 2010, Kolinko et al 2012). Kolinko and colleagues provided further insight into the function of OP3 by combining targeted sampling techniques, genomic sequencing and microscopic imaging (Kolinko et al 2012). They isolated OP3 cells from magnetically collected multi-species assemblages of magnetotactic bacteria (MTB) and confirmed the presence of magnetosomes microscopically. Magnetotaxis is driven by specific organelles, called the magnetosomes that support orientation along Earth geomagnetic fields. It has been suggested that Bacteria thriving in microaerophilic, chemically stratified environments use magnetotaxis for navigation along redox gradients (Flies et al 2005). Microscopic and or genomic analyses of the RTZ and monimolimnion populations of OP3 in Sakinaw Lake will be required to identify whether RTZ associated and/or monimolimnion associated populations of OP3 in Sakinaw Lake are magnetotactic as well. Next to the evidence of magnetotaxis, Kolinko and colleagues also discovered Fe and S inclusions with microscopic imaging analyses, supporting hypotheses for Fe and S respiration (Kolinko et al 2012).

Phylogenetic anchoring using the MDM database for Sakinaw Lake metagenomes did not only recover Fe-S oxidoreductases associated with OP3, but also with numerous other candidate divisions including OD1, OP11, OP8, OP9, and WWE1. As mentioned above, Fe-S oxidoreductases have been linked to pyrite oxidation. Microbial pyrite oxidation has recently been suggested as important mediator of Fe and S cycles during the Archaean Eon (~ 3.5 billion years ago), when atmospheric oxygen was thought to be $< 10^{-5}$ of current levels (Wacey et al 2011). The presence of pyrite in the anoxic-sulfidic waters of Sakinaw Lake has previously been described (Perry 1990). To date microbial mediated pyrite oxidation has only been shown in the presence of TEA NO_3^- and SO_4^{2-} , which are absent from the deep waters of Sakinaw Lake.

Previous analysis in Sakinaw Lake revealed presence of manganese (Mn) throughout the anoxic-sulfidic part of the water column (Gies et al 2014), it is however not known whether Mn could function as TEA in microbial mediated pyrite oxidation. Next to Fe-S oxidoreductases, genomic analysis anchored ORFs that encoded for polysulfide reductases affiliated with candidate divisions EM19 and KSB1, which is consistent with pathway reconstruction of metagenomic reads that identified polysulfide reduction as prevalent throughout the anoxic sulfidic part of the water column. Given the high polysulfide and H₂S concentrations in Sakinaw Lake (Perry 1990), polysulfide reduction to H₂S would be a feasible energy generation pathway in the RTZ.

4.5 Conclusion

Taken together, this chapter provided deep insights into the diversity, and population structure of MDM in Sakinaw Lake. Phylogenetic and genomic analyses indicated the division of candidate division OP3 into two populations with divergent metabolic potential. Hierarchical cluster analyses of MDM OTUs suggested that most candidate divisions diverged along the stratified watercolumn, including OP11, OD1 and OP9. Single cell genome analyses or high throughput sequencing and consecutive assembly of population genomes will be needed for comparative analysis of water compartment-specific genotypes. The high relative abundances, and the distinct stratification of MDM OTUs along the water column, provide an ideal environment to study metabolic and genomic versatility in candidate divisions. Furthermore, the extraordinary S chemistry in Sakinaw Lake, combined with metabolic potential for polysulfide reduction and Fe-S respiration provide a unique opportunity to study MDM in the context of Fe and S cycling now and through the deep expanse of evolutionary time.

Chapter 5: Phylogeny and function of novel and known Archaea in Sakinaw Lake

5.1 Synopsis

Over the past two decades, cultivation-independent methods have vastly improved our understanding of archaeal diversity and expanded our perception of this domain beyond extreme environmental settings. Indeed, single-cell genomic sequencing recently led to the proposal of two superphyla: the TACK (Thaumarchaeota-Aigarchaeota-Crenarchaeota-Korarchaeota) and the DPANN (Diapherotrites-Parvarchaeota-Nanoarchaeota-Nanohaloarchaeota) with global distribution patterns. Despite these improvements, archaeal population structure and coding potential remain poorly constrained in many natural and engineered ecosystems. In this study a combination of CARD-FISH, phylogenetic and functional anchors including 16S rRNA gene and methyl coenzyme m reductase alpha subunit (*mcrA*), as well as meta- and single-cell genomic sequencing were used to illuminate the phylogeny and coding potential of previously described “unassigned” Archaea in Sakinaw Lake (Table 5.1). Most of these Archaea turned out to be affiliated with the novel superphylum DPANN including the *Nanoarchaeota*, *Aenigmarchaeota* and *Diapherotrites*. While recruitment of shotgun reads was generally in agreement with 16S rRNA gene profiles, additional groups including a novel class of *Geoarchaea* within the *Crenarchaeota*, methanogenic *Methanomicrobiales* and anaerobic methane oxidizing (ANME) were identified. Taken together, the Sakinaw Lake archaeal community manifests an extraordinary diversity and abundance of archaeal candidate divisions, including known and novel groups with important implications for methane cycling in aquatic ecosystems.

5.2 Material and methods

5.2.1 Study site and sampling

Sakinaw Lake water samples were taken on three different field campaigns on June 6th 2007, May 24th 2011, and June 6th 2013 at deep basin station S1 (49 40.968 N 124 00.119 W) with a combination of 12L Niskin and 8L go-flow bottles. Water samples for physicochemical parameter and cell counts with flow cytometry were collected as described previously (Gies et al 2014). For DNA extraction water was kept at 4°C, and cells were harvested on day of sample collection by filtration through a 0.22 µm sterivex-GV-filter (Millipore) without pre-filter using a Masterflex L/S 7553–70 peristaltic pump (Cole-Parmer). Sample procession for environmental DNA extraction was done as described previously (chapter 3 and (Gies et al 2014)). Samples for the generation of *mcrA* and 16S rRNA clone libraries were collected on June 6th 2007 from 50 m and from 25 m, 30 m, 33 m, 40 m, 50 m and 120 m respectively. Metagenomes were generated from DNA samples collected May 24th 2011 from 33 m, 36 m, 50 m, 60 m, 120 m. CARD-FISH samples were collected on June 6th 2013 from 5 m, 20 m, 30 m, 33 m, 36 m, 40 m, 45 m, 50 m, 55 m, 60 m, and 80 m. Water was filled into 200 ml, acid washed and autoclaved serum bottles and fixed with particle free paraformaldehyde (PFA) at 2% final concentration directly after sample collection and kept at 4°C in the dark until filtration in the laboratory ~ 10h after sample collection. Per depth, five mL water was filtered in triplicates onto white polycarbonate filter (Millipore) with 22 mm diameter and 0.2 µM pore size and stored in the dark at -80°C until further processing.

5.2.2 Methods to infer taxonomic composition based on the 16S rRNA gene

For a comprehensive representation of the Archaeal community in Sakinaw Lake we combined several well-established cultivation independent methods to assess the taxonomic composition based on 16S rRNA gene phylogeny (Table 5.1). Individual methods are described in the following sections.

Table 5.1 Methods and primers used to infer taxonomic composition based on the 16S rRNA gene

	Probe/primer	Probe/primer sequence	# of recovered 16S sequences	original probe/primer reference
CARD-Fish clone libraries	Arch915	5'- GTG CTC CCC CGC CAA TTC CT -3'	NA	Stahl D. A. 1991
	20F	5'-TTC CGG TTG ATC CYG CCR G -3'		
(Sanger sequencing)	958R	5'-YCC GGC GTT GAM TCC AAT T -3'	108	DeLong et al, 1992
454 (V8-V8 region of the 16S rRNA gene)	926F	5'-AAA CTY AAA KGA ATT GRC GG-3'		Engelbrektsen, Kunin et al. 2010
	1392R	5'-ACG GGC GGT GTG TRC-3'	194,547	
454 (genomic shotgun)	NA	NA	14	NA
Single amplified genomes	NA	NA	2	NA

5.2.3 Quantification of Archaea with CARD-FISH

Archaeal cell numbers were enumerated using previously published domain specific probe Arch915 (5'- GTG CTC CCC CGC CAA TTC CT -3') (Stahl D. A. 1991). Incubation with HRP-labeled probes was done as previously described (Pernthaler et al 2002). In short, filters were cut into four equal sized sections and subsequently embedded in low gelling point 0.1% Agarose (Sea Kem LE Agarose, Cambrex). Cells were permeabilized with Lysozyme (chicken egg, Fluka) at 37°C for 60 min. For hybridization, 400 µL hybridization buffer (35% formamide) was supplemented with 2 µl HRP probe at a final concentration of 50 pmol. Hybridization was carried out for 8 h to 12 h in a rotary incubator in the dark. Filter sections were then washed in pre warmed (37°C washing buffer, 35% NaCl) for 15 min. Tyramide signal amplification was carried out in amplification buffer in the dark at 37°C for 20 min. Filters were then washed in PBS, dried with 90% ethanol and mounted on cover slips. For counter staining of all cells,

mounting solution (Vectashield, Linaris) was supplemented with DAPI (1 $\mu\text{g ml}^{-1}$ final concentration, Sigma). Filter sections were evaluated using an epifluorescence Microscope (Axioscope 2 mot plus, Zeiss) with an Apochromate oil objective lense (100x, Zeiss). Arch915 positive cells were counted in 10 randomly distributed fields of view (500 - 1000 cells/sample), and counts were complemented with total microbial cell counts inferred from the DAPI staining.

5.2.4 PCR amplification of 16S rRNA gene, clone library construction and sequencing

Archaeal small subunit (SSU) ribosomal RNA sequences were amplified via PCR using universal primers 20F (5'TTCCGGTTGATCCYGCCRG) and 958R (5'YCCGGCGTTGAMTCCAATT) using the following PCR profile: a 2 min hot start at 95°C followed by 30 cycles of 95°C for 20 s, 55°C for 20 s, 72°C for 1 min and a final extension at 72°C for 3 min. For PCR amplification 50 μL reactions were prepared using 1 μL of template DNA, 10 μL of Hercules II 5x buffer, 1 μL of Hercules II proofreading polymerase (Stratagene, La Jolla, CA), 5 μL deoxynucleotides (10mM), 1.25 μL of reverse and forward primer (10 μM), and 30.5 μL nuclease free, sterile water. The quality of the PCR product was checked on 0.8% agarose gel with 1x TBE buffer and subsequently purified using a Purelink PCR purification kit according to the manufacturer's instructions. DNA concentrations were estimated on a Nanodrop and samples were concentrated to > 65 ng/ μL using a SpeedVac. Cloning, subsequent PCR amplification, digestion of the inserts and sequencing was done as described for *mcrA* clone library. Sequences were edited manually using Sequencher software V4.1.2. (Gene Codes Corporation, Ann Arbor, MI). Sequencher software editing includes assembly of contigs from the reverse and forward reads, trimming of ends, and removal of the vector and the universal

primer. Chimeric sequences were removed using the software tool Mallard (<http://www.bioinformatics-toolkit.org/Mallard/index.html>).

5.2.5 Phylogenetic analysis and matching of Arch915 probe

A total 108 non-chimeric Sakinaw Lake 16S rRNA gene sequences were used for the generation of a phylogenetic tree. Cluster representatives were determined at 99% identity, using the open source software mothur v.1.19.0 (Schloss, Westcott et al. 2009), resulting in the selection of 34 OTU representatives for downstream analysis. Sequences were aligned using online aligning software “Sina” (<http://www.arb-silva.de/aligner/>), and subsequently imported into the ARB software package using the ARB parsimony tool at default settings ((Ludwig et al 2004), database v 115)). To test whether universal Archaeal CARD-FISH probe Arch915 matches all Sakinaw Lake sequence representatives, *in silico* probe matching was performed using the probe matching tool implemented in the ARB software, allowing 2 mismatches. A maximum likelihood tree was generated using open source software PHYML (Guindon et al., 2005). Positioning of the sequences was inferred with GTR+4G+I model of nucleotide evolution. The parameter for gamma distribution, and the transition/transversion ratio were estimated, and assembly of a consensus tree was done from 100 bootstrap replicates.

5.2.6 Metagenome assembly and analysis

Unassembled reads from six 454 metagenome samples collected on May 24th 2011, was done using the Metapathways pipeline (v.2.4.1) (Konwar et al 2013). Samples were analyzed using the Metapathways pipeline with standard settings (Konwar et al 2013). Unassembled sequences were loaded into the program and sequences shorter than 200 bp were removed. On average 255,193

sequences per sample passed this quality control step. Open reading frames were predicted using the Prokaryotic Dynamic Programming Gene-finding Algorithm (Prodigal) (Hyatt et al 2010), resulting in an average of 300,707 predicted ORFs per sample. A quality control step removed ORFs shorter than 180 bp (60 amino acids) leading to an average of 207,747 ORFs for downstream analysis. For annotation, translated ORFs were queried against protein databases using the protein BLAST or optimized LAST algorithm (Altschul et al 1990, Kielbasa et al 2011). Annotation of ORFs was done using the following seven publicly available databases (release date in parentheses): metacyc-v4 (2011-07-16), InnateDB mus (2010-7-16), InnateDB human (2010-07-16), refseq-nr (2014-01-18), COG (2013-12-27), kegg-pep (2011-06-18), seed (2014-01-30). Moreover, a custom database with proteins predicted in 201 recently published single cell genomes (Rinke et al 2013) was also queried. MetaPathways BLAST/LAST searches were run on the WestGrid compute grid grex on the Compute/Comsol Canada Consortium (<https://www.westgrid.ca/>).

For recovery of near complete archaeal 16S rRNA gene sequences, raw reads were assembled with open source software Genovo assembler v0.4. Assembly was done by sampling 30 iterations under default error model settings (Laserson et al 2011). Assembled contigs shorter than 500 bp were dropped from further analysis. For annotation, identification and extraction of archaeal 16S rRNA genes, assembled contigs were analyzed with Metapathways using identical settings as for unassembled reads.

5.2.7 *mcrA* clone library

Archaeal methyl coenzyme M reductase subunit A (*mcrA*) was amplified using previously published primers ME1 (5'-GCMATGCARATHGGWATGTC) and ME2 (5'-

TCATKGCRTAGTTDGGRTAGT) (Hales et al 1996). For PCR amplification 50 μ l reactions were prepared using 1 μ l of template DNA, 1 μ l each forward and reverse primer (10 μ M), 2.5 U Taq-polymerase (Qiagen), 5 μ l deoxynucleotides (10 mM), and 41.5 μ l Taq-buffer (Qiagen Taq Polymerase Kit). The protocol for PCR amplification was as follows: 3 min at 94°C followed by 35 cycles of 94°C for 40s, 52°C for 1.5 min, 72°C for 2 min and a final extension of 10 min at 72°C. For quality control, amplicons were visualized on 1% agarose gels in 1X Tris-acetate-EDTA (TAE), and purified using a PCR Purification Kit (MiniElute, Qiagen, CA) according manufacturer's instructions. Approximately 4 μ l purified product was cloned into pCR4-TOPO vector using TOPO TA cloning kit (Invitrogen, Carlsbad CA) and transformed by chemical transformation into Mach-1-T1R cells, following manufacturer's instructions. Transformed cells were transferred into four 96-well plates, which contained 180 μ l Lb medium, 50 μ g/ml kanamycin and 10% glycerol to grow cells at 37°C overnight prior to storage at -80°C. One 96-well plate was selected for DNA fingerprint screening using RSA I (Invitrogen, CA) 4-base cutter. For fingerprint analysis inserts were amplified directly from glycerol stocks with M13F (5'- GTAAAACGACGGCCAG) and M13R (5' - CAGGAAACAGCTATGAC) primers using above-mentioned PCR profile. To ensure adequate cell lysis samples were incubated for 10 min at 94°C prior to the start of the protocol. The reaction mixture consisted of 2 μ l 10x React 1 buffer, 5 μ l of M13 amplified PCR product, 16 μ l sterile water and 10U RSA I. Digestion reactions incubated for 2 hours at 37°C followed by RSA I inactivation at 65°C for 10 min. The digested fragments were visualized by running 5 μ l of each RSA I reaction for 90 min on a 2% agarose gel at 120 V (20 cm in width) using 1X TAE buffer. Restriction patterns were visually inspected and unique patterns selected and run against one another for confirmation.

Twenty-four unique restriction patterns were selected for Sanger sequencing through the McGill University and Genome Quebec Innovation Centre (Montreal, Quebec, Canada). Sequence data was collected on an ABI Prism 3100 DNA sequencer (Applied Biosystems Inc, Foster, CA) using Big Dye™ chemistry (PE Biosystems, Foster, CA) according to manufacturer's instructions. Plasmids were sequenced bidirectional with M13F and M13R primers. Sequences were quality controlled and edited manually using Sequencher software V4.1.2 (Gene Codes Corporation, Ann Arbor, MI). For the calculation of a phylogenetic tree, the 24 *mcrA* sequences were translated into amino acid sequences using GeneWise software (Birney et al 2004), implemented in the software package MLtreemap (Stark et al 2010). Sequences were then sorted by length and clustered at 99% identity using open source software uclust (Edgar 2010). Sequence alignment was done using the online tool muscle, which was set at clustalW (<http://www.ebi.ac.uk/Tools/msa/muscle/>) (Edgar 2004). A maximum likelihood tree was generated using open source software PHYML (Guindon et al., 2005). Positioning of the sequences was inferred with the WAG model of amino acid evolution. The proportion of invariable sites was estimated, the number of substitution rate categories was 4, gamma distribution parameters were estimated, and assembly of a consensus tree was done from 100 bootstrap replicates.

5.2.8 SAG analysis

The single amplified genome of a Euryarchaeota Archaeon was generated but not curated from a 120 m sample collected January 5th 2010 in Sakinaw Lake, and published in a study by Rinke and colleagues (Rinke et al., 2013). The sequence available on the IMG server (<http://img.jgi.doe.gov>) can also be found under the ID SCGC AAA252-I15 (SAK_001_55).

Details about cell sorting and lysis, single-cell whole genome amplification, 16S rRNA amplicon sequencing, SAG sequencing and assemblies as well as decontamination have been described (Rinke et al., 2013). My analysis was focused on curation, specifically metabolic reconstruction using functional annotations present on the IMG platform, and was conducted using results from KEGG (Kyoto Encyclopedia of Genes and Genomes) and Metacyc databases (Caspi et al 2008, Kanehisa 2002). For the prediction of proteases, peptidases and protease inhibitors Blastp analysis against the Merops database was performed using the Merops online service (<http://merops.sanger.ac.uk/>(Rawlings et al 2006)). Gene duplications were identified by Blastp analysis running proteins of the SAG as subject and query. Transporter identification was implemented using IMG and the transporter classification database (TCDB) (Saier et al 2006). Protein subcellular localizations were identified using the online tool PSORTb V 3.0.2 (<http://www.psort.org> (Nancy et al 2010)).

5.3 Results

5.3.1 Archaeal cell counts

To estimate total microbial cell numbers along the water column of Sakinaw Lake, samples were collected for automated cell counts using flow cytometry as well as for microscopic evaluation of DAPI positive cells. For DAPI counts biomass was filtered directly onto 0.2 μm polycarbonate filter. Comparison of the flow cytometric and DAPI count profiles revealed cell numbers within the same order of magnitude, but discrepancies in the shape of the abundance curve (Figure 5.1). DAPI counts revealed a peak of 1.48×10^5 cells/mL at 20 m depth and a second peak of 2.7×10^5 cells/mL at 55 m depth, whereas flow cytometry counts were at 7.72×10^4 cells/mL and 1.69×10^5 cells/mL respectively. In contrast, using flow cytometry a peak of 2.51×10^5 cells/mL was

measured at 36 m, with corresponding DAPI counts at 1.49×10^5 cells/mL. Microscopic evaluation with DAPI staining uncovered high diversity in cell shape and size, which might be the underlying reason for discrepancies between methods, and potentially skewed automated counting with flow cytometry.

Based on 454 tag sequencing, it was estimated that Archaea belong to the rare biosphere in the mixolimnion, becoming increasingly abundant in anoxic-sulfidic waters (Gies et al 2014). To test whether CARD-FISH imaging using a previously published universal probe for Archaea (Arch915) reveals a similar abundance profile (Raskin et al 1994), biomass was filtered directly onto 0.2 μm polycarbonate filter and prepared for microscopic evaluation using previously described methods (Pernthaler et al 2002). Arch915 counts revealed a peak of 3.7×10^4 cells/mL at 36 m depth and in the remaining part of the water column Archaea were estimated between 4.5×10^3 and 3.1×10^4 cells (Figure 5.1 A).

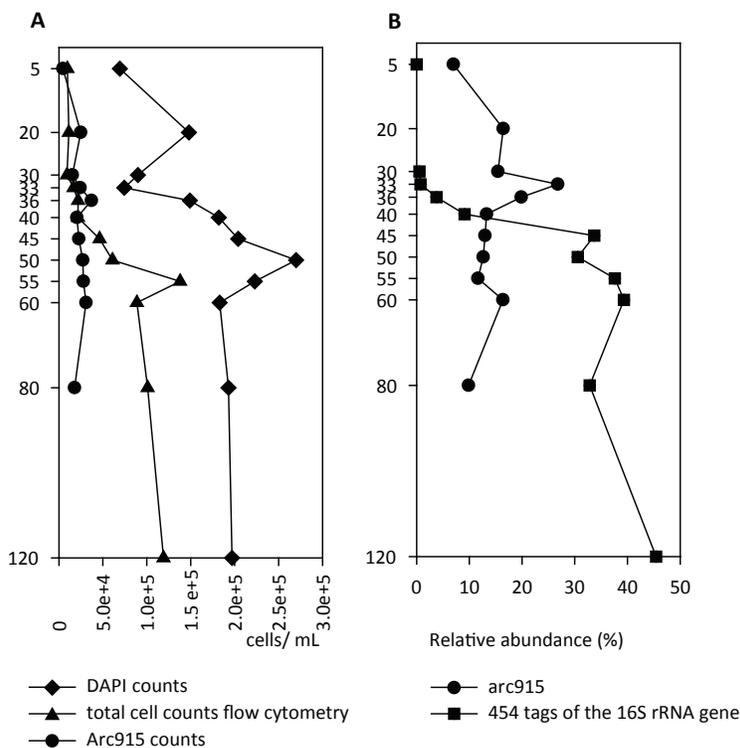


Figure 5.1 Total cell counts and relative abundance of Archaea.

Samples for total cell counts with flow cytometry were unfiltered, and water samples for DAPI and Archaeal cell counts were filtered onto membrane filter with 0.22 μm pore size. (A) DAPI counts, Arch915 counts, and total cell count using flow cytometry. (B) Proportion of Arch915 positive DAPI stained cells, compared to relative abundance of Archaea inferred with 454 sequencing of the 16S rRNA gene using a universal three-domain primer.

Relativization towards DAPI positive counts, revealed that approximately 6% to 26% of the microbial community belong to the Archaea. Comparison of these microscopic results to results from 454 tag sequencing revealed significant discrepancies between methods: while less than 1% of the 454 tag sequences from the mixolimnion were affiliated with Archaea, 16% of the DAPI positive cells hybridized with the Arch915 probe. In contrast, in anoxic-sulfidic waters below 33 m more than 40% of tag sequences were affiliated with Archaea, while only 16% of the DAPI positive cells hybridized with the Arch915 probe.

5.3.2 Archaeal phylogeny

Robust phylogenetic trees, using 93 near complete 16S rRNA sequences generated from clone libraries, and 14 assembled shotgun sequences were constructed. This enabled more accurate phylogenetic placement without potential primer bias in the case of assembled shotgun sequences. Moreover, one 16S rRNA sequence from a previously published archaeal Sakinaw Lake SAG (CICYyyyy) was added as well, to gain insight into its phylogenetic relationship (Rinke et al 2013). Since the SAG itself did not have coverage for the 16S rRNA gene (see table 5.2), a partial sequence (504 bp) obtained from 454 sequencing of the 16SrRNA gene of FACS sorted cells prior to whole genome amplification was used. Altogether, a total of 108 sequences were clustered at 99% identity resolving 34 operational taxonomic units (OTUs) used for downstream phylogenetic analysis. To infer the relationship of Archaea in Sakinaw Lake with recently described archaeal candidate phyla, 14 16S rRNA sequences extracted from archaeal SAGs generated by Rinke and colleagues were included in the phylogenetic analysis (Rinke et al 2013). To provide a comprehensive view of the biogeography for relatives of the Sakinaw Lake archaeal population, reference sequences from natural and engineered ecosystems were also included resulting in 152 sequences used in tree construction.

The phylogenetic diversity of *Thaumarchaeota* inferred from clone libraries generated with the Archaea specific primer pair (20F, 958R) was overall in agreement with results from previous analysis of 454 tag sequences generated with universal three domain primer (926F, 1392R) (Gies et al 2014) (Figure 5.2). The *Miscellaneous Crenarcheotic Group* (MCG) was represented by seven OTUs that formed three different clades. Moreover, assembly of shotgun sequences recovered an additional MCG 16S rRNA gene. Members of the MCG were absent from the mixolimnion and increased in abundance in the lower part of the RTZ and

monimolimnion, making up 6.5% of the archaeal population at 120 m. Affiliates of the *Marine Group 1* (MG1) were represented by only one OTU, that was abundant in the mixolimnion and the upper part of the RTZ, making up 6.5% at 33 m. Marine Group 1 Archaea were absent from the monimolimnion. The thaumarchaeal class C3 was represented by two OTUs recovered at 40 m and 50 m with 2.1% and 1.1% relative abundance respectively. Results from 454 tag sequencing suggested that *Thaumarchaeota* belonged to the rare biosphere (< 1%) (chapter 3, Figure 3.6 C), while 16S clone library generation revealed that more than 20% of the sequences belonged to the *Thaumarchaeota*. Although a direct comparison of relative abundances inferred from 454 tag sequencing with relative abundance inferred from clone libraries is not possible, the diversity of thaumarchaeal OTUs represented in the clone libraries and the ability to recover an assembled 16S rRNA sequence from the 454 genomic shotgun sequencing dataset suggests the presence of a respectable thaumarchaeal population size. Moreover, *in silico* matching of the CARD-FISH probe further reveals matches for all Sakinaw Lake *Thaumarchaeota* recovered in the clone library, suggesting that the relatively high archaeal abundances in the mixolimnion inferred by CARD-FISH analysis might represent the MG1 (Figure 5.1).

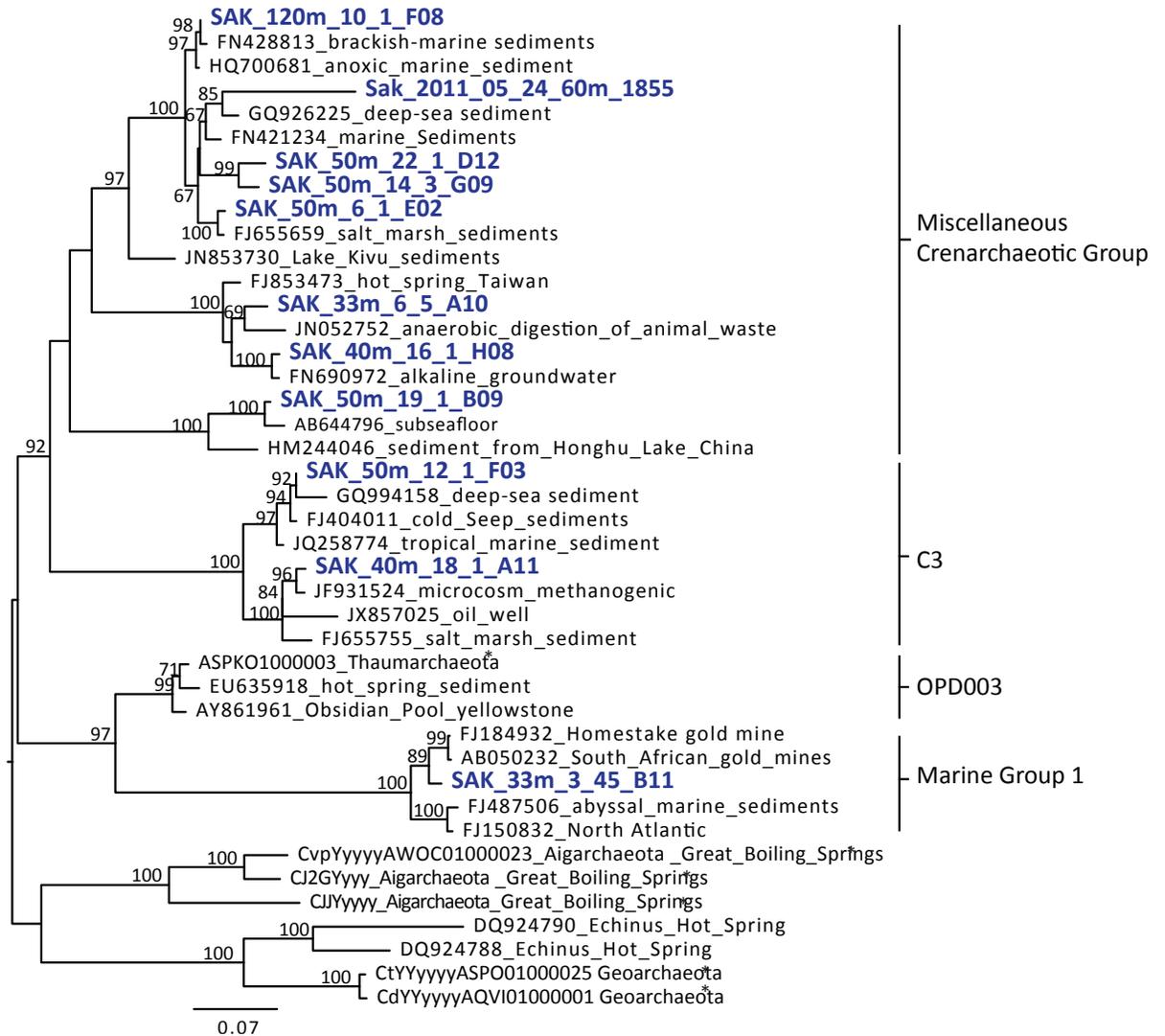


Figure 5.2 Phylogenetic tree of Thaumarchaeota in Sakinaw Lake.

Phylogenetic tree generated with near full-length 16S rRNA gene sequences from gene clone libraries and full-length 16S rRNA sequences from 454 shotgun sequence assembly. The tree was calculated using maximum likelihood estimation and generalized time reversible (GTR) models implemented in open source software PhyML. The proportion of invariable sites was estimated and assembly of a consensus tree was done from 100 bootstrap replicates. Sakinaw Lake cluster representatives (clustered at 99% identity) are indicated in bold face. Sequences with *in silico* matches to CARD-FISH probe Arch915 are indicated in blue. Cluster representatives with 2011_05_24 designations were recovered from assembled shotgun sequences.

The phylogenetic diversity of DPANN recovered from the clone library and genome assembly could not be compared to previous results from 454 sequencing, because DPANN representatives were not included in the Silva v111 database used for previous analyses (chapter 3 (Gies et al 2014)). Clone library sequences were affiliated with *Aenigmarchaeota* (up to 4.3%) that were abundantly present throughout the water column. The population of *Aenigmarchaeota* revealed to be relatively diverse and was represented by eight OTUs that formed three clades (Figure 5.3). Moreover, affiliates of the *Diapherotrites* and *Nanoarchaeota* were recovered from assembled genomic sequences represented by one and three OTUs respectively. *In silico* matching of the CARD-FISH probe Arch915 showed strong biases against representative of the DPANN superphylum. Only one OTU of the *Nanoarchaeota* matched the probe.

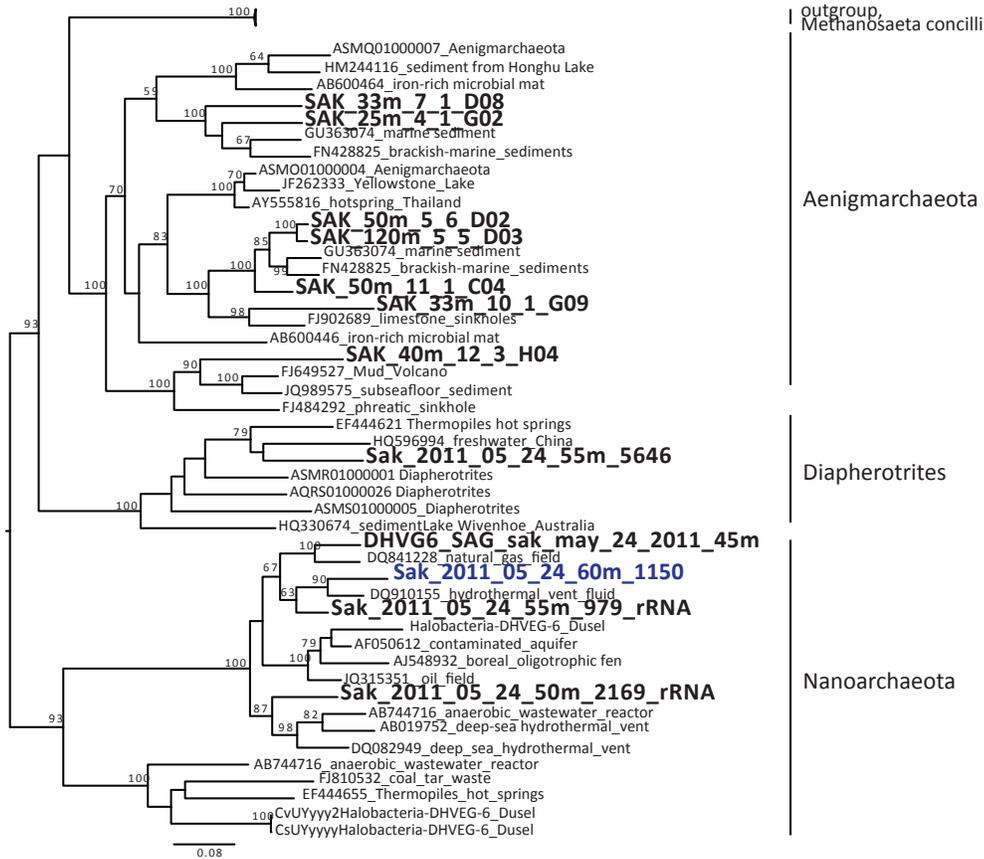


Figure 5.3 Phylogenetic tree of the DPANN.

Phylogenetic tree generated with near full-length 16S rRNA gene sequences from gene clone libraries and full-length 16S rRNA sequences from 454 shotgun sequence assembly. The tree was calculated using maximum likelihood estimation and generalized time reversible (GTR) models implemented in open source software PhyML. The proportion of invariable sites was estimated and assembly of a consensus tree was done from 100 bootstrap replicates. Sakinaw Lake cluster representatives (clustered at 99% identity) are indicated in bold face. The sequences affiliated with *Methanosaeta concilli* (accession numbers X51423, AB679168, and CP002565) functioned as outgroup. Sequences with *in silico* matches to CARD-FISH probe Arch915 are indicated in blue. Cluster representatives with 2011_05_24 designations were recovered from assembled shotgun sequences.

The phylogenetic diversity of *Euryarchaeota* recovered from the clone library was overall similar to observations using universal tag sequencing primers (Figure 5.4). However, clone libraries recovered three *Euryarchaeota* that were absent from the 454 sequencing dataset, each represented by one OTU: the *TMEG* (3.2% at 40 m), the *Marine Benthic Group* (1.1% at 25 m) and the *WCHAI* (2.1% at 40 m). *Methanosarcinales* was represented by eight OTUs that formed three different clades with maximum relative abundance of 5.5% at 20 m, while *Methanomicrobiales* was represented by only one OTU reaching 2.1% relative abundance at 40 m. One additional *Methanomicrobiales* sequence, which was distinctly diverging from the clone library representative, was recovered from the assembled shotgun sequences. Moreover, sequences affiliated with ANME Archaea were recovered from clone libraries and from assembled shotgun sequences reaching 2.5% relative abundance at 50 m and 120 m. The majority of the euryarchaeal sequences matched the Arch915 probe. Exceptions were assembled sequence representatives of the *Methanomicrobiales* and the ANME. The Sakinaw Lake SAG CICYyyyy was suggested to be a putative novel methanogen or methanotroph ((Rinke et al 2013) supplementary material), however phylogenetic analysis placed the novel lineage in a cluster that distinctly diverged from known methanogens and methanotrophs.

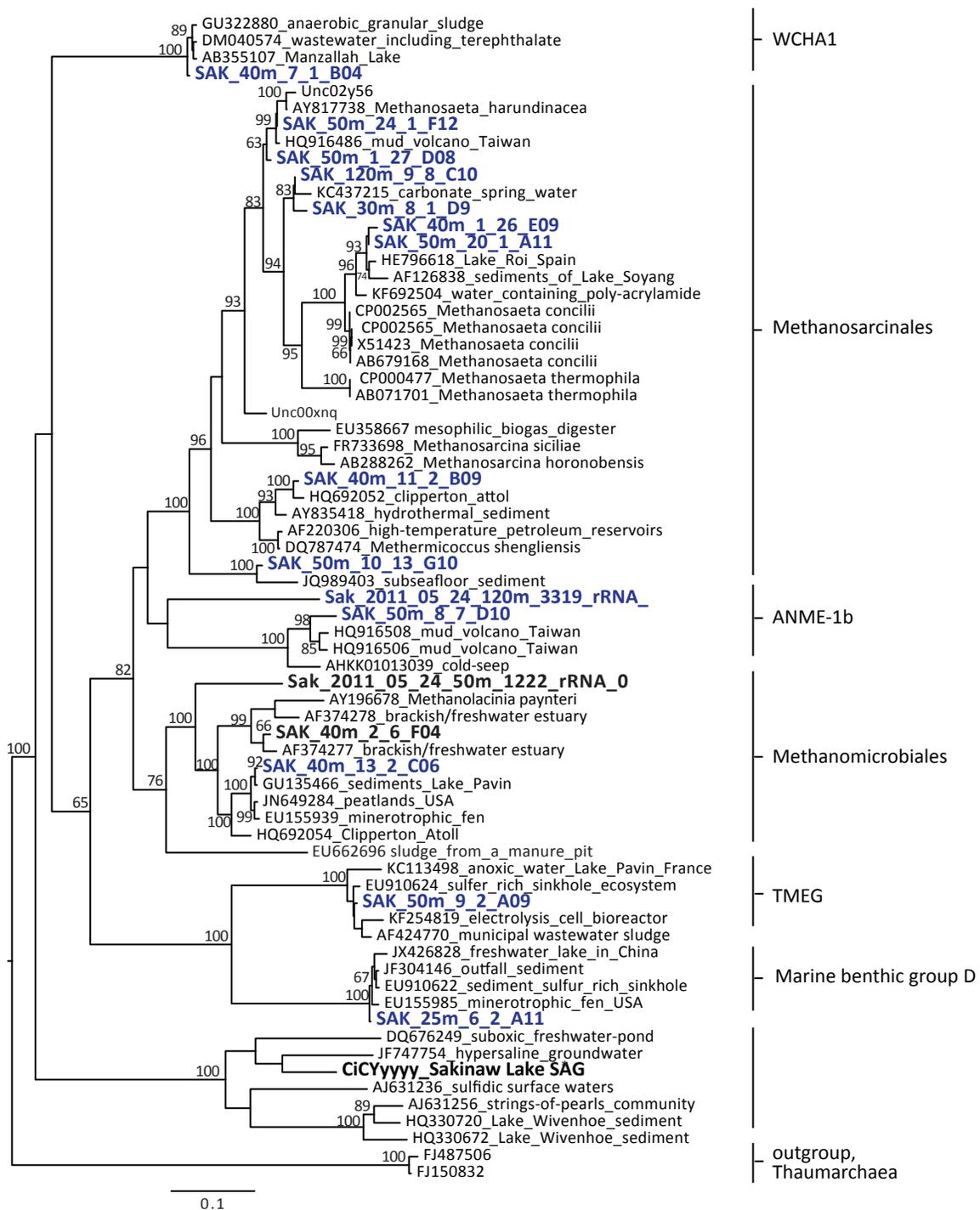


Figure 5.4 Phylogenetic tree of Euryarchaeota in Sakinaw Lake.

legend on following page

Figure 5.4 Phylogenetic tree of Euryarchaeota in Sakinaw Lake.

Phylogenetic tree generated with near full-length 16S rRNA gene sequences from gene clone libraries and full-length 16S rRNA sequences from 454 shotgun sequence assembly. The tree was calculated using maximum likelihood estimation and generalized time reversible (GTR) models implemented in open source software PhyML. The proportion of invariable sites was estimated and assembly of a consensus tree was done from 100 bootstrap replicates. Additionally included is the 16S rRNA gene sequences from the previously published Sakinaw Lake SAG. Sakinaw Lake cluster representatives (clustered at 99% identity) are indicated in bold face. Sequences with *in silico* matches to CARD-FISH probe Arch915 are indicated in blue. Cluster representatives with 2011_05_24 designations were recovered from assembled shotgun sequences.

To follow up on the potential existence of a putative novel methanogenic or methanotrophic archaeal lineage in Sakinaw Lake, a clone library of the methyl coenzyme M reductase subunit A (*mcrA*) marker gene for methanogenesis and methanotrophy was constructed from 50 m sample depth and used for tree making. This depth was chosen based on 16S rRNA information indicating highest water column abundance and diversity of potential methane cycling Archaea. A total of 24 restriction patterns were selected for sequencing. Clustering of sequences at 99% revealed 11 representatives that were used for downstream analysis. The consensus tree was calculated using a database of 191 *mcrA* sequences recovered from cultured and uncultured representatives (Figure 5.4). All 11 *mcrA* genes recovered from Sakinaw Lake fell within the *Methanomicrobiales*, contrasting the phylogenetic distribution of potentially methane cycling Archaea as inferred by 16S rRNA sequencing.

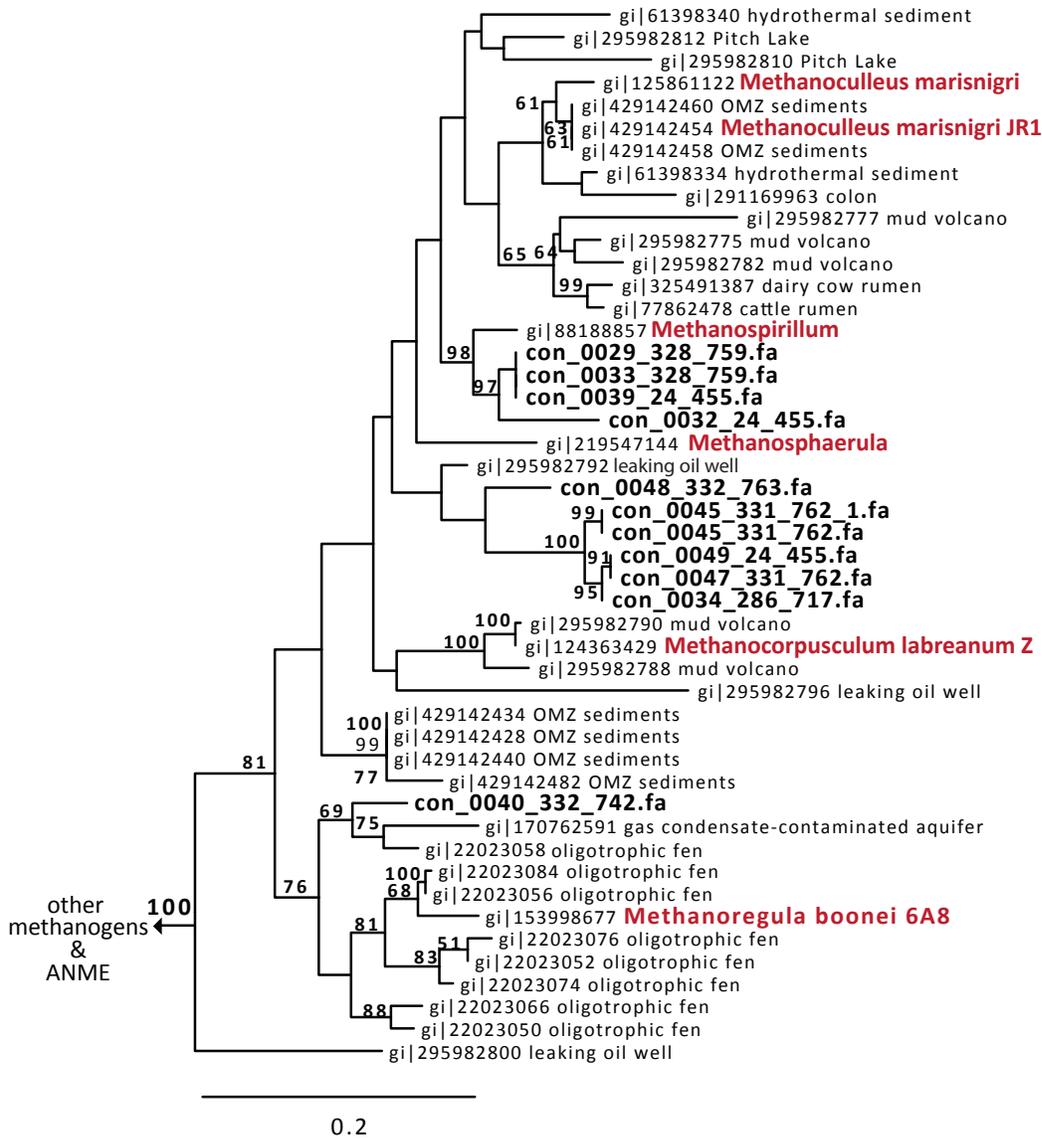


Figure 5.5 Phylogenetic analysis of Sakinaw Lake *mcrA* sequences.

Phylogenetic tree with midpoint root, based on *mcrA* clone library sequences. Positioning of the sequences was inferred with the WAG model of amino acid evolution. The proportion of invariable sites was estimated and assembly of a consensus tree was done from 100 bootstrap replicates. The diversity of *mcrA* sequences is low, and all representatives were affiliated with *Methanomicrobiales*. Sequences with cultured representatives are indicated with red, bold face. Sequences originating from Sakinaw Lake *mcrA* clones are indicated in black bold face.

5.3.3 Assigning the unassigned

To be able to infer the taxonomic affiliation of previously unassigned Archaea (chapter 3 (Gies et al 2014)), 16S rRNA sequences used to construct the phylogenetic trees in Figures 5.2 - 5.4 were used to generate a Sakinaw Lake-specific query database. Although, the 16S rRNA sequences generated with the 20F and 958R primer pair did not overlap with the universal V6-V8 primers used for 454 tag sequencing, full length reference sequences from other environments and near complete 16S rRNA genes recovered from assembled shotgun sequences bridged the two datasets. With the new database, the majority of previously unassigned archaeal reads could be affiliated with a cognate phylum (Figure 5.6 A). The majority (>80%) of the previously unassigned Archaea were affiliated with the *Nanoarchaeota*. Moreover, one previously unassigned and abundant OTU (2.5% relative abundance) generated from 454 tag sequences was affiliated with *Methanomicrobiales*. Another OTU with more than 3% relative abundance in the 454 tag sequences was previously assigned to *Methanosarcinales*, but revealed higher similarity to near complete 16S rRNA gene sequences affiliated with ANME-1b. Additionally, at 120 m 2.5% of previously unassigned OTUs were affiliated with the archaeal SAG CICYyyyy identified in Rinke's study (Rinke et al 2013).

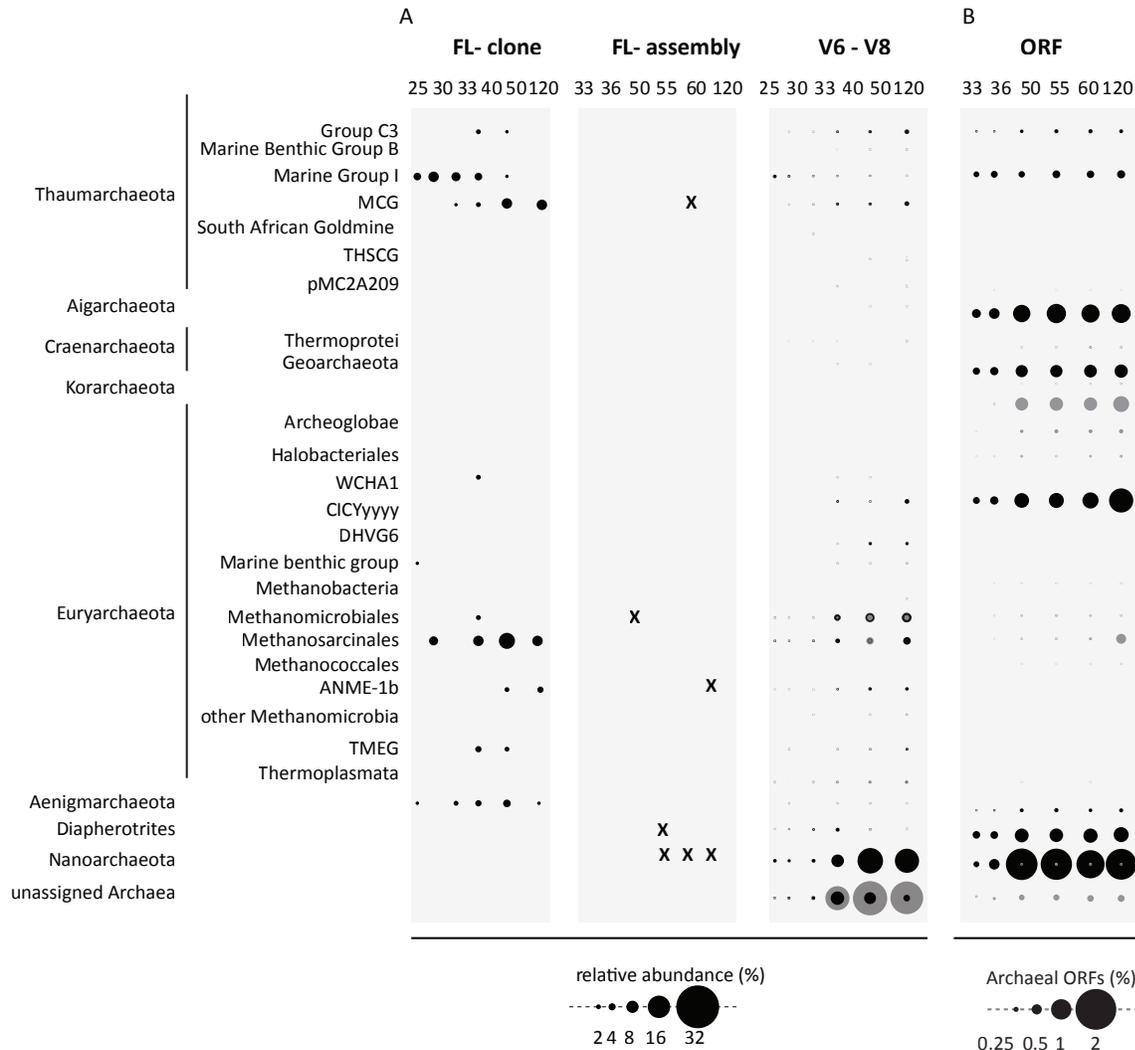


Figure 5.6 Assigning the unassigned.

(A) Using sequences from phylogenetic trees as a query database, previously unassigned 454 tag sequences (V6-V8) could be assigned to specific archaeal phyla. Grey shaded dots represent previously published phylogenetic assignments using Silva v111 as query database. (FL = full length) (B) Phylogenetic anchoring of archaeal open reading frames (ORFs) from a genomic shotgun sequence dataset. ORFs were queried against the NCBI RefSeq database (grey dots) and a custom database that included representatives of the DPANN and TACK superphyla, which are currently underrepresented in NCBI.

5.3.4 Inferring metabolic potential of Archaea in Sainkaw Lake

Archaeal genomic coding potential was determined using 454 shotgun sequencing data generated from samples collected on May 24th 2011 at 6 depths traversing the redox transition zone (RTZ) (33 m, 36 m, 50 m, 55 m) and the monimolimnion (60 m and 120 m). Sequencing yielded an average of 310,636 reads per sample, with an average length of 336 bp. Sequence assembly using the open source software Genovo (Laserson et al 2011), produced an average of 28,428 contigs at 812 bp per sample. Attempts to improve assembly of longer contigs with alternative software (Newbler and Velvet (Zerbino and Birney 2008)) were unsuccessful. Open reading frame prediction produced approximately 36,782 ORFs per sample. In total, 55% of these ORFs could be annotated with product descriptions.

To identify sequences affiliated with novel archaeal phyla, a custom database including protein-coding sequences from 201 recently published single amplified genomes (SAGs) (Rinke et al 2013), was used. Among these 201 MDM SAGs were representatives of eight currently in the NCBI database highly underrepresented archaeal lineages (e.g. *Diapherotrites*, *Parvarchaeota*, *Aenigmarchaeota*, *Nanoarchaeota*, *Nanohaloarchaeota*, *Aigarchaeota*, *Geoarchaeota*, and one undescribed lineage of the *Euryarchaeota*). Given the poor assemblies described above, this database was queried with unassembled reads. A total of 237,061 ORFs were identified and 49% could be annotated with product descriptions. BLAST/LAST output tables from the analysis with the MDM database as well as from the NCBI non-redundant database were searched for ORFs with archaeal origin (Figure 5.6 B). Using the NCBI database, a high number of ORFs could not be identified beyond the domain level (450 ORFs). Moreover, the majority of ORFs could not be classified beyond the phylum level and were assigned to *Euryarchaeota* (2,637 ORFs), *Aigarchaeota* (7 ORFs), *Nanoarchaeota* (171 ORFs),

Korarchaeota (101 ORFs), and *Crenarchaeota* (84 ORFs). Higher-level taxonomic assignments could be achieved for *Halobacteriales* (127 ORFs), *Methanobacteriales* (17), *Methanomicrobiales* (95 ORFs), *Methanosarcinales* (543 ORFs), *Methanococci* (18), *Thermoplasmatales* (6), *Thermococci* (35). Using the in-house generated MDM database, members of the DPANN superphylum were abundantly recovered, represented by *Nanoarchaeota* (7,041 ORFs), *Diapherotrites* (3,354 ORFs) and *Aenigmarchaeota* (452 ORFs). Moreover, representatives of the TACK superphylum could be recovered abundantly as well, and were represented by *Thaumarchaeota* (452 ORFs), the crenarchaeal classes *Geoarchaeota* (3,029 ORFs) and *MGI* (1,713 ORFs), and *Aigarchaeota* (4,551 ORFs). Finally, ORFs affiliated with the Sakinaw Lake euryarchaeon CICYyyyy (4,100 ORFs), were abundantly recovered as well.

To gain insight into the metabolic potential of Archaea in Sakinaw Lake, ORFs with archaeal origing were extracted from the genomic dataset, and analyzed separately using the Metapathways software. Pathway reconstruction yielded 174 complete or near complete (one step missing) pathways. The majority of the energy acquisition pathways was present throughout the RTZ and monimolimnion and represented anaerobic metabolisms, e.g. methanogenesis related pathways (METHANOGENESIS-PWY, METHFORM-PWY, PWY-6613 (tetrahydrofolate salvage), acetate metabolism (PWY0-1313, PWY0-1312, PWY-5535), and hydrogen production (PWY-6744, PWY-6758, PWY-6759, PWY-6780). However, ORFs distinctly associated with the upper part of the RTZ (uRTZ) revealed potential for aerobic respiration (PWY-3781, PWY-5046).

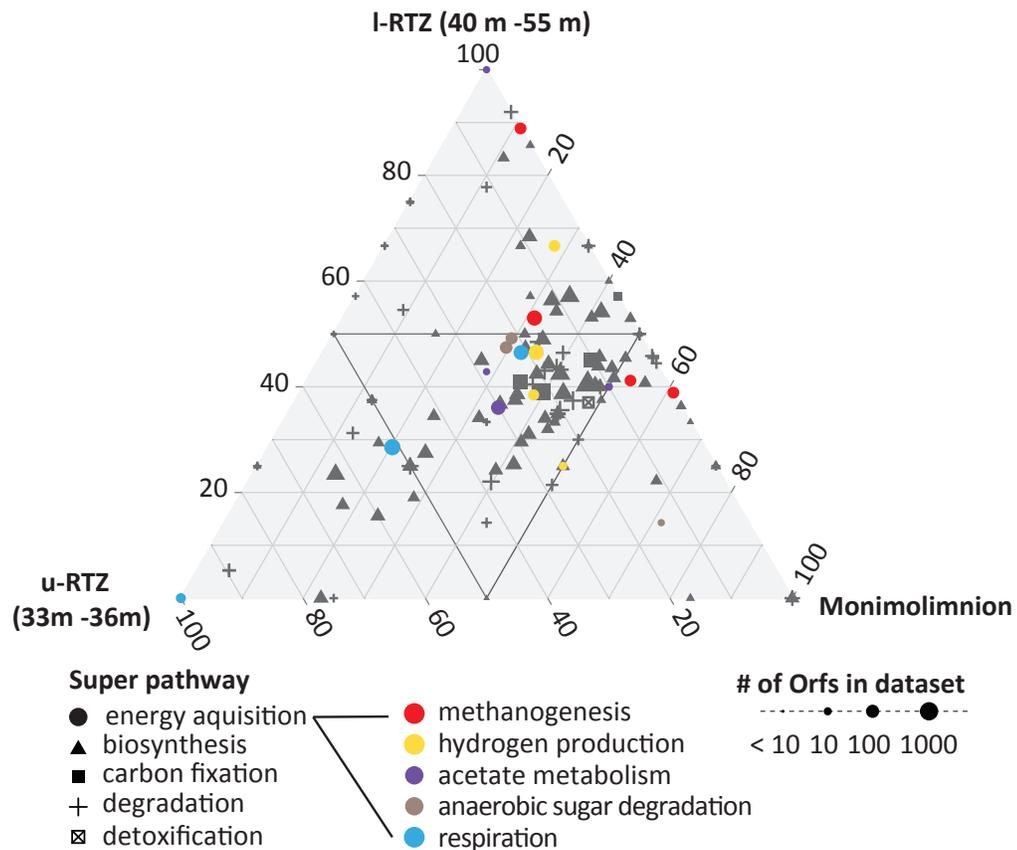


Figure 5.7 Metabolic potential of the archaeal community.

Simplex plot depicting metabolic pathways reconstructed from ORFs of archaeal origin using open source software Metapathways. Pathways included in this plot had complete coverage or one step missing. Size of the data points corresponds to number of ORFs detected in the given pathway..

5.3.5 Single cell genome analysis

To gain insight into the putative function of the novel *Euryarchaeote* class discovered by Rinke and colleagues, the metabolic coding potential of the CICYyyy SAG was analyzed using previously published genomic data on the IMG platform. To be noted, Rinke and colleagues hypothesized a potential role in the methane cycling based on a general screening of metabolic

marker genes (supplementary material (Rinke et al 2013)), but never reconstructed metabolic pathways. General features of the partial euryarchaeal genome are listed in Table 1. The estimated genome size was 1.8 Mbp and the completeness of the genome was predicted at 71%. Average gene length was 850 bp, the coding density was high and the occurrence of gene duplication was estimated as low.

Table 5.2 General genomic features of the Sakinaw Lake SAG.

Genome size (bp)	1,325,681
Genome completeness (%)	71
Estimated size (bp)	1,867,156
GC %	45
% Noncoding	7.5
Gene duplications (%)	3.9
Average gene length, bp	816
<i>RNA genes</i>	
5S rRNA count	0
16S rRNA count	0
23S rRNA count	1
tRNA count	20
Number of CDS	1,504
With function prediction	958
Without function prediction	546
Number of paralog clusters	198
<i># Of proteins sublocalized to</i>	
Cytoplasm	1,012
Cytoplasmic membrane	272
Extracellular milieu	19

It has previously been established that genomic features and the distribution pattern of clusters of orthologous genes (COG) correlates to metabolic capacities (Lauro et al 2009, Youssef et al 2014). To gain insight into a putative trophic lifestyle of CICYyyyy I used principal component analysis (PCA) to compare the putative novel lineage of *Euryarchaeota* with known representatives of the *Euryarchaeota* (*Halobacteriales*, *Archaeoglobus*, *Methanomicrobiales*, *Methanosarcinales*, *Methanopyrales*, *Methanococci*, *Methanocellales*, ANME, *Thermoplasmatales*, *Thanumarchaeota*) and members of the novel superphylum

DPANN (*Diapherotrites*, *Parvarchaeota*, *Nanoarchaeota*, *Nanohaloarchaeota*). Representative complete or near complete genomes were downloaded from the IMG server. The analysis revealed that genomes clustered into three distinct groups: (1) the *Halobacteriales*, (2) Methanogens, *Archaeoglobus*, *Thaumarchaeota* and ANME, and (3) *Parvarchaeota*, *Nanohaloarchaeota*, *Nanoarchaeota* and *Thermoplasmatales*. The Diapherotrites and the Sakinaw Lake CICYyyyy genome did not cluster with any of these groups. The positioning of CICYyyyy as an outlier of the group could be caused by the incompleteness of the SAG (71%). Closest similarity of COG category distribution was recovered for members of group (2) including methanogens.

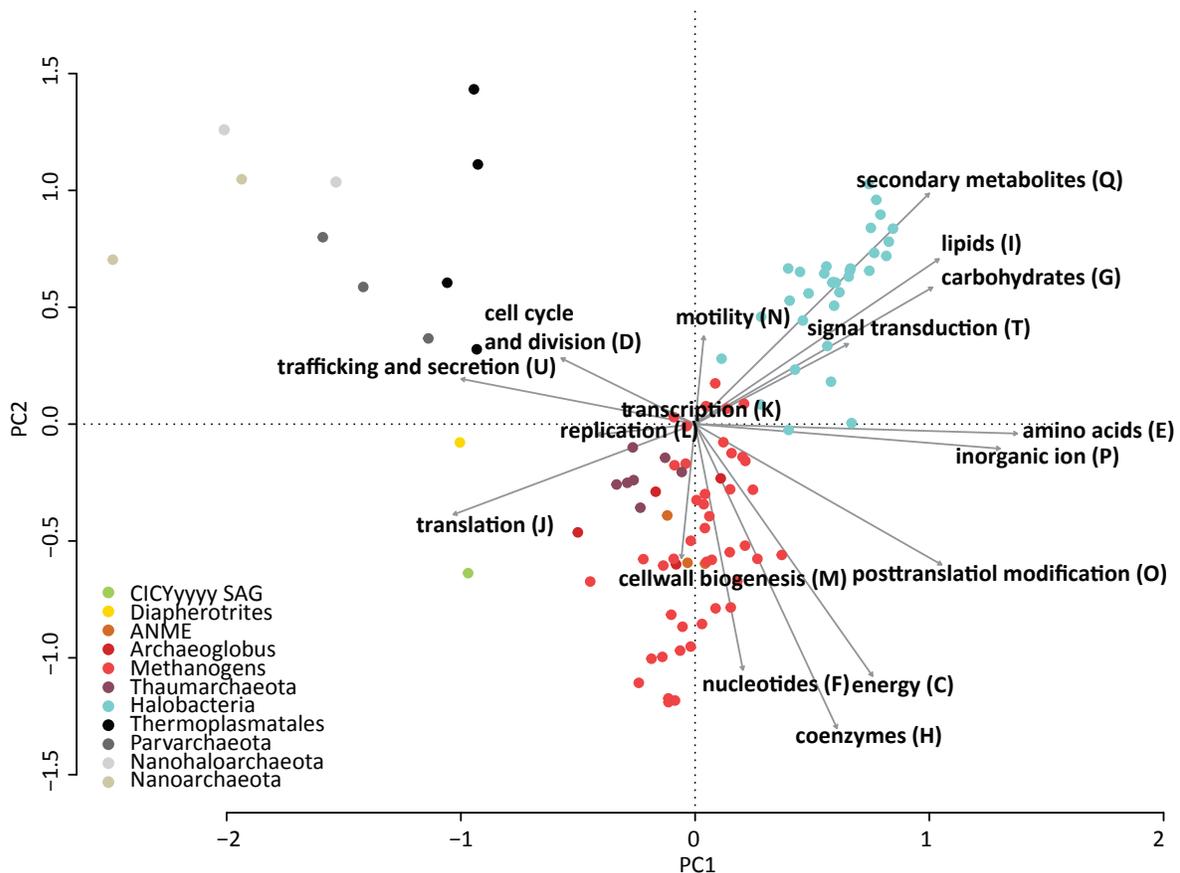


Figure 5.8 PCA analysis of euryarchaeal and DPANN genomes

(legend on following page)

Figure 5.8 PCA analysis of euryarchaeal and DPANN genomes.

PCA biplot of COG category distribution in genomes of DPANN and *Euryarchaeota*. Genomes are represented by colour according to their phylogenetic affiliation. Arrows represent COG categories used for comparison. The arrow direction follows the maximal abundance and their lengths are proportional to the maximal rate of change. The Sakinaw Lake SAG did neither cluster with the Euryarchaeota nor with members of the DPANN. The unique positioning in the biplot was predominantly caused by an expansion in translation related proteins.

To further elucidate a putative potential for methanogenesis in CICYyyyy, I generated a metabolic model using KEGG and Metacyc annotations. Reconstruction of metabolic pathways suggested that Isoprenoid synthesis for the Archaeal cell wall might be catalyzed via the recently described archaeal mevalonate pathway, which was completely covered (Grochowski et al 2006) (Figure 5.9). Inspection of transmembrane protein coding genes revealed, among others, the presence of type IV pilus assembly genes, which suggested that the Archaeon might be mobile. Osmoregulation seems to be achieved via an osmoregulated proline transporter (Na^+ /proline symporter), as well as through the synthesis of the osmolyte Myo-inositol-1-phosphate. The presence of an ammonia transporter gene and coverage of the glutamine synthetase (GS)/glutamate synthase (glutamine:2-oxoglutarate aminotransferase, or GOGAT) suggests nitrogen assimilation from ammonia, regulated by PII nitrogen regulation proteins.

Ribose-5-phosphate generation might be catalyzed by the ribulose monophosphate pathway (RuMP), which was completely covered, and has previously been described as pathway for Ribose-5-phosphate generation in methanogenic Archaea (Grochowski et al 2005). Alternatively, ribose could be activated by ribokinase and then enter the non-oxidative branch of the pentose phosphate cycle. The presence of AMP phosphorylase and type III ribulose-1,5-bisphosphate carboxylase (Rubisco) reveals genomic potential for the recently described

adenosine monophosphate (AMP) salvage pathway (Aono et al. (2012)). The presence of genes that encode for the AMP salvage pathway in combination with the presence of ribokinases could mediate 3-phosphoglycerate generation for serine biosynthesis. Amino acid synthesis was partly covered for eight of the 20 amino acids. The presence of extracellular and intracellular peptidases, and peptide transporters suggests that the remaining amino acids might be generated through breakdown of peptides. The genome shows further evidence for possible utilization of aspartate as a carbon source. An aspartate transaminase gene was recovered, which has the potential to convert aspartate to oxaloacetate (OAA), and coverage for a gene encoding for malate dehydrogenase suggests further the potential to mediate the conversion of OAA to pyruvate for anabolic usage.

The genomic potential points towards the ability to produce acetyl-coA from CO₂ via the Wood-Ljungdahl pathway, based on the presence of genes that encode subunits of the CO dehydrogenase/acetyl-coA synthase pathway, a metabolic feature commonly observed in methanogenic Archaea, acetogens and acetate oxidizers. The first two steps and the second-last step for methanogenesis, from CO₂ were covered. However, methenyltetrahydromethanopterin cyclohydrolase, H₂-forming methenyltetrahydromethanopterin dehydrogenase, F₄₂₀H₂:Heterodisulfide oxidoreductase, and methyl co-enzyme M reductase (*mcr*), were not recovered. Methanogenic Archaea can generate energy using an electron transport chain with or without cytochromes (Thauer et al 2008). Presence of genes that encode for cytochrome c biogenesis suggests that the Archaeon, if it is a methanogen, belongs to the latter type. The pathway for co-enzyme B synthesis, which is a vital co-factor for methanogenesis, revealed to be partly covered. Vitamin B synthesis is catalyzed through a series of reactions in α -ketoacid chain elongation that involves three basic reactions also found in the tri-carboxylic acid cycle: a

condensation between an α -ketodiacid and acetyl-CoA to form citrate, isomerization of the citrate via acetonitrate to isocitrate, and the oxidative decarboxylation of isocitrate to α -ketodiacid (Howell et al 1998). Coenzyme M synthesis was incomplete as well, however it is not uncommon that methanogens rely on an exogenous supply of co-enzyme M (Graham et al 2009). Coverage for coenzyme F390 synthetase gene suggests the presence of the methanogen specific system to sense the reduction and oxidation potential of the cell (Vermeij et al 1996). Most methanogens are sensitive to sulfite, and the coverage of dissimilatory sulfite reductases genes suggests it as a protective mechanism to detoxify the cell (Balderston and Payne 1976). The IMG genome analysis pipeline asserted sulfur reduction potential based on the presence of 2-polyprenylphenol hydroxylase genes that were annotated as sulfide dehydrogenases subunits A and B (SudA and SudB). Inspection of genomes from other methanogenic Archaea deposited on the IMG server revealed that this is not an uncommon feature, however the metabolic relevance in methanogens currently not known.

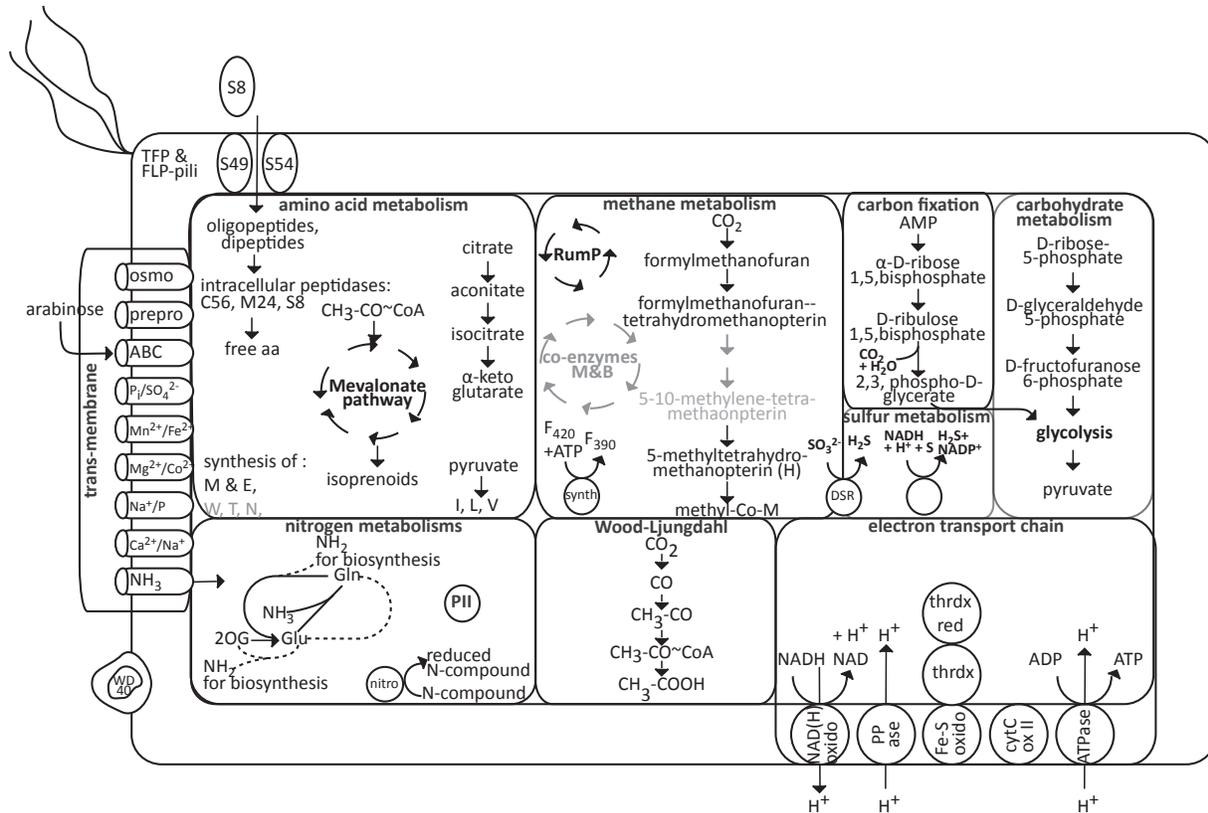


Figure 5.9 Metabolic model for archaeal Sakianw Lake SAG.

Reconstruction of the metabolic potential for Sakianw Lake SAG generated from a 120 m sample. The SAG was generated in a previous study by Rinke and colleagues and identified as belonging to a novel class of Euryarchaeota (Rinke et al 2013). (synth = synthasase, DSR = dissimilatory sulfite reductase, PII = nitrogen regulating protein PII, thrdx red = thioredoxin-disulfide reductase, thrdx= thioredoxin, PPase= pyrophosphatase, NAD(H) oxido = NADH:ubiquinone oxidoreductase, FE-S oxido = Fe-S-cluster oxidoreductase, cytC = cytochrome c, ATPase = Archaeal/vacuolar-type H⁺-ATPase, TFP = type IV pili, FLP= FLP subfamily of type IVb pili, RumP = ribulose monophosphate pathway).

5.4 Discussion

Using a combination of gene clone libraries (16S rRNA and *mcrA*), 454 sequencing (16S rRNA and genomic), CARD-FISH and single cell genomics, led to the discovery of a previously unrecognized archaeal diversity and population structure in meromictic Sakianw Lake. The

combination of these well-established cultivation independent methods captured a view of the archaeal community that could not have been uncovered by using just solely one of these methods. The initial motivation for this comprehensive study was given by the inability to assign taxonomy to abundant Archaea in a previous study that used 454 tag amplified sequences of the V6-V8 region of the 16S rRNA gene for taxonomic analysis (chapter 3 and (Gies et al 2014)). To improve taxonomic classification of unassigned Archaea, a Sakinaw Lake specific archaeal database was generated from near full-length 16S rRNA from clone libraries and assembled genomic sequences. Using this custom made database, previously unassigned 454 reads could be successfully assigned, and Sakinaw Lakes enigmatic Archaea revealed to be affiliated with members of the novel superphylum DPANN, namely the *Nanoarchaeota*, *Aenigmarchaeota* and *Diapherotrites*. In addition to sequencing based analyses, CARD-FISH was used to estimate the population structure and sizes using a previously established universal Archaeal probe (Arch915). The probe sequence was further evaluated *in silico* for its ability to detect novel archaeal groups, and significant biases could be uncovered. Furthermore, to complement taxonomic analysis of the 16S rRNA gene with a method that bypasses primer and or probe biases, genomic sequences were analyzed and ORFs were taxonomically assigned using a LCA and LAST based methods. The results from this analysis were overall consistent with previous results from 16S rRNA gene surveys. However, using genomic sequencing data, *Crenarchaeota*, including members of the novel class *Geoarchaeota*, and *Aigarchaeota* could be recovered as abundant members of the community as well. Finally analysis of genomic sequencing data and one previously published single cell genome gave first insights into the metabolic potential of this unique archaeal community.

CARD-FISH analysis revealed high relative abundance (>16%) of Archaea in the mixolimnion, while 454 tag sequencing using universal three domain primer suggested that Archaea in the mixolimnion belong to the rare biosphere (<1%). Phylogenetic analysis of clone library sequences was consistent with observations made with CARD-FISH, and recovered MG1 as abundant member of the mixolimnion. *In silico* matching of the Arch915 probe with the MG1 OTU was positive, and suggested that abundance estimates with CARD-FISH were true. Archaea affiliated with MG1 have been found in other lakes as well, for example in a survey of Great Lakes of North America, Eurasia and Africa, (Keough et al 2003), and in the oxycline of two meromictic lakes in the high Arctic (Pouliot et al 2009). Members of the MG1 have been shown to be ammonium (NH_4^+) oxidizers (Konneke et al 2005). Ammonium levels in the Sakinaw Lake surface waters are low, however NH_4^+ concentrations in the monimolimnion reach 8 to 10 mM, and upwards diffusion might provide enough NH_4^+ for the metabolic needs of MG1. The poor assembly of the 454 genome sequences and inadequate sequence depth prevented the reconstruction of metabolic pathways for Archaea in Sakinaw Lake, and genes encoding ammonia monooxygenase were therefore not recovered. *Marine Group 1* Archaea were not the only Thaumarchaeota that were under-represented in the 454 tag dataset: diversity and abundance of the MCG recovered from the clone library surpassed estimates from the 454 tag sequence data by far. The metabolic potential of MCG remains elusive, however a study by Biddle and colleagues suggested that they might play a role in the methane cycling of marine sediments (Biddle et al 2006). A potential role of MCG in the methane cycling in Sakinaw Lake would be in agreement with the increasing abundance of MCG affiliates in the methane-saturated monimolimnion.

Abundant members of the novel superphylum DPANN further marked the archaeal community in the monimolimnion. *Aenigmarchaeota*, were recovered in both 16S rRNA sequencing datasets, with fairly equal abundance estimates. The metabolic potential for *Aenigmarchaeota* remains so far completely unknown, but their presence in oxygenated surface waters as well as in the anoxic-sulfidic and CH₄-saturated waters of Sakinaw Lake suggests a diversity of metabolic abilities within the phylum. Archaea specific primers used for clone library generation were biased against *Diapherotrites* and *Nanoarchaeota*, but genomic sequencing data and 454 tag amplicons revealed high relative abundances for both phyla. The metabolic potential of *Diapherotrites* was recently eluded in a SAG study by Youssef and colleagues, which described them as free-living Archaea with the ability to metabolize substrates, such as ribose, polyhydroxybutyrate and several amino acids (Youssef et al 2014). *Nanoarchaeota* are small celled Archaea with strongly reduced genome sizes (~ 500 kb) that thrive as ectosymbionts on other Archaea (e.g. methanogens) (Waters et al 2003).

The relative high proportion of *Nanoarchaeota* (~60%) in the archaeal population of the anoxic-sulfidic methane-saturated waters could explain the discrepancy between CARD-FISH and tag sequencing estimates of the relative archaeal abundance. The CARD-FISH probe Arch915 was shown to hybridize with only one of the three *Nanoarchaeal* species recovered from the dataset, suggesting an underestimation of their presence. Another potential caveat that might have compromised the detection of *Nanoarchaeota* by CARD-FISH is that the microscopic resolution might not have been well enough to distinguish between *Nanoarchaea* and their hosts or that the Archaea might not have been active enough, and hence did not produce enough ribosomes for their detection with CARD-FISH. Incomplete permeabilization of the cell wall could however also have been reason for a biased estimate of the archaeal population using

CARD-FISH. Next to the bias against two of the *Nanoarchaea* lineages other members of the DPANN phylum did not match the probe either. To overcome this bias, one could combine probes that target DPANN with Arch915, for example the recently published probe Darch335 (Kraft et al 2014). The combination of probes is an approach has previously been shown as effective for the universal bacterial probe (Daims et al 1999).

Our analysis further uncovered biases in Archaea specific 16S rRNA primer and *mcrA* primer against *Methanomicrobiales* and *Methanosarcinales*, resulting in a vastly skewed representation of these methanogens in studies using either of the two. This bias could be quite relevant, and might explain why a recent study in the methane saturated waters of meromictic Lake Kivu, did not recover hydrogenotrophic methanogens (e.g. *Methanomicrobiales*), and only low abundances of acetoclastic methanogens (Bhattarai et al 2012). Moreover, genomic sequences retrieved two novel orders of the ANME and *Methanomicrobiales*, which were not recovered in clone libraries. One previously unassigned abundant OTU from the 454 tag dataset revealed to be affiliated with the novel *Methanomicrobiales* species, and one previously as *Methanosarcinales* assigned sequences was closer related to a novel ANME species. Relatives of these new lineages are currently represented by less than 5 sequences in the Silva115 database, suggesting that they belong to an order of the *Euryarchaeota* that might have been overseen in environmental surveys due to primer biases. The relatively low recruitment of metagenomic reads to known methanogens raises more questions with regards to (1) the relative abundance of methanogenic Archaea in the water column of Sakinaw Lake, and (2) how closely they are related to cultured representatives with sequenced genomes. Finally, our analysis recovered affiliates of a novel class of *Euryarchaeota* recently described in a SAG study from Rinke and colleagues as relatively abundant members in the 454 tag dataset (Rinke et al 2013). The

distribution of this novel *Euryarchaeon* was highly similar to the population profile of the ANME. Genomic analysis revealed overlap in pathway coverage typically found in methanogenic Archaea: for example, partial coverage for methanogenesis from CO₂, complete coverage for the Wood Ljungdahl pathway, complete coverage for the RumP pathway, and complete coverage for the AMP salvage pathway. The marker gene for methanogenesis, *mcrA*, was however not recovered. It remains therefore elusive whether this novel *Euryarchaeon* contributes to the methane cycling in Sakinaw Lake.

5.5 Conclusion

Combining 454 tag sequencing, clone library generation, and genomic shotgun sequencing allowed a comprehensive phylogenetic presentation of Archaea in Sakinaw Lake. Moreover, the phylogenetic resolution achieved by using assembled full-length sequences, and partial clone library sequences surpassed the resolution gained by sole analysis of 454 tag sequences, particularly with regards to the *Thaumarchaeota* and the *Euryarchaeota*. This comprehensive study of Archaea provided valuable insight into limitations of well-established methods, and supported results from previous studies that advised to be cautious of primer biases when estimating abundance and diversity. Meromictic lakes have frequently been described as environments that host abundant “unassigned” and “uncultivated” Archaea (Borrel et al 2012, Comeau et al 2012), however our study was the first to tackle primer biases and discrepancies in out-dated databases to give a detailed view on the archaeal community in a permanently stratified lake. Our analyses thereby uncovered an unprecedented diversity of archaeal candidate divisions, which is on par with the recently described abundance and diversity of bacterial candidate divisions (Gies et al 2014). Moreover, our analysis suggests that methanogenic and

methanotrophic members of the community differ from cultured representatives, which makes Sakinaw Lake a promising model-ecosystem for the study of novel microbial species with the potential to fuel biotechnological inventions for methane production.

Chapter 6: Concluding chapter

Microorganisms represent the invisible majority of living things on Earth. Interconnected microbial communities drive matter and energy transformations integral to ecosystem functions and services through distributed metabolic networks innovated over 3.5 billion years of evolution. By studying microbial community structure and function we perceive ecological design principles that can inform numerical models to predict responses to perturbation and to engineer biological systems to solve vexing human issues including wastewater treatment, remediation of contaminated ecosystems, and biofuel generation. However, the vast majority of microbial diversity cannot be brought into pure culture, due in part to unknown metabolic interdependencies, and cultivation-independent methods are needed to evaluate their ecological roles, and discover biotechnological potential. Although next generation sequencing technologies are rapidly transforming our ability to perceive the microcosmos, fundamental questions about the function and phylogeny of uncultivated candidate divisions, so-called MDM remain.

Meromictic lakes are permanently stratified ecosystems that provide relatively stable conditions in which to study MDM. The water column partitions along defined redox gradients into distinct ecological zones. In this thesis I used meromictic Sakinaw Lake as a model ecosystem to gain insight into population structure and metabolic potential of a microbial community that is dominated by archaeal and bacterial candidate divisions. I leveraged a potent combination of CARD-FISH, phylogenetic and functional anchors including 16S rRNA gene and methyl coenzyme m reductase alpha subunit (*mcrA*), metagenomic and single-cell genomic sequencing, network analysis, multivariate statistics and metabolic prediction to illuminate the phylogeny and coding potential of MDM inhabiting the Sakinaw Lake water column.

6.1 Meromictic Sakinaw Lake: remarkable stability of the salinity gradient and remaining questions about the sulfur chemistry

Chapter 2 presented physicochemical data from Sakinaw Lake samples collected between 2007 and 2013, which were compared to data from a previous doctoral study that focused on the S chemistry in the lake. I performed the first multi-year survey of the Sakinaw Lake biogeochemistry, providing valuable insight into the spatiotemporal dynamics of the water column. Salinity, temperature, and O₂ profiles were remarkably stable, sustaining water column stratification over time, and partitioning the microbial community into stable ecological zones or compartments. However, the survey also uncovered significant variations of H₂S, NH₄⁺ and PO₃⁻ concentrations in the RTZ and monimolimnion of the lake. It remains to be determined whether the detected variations are a result of inaccuracy in the methods, or whether there is an unrecognized source of S and nutrients that episodically enter the lake. Karen Perry underlined in her study that the Sakinaw Lake S and P chemistry is rather unusual (Perry 1990), and does not compare to S and P chemistry of meromictic Powell Lake, which is located only a few kilometers north of Sakinaw Lake. She suggested that occasional intrusion of SO₄³⁻ rich seawater might be a plausible reason for the unusual S chemistry. After inspection of a dam that regulates water levels in the lake, it is dubious that seawater has intruded in recent years. If the observed fluctuations of H₂S, PO₃⁻ and NH₄⁺ are indeed valid, the microbial community in the monimolimnion of the lake will experience more variability in the availability of resources than initially anticipated. This could directly affect the community composition and metabolic activity, which could eventually lead to an increased efflux of gases to the surface waters, with resulting feedback on ecosystem functions.

6.2 The Sakinaw Lake microbial community is dominated by bacterial and archaeal candidate divisions

Chapter 3 presented a survey of the microbial community composition and structure using a 454 tag sequencing approach with three-domain resolution. Statistical analyses gave a thorough view on how environmental parameter shape the microbial community, and how microbial community members partition between water column compartments with distinct physicochemical features. Bacterial candidate divisions and Archaea that at the time, could not be taxonomically assigned dominate the microbial community. The abundance and diversity of this MDM was unprecedented, and underlined the potential utility of Sakinaw Lake as a model ecosystem in which to illuminate their phylogeny and genomic coding potential. I further identified statistically significant co-occurrence patterns between abundant indicator species traversing water column compartments. Specific associations between candidate divisions OP9/JS-1, OP8, and WWE1 with methanogenic Archaea in the anoxic-sulfidic and methane-saturated waters of the monimolimnion allowed me to pose hypotheses about their metabolic function. Moreover, for the first time I was able to inform a co-occurrence network with single-cell genomic sequencing data leading to the realization that candidate divisions OP8 and OP9/JS-1 might be syntrophic acetate oxidizers with the capacity to shuttle hydrogen to methanogens affiliated with the *Methanomicrobiales*. Moreover, the metabolic potential of WWE1 alluded to a potential role in providing acetate for OP8 and OP9/JS1 further resolving syntrophic growth modes under anaerobic conditions. Global occurrence of candidate divisions OP8 and OP9/JS-1 in for example marine sediments, freshwater lakes and hotsprings, suggest SAO might be a previously underestimated pathway for methanogenesis. Moreover, similarity between the community composition of Sakinaw Lake and methanogenic bioreactors provided first evidence that

Sakinaw Lake might be a valuable source to inform the set up and operation of engineered biofuel producing ecosystems. In addition to similarities between the Sakinaw Lake microbial community and traditional methanogenic bioreactors, I also identified a striking similarity to a terephthalate degrading methane-producing bioreactor. Terephthalate and other plastic polymers have been identified as concerning environmental pollutants with increasing abundance in aquatic ecosystems. Water from the monimolimnion of Sakinaw Lake could be a potent inoculant for novel bioreactors to optimize treatment of wastewater. Taken together, by combining meta- and single-cell genomic sequence analyses with physicochemical data, complexity theory and statistical methods I provided a framework that will inform experimental design for process oriented studies in Sakinaw Lake and other MDM enriched environments for many years to come.

6.3 Metabolic potential for dissimilatory nitrate reduction, iron and sulfur cycling in RTZ and monimolimnion

Chapter 4 presented a taxonomic, phylogenetic and genomic analysis of the Sakinaw Lake water column using samples collected over a five year period, to identify putative anaerobic respiring populations of MDM and elucidate their population structure and potential metabolic interactions. Candidate division RF3 and subpopulations of OP3, OP9, OP11 and OD1 were identified as distinctly associated with the RTZ, and phylogenetic analysis of full length 16S rRNA genes confirmed the existence of RTZ specific clades for the OD1 and OP3. Inspection of metagenomic and single-cell genomic sequencing data revealed the potential for respiratory NO_3^- reduction in the RTZ associated OP3 population. Moreover, while not detected in 16S rRNA surveys, metagenomic and single-cell genomic sequencing data revealed high abundance of

ORFs that were affiliated with candidate division EM 19, and uncovered potential for S_n^{2-} reduction. In addition, numerous ORFs encoding Fe-S oxidoreductases were found throughout the water column, and were affiliated with rare and abundant candidate divisions, including OP9/JS1, OP8 and WS3. The combined analysis of RTZ associated phyla revealed a detailed perspective on the metabolic and phylogenetic diversity of candidate divisions inhabiting a RTZ and provided a first glimpse at their putative ecological roles. This analysis further underlined that immense power of using single-cell genomics to bin and annotate metagenomic sequencing data and to reveal the population structure of uncultivated microorganisms. The distinct partitioning of the Sakinaw Lake water column and the distinct partitioning of MDM along the watercolumn provides an ideal environment for more target population studies to illuminate metabolic and phylogenetic diversification of MDM.

6.4 The archaeal community

Chapter 5 presented a comprehensive phylogenetic analysis of the archaeal community and gave first insights into the metabolic potential of a novel class of Euryarchaeota. The motivation for this study was drawn from results of chapter 3, which left 80% of the Archaea unassigned. To be able to fully capture phylogenetic diversity of Archaea, and to estimate their relative abundance, I combined CARD-FISH, anchor screening including *mcrA* and 16S rRNA clone and tag sequences with metagenomic and single-cell genomic sequencing. The combined approach resulted in a comprehensive view on the archaeal population and underlined the need to use multiple independent methods when assessing the numerical abundance and population structure of MDM.

With lower costs of high throughput sequencing, and sophisticated software for the analysis of short sequences, phylogenetic analysis using full-length 16S rRNA genes has become less frequent. However, my results reinforce the power of full-length sequences in providing additional more robust phylogenetic trees with high bootstrap support that in turn improved the accuracy of diversity estimates. Moreover, using 16S rRNA genes from assembled shotgun sequences overcame observed primer biases, and recovered sequences affiliated with MCG, *ANME*, *Methanomicrobiales*, *Nanoarchaeaota*, and *Diapherotrites*. The analysis suggested that in public databases under represented OTUs affiliated with *ANME* and *Methanomicrobiales* might be common inhabitants of meromictic lake ecosystems. Moreover, previously unassigned 454 tag sequences were identified using a database of phylogenetically anchored sequences recovered from the Sakinaw Lake water column. Lastly, I was able to infer abundance and coding potential of a novel class of Euryarchaeota, which thrives in the monimolimnion of Sakinaw Lake and has previously been identified but not investigated in a study by Rinke and colleagues (Rinke et al 2013). Analysis of one near complete SAG pointed towards a potential role in methane cycling.

Taken together, my analyses uncovered a diverse archaeal population that is dominated by candidate phyla of the superphylum DPANN. To date little is known about the ecological role of most members of the DPANN, and previously discussed biases in cultivation independent methods led to a patchy view on their global distribution, necessitating improved efforts and establishment of validated standard methods to capture archaeal diversity more accurately. Furthermore, this study reinforces the observation that Sakinaw Lake has a unique population of CH₄ cycling microorganisms, and provides an untapped reservoir of microbial metabolic potential to generate sustainable energy and materials from biomass.

6.5 Future directions

To improve understanding of the biogeochemistry and microbial ecology of Sakinaw Lake, it will be essential to gain detailed insight into the hydrology of the lake, and it will be required to determine all sources of inflowing water. Additionally, it will be required to assess potential perturbations through residents and recreational use of the lake. Sakinaw Lake is a popular site for recreational fishing, and the Department of Fisheries and Oceans regularly stocks its waters with juvenile fish. Occasionally, stocking is badly timed, leading to mass die-offs (personal conversation with lake residents), which provides additional sources of C, N and S for microbial processes in the monimolimnion of the lake. Moreover, multiple vacation homes mark the shoreline, and residents often set up “custom made” sewage systems, with the potential to impact lake processes as well.

My research uncovered unusually high and variable H_2S concentrations in the monimolimnion of the lake, but also predicted metabolic potential for S_n^{2-} reduction to H_2S for candidate division EM19, uncovering a previously unrecognized contribution to S transformation. It is imperative to follow up on S_n^{2-} in Sakinaw Lake. Moreover, Sakinaw Lake sediments have not been studied, and sedimentary SO_4^{2-} and H_2S profiles, as well as a survey of the microbial community structure might give valuable insights into sedimentary contributions to potential changes in H_2S concentrations of the water column during the evolution of the lake ecosystem. Furthermore, it will be of high interest to see whether MDM in the sediments is as abundant and diverse as in the water column.

Multiple process oriented approaches will be required to test the hypothesis for SAO driven by associations between *Methanomicrobiales* and candidate divisions OP8 and OP9/JS-1, and to infer the rate of methane production. NanoSIMS would give direct insight into putative

uptake of acetate by OP9/JS-1 and OP8, and could further give insight into potential consortia forming with methanogenic Archaea. For this, CARD-FISH probes for candidate divisions OP8 and OP9/JS1 will be needed to identify target cells for NanoSIMS measurements. In a collaborative effort with the Baltic Sea Institute Warnemünde, Germany I have already developed and tested probes for the Sakinaw Lake population of OP9/JS-1, and preliminary tests for NanoSIMS measurements have been done in collaboration with the Pacific Northwest National Laboratories (PNNL), Washington (see Appendix A). Incubation studies with [2-¹⁴C]-acetate can further provide insight into the quantitative contribution of SAO to overall rates of methanogenesis. Methane produced via SAO, results in the accumulation of ¹³CH₄ and ¹³CO₂, while aceticlastic methanogenesis only produces ¹³CH₄. This difference in acetate breakdown allows to distinguish between the two pathways in incubation studies with [2-¹⁴C]-acetate (Nusslein et al 2001). Incubation experiments could further be used to confirm a potential role for the novel Euryarchaeote in the CH₄ cycle of Sakinaw Lake. As most methanogens and methanotrophs assimilate some of the substrate that they use for their energetic needs, RNA or DNA-stable isotope probing using ¹³C labeled substrate (acetate, bicarbonate and CH₄) might be a promising approach. Moreover, given the absence of SO₄²⁻ and the presence of transition metals Fe and Mn in the monimolimnion of the lake, it would be imperative to test whether the Sakinaw Lake population of ANME uses either of the two metals as terminal electron acceptor during the anaerobic oxidation of CH₄.

To date little is known about N cycling in meromictic lakes, and the discovery of diverse and abundant members of the Thaumarchaeota suggests that NH₄⁺ oxidation might be driven by Archaea. Genome analyses, quantifications for NH₄⁺ oxidation and nitrification will be required to better understand the ecological role of Thaumarchaeota in Sakinaw Lake. Further, to

validate a potential role for Thaumarchaeota as well as candidate division OP3 in the N-cycling of Sakinaw Lake, NanoSIMS might again be a useful tool. Previous studies have already developed CARD-FISH probes for OP3 (Rotaru et al 2012), however *in silico* matching with Sakinaw Lake sequences revealed that they were not compatible with the Sakinaw Lake population, and it will be required to design and test a new OP3 CARD-FISH probe. Population genomics will further be needed to elucidate differences in the metabolic potential and ecological function for OP3 members that thrive in the RTZ and monimolimnion.

Furthermore, metagenomic and single-cell genomic sequencing with greater coverage depth will be needed to identify the metabolic function of RF3. Expression studies would also be invaluable to validate predicted metabolic pathways and infer metabolic linkages among and between MDM populations in different water column compartments. Assembly of shotgun data in this study were too poor to recover RF3 contigs, and current databases lack reference genomes for RF3, which made phylogenetic anchoring of metagenomic reads impossible. The distinct restriction of RF3 to the RTZ of the lake in combination with their relative high abundance (>6%) provides a promising environment to recover complete genomes using high throughput platforms (e.g. Illumina). Given the close similarity to the population profile of green sulfur Bacteria affiliated with the *Chlorobiales* it might be worthwhile to perform an exhaustive study of Bchl types in the RTZ. Bacteriochlorophyll *e* could already be recovered from the RTZ, and concentrations were in agreement with inferred relative abundances of *Chlorobiales* using tag-sequencing data. However, anoxygenic phototrophs affiliated with the *Heliobacter* use Bchl *g* to harvest light. The RF3 and *Heliobacter* belong both to the *Firmicutes*, and it might therefore be possible that RF3 use Bchl *g* as well, that is if RF3 perform phototrophy.

6.6 Closing remarks

In this thesis, I presented a comprehensive view of the microbial community in meromictic Sakinaw Lake, which was dominated by bacterial and archaeal candidate divisions. Phylogenetic and genomic analyses revealed highly promising metabolic potential in many of these novel lineages indicating the ability to produce biofuel and breakdown pollutants. Fuel and clean water are fundamental for the functioning of our society, and my work might be instrumental for improvements in the design and operation of engineered ecosystems that produce biofuel and treat wastewater. Next to the economic potential, I uncovered a remarkable biodiversity and provided insights into metabolic potential integral for the ecosystem function. Given the widespread distribution of many candidate divisions, these findings are of global importance. The unique biogeochemical setting in the anoxic monimolimnion suggests that the interconnected microbial community might be similar to ancestral microbes that thrived during early Earth history, when the atmosphere was devoid of oxygen and Fe and S were abundant. Overall, my work in Sakinaw Lake will be instrumental for future endeavors to shed a light on the genomic and economic potential, as well as on the evolution of microbial dark matter.

Bibliography

- Airs RL, Atkinson JE, Keely BJ (2001). Development and application of a high resolution liquid chromatographic method for the analysis of complex pigment distributions. *J Chromatogr A* **917**: 167-177.
- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat Biotechnol* **31**: 533-+.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *J Mol Biol* **215**: 403-410.
- Amann RI, Ludwig W, Schleifer KH (1995). Phylogenetic Identification and in-Situ Detection of Individual Microbial-Cells without Cultivation. *Microbiol Rev* **59**: 143-169.
- American Public Health Association (APHA) AWWAA, Water Environment Federation (WEF) (1999). *Standard Methods for the Examination of Water and Wastewater*. American Public Health Association: Washington.
- Anderson GC (1958). Some Limnological Features of a Shallow Saline Meromictic Lake. *Limnol Oceanogr* **3**: 259-270.
- Anderson MJ, Ellingsen KE, McArdle BH (2006). Multivariate dispersion as a measure of beta diversity. *Ecology Letters* **9**: 683-693.
- Andrei AS, Banciu HL, Oren A (2012). Living with salt: metabolic and phylogenetic diversity of archaea inhabiting saline ecosystems. *FEMS Microbiology Letters* **330**: 1-9.
- Baker BJ, Comolli LR, Dick GJ, Hauser LJ, Hyatt D, Dill BD *et al* (2010). Enigmatic, ultrasmall, uncultivated Archaea. *P Natl Acad Sci USA* **107**: 8806-8811.
- Balderston WL, Payne WJ (1976). Inhibition of methanogenesis in salt marsh sediments and whole-cell suspensions of methanogenic bacteria by nitrogen oxides. *Appl Environ Microbiol* **32**: 264-269.
- Barns SM, Delwiche CF, Palmer JD, Pace NR (1996). Perspectives on archaeal diversity, thermophily and monophyly from environmental rRNA sequences. *P Natl Acad Sci USA* **93**: 9188-9193.
- Beal EJ, House CH, Orphan VJ (2009). Manganese- and iron-dependent marine methane oxidation. *Science* **325**: 184-187.

- Berg JS, Schwedt A, Kreutzmann AC, Kuypers MM, Milucka J (2014). Polysulfides as intermediates in the oxidation of sulfide to sulfate by *Beggiatoa* spp. *Appl Environ Microbiol* **80**: 629-636.
- Bhattarai S, Ross KA, Schmid M, Anselmetti FS, Burgmann H (2012). Local conditions structure unique archaeal communities in the anoxic sediments of meromictic Lake Kivu. *Microb Ecol* **64**: 291-310.
- Biddle JF, Lipp JS, Lever MA, Lloyd KG, Sorensen KB, Anderson R *et al* (2006). Heterotrophic Archaea dominate sedimentary subsurface ecosystems off Peru. *P Natl Acad Sci USA* **103**: 3846-3851.
- Bidre-Petit C, Boucher D, Kuever J, Alberic P, Jézéquel D, Chebance B *et al* (2011a). Identification of Sulfur-Cycle Prokaryotes in a Low-Sulfate Lake (Lake Pavin) Using *aprA* and 16S rRNA Gene Markers. *Microb Ecol* **61**: 313-327.
- Bidre-Petit C, Jezequel D, Dugat-Bony E, Lopes F, Kuever J, Borrel G *et al* (2011b). Identification of microbial communities involved in the methane cycle of a freshwater meromictic lake. *FEMS Microbiology Ecology* **77**: 533-545.
- Binga EK, Lasken RS, Neufeld JD (2008). Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *Isme J* **2**: 233-241.
- Birney E, Clamp M, Durbin R (2004). GeneWise and Genomewise. *Genome Research* **14**: 988-995.
- Blainey PC, Mosier AC, Potanina A, Francis CA, Quake SR (2011). Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS One* **6**: e16626.
- Boehrer B, Schultze M (2008). Stratification of lakes. *Rev Geophys* **46**: -.
- Bonhomme C, Poulin M, Vincon-Leite B, Saad M, Groleau A, Jezequel D *et al* (2011). Maintaining meromixis in Lake Pavin (Auvergne, France): The key role of a sublacustrine spring. *Cr Geosci* **343**: 749-759.
- Borrel G, Lehours AC, Bardot C, Bailly X, Fonty G (2010). Members of candidate divisions OP11, OD1 and SR1 are widespread along the water column of the meromictic Lake Pavin (France). *Arch Microbiol* **192**: 559-567.
- Borrel G, Lehours AC, Crouzet O, Jezequel D, Rockne K, Kulczak A *et al* (2012). Stratification of Archaea in the deep sediments of a freshwater meromictic lake: vertical shift from methanogenic to uncultured archaeal lineages. *PloS one* **7**: e43346.

- Bosshard PP, Santini Y, Gruter D, Stettler R, Bachofen R (2000). Bacterial diversity and community composition in the chemocline of the meromictic alpine Lake Cadagno as revealed by 16S rDNA analysis. *FEMS Microbiology Ecology* **31**: 173-182.
- Camacho A, Erez J, Chicote A, Florin M, Squires MM, Lehmann C *et al* (2001). Microbial microstratification, inorganic carbon photoassimilation and dark carbon fixation at the chemocline of the meromictic Lake Cadagno (Switzerland) and its relevance to the food web. *Aquat Sci* **63**: 91-106.
- Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK *et al* (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-336.
- Caspi R, Foerster H, Fulcher CA, Kaipa P, Krummenacker M, Latendresse M *et al* (2008). The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **36**: D623-D631.
- Caspi R, Altman T, Billington R, Dreher K, Foerster H, Fulcher CA *et al* (2014). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res* **42**: D459-471.
- Cavaleiro AJ, Sousa DZ, Alves MM (2010). Methane production from oleate: assessing the bioaugmentation potential of *Syntrophomonas zehnderi*. *Water Res* **44**: 4940-4947.
- Cho JC, Vergin KL, Morris RM, Giovannoni SJ (2004). *Lentisphaera araneosa* gen. nov., sp. nov., a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, Lentisphaerae. *Environmental Microbiology* **6**: 611-621.
- Chomczynski P, Sacchi N (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162**: 156-159.
- Clague JJ, James TS (2002). History and isostatic effects of the last ice sheet in southern British Columbia. *Quaternary Sci Rev* **21**: 71-87.
- Cline JD (1968). Spectrophotometric Determination of Hydrogensulfide in Natural Waters. *Limnology and Oceanography* **14**: 454-458.
- Comeau AM, Harding T, Galand PE, Vincent WF, Lovejoy C (2012a). Vertical distribution of microbial communities in a perennially stratified Arctic lake with saline, anoxic bottom waters. *Sci Rep* **2**: 604.
- Conrad R (1999). Contribution of hydrogen to methane production and control of hydrogen concentrations in methanogenic soils and sediments. *FEMS Microbiology Ecology* **28**: 193-202.

Coolen MJL, Muyzer, G. , Schouten, S., Volkman, J.K., Sinninghe Damste J.S. (2006). Sulfur and Methane Cycling during the Holocene in Ace Lake (Antarctica) revealed by Lipid And Stratigraphy. *Past and present Water Coloumn Anoxia* **64**: 41-65.

Cordero OX, Ventouras LA, DeLong EF, Polz MF (2012). Public good dynamics drive evolution of iron acquisition strategies in natural bacterioplankton populations. *Proc Natl Acad Sci U S A* **109**: 20059-20064.

Costa KC, Navarro JB, Shock EL, Zhang CL, Soukup D, Hedlund BP (2009). Microbiology and geochemistry of great boiling and mud hot springs in the United States Great Basin. *Extremophiles* **13**: 447-459.

Crowe SA, Katsev S, Leslie K, Sturm A, Magen C, Nomosatryo S *et al* (2011). The methane cycle in ferruginous Lake Matano. *Geobiology* **9**: 61-78.

Daims H, Bruhl A, Amann R, Schleifer KH, Wagner M (1999). The domain-specific probe EUB338 is insufficient for the detection of all Bacteria: development and evaluation of a more comprehensive probe set. *Syst Appl Microbiol* **22**: 434-444.

De Caceres M, Legendre P, Moretti M (2010). Improving indicator species analysis by combining groups of sites. *Oikos* **119**: 1674-1684.

DeLong EF, Preston CM, Mincer T, Rich V, Hallam SJ, Frigaard NU *et al* (2006). Community genomics among stratified microbial assemblages in the ocean's interior. *Science* **311**: 496-503.

Derakshani M, Lukow T, Liesack W (2001). Novel bacterial lineages at the (sub)division level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms. *Appl Environ Microbiol* **67**: 623-631.

DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K *et al* (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microb* **72**: 5069-5072.

Dodsworth JA, Blainey PC, Murugapiran SK, Swingley WD, Ross CA, Tringe SG *et al* (2013a). Single-cell and metagenomic analyses indicate a fermentative and saccharolytic lifestyle for members of the OP9 lineage. *Nat Commun* **4**: 1854.

Dodsworth JA, Nobu MK, Murugapiran SK, Rinke C, Gies EA, Webster G *et al* (2015). Phylogeny and physiology of candidate phylum 'Atribacteria' (OP9/JS1) inferred from cultivation-independent genomics. *Isme J* **accepted**.

Dojka MA, Hugenholtz P, Haack SK, Pace NR (1998). Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl Environ Microb* **64**: 3869-3877.

- Drews G (2000). The roots of microbiology and the influence of Ferdinand Cohn on microbiology of the 19th century. *FEMS Microbiol Rev* **24**: 225-249.
- Dufrene M, Legendre P (1997). Species assemblages and indicator species: The need for a flexible asymmetrical approach. *Ecol Monogr* **67**: 345-366.
- Dupont CL, Rusch DB, Yooseph S, Lombardo MJ, Richter RA, Valas R *et al* (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *Isme J* **6**: 1186-1199.
- Edgar RC (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797.
- Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.
- Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H *et al* (2010). Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *Isme J* **4**: 642-647.
- Fenchel T. FBJ (1996). Ecology and Evolution in Anoxic Worlds. *Journal of Evolutionary Biology* **9**: 259-260.
- Ferris K-TCaDH (1996). Martinus Willem Beijerinck (1851-1931): Pioneer of general microbiology. *ASM News (Washington, DC: American Society For Microbiology)* **62**: 539-543.
- Field EK, Sczyrba A, Lyman AE, Harris CC, Woyke T, Stepanauskas R *et al* (2014). Genomic insights into the uncultivated marine Zetaproteobacteria at Loihi Seamount. *Isme J*.
- Findenegg I (1935). Limnologische Untersuchungen im Kärntner Seengebiet. Ein Beitrag zur Kenntnis des Sauerstoffhaushalts in Alpenseen. *Internationale Revenue der gesamten Hydrobiologie* **32**: 369-423.
- Flies CB, Jonkers HM, de Beer D, Bosselmann K, Bottcher ME, Schuler D (2005). Diversity and vertical distribution of magnetotactic bacteria along chemical gradients in freshwater microcosms. *FEMS Microbiology Ecology* **52**: 185-195.
- Fujii M, Kojima H, Iwata T, Urabe J, Fukui M (2012). Dissolved organic carbon as major environmental factor affecting bacterioplankton communities in mountain lakes of eastern Japan. *Microb Ecol* **63**: 496-508.
- Gies EA, Konwar KM, Beatty JT, Hallam SJ (2014a). Illuminating microbial dark matter in meromictic sakinaw lake. *Appl Environ Microb* **80**: 6807-6818.

- Gihring TM, Green SJ, Schadt CW (2012). Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes. *Environmental Microbiology* **14**: 285-290.
- Gittel A, Sorensen KB, Skovhus TL, Ingvorsen K, Schramm A (2009). Prokaryotic community structure and sulfate reducer activity in water from high-temperature oil reservoirs with and without nitrate treatment. *Appl Environ Microb* **75**: 7086-7096.
- Glöckner J, Kube M, Shrestha PM, Weber M, Glöckner FO, Reinhardt R *et al* (2010). Phylogenetic diversity and metagenomics of candidate division OP3. *Environmental Microbiology* **12**: 1218-1229.
- Gogarten JP, Doolittle WF, Lawrence JG (2002). Prokaryotic evolution in light of gene transfer. *Mol Biol Evol* **19**: 2226-2238.
- Graham DE, Taylor SM, Wolf RZ, Namboori SC (2009). Convergent evolution of coenzyme M biosynthesis in the Methanosarcinales: cysteate synthase evolved from an ancestral threonine synthase. *Biochem J* **424**: 467-478.
- Gregersen LH, Habicht KS, Peduzzi S, Tonolla M, Canfield DE, Miller M *et al* (2009). Dominance of a clonal green sulfur bacterial population in a stratified lake. *Fems Microbiology Ecology* **70**: 30-41.
- Grochowski LL, Xu HM, White RH (2005). Ribose-5-phosphate biosynthesis in *Methanocaldococcus jannaschii* occurs in the absence of a pentose-phosphate pathway. *J Bacteriol* **187**: 7382-7389.
- Grochowski LL, Xu H, White RH (2006). *Methanocaldococcus jannaschii* uses a modified mevalonate pathway for biosynthesis of isopentenyl diphosphate. *J Bacteriol* **188**: 3192-3198.
- Grossart HP, Frindte K, Dziallas C, Eckert W, Tang KW (2011). Microbial methane production in oxygenated water column of an oligotrophic lake. *Proceedings of the National Academy of Sciences of the United States of America* **108**: 19657-19661.
- Gugliandolo C, Lentini V, Maugeri TL (2011). Distribution and Diversity of Bacteria in a Saline Meromictic Lake as Determined by PCR-DGGE of 16S rRNA Gene Fragments. *Curr Microbiol* **62**: 159-166.
- Guler C, Thyne GD, McCray JE, Turner AK (2002). Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. *Hydrogeol J* **10**: 455-474.
- Guy L, Ettema TJ (2011). The archaeal 'TACK' superphylum and the origin of eukaryotes. *Trends Microbiol* **19**: 580-587.

Habicht KS, Miller M, Cox RP, Frigaard NU, Tonolla M, Peduzzi S *et al* (2011). Comparative proteomics and activity of a green sulfur bacterium through the water column of Lake Cadagno, Switzerland. *Environmental Microbiology* **13**: 203-215.

Hakala A (2004). Meromixis as a part of lake evolution - observations and a revised classification of true meromictic lakes in Finland. *Boreal Environ Res* **9**: 37-53.

Hales BA, Edwards C, Ritchie DA, Hall G, Pickup RW, Saunders JR (1996). Isolation and identification of methanogen-specific DNA from blanket bog peat by PCR amplification and sequence analysis. *Appl Environ Microb* **62**: 668-675.

Hallam SJ, Girguis PR, Preston CM, Richardson PM, DeLong EF (2003). Identification of methyl coenzyme M reductase A (*mcrA*) genes associated with methane-oxidizing archaea. *Appl Environ Microb* **69**: 5483-5491.

Hallam SJ, Putnam N, Preston CM, Detter JC, Rokhsar D, Richardson PM *et al* (2004). Reverse methanogenesis: testing the hypothesis with environmental genomics. *Science* **305**: 1457-1462.

Halm H, Musat N, Lam P, Langlois R, Musat F, Peduzzi S *et al* (2009). Co-occurrence of denitrification and nitrogen fixation in a meromictic lake, Lake Cadagno (Switzerland). *Environmental Microbiology* **11**: 1945-1958.

Hamilton TL, Bovee RJ, Thiel V, Sattin SR, Mohr W, Schaperdoth I *et al* (2014). Coupled reductive and oxidative sulfur cycling in the phototrophic plate of a meromictic lake. *Geobiology* **12**: 451-468.

Handelsman J, Rondon MR, Brady SF, Clardy J, Goodman RM (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chem Biol* **5**: R245-249.

Handelsman J (2004). Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**: 669-685.

Harris JK, Kelley ST, Pace NR (2004). New perspective on uncultured bacterial phylogenetic division OP11. *Appl Environ Microb* **70**: 845-849.

Haruta S, Kato S, Yamamoto K, Igarashi Y (2009). Intertwined interspecies relationships: approaches to untangle the microbial network. *Environ Microbiol* **11**: 2963-2969.

Hatamoto M, Imachi H, Yashiro Y, Ohashi A, Harada H (2007). Diversity of anaerobic microorganisms involved in long-chain fatty acid degradation in methanogenic sludges as revealed by RNA-based stable isotope probing. *Appl Environ Microb* **73**: 4119-4127.

Hedlund BP, Dodsworth JA, Murugapiran SK, Rinke C, Woyke T (2014). Impact of single-cell genomics and metagenomics on the emerging view of extremophile "microbial dark matter". *Extremophiles* **18**: 865-875.

Henrici AT (1933). Studies of Freshwater Bacteria: I. A Direct Microscopic Technique. *J Bacteriol* **25**: 277-287.

Hongve D (2002). Seasonal mixing and genesis of endogenic meromixis in small lakes in Southeast Norway. *Nord Hydrol* **33**: 189-206.

Howell DM, Harich K, Xu HM, White RH (1998). alpha-keto acid chain elongation reactions involved in the biosynthesis of coenzyme B (7-mercaptoheptanoyl threonine phosphate) in methanogenic archaea. *Biochemistry* **37**: 10108-10117.

Huber H, Hohn MJ, Rachel R, Fuchs T, Wimmer VC, Stetter KO (2002). A new phylum of Archaea represented by a nanosized hyperthermophilic symbiont. *Nature* **417**: 63-67.

Hug LA, Castelle CJ, Wrighton KC, Thomas BC, Sharon I, Frischkorn KR *et al* (2013). Community genomic analyses constrain the distribution of metabolic traits across the Chloroflexi phylum and indicate roles in sediment carbon cycling. *Microbiome* **1**: 22.

Hugenholtz P, Pitulle C, Hershberger KL, Pace NR (1998). Novel division level bacterial diversity in a Yellowstone hot spring. *J Bacteriol* **180**: 366-376.

Humayoun SB, Bano N, Hollibaugh JT (2003). Depth distribution of microbial diversity in Mono Lake, a meromictic soda lake in California. *Appl Environ Microb* **69**: 1030-1042.

Hutchinson GE (1957). A Treatise on Limnology: Geography, physics, and chemistry. *pt. I. Geography and physics of lakes*. Wiley.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ (2010). Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* **11**: 119.

IOS IOC (2010). The international thermodynamic equation of seawater - 2010: Calculation and use of thermodynamic properties **Manuals and Guides 56**.

Jannasch HW, Wirsén CO, Molyneux SJ, Langworthy TA (1992). Comparative Physiological Studies on Hyperthermophilic Archaea Isolated from Deep-Sea Hot Vents with Emphasis on Pyrococcus Strain GB-D. *Appl Environ Microb* **58**: 3472-3481.

Jansson M (1993). Uptake, exchange, and excretion of orthophosphate in phosphate-starved *Scenedesmus quadricauda* and *Pseudomonas K7*. *American Society of Limnology and Oceanography* **38**: 1162-1178.

- Jorgensen BB, Kuenen JG, Cohen Y (1979). Microbial Transformations of Sulfur-Compounds in a Stratified Lake (Solar Lake, Sinai). *Limnology and Oceanography* **24**: 799-822.
- Junier P, Kim OS, Eckert W, Casper P, Imhoff JF, Witzel KP *et al* (2010). Methane- and ammonia-oxidizing bacteria at the chemocline of Lake Kinneret (Israel). *Aquat Microb Ecol* **58**: 241-248.
- Kallistova A, Kevbrina MV, Pimenov NV, Rusanov, II, Rogozin D, Wehrli B *et al* (2006). [Sulfate reduction and methanogenesis in the Shira and Shunet meromictic lakes (Khakass Republic, Russia)]. *Mikrobiologiya* **75**: 828-835.
- Kanehisa M (2002). The KEGG database. *Silico Simulation of Biological Processes* **247**: 91-103.
- Karakashev D, Batstone DJ, Trably E, Angelidaki I (2006). Acetate oxidation is the dominant methanogenic pathway from acetate in the absence of Methanosaetaceae. *Appl Environ Microb* **72**: 5138-5141.
- Karl DM, Beversdorf L, Bjorkman KM, Church MJ, Martinez A, Delong EF (2008). Aerobic production of methane in the sea. *Nature Geosci* **1**: 473-478.
- Karner MB, DeLong EF, Karl DM (2001). Archaeal dominance in the mesopelagic zone of the Pacific Ocean. *Nature* **409**: 507-510.
- Keeling PJ, Palmer JD (2008). Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet* **9**: 605-618.
- Kent AD, Jones SE, Yannarell AC, Graham JM, Lauster GH, Kratz TK *et al* (2004). Annual patterns in bacterioplankton community variability in a humic lake. *Microb Ecol* **48**: 550-560.
- Kent AD, Jones SE, Lauster GH, Graham JM, Newton RJ, McMahon KD (2006). Experimental manipulations of microbial food web interactions in a humic lake: shifting biological drivers of bacterial community structure. *Environmental Microbiology* **8**: 1448-1459.
- Kent AD, Yannarell AC, Rusak JA, Triplett EW, McMahon KD (2007). Synchrony in aquatic microbial community dynamics. *Isme J* **1**: 38-47.
- Keough BP, Schmidt TM, Hicks RE (2003). Archaeal nucleic acids in picoplankton from great lakes on three continents. *Microb Ecol* **46**: 238-248.
- Kielbasa SM, Wan R, Sato K, Horton P, Frith MC (2011). Adaptive seeds tame genomic sequence comparison. *Genome Research* **21**: 487-493.
- Klepac-Ceraj V, Hayes CA, Gilhooly WP, Lyons TW, Kolter R, Pearson A (2012). Microbial diversity under extreme euxinia: Mahoney Lake, Canada. *Geobiology* **10**: 223-235.

- Knittel K, Boetius A (2009). Anaerobic oxidation of methane: progress with an unknown process. *Annu Rev Microbiol* **63**: 311-334.
- Kojima H, Nakajima T, Fukui M (2007). Carbon source utilization and accumulation of respiration-related substances by freshwater Thioploca species. *FEMS Microbiol Ecol* **59**: 23-31.
- Kolinko S, Jogler C, Katzmann E, Wanner G, Peplies J, Schuler D (2012). Single-cell analysis reveals a novel uncultivated magnetotactic bacterium within the candidate division OP3. *Environmental Microbiology* **14**: 1709-1721.
- Konneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437**: 543-546.
- Konwar KM, Hanson NW, Page AP, Hallam SJ (2013). MetaPathways: a modular pipeline for constructing pathway/genome databases from environmental sequence information. *BMC Bioinformatics* **14**: 202.
- Kozubal MA, Romine M, Jennings R, Jay ZJ, Tringe SG, Rusch DB *et al* (2013). Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *Isme J* **7**: 622-634.
- Kraft B, Tegetmeyer HE, Meier D, Geelhoed JS, Strous M (2014). Rapid succession of uncultured marine bacterial and archaeal populations in a denitrifying continuous culture. *Environmental Microbiology* **16**: 3275-3286.
- Kunin V, Engelbrekton A, Ochman H, Hugenholtz P (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environmental Microbiology* **12**: 118-123.
- Lanigan C (1989). Lake Nyos disaster. *BMJ* **299**: 183.
- Lanzen A, Simachew A, Gessesse A, Chmolowska D, Jonassen I, Ovreas L (2013). Surprising Prokaryotic and Eukaryotic Diversity, Community Structure and Biogeography of Ethiopian Soda Lakes. *PloS one* **8**.
- Laserson J, Jojic V, Koller D (2011a). Genovo: De Novo Assembly for Metagenomes. *J Comput Biol* **18**: 429-443.
- Lasken RS, Stockwell TB (2007). Mechanism of chimera formation during the Multiple Displacement Amplification reaction. *BMC Biotechnol* **7**: 19.
- Lauro FM, McDougald D, Thomas T, Williams TJ, Egan S, Rice S *et al* (2009). The genomic basis of trophic strategy in marine bacteria. *Proc Natl Acad Sci U S A* **106**: 15527-15533.

- Lauro FM, DeMaere MZ, Yau S, Brown MV, Ng C, Wilkins D *et al* (2011). An integrative study of a meromictic lake ecosystem in Antarctica. *Isme Journal* **5**: 879-895.
- Laverman AM, Blum JS, Schaefer JK, Phillips E, Lovley DR, Oremland RS (1995). Growth of Strain SES-3 with Arsenate and Other Diverse Electron Acceptors. *Appl Environ Microbiol* **61**: 3556-3561.
- Lehours AC, Bardot C, Thenot A, Debroas D, Fonty G (2005). Anaerobic microbial communities in Lake Pavin, a unique meromictic lake in France. *Appl Environ Microb* **71**: 7389-7400.
- Lehours AC, Evans P, Bardot C, Joblin K, Gerard F (2007). Phylogenetic diversity of archaea and bacteria in the anoxic zone of a meromictic lake (Lake Pavin, France). *Appl Environ Microbiol* **73**: 2016-2019.
- Leplae R, Lima-Mendez G, Toussaint A (2010). ACLAME: a CLAssification of Mobile genetic Elements, update 2010. *Nucleic Acids Res* **38**: D57-61.
- Li L, Kato C, Horikoshi K (1999). Bacterial diversity in deep-sea sediments from different depths. *Biodiversity and Conservation* **8**: 659-677.
- Li W, Godzik A (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658-1659.
- LI Wan-chun YX-d, YIN Yu, JI Jiang, LI Shi-jie, PU Pei-min (2005). Thermal Stratification in lake Zige Tagco, Central Tibetan Plateau. *Wuhan University Journal of Natural Sciences* **10**: 689-693.
- Liu Y, Beer LL, Whitman WB (2012). Methanogens: a window into ancient sulfur metabolism. *Trends Microbiol* **20**: 251-258.
- Lovley DR, Holmes DE, Nevin KP (2004). Dissimilatory Fe(III) and Mn(IV) reduction. *Adv Microb Physiol* **49**: 219-286.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar *et al* (2004). ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363-1371.
- Lykidis A, Chen CL, Tringe SG, McHardy AC, Copeland A, Kyrpides NC *et al* (2011). Multiple syntrophic interactions in a terephthalate-degrading methanogenic consortium. *Isme J* **5**: 122-130.
- Macintyre S, Melack JM (1982). Meromixis in an Equatorial African Soda Lake. *Limnol Oceanogr* **27**: 595-609.

Magadza CHD (1979). Physical and Chemical Limnology of 6 Hydroelectric Lakes on the Waikato River, 1970-72. *New Zeal J Mar Fresh* **13**: 561-572.

Mallet C, Charpin MF, Devaux J (1998). Nitrate reductase activity of phytoplankton populations in eutrophic Lake Aydat and meso-oligotrophic Lake Pavin: a comparison. *Hydrobiologia* **373-374**: 135-148.

Manefield M, Whiteley AS, Griffiths RI, Bailey MJ (2002). RNA stable isotope probing, a novel means of linking microbial community function to Phylogeny. *Appl Environ Microb* **68**: 5367-5373.

Manske AK, Glaeser J, Kuypers MAM, Overmann J (2005). Physiology and phylogeny of green sulfur bacteria forming a monospecific phototrophic assemblage at a depth of 100 meters in the Black Sea. *Appl Environ Microb* **71**: 8049-8060.

Marcy Y, Ishoey T, Lasken RS, Stockwell TB, Walenz BP, Halpern AL *et al* (2007a). Nanoliter reactors improve multiple displacement amplification of genomes from single cells. *PLoS Genet* **3**: 1702-1708.

Marcy Y, Ouverney C, Bik EM, Losekann T, Ivanova N, Martin HG *et al* (2007b). Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. *P Natl Acad Sci USA* **104**: 11889-11894.

Mc Cune BaG, J. B. (2002). *Analysis of Ecological Communities*. MjM Software Design: Oregon 97388.

Metcalf WW, Griffin BM, Cicchillo RM, Gao J, Janga SC, Cooke HA *et al* (2012). Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. *Science* **337**: 1104-1107.

Meyer KM, Macalady JL, Fulton JM, Kump LR, Schaperdoth I, Freeman KH (2011). Carotenoid biomarkers as an imperfect reflection of the anoxygenic phototrophic community in meromictic Fayetteville Green Lake. *Geobiology* **9**: 321-329.

Mizoguchi T, Yoshitomi T, Harada J, Tamiaki H (2011). Temperature- and time-dependent changes in the structure and composition of glycolipids during the growth of the green sulfur photosynthetic bacterium *Chlorobaculum tepidum*. *Biochemistry* **50**: 4504-4512.

Moreira S, Boehrer B, Schultze M, Dietz S, Samper J (2011). Modeling Geochemically Caused Permanent Stratification in Lake Waldsee (Germany). *Aquat Geochem* **17**: 265-280.

Morris JJ, Lenski RE, Zinser ER (2012). The Black Queen Hypothesis: Evolution of Dependencies through Adaptive Gene Loss. *Mbio* **3**.

- Muller B, Sun L, Schnurer A (2013). First insights into the syntrophic acetate-oxidizing bacteria-a genetic study. *Microbiologyopen* **2**: 35-53.
- Musat N, Halm H, Winterholler B, Hoppe P, Peduzzi S, Hillion F *et al* (2008). A single-cell view on the ecophysiology of anaerobic phototrophic bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **105**: 17861-17866.
- Muyzer G, Stams AJ (2008). The ecology and biotechnology of sulphate-reducing bacteria. *Nature Reviews Microbiology* **6**: 441-454.
- Nancy YY, Wagner JR, Laird MR, Melli G, Rey Sb, Lo R *et al* (2010). PSORTb 3.0: improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* **26**: 1608-1615.
- Natacha Pasche CD, Beat Müller, Martin Schmid, Alfred Wüest, Bernhard Wehrli, (2009). Physical and biogeochemical limits to internal nutrient loading of meromictic Lake Kivu. *Limnology and Oceanography* **54**: 1863-1873.
- Newton RJ, Jones SE, Eiler A, McMahon KD, Bertilsson S (2011). A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiol Mol Biol R* **75**: 14-49.
- Northcote TG, Johnson WE (1964). Occurrence and Distribution of Seawater in Sakinaw Lake, British Columbia *Canadian Journal of Fisheries and Aquatic Sciences* **21**: 1321-1324.
- Nunoura T, Takaki Y, Kakuta J, Nishi S, Sugahara J, Kazama H *et al* (2011). Insights into the evolution of Archaea and eukaryotic protein modifier systems revealed by the genome of a novel archaeal group. *Nucleic Acids Res* **39**: 3204-3223.
- Nusslein B, Chin KJ, Eckert W, Conrad R (2001). Evidence for anaerobic syntrophic acetate oxidation during methane production in the profundal sediment of subtropical Lake Kinneret (Israel). *Environmental Microbiology* **3**: 460-470.
- O'Day PA, Vlassopoulos D, Root R, Rivera N (2004). The influence of sulfur and iron on dissolved arsenic concentrations in the shallow subsurface under changing redox conditions. *Proc Natl Acad Sci U S A* **101**: 13703-13708.
- Olsen GJ, Pace NR, Nuell M, Kaine BP, Gupta R, Woese CR (1985). Sequence of the 16S rRNA gene from the thermoacidophilic archaeobacterium *Sulfolobus solfataricus* and its evolutionary implications. *J Mol Evol* **22**: 301-307.
- Oremland JRLaRS (2006). Microbial transformations of arsenic in the environment; from soda lakes to aquifers (in Arsenic). *Elements* **2**: 85-90.
- Oremland RS, Stolz JF (2003). The ecology of arsenic. *Science* **300**: 939-944.

Orphan VJ, House CH, Hinrichs KU, McKeegan KD, DeLong EF (2001). Methane-consuming archaea revealed by directly coupled isotopic and phylogenetic analysis. *Science* **293**: 484-487.

Overmann J (1997). Mahoney Lake: A case study of the ecological significance of phototrophic sulfur bacteria. *Advances in Microbial Ecology* **15**: 251-288.

Overmann J (2008). Ecology of Phototrophic Sulfur Bacteria. In: Hall RD, C.; Knaff, D.B.; Leustek, Th. (ed). *Advances in Photosynthesis and Respiration*. Springer. p 516.

Overmann Jr, Thomas Beatty J, Hall KJ (1994). Photosynthetic activity and population dynamics of *Amoebobacter purpureus* in a meromictic saline lake. *Fems Microbiology Ecology* **15**: 309-319.

Pace N, Stahl D, Lane D, Olsen G (1986). The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences. In: Marshall KC (ed). *Advances in Microbial Ecology*. Springer US. pp 1-55.

Pace NR (1995). Opening the door onto the natural microbial world: molecular microbial ecology. *Harvey Lect* **91**: 59-78.

Parkin TB, Winfrey, M. R. , Brock, T. D. (1980). The physical and chemical limnology of a Wisconsin meromictic lake. *Transactions of the Wisconsin Academy of Sciences, Arts and Letters* **68**: 111-125.

Pasche N, Dinkel C, Muller B, Schmid M, Wuest A, Wehrli B (2009). Physical and biogeochemical limits to internal nutrient loading of meromictic Lake Kivu. *Limnology and Oceanography* **54**: 1863-1873.

Pasche N, Schmid M, Vazquez F, Schubert CJ, Wuest A, Kessler JD *et al* (2011). Methane sources and sinks in Lake Kivu. *J Geophys Res-Bioge* **116**.

Paul K, Nonoh JO, Mikulski L, Brune A (2012). "Methanoplasmatales," Thermoplasmatales-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl Environ Microb* **78**: 8245-8253.

Peduzzi S, Tonolla M, Hahn D (2003). Vertical distribution of sulfate-reducing bacteria in the chemocline of Lake Cadagno, Switzerland, over an annual cycle. *Aquat Microb Ecol* **30**: 295-302.

Pelletier E, Kreimeyer A, Bocs S, Rouy Z, Gyapay G, Chouari R *et al* (2008). "Candidatus Cloacamonas acidaminovorans": genome sequence reconstruction provides a first glimpse of a new bacterial division. *J Bacteriol* **190**: 2572-2579.

Pernthaler A, Pernthaler J, Amann R (2002). Fluorescence in situ hybridization and catalyzed reporter deposition for the identification of marine bacteria. *Appl Environ Microb* **68**: 3094-3101.

Perry KA (1990). The Chemical Limnology of two meromictic Lakes with Emphasis on Pyrite Formation. *PhD Thesis, submitted at the University of British Columbia.*

Pett-Ridge J, Weber PK (2012). NanoSIP: NanoSIMS applications for microbial biology. *Methods Mol Biol* **881**: 375-408.

Peura S, Eiler A, Bertilsson S, Nykanen H, Tiirola M, Jones RI (2012). Distinct and diverse anaerobic bacterial communities in boreal lakes dominated by candidate division OD1. *Isme J* **6**: 1640-1652.

Plugge CM, Zhang W, Scholten JC, Stams AJ (2011). Metabolic flexibility of sulfate-reducing bacteria. *Front Microbiol* **2**: 81.

Pouliot J, Galand PE, Lovejoy C, Vincent WF (2009). Vertical structure of archaeal communities and the distribution of ammonia monooxygenase A gene variants in two meromictic High Arctic lakes. *Environmental Microbiology* **11**: 687-699.

Prokopkin IG, Barkhatov YV, Khromechek EB (2014). A one-dimensional model for phytoflagellate distribution in the meromictic lake. *Ecological Modelling* **288**: 1-8.

Pronk JT, de Bruyn JC, Bos P, Kuenen JG (1992). Anaerobic Growth of Thiobacillus ferrooxidans. *Appl Environ Microb* **58**: 2227-2230.

Qiu YL, Muramatsu M, Hanada S, Kamagata Y, Guo RB, Sekiguchi Y (2013). *Oligosphaera ethanolica* gen. nov., sp. nov., an anaerobic, carbohydrate-fermenting bacterium isolated from methanogenic sludge, and description of *Oligosphaeria classis* nov. in the phylum Lentisphaerae. *International Journal of Systematic and Evolutionary Microbiology* **63**: 533-539.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P *et al* (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590-596.

Radajewski S, Ineson P, Parekh NR, Murrell JC (2000). Stable-isotope probing as a tool in microbial ecology. *Nature* **403**: 646-649.

Ramirez-Moreno S, Martinez-Alonso M, Mendez-Alvarez S, Gaju N (2005). Seasonal microbial ribotype shifts in the sulfurous karstic lakes Ciso and Vilar, in northeastern Spain. *Int Microbiol* **8**: 235-242.

Rappe MS, Giovannoni SJ (2003). The uncultured microbial majority. *Annu Rev Microbiol* **57**: 369-394.

Raskin L, Stromley JM, Rittmann BE, Stahl DA (1994). Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. *Appl Environ Microbiol* **60**: 1232-1240.

Rawlings ND, Morton FR, Barrett AJ (2006). MEROPS: the peptidase database. *Nucleic Acids Res* **34**: D270-D272.

Richter K, Schicklberger M, Gescher J (2012). Dissimilatory reduction of extracellular electron acceptors in anaerobic respiration. *Appl Environ Microb* **78**: 913-921.

Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng JF *et al* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**: 431-437.

Riviere D, Desvignes V, Pelletier E, Chaussonnerie S, Guermazi S, Weissenbach J *et al* (2009). Towards the definition of a core of microorganisms involved in anaerobic digestion of sludge. *Isme J* **3**: 700-714.

Rogozin DY, Zykov VV, Chernetsky MY, Degermendzhy AG, Gulati RD (2009). Effect of winter conditions on distributions of anoxic phototrophic bacteria in two meromictic lakes in Siberia, Russia. *Aquat Ecol* **43**: 661-672.

Rohini Kumar M, Saravanan VS (2010). Candidate OP Phyla: Importance, Ecology and Cultivation Prospects. *Indian J Microbiol* **50**: 474-477.

Rotaru AE, Schauer R, Probian C, Musmann M, Harder J (2012). Visualization of candidate division OP3 cocci in limonene-degrading methanogenic cultures. *J Microbiol Biotechnol* **22**: 457-461.

Rybak M, Dickman M (1988). Paleoecological Reconstruction of Changes in the Productivity of a Small, Meromictic Lake in Southern Ontario, Canada. *Hydrobiologia* **169**: 293-306.

Saier MH, Tran CV, Barabote RD (2006). TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucleic Acids Res* **34**: D181-D186.

Salcher MM (2013). Same same but different: ecological niche partitioning of planktonic freshwater prokaryotes. *Journal of Limnology* **73**.

Schink B (1997). Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol Mol Biol R* **61**: 262-280.

Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB *et al* (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl Environ Microbiol* **75**: 7537-7541.

Seyler P, Martin JM (1989). Biogeochemical processes affecting arsenic species distribution in a permanently stratified lake. *Environmental Science & Technology* **23**: 1258-1263.

Shade A, Jones SE, McMahon KD (2008). The influence of habitat heterogeneity on freshwater bacterial community composition and dynamics. *Environmental Microbiology* **10**: 1057-1067.

- Shade A, Read JS, Welkie DG, Kratz TK, Wu CH, McMahon KD (2011). Resistance, resilience and recovery: aquatic bacterial dynamics after water column disturbance. *Environmental Microbiology* **13**: 2752-2767.
- Shade A, Read JS, Youngblut ND, Fierer N, Knight R, Kratz TK *et al* (2012). Lake microbial communities are resilient after a whole-ecosystem disturbance. *Isme J*.
- Shade A, Jones SE, Caporaso JG, Handelsman J, Knight R, Fierer N *et al* (2014). Conditionally rare taxa disproportionately contribute to temporal changes in microbial diversity. *MBio* **5**.
- Sieber JR, McInerney MJ, Gunsalus RP (2012). Genomic Insights into Syntrophy: The Paradigm for Anaerobic Metabolic Cooperation. *Annu Rev Microbiol* **66**: 429-452.
- Sorensen J, Jorgensen BB, Revsbech NP (1979). A comparison of oxygen, nitrate, and sulfate respiration in coastal marine sediments. *Microb Ecol* **5**: 105-115.
- Stahl D. A. AR (1991). Development and application of nucleic acid probes. . In: Stackebrandt E. GM (ed). *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons, Inc.: New York. pp 205–248.
- Stams AJM, Plugge CM (2009). Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nature Reviews Microbiology* **7**: 568-577.
- Stark M, Berger SA, Stamatakis A, von Mering C (2010). MLTreeMap--accurate Maximum Likelihood placement of environmental DNA sequences into taxonomic and functional reference phylogenies. *BMC Genomics* **11**: 461.
- Stein JL, Marsh TL, Wu KY, Shizuya H, DeLong EF (1996). Characterization of uncultivated prokaryotes: isolation and analysis of a 40-kilobase-pair genome fragment from a planktonic marine archaeon. *J Bacteriol* **178**: 591-599.
- Strachan CR, Singh R, VanInsberghe D, Ievdokymenko K, Budwill K, Mohn WW *et al* (2014). Metagenomic scaffolds enable combinatorial lignin transformation. *Proc Natl Acad Sci U S A* **111**: 10143-10148.
- Sundh I (1992). Biochemical composition of dissolved organic carbon derived from phytoplankton and used by heterotrophic bacteria. *Appl Environ Microb* **58**: 2938-2947.
- Suzuki I, Takeuchi TL, Yuthasastrakosol TD, Oh JK (1990). Ferrous Iron and Sulfur Oxidation and Ferric Iron Reduction Activities of Thiobacillus ferrooxidans Are Affected by Growth on Ferrous Iron, Sulfur, or a Sulfide Ore. *Appl Environ Microb* **56**: 1620-1626.
- Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D *et al* (2011). Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. *Science* **333**: 1296-1300.

Tang YQ, Shigematsu T, Morimura S, Kida K (2007). Effect of dilution rate on the microbial structure of a mesophilic butyrate-degrading methanogenic community during continuous cultivation. *Appl Microbiol Biotechnol* **75**: 451-465.

Tang YQ, Ji P, Hayashi J, Koike Y, Wu XL, Kida K (2011). Characteristic microbial community of a dry thermophilic methanogenic digester: its long-term stability and change with feeding. *Appl Microbiol Biotechnol* **91**: 1447-1461.

Tedersoo L, Nilsson RH, Abarenkov K, Jairus T, Sadam A, Saar I *et al* (2010). 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytol* **188**: 291-301.

Teske A, Sorensen KB (2008). Uncultured archaea in deep marine subsurface sediments: have we caught them all? *Isme J* **2**: 3-18.

Thauer RK, Kaster AK, Seedorf H, Buckel W, Hedderich R (2008). Methanogenic archaea: ecologically relevant differences in energy conservation. *Nature Reviews Microbiology* **6**: 579-591.

Tonolla M, Peduzzi S, Demarta A, Peduzzi R, Hahn aD (2004). Phototropic sulfur and sulfate-reducing bacteria in the chemocline of meromictic Lake Cadagno, Switzerland. *Journal of Limnology* **63**: 161-170.

Tonolla M, Peduzzi R, Hahn D (2005). Long-term population dynamics of phototrophic sulfur bacteria in the chemocline of Lake Cadagno, Switzerland. *Appl Environ Microb* **71**: 3544-3550.

Tourna M, Stieglmeier M, Spang A, Konneke M, Schintlmeister A, Urich T *et al* (2011). Nitrososphaera viennensis, an ammonia oxidizing archaeon from soil. *P Natl Acad Sci USA* **108**: 8420-8425.

Tripathi BM, Kim M, Lai-Hoe A, Shukor NA, Rahim RA, Go R *et al* (2013). pH dominates variation in tropical soil archaeal diversity and community structure. *FEMS Microbiology Ecology* **86**: 303-311.

Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Sci Transl Med* **1**: 6ra14.

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM *et al* (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428**: 37-43.

Vagle S, Hume J, McLaughlin F, MacIsaac E, Shortreed K (2010). A methane bubble curtain in meromictic Sakinaw Lake, British Columbia. *Limnology and Oceanography* **55**: 1313-1326.

Valero-Garcés BL, Kelts KR (1995). A sedimentary facies model for perennial and meromictic saline lakes: Holocene Medicine Lake Basin, South Dakota, USA. *J Paleolimnol* **14**: 123-149.

van Gemerden H, Mas J (1995). Ecology of Phototrophic Sulfur Bacteria. In: R. E. Blankenship MTMaCEB (ed). *Anoxygenic Photosynthetic Bacteria*. Kluwer Academic Publishers. pp 49–85.

van Passel MW, Kant R, Palva A, Lucas S, Copeland A, Lapidus A *et al* (2011). Genome sequence of *Victivallis vadensis* ATCC BAA-548, an anaerobic bacterium from the phylum Lentisphaerae, isolated from the human gastrointestinal tract. *J Bacteriol* **193**: 2373-2374.

Vermeij P, van der Steen RJ, Keltjens JT, Vogels GD, Leisinger T (1996). Coenzyme F390 synthetase from *Methanobacterium thermoautotrophicum* Marburg belongs to the superfamily of adenylate-forming enzymes. *J Bacteriol* **178**: 505-510.

Vila-Costa M, Barberan A, Auguet JC, Sharma S, Moran MA, Casamayor EO (2013). Bacterial and archaeal community structure in the surface microlayer of high mountain lakes examined under two atmospheric aerosol loading scenarios. *FEMS Microbiology Ecology* **84**: 387-397.

Vincent AC, Mueller DR, Vincent WF (2008). Simulated heat storage in a perennially ice-covered high Arctic lake: Sensitivity to climate change. *J Geophys Res-Oceans* **113**.

Wacey D, Saunders M, Brasier MD, Kilburn MR (2011). Earliest microbially mediated pyrite oxidation in ~3.4 billion-year-old sediments. *Earth and Planetary Science Letters* **301**: 393-402.

Walker KF, Likens GE (1975). Meromixis and a reconsidered typology of lake circulation patterns. *Verh Internat Ver Limnol* **19**: 442-458.

Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M *et al* (2003). The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism. *P Natl Acad Sci USA* **100**: 12984-12988.

Webster G, Watt LC, Rinna J, Fry JC, Evershed RP, Parkes RJ *et al* (2006). A comparison of stable-isotope probing of DNA and phospholipid fatty acids to study prokaryotic functional diversity in sulfate-reducing marine sediment enrichment slurries. *Environmental Microbiology* **8**: 1575-1589.

Wetzel RG (2001). *Limnology Lake and River Ecosystem*, Third Edition edn. Elsevier Academic press.

Wildenauer FX, Winter J (1986). Fermentation of Isoleucine and Arginine by Pure and Syntrophic Cultures of *Clostridium-Sporogenes*. *FEMS Microbiology Ecology* **38**: 373-379.

William M. Last a JDb, Karen Greengrass, Sunita Sukhan (2002). Re-examination of the recent history of meromictic Waldsea Lake, Saskatchewan, Canada. *Sedimentary Geology* **148**: 147–160.

Winfrey MR, Zeikus JG (1979). Microbial methanogenesis and acetate metabolism in a meromictic lake. *Appl Environ Microbiol* **37**: 213-221.

Winkler L (1888a). Die Bestimmung des im Wasser Gelösten Sauerstoffes. *Berichte der Deutschen Chemischen Gesellschaft* **21**: 2843–2855.

Woese CR, Kandler O, Wheelis ML (1990). Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *P Natl Acad Sci USA* **87**: 4576-4579.

Worm P, Müller N, Plugge C, Stams AM, Schink B (2010). Syntrophy in Methanogenic Degradation. In: Hackstein JHP (ed). *(Endo)symbiotic Methanogenic Archaea*. Springer Berlin Heidelberg. pp 143-173.

Woyke T, Sczyrba A, Lee J, Rinke C, Tighe D, Clingenpeel S *et al* (2011). Decontamination of MDA reagents for single cell whole genome amplification. *PloS one* **6**: e26161.

Wright JJ, Lee S, Zaikova E, Walsh DA, Hallam SJ (2009). DNA extraction from 0.22 micromM Sterivex filters and cesium chloride density gradient centrifugation. *J Vis Exp*.

Wrighton KC, Thomas BC, Sharon I, Miller CS, Castelle CJ, VerBerkmoes NC *et al* (2012). Fermentation, Hydrogen, and Sulfur Metabolism in Multiple Uncultivated Bacterial Phyla. *Science* **337**: 1661-1665.

Wrighton KC, Castelle CJ, Wilkins MJ, Hug LA, Sharon I, Thomas BC *et al* (2014). Metabolic interdependencies between phylogenetically novel fermenters and respiratory organisms in an unconfined aquifer. *Isme J* **8**: 1452-1463.

Youssef NH, Blainey PC, Quake SR, Elshahed MS (2011). Partial genome assembly for a candidate division OP11 single cell from an anoxic spring (Zodletone Spring, Oklahoma). *Appl Environ Microb* **77**: 7804-7814.

Youssef NH, Ashlock-Savage KN, Elshahed MS (2012). Phylogenetic diversities and community structure of members of the extremely halophilic Archaea (order Halobacteriales) in multiple saline sediment habitats. *Appl Environ Microb* **78**: 1332-1344.

Youssef NH, Rinke C, Stepanauskas R, Farag I, Woyke T, Elshahed MS (2014). Insights into the metabolism, lifestyle and putative evolutionary history of the novel archaeal phylum 'Diapherotrites'. *Isme J*.

Zaikova E, Walsh DA, Stilwell CP, Mohn WW, Tortell PD, Hallam SJ (2010). Microbial community dynamics in a seasonally anoxic fjord: Saanich Inlet, British Columbia. *Environmental Microbiology* **12**: 172-191.

Zerbino DR, Birney E (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research* **18**: 821-829.

Zhang J, Hudson J, Neal R, Sereda J, Clair T, Turner M *et al* (2010). Long-term patterns of dissolved organic carbon in lakes across eastern Canada: Evidence of a pronounced climate effect. *Limnology and Oceanography* **55**: 30-42.

Appendices

Appendix A Design and testing of CARD-FISH probe targeting candidate division

OP9/JS-1

A.1 Motivation for the project

In chapter 3 I elucidated how I identified statistically significant co-occurrences between candidate divisions OP9/JS-1, OP8, and WWE1 with methanogenic Archaea in the anoxic-sulfidic and methane-saturated waters of the monimolimnion. This allowed me to pose specific hypotheses about their metabolic potential. Using a co-occurrence network in combination with meta- and single-cell genomic sequencing data I posited that candidate divisions OP8 and OP9/JS-1 might be syntrophic acetate oxidizers that shuttle hydrogen to methanogens affiliated with the *Methanomicrobiales*. To test this hypothesis and to link phylogeny to function, incubation experiments with stable-isotope labeled substrates can be instrumental (Manefield et al 2002, Radajewski et al 2000). A recently developed method to investigate the incorporation of labeled substrates into microbial cells is called Secondary Ion Mass Spectrometry (SIMS), recently reviewed by Pett-Ringe and colleagues (Pett-Ridge and Weber 2012). The resolution of NanoSIMS instruments is at the submicron scale enabling measurements of metabolic activities with single-cell resolution. However, to identify and locate cells of interest NanoSims needs to be combined with chemotaxonomic visualization methods such as catabolized reporter deposition fluorescent in situ hybridisation (CARD-FISH). The first successful use of NanoSims as a tool in environmental microbiology monitored CH₄ consumption in a natural consortium of Archaea and Bacteria (Orphan et al 2001). To be able to specifically address the metabolic function and putative syntrophic interactions of candidate phylum JS-1 with NanoSIMS, I

designed a molecular probe for CARD-FISH and optimized experimental procedures for probe hybridisation.

A.2 Methods

Study site and sampling

Sakinaw Lake water samples were taken on June 6th 2013 at deep basin station S1 (49 40.968 N 124 00.119 W) with 12L Niskin bottles. Samples were taken from previously described water compartments epilimnion (5 m, 20 m, 30 m), upper part of the chemocline (33 m, 36 m, 40 m, 45 m), lower part of the chemocline (50 m, 55 m) and monimolimnion (60 m, 80 m, 120 m). Conductivity Temperature and Depth (CTD) were measured with a Seabird SBE19 (Sea-Bird Electronics Inc., Bellevue USA). Samples for CARD-FISH were taken aboard the research vessel and water was filled into 200 ml, acid washed and autoclaved serum bottles and fixed with particle free paraformaldehyde (PFA) at 2% final concentration directly after sample collection and kept at 4°C in the dark until filtration. Samples were filtered 10 h after sample collection in volumes of 5 mL and 2.5 mL onto white polycarbonate filter (Millipore) with 22 mm diameter and 0.2 µM pore size and stored in the dark at -80°C until further processing.

Probe Design

A specific probe against candidate division JS-1/OP9 5'-GACTATCCTCAAGTCTTAAAGTT - 3' was designed using the PROBE_FUNCTION tool of the ARB package. The probe is named S*-JS1-640-a-A-23 according to the oligonucleotide probe database nomenclature (Alm et al 1996), and is referred to as JS1640 from now on. Probe specificity against JS-1 was validated by BLAST and the PROBE_MATCH tool in Arb using Silva database v114 including a total of 263

full length Sakinaw Lake 16S sequences whereof 25 were affiliated with the JS-1. The 25 JS-1 sequences were clustered using mother, which lead to 8 representative sequences at 99% identity. PROBE_MATCH predicted that the JS1640 probe would hybridize to all JS-1 cluster representatives and to no sequence outside the JS-1 cluster. Allowing three mismatches, the probe was predicted to hybridize with sequence outside the JS-1 cluster. Therefore, the probe was considered specific. The horseradish peroxidase (HRP)-labeled JS1640 probe was synthesized by Biomers (Germany), and hybridization conditions were evaluated experimentally. Per definition candidate divisions are uncultured and therefore positive control-strains were not available. 454 Sequencing of the 16S rRNA gene of more than 66 samples taken in Sakinaw Lake throughout the water column at 6 different time points showed consistently that JS-1 was completely absent from surface waters. Surface water samples (5 m) were therefore taken as negative control. Hybridization conditions were tested empirically with gradually increasing formamide concentrations in the hybridization buffer (40%, 45%, 50%, 55% and 60%) and incubation temperatures of 32°C and 35°C. It was found that a formamide concentration of 50% and an incubation temperature of 32°C produced the strongest specific signals.

A.3 Results and discussion

Microscopic evaluation revealed an abundance of ~ 9% in Sakinaw Lake samples below 33 m and 0% in surface samples, which is in agreement with 454 sequencing data of the 16S rRNA gene (Fig. A.1).

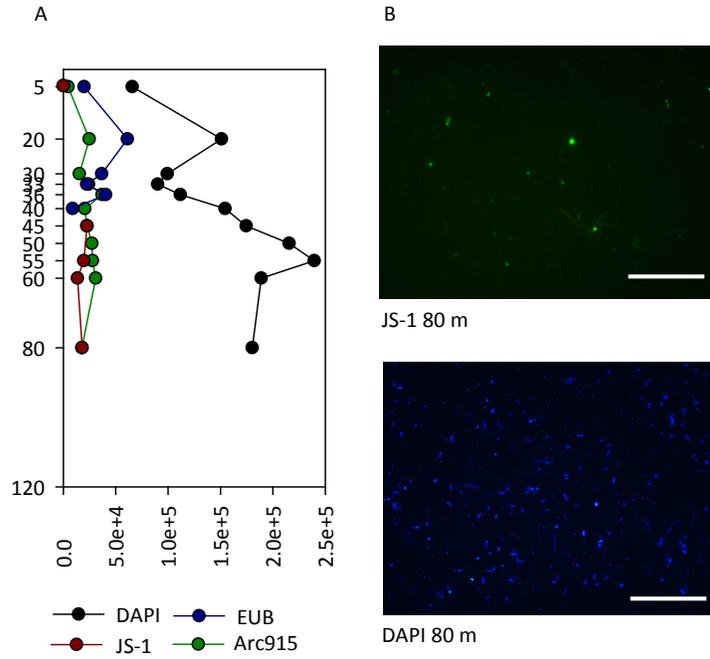


Figure A.1 Microscopic evaluation of CARD-FISH probe JS1-640-a-A-23.

Relative abundances for total Bacteria (EUB I-III), total Archaea (Arc915) and JS-1 estimated with CARD-FISH and total cell counts with DAPI. B Image of positively hybridized JS-1 cells in a 80m sample from Sakinaw Lake.

Appendix B Phylogeny of Bacteria in Sakinaw Lake

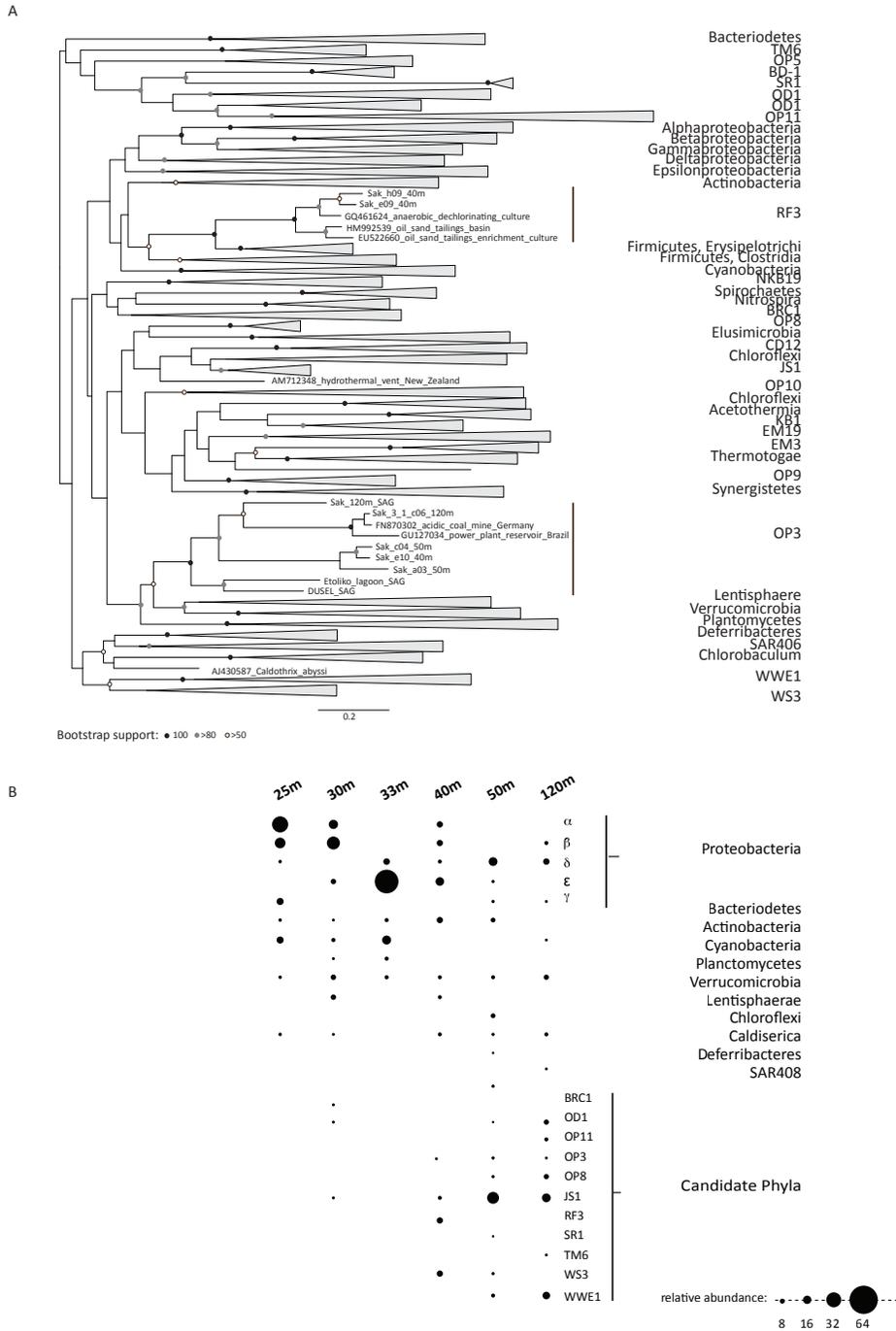


Figure B.1 Phylogeny of Bacteria in Sakinaw Lake

(legend on following page)

Figure B.1 Phylogeny of Bacteria in Sakianw Lake

Phylogenetic tree of 174 full length Sakinaw Lake 16S rRNA sequences and 318 reference sequences places Candidate Division RF3 within the *Firmicutes* (bootstrap support 100). B Phylogenetic affiliation and relative abundance of bacterial 16S rRNA clones retrieved from Sakinaw Lake samples.