METAGENOMIC AND GENOMIC ANALYSES OF MODERN FRESHWATER MICROBIALITES: UNMASKING A COMMUNITY OF COMPLEX METABOLIC POTENTIAL

by

RICHARD ALLEN WHITE III

Bachelors of Science, California State University East Bay, Hayward, California USA, 2007
Masters of Science, California State University East Bay, Hayward, California USA, 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)

September 2014

© Richard Allen White III, 2014
Abstract

Microbialites represent the oldest known persistent ecosystems and potentially the earliest evidence of life on the planet, having existed for ~85% of the geologic history of Earth (Dupraz et al., 2009). Despite over one hundred years of active research, little is known about modern freshwater microbialite ecosystems with regards to metabolic potential and microbialite-specific community structure. We performed metagenomic analysis of freshwater thrombolithic clotted microbialites from Pavilion Lake (British Columbia, Canada) and Clinton Creek (Yukon, Canada). In addition, metagenomes were obtained from the surrounding water and sediments to sort out which members of the microbial community were microbialite-specific. Pavilion Lake microbialites are distinct from the surrounding environments in microbial community structure and metabolic potential. The microbialites are dominated by heterotrophic processes with high abundances of heavy metal, antibiotic resistance, and alcohol fermentation pathways from the numerically dominant Proteobacteria. Clinton Creek houses the northern-most and fastest growing microbialites, which have a high proportion of photosynthetic genes, supporting isotopic data that photosynthesis drives microbialite formation. Clinton Creek has distinct communities, with microbialites dominated by Alphaproteobacteria (photoheterotrophs) and sediments dominated by Gammaproteobacteria (mainly heterotrophic nitrogen-fixers). To complement the metagenomic study of Pavilion Lake, a culturing based study was performed that yielded over one hundred new bacterial isolates. The new bacterial isolates were further screened for pigment containing strains that were non-photosynthetic. Amongst these pigment containing bacteria two new isolates were found and designated as an Exiguobacterium and an Agrococcus. Polyphasic analysis revealed that both are new species, which were named Agrococcus pavilionensis strain RW1 and Exiguobacterium pavilionensis strain RW2. Genome sequencing of both strain RW1 and
RW2 was completed and a comparative genomic and phylogenetic study was performed to evaluate their evolutionary placement and metabolic potential. Both isolates have low abundance in the Pavilion Lake microbialites, although they contribute heavy metal resistance genes that are found amongst the microbialite metagenomes. Hypothetical carotenoid biosynthesis pathways are also described which may be responsible for the coloration in *Agrococcus* and *Exiguobacterium* and may be related to photo-protection.
Preface

Chapter 2 is based on a manuscript in preparation:

White III RA, Suttle CA. Modern freshwater microbialites reveals distinct microbial community structure and metabolic potential that differs from the surrounding environment.

CAS, RAWIII and AMC designed the project. RAWIII, AMC and CAS conducted field work. RAWIII did the laboratory work and data analysis. RAWIII and CAS wrote the manuscript.

Chapter 3 is based on a manuscript in preparation:


CAS, GMD, RAWIII and IMP envisioned the project. RAWIII and IMP designed the experimental approach. IMP did the water chemistry and field sampling. RAWIII processed the genomic samples and did the data analysis with help from NWH and SJH. RAWIII, IMP and CAS wrote the paper.

Chapter 4 is based on a published manuscript and manuscript in preparation:


RAWIII and CAS initiated the project. RAWIII designed and conducted the experiments. EG contributed to experiments on growth and characterization. CJG completed the first draft assembly of the genome (published) and advised on *de novo* assembly and genome finishing. RAWIII revised the assembly, finished the genome and completed annotation. SAS and GFS completed all PLFA experiments and data analysis. GG did the electron microscopy of the isolate. RAWIII, CAS, GS, SAS and GG contributed to writing the manuscript in preparation.

**Chapter 5** is based on a published manuscript and manuscript in preparation:


RAWIII and CAS initiated the project. RAWIII designed and conducted the experiments. EG contributed to experiments on growth and characterization. CJG completed the first draft assembly of the genome (published) and advised on *de novo* assembly and genome finishing. RAWIII revised the assembly, finished the genome and completed annotation. SAS and
GFS completed all PLFA experiments and data analysis. GG did the electron microscopy of the isolate. RAWIII, CAS, GS, SAS and GG contributed to writing the manuscript in preparation.
# Table of Contents

Abstract ........................................................................................................................................... ii

Preface............................................................................................................................................. iv

Table of Contents ............................................................................................................................ vii

List of Tables .................................................................................................................................... xiv

List of Figures ................................................................................................................................. xv

List of Abbreviations ....................................................................................................................... xix

List of Keywords ........................................................................................................................... xxi

Acknowledgments .......................................................................................................................... xxiii

Dedication ........................................................................................................................................ xxiv

Chapter 1: Introduction .................................................................................................................. 1

1.1 What is a microbialite? .............................................................................................................. 1

1.1.1 Physical classification and brief history .............................................................................. 1

1.2 Microbialites as the oldest persistent ecosystems ................................................................. 2

1.3 Global distribution of microbialites ....................................................................................... 2

1.4 Proposed microbialite formation hypotheses ......................................................................... 5

1.4.1 Abiotic (spectator hypothesis) .......................................................................................... 5

1.4.2 Abiotic combined biotic hypothesis .................................................................................. 6
1.5 Major microbial taxa and their metabolic pathways that contribute to microbialite formation

1.5.1 *Alphaproteobacteria* ........................................................................................................ 8

1.5.2 *Cyanobacteria* ............................................................................................................... 9

1.5.2.1 Planktonic cyanobacteria .............................................................................................. 9

1.5.2.2 Filamentous cyanobacteria, mat-builders ...................................................................... 9

1.5.3 Sulfur-oxidizing bacteria (SOBs) ..................................................................................... 10

1.5.4 Sulfate-reducing bacteria (SRBs) .................................................................................... 11

1.5.5 *Firmicutes* .................................................................................................................... 13

1.5.6 *Actinobacteria* .............................................................................................................. 14

1.5.7 Eukaryotes...................................................................................................................... 15

1.5.8 Viruses .......................................................................................................................... 16

1.6 Previously studied microbialite-forming ecosystems ....................................................... 17

1.6.1 Shark Bay microbialites .................................................................................................. 17

1.6.2 Highbourne Cay microbialites ....................................................................................... 18

1.6.3 Cuatro Cienegas microbialites: Pozas Azules II and Rios Mesquites ......................... 20

1.7 Microbialite-forming ecosystems described in this thesis ................................................. 21

1.7.1 Pavilion Lake ................................................................................................................. 21

1.7.2 Clinton Creek ................................................................................................................ 23

1.8 Thesis hypotheses ............................................................................................................. 24
Chapter 2: Modern freshwater microbialite microbial community structure and metabolic potential is distinct from the surrounding environment .......................................................... 32

2.1 Introduction .............................................................................................................. 32

2.2 Materials and methods .......................................................................................... 33

2.2.1 Sample collection ................................................................................................. 33

2.2.2 DNA extraction ...................................................................................................... 34

2.2.3 Metagenomic library preparation (454 FLX Titanium and Illumina) ................. 34

2.2.4 Metagenomic data assembly and analysis ............................................................ 35

2.2.5 Recruitment analysis to draft genomes ................................................................. 37

2.2.6 Metagenomic data depositing ............................................................................... 37

2.3 Results and discussion .......................................................................................... 38

2.3.1 Metagenomic assembly statistics ......................................................................... 38

2.3.2 Microbialites communities differ from the surrounding environment ............... 38

2.3.3 Pavilion Lake microbialite core microbial community and related pathways ....... 40

2.3.4 Low abundance members of Pavilion Lake microbialite microbial community ..... 45

2.3.5 Metabolic potential relating to photosynthetic or heterotrophic processes ......... 48

2.3.6 Viral community and viral defense ....................................................................... 48

2.4 Conclusion ............................................................................................................... 50

2.5 Tables and figures .................................................................................................. 52
Chapter 3: Comparative analysis of microbiome of rapidly forming modern microbialites from the Canadian Yukon

3.1 Introduction

3.2 Materials and methods

3.2.1 Site description and water chemistry

3.2.2 Sampling, DNA extraction, purity and concentration measurements

3.2.3 Illumina HiSeq/MiSeq library construction quality control and quantification

3.2.4 Analysis of Illumina sequencing data

3.2.5 Comparative metagenomics

3.2.6 Metagenomic data depositing

3.3 Results and discussion

3.3.1 Metagenomic assembly and contig properties

3.3.2 Community composition

3.3.3 Metabolic potential

3.3.4 Comparative metagenomic analysis

3.4 Conclusion

3.5 Tables and figures

Chapter 4: The genome of Agrococcus sp. strain RW1 isolated from modern microbialites reveals genetic promiscuity and elements responsive to environmental stresses

4.1 Introduction
4.2 Materials and methods ............................................................................................................................................. 100

4.2.1 Isolation, growth conditions, phage induction, biochemical and antibiotic susceptibility tests ........................................... 100

4.2.2 Phospholipid fatty acid analysis .......................................................................................................................... 101

4.2.3 Light and scanning electron microscopy (SEM) ........................................................................................................ 102

4.2.4 DNA extraction, PCR and Illumina library construction .......................................................................................... 103

4.2.5 Phylogenetic analysis ........................................................................................................................................... 103

4.2.6 Whole genome assembly and genome finishing ................................................................................................... 104

4.3 Results and discussion .............................................................................................................................................. 106

4.3.1 Morphology and growth characteristics ............................................................................................................. 106

4.3.2 PLFA characterization and comparative analysis .................................................................................................. 107

4.3.3 Evolutionary placement of Agrococcus sp. RW1 within the genus ................................................................. 108

4.3.4 Biochemical properties and antibiotic susceptibility ........................................................................................... 110

4.3.5 Mobile DNA and viral elements in Agrococcus sp. RW1 ............................................................................... 111

4.3.6 Agrococcus sp. RW1 life in a cold microbialite mat ............................................................................................ 113

4.3.7 Carotenoid biosynthesis ...................................................................................................................................... 115

4.4 Conclusions ............................................................................................................................................................. 117

4.5 Tables and figures .................................................................................................................................................... 119

Chapter 5: Illumination of an Exiguobacterium through the comparative lenses of classical characterization and modern genomic analysis .................................................................................................................. 131
5.1 Introduction .................................................................................................................. 131

5.2 Materials and methods ............................................................................................... 133
   5.2.1 Isolation, growth conditions, biochemical and antibiotic susceptibility tests ...... 133
   5.2.2 Phospholipid fatty acid analysis (PLFA) ............................................................ 134
   5.2.3 Light and scanning electron microscopy (SEM) ................................................. 135
   5.2.4 DNA extraction, PCR and Illumina library construction .................................... 135
   5.2.5 Whole genome assembly, assembler comparison, genome finishing and annotation ................................................................. 136
   5.2.6 Phylogenetic analysis and multiple locus sequence typing (MLST) analysis ...... 137

5.3 Results and discussion ............................................................................................... 138
   5.3.1 Morphology and growth characteristics............................................................. 138
   5.3.2 Biochemical properties and antibiotic susceptibility .......................................... 140
   5.3.3 PLFA characterization and comparative analysis ............................................... 142
   5.3.4 Evolutionary placement of *Exiguobacterium* sp. RW2 within the genus ............ 142
   5.3.5 Comparative genome analysis ........................................................................... 144
   5.3.6 Carbohydrate metabolism .................................................................................. 146
   5.3.7 Amino acid biosynthesis and catabolism ........................................................... 147
   5.3.8 Motility and flagellum biosynthesis .................................................................... 148
   5.3.9 Stress response temperature, heavy-metals and salinity ..................................... 149
   5.3.10 Carotenoid biosynthesis ................................................................................... 151
List of Tables

Table 2.1: Pavilion Lake metagenome assembly statistics ................................................................. 52
Table 2.2: Taxonomic classes (RefSeq) that are overrepresented in microbialites relative to the surrounding environment as determined by ANOVA using STAMP ........................................ 53
Table 2.3: Functional annotations (SEED subsystem level I) that are overrepresented in microbialites relative to the surrounding environment as determined by ANOVA using STAMP ........................................................................................................................................... 54
Table 3.1: Assembly statistics for Clinton Creek metagenomes .......................................................... 88
Table 3.2: Domain classification of the microbial community in Clinton Creek (%) ...................... 89
Table 4.1: Comparison of physiological properties of Agrococcus selected strains .................... 119
Table 4.2: Selected cellular phospholipid fatty acids of Agrococcus selected strains (%) ....... 120
Table 4.3: Agrococcus spp. final assembly statistics ........................................................................ 121
Table 4.4: Biochemical properties of Agrococcus spp. selected strains ........................................ 122
Table 4.5: Antibiotic susceptibility of Agrococcus spp. selected strains ....................................... 123
Table 5.1: Names and isolation location of Exiguobacterium species ........................................... 154
Table 5.2: Properties of Exiguobacterium selected strains ............................................................. 156
Table 5.3: Biochemical tests for selected strains of Exiguobacterium spp..................................... 157
Table 5.4: Antibiotic susceptibility for selected strains Exiguobacterium spp.............................. 158
Table 5.5: Selected phospholipid fatty acids from Exiguobacterium spp. strains (%) ................ 159
Table 5.6: Exiguobacterium spp. assembly and annotation statistics ............................................. 160
Table 5.7: Exiguobacterium spp. carotenoid biosynthesis genes ..................................................... 161
List of Figures

Figure 1.1: Microbialite morphology and carbonate macrofabric layout ........................................... 26
Figure 1.2: Microbialite morphology examples ................................................................................ 27
Figure 1.3: Timeline of microbialite and global atmospheric oxygen concentration over geological time ................................................................................................................................. 28
Figure 1.4: Generalized equations for microbialite formation and dissolution ............................... 29
Figure 1.5: Pavilion microbialite morphologies .............................................................................. 30
Figure 1.6: Pavilion Lake nodular cyanobacterial morphologies ................................................... 31
Figure 2.1: Pavilion Lake microbialite morphology with GC content and protein isoelectric points (pI) ........................................................................................................................................... 55
Figure 2.2: Pavilion Lake microbial community composition .......................................................... 56
Figure 2.3: PCA and scatter plots using SEED subsystems based on ANOVA in STAMP for Pavilion Lake metagenomes ......................................................................................................................... 57
Figure 2.4: MetaCyc pathway annotations for Pavilion Lake metagenomes .................................. 58
Figure 2.5: Box and whisker plots for significant taxa represented in water filters and sediments by RefSeq classification using ANOVA in STAMP based on multiple groups .................................. 59
Figure 2.6: Post-havoc confidence interval plots functional classification for represented metabolic potential in surrounding environment (water filters and sediments) ......................... 60
Figure 2.7: Post-havoc confidence interval plots for Alphaproteobacteria using RefSeq classification .................................................................................................................................................. 61
Figure 2.8: Post-havoc confidence interval plots for Deltaproteobacteria using RefSeq classification .................................................................................................................................................. 62
Figure 2.9: Post-havoc confidence interval plots for Cyanobacteria using RefSeq classification 63
Figure 2.10: Post-havoc confidence interval plots for cyanobacterial based functional genes using SEED subsystem classification and urealytic pathways ................................................................. 64

Figure 2.11: Box and whisker plots for novel functional classifications (SEED subsystem) represented in microbialites based on ANOVA in STAMP using multiple groups ............... 65

Figure 2.12: Box and whisker plots for novel Archaeal based functional classifications (SEED subsystem) represented in microbialites based ANOVA in STAMP using multiple groups ...... 66

Figure 2.13: *Exiguobacterium* presence in Pavilion Lake microbialite metagenome ............ 67

Figure 2.14: Metagenomic recruitment plot for Pavilion Lake microbialite against *Agrococcus* 68

Figure 2.15: Microbialite metagenome functional annotation ranking using SEED subsystem (Level I) ......................................................................................................................................................... 69

Figure 2.16: Microbialite metagenome carbohydrate related SEED subsystem (Level II) functional annotations ........................................................................................................................................................................... 70

Figure 2.17: Box and whisker plots for viruses and CRISPRs in Pavilion Lake .................... 71

Figure 2.18: Post-havoc confidence interval plots for viral taxa using RefSeq and viral specific functional genes in SEED ................................................................................................................................................. 72

Figure 3.1: Clinton Creek sample site and examples of microbialite morphology .................. 90

Figure 3.2: Molecular properties of the Clinton Creek microbialite and sediment contigs........ 91

Figure 3.3: Microbial community structure of Clinton Creek metagenomes ....................... 92

Figure 3.4: Extended error plots for functional gene annotations for Clinton Creek metagenomes in STAMP using SEED subsystems .................................................................................................................. 93

Figure 3.5: MetaCyc pathway annotations for Clinton Creek metagenomes ....................... 94

Figure 3.6: Functional gene comparative PCA plot for Clinton Creek metagenomes ............ 95

Figure 3.7: Scatter plots of functional gene annotations using SEED subsystem level III ....... 96
Figure 3.8: Comparative functional gene PCA plot for Clinton Creek against whale fall metagenomes

Figure 3.9: Comparative functional gene PCA plot for Clinton Creek against coral metagenomes

Figure 4.1: Agrococcus sp. strain RW1 microscopy

Figure 4.2: Maximum-likelihood tree of 16S rRNA gene sequences for Agrococcus spp. and related taxa (~1409 bp)

Figure 4.3: MLST (multiple locus sequence typing) maximum-likelihood tree of Microbacteriaceae (~2939 amino acids)

Figure 4.4: Genome plot (~2.6Mb) of Agrococcus sp. strain RW1

Figure 4.5: Genome synteny and Venn diagrams of Agrococcus sp. RW1 vs. A. lahaulensis K22-21

Figure 4.6: Agrococcus sp. RW1 LC-RRW783 plasmid genomic plot

Figure 4.7: Proposed carotenoid biosynthetic pathway for Agrococcus

Figure 5.1: Exiguobacterium sp. strain RW2 microscopy

Figure 5.2: Exiguobacterium maximum-likelihood tree based on 16S rRNA gene sequences (~1353 bp)

Figure 5.3: Exiguobacterium combined ribosomal operon maximum-likelihood tree

Figure 5.4: MLST maximum-likelihood phylogenetic tree

Figure 5.5: Genome plot (~3 Mb) of Exiguobacterium sp. strain RW2

Figure 5.6: Genome synteny and functional Venn diagrams (SEED) with high amino acid identities (AAIr >80%) to Exiguobacterium sp. RW2
Figure 5.7: Genome synteny and functional Venn diagrams (SEED) with low amino acid identities (AAIr <65%) to *Exiguobacterium* sp. RW2 ................................................................. 168

Figure 5.8: Genome synteny and functional Venn diagrams (SEED) with low amino acid identities (AAIr <65%) to *Exiguobacterium* sp. RW2 ................................................................. 169

Figure 5.9: MetaCyc pathway annotations for *Exiguobacterium* genomes........................................ 170

Figure 5.10: Proposed carotenoid biosynthesis pathway in the genus *Exiguobacterium* ............ 171
List of Abbreviations

ANOVA: Analysis of variance.
BLAST: Basic local alignment search tool. It finds local similarity between two sequences which can be either protein or nucleotide sequences or a search across these groups.
BLAT: Blast-like alignment tool. Similar to blast but is faster by keeping the entire index of the database. Instead of a set of database sequences it searches against an index derived from the database in smaller sequences or non-overlapping n-mers.
CRISPR: Clustered regularly interspaced short palindromic repeats. These short repetitions of DNA sequences are imprints of prior viral exposures. CRISPRs are associated with cas genes which act combined as a prokaryotic immune system.
CTAB: Cetyl trimethylammonium bromide which is used for nucleic acid extraction for samples with high polysaccharides (e.g. plants and cyanobacteria).
DIC: Dissolved inorganic carbon.
DOC: Dissolved organic carbon.
EPS: Extracellular polymeric substances containing exopolysaccharide formed by filamentous cyanobacteria.
FAMEs: Fatty acid methyl ester analysis.
GC content: Analysis of guanine and cytosine content, often a percentage in a collection of DNA sequences.
Gya: One gigayear ago or $10^9$ years ago, which is a billion years ago. See Gyr.
Gyr: One gigayear or $10^9$ years, which is a billion years.
LAST: Local alignment search tool. LAST is a blast-like tool that is faster than standard blast; LAST uses adaptive seeds to align and assign similarities between two sequences.
MDS: Multidimensional scaling.
MG-RAST: Metagenomic Rapid Annotations using Subsystems Technology. MG-RAST is an online server for metagenomic annotation and analysis.
MLST: Multiple locus sequence typing commonly used for genotyping clinical bacterial isolates but used here as multiple gene phylogenetic analysis.
Mya: One million ago or $10^6$ years ago. See Myr.
Myr: One million years or $10^6$ years.
N50: Genomic assembly statistic. Defined as the length for which all contigs of that length or longer contain at least 50% of the total of the lengths of the contigs.
N90: Genomic assembly statistic. Defined as the length for which all contigs of that length or longer contain at least 90% of the total of the lengths of the contigs.
NMDS: Nonmetric multidimensional scaling.
ORFs: Open reading frame related to a protein coding gene.
PCA: Principal component analysis.
pI: Predicted protein isoelectric point.
PAL%: Present atmospheric levels, where 100% PAL equals Earth’s atmospheric oxygen at the present day.
PLFAs: Phospholipid fatty acids analysis.
qPCR: Quantitative real-time polymerase chain reaction.
R: Statistical programming language.
RAST: Rapid annotations using subsystems technology. It is a web tool that annotates a bacterial, plastid or archaea genome using SEED subsystem classifications.

SEED: A protein sequence database that uses subsystem classifications for metabolic pathways. Each subsystem are functional classifications labeled as levels, with level I the highest followed by functions the lowest level.

SEM: Scanning electron microscopy.

SOB: Sulfur oxidizing bacteria using aerobic sulfur oxidation pathway called Sox.

SRB: Sulfate reducing bacteria using anaerobic dissimilatory sulfate reduction via the dsr pathway.

STAMP: Statistical Analysis of Metagenomic Profiles. It is a tool used to analyze metagenomic data, in this case from MG-RAST.
List of Keywords

454 FLX Titanium: Roche DNA pyrosequencing technology with read lengths >300bp
Abiotic: A process without biology, meaning geological in this context.
AbySS: A whole genome de novo assembler that uses De bruijn graph based method.
Anoxic or Anoxygenic: Not containing or using molecular oxygen. Used to describe photosynthetic processes.
Biotic: A process with biology, in this context biogeochemical.
Celera: A whole genome de novo using overlap layout consensus (OLC) following a Hamiltonian path. It attempts to combine overlapping sequences in the nodes into larger contigs.
Clinton Creek: A microbialite containing ecosystem in Yukon, Canada.
Complete genome: A collection of DNA genomic sequence data from a single organism that for bacteria contains one chromosome (usually circular) with possibly many plasmids (usually fewer then five contigs in total).
Contig: A set of overlapping DNA sequence fragments assembled to form a single sequence.
Draft genome: A collection of DNA genomic sequence data from a single organism that is presented as many contigs and scaffolds.
Illumina: DNA sequencing based on reversible chain terminators.
Kelly Lake: A microbialite-forming ecosystem in British Columbia, Canada near Pavilion Lake.
K-mer: A contiguous DNA sequence of n length. For example, a K-mer of 39 would contain a unique DNA sequence of 39 base pairs in length.
Leiolites: A type of microbialite that does not fit in any direct classification due to its arranged, not-laminated, and amorphous properties.
Macrofabric: The arrangement of the carbonate in a microbialite.
MetaCyc: A metabolic pathway database.
Metagenomics: Whole community shotgun sequencing as a field or approach.
Metagenome: Whole community shotgun sequencing of a single community.
Metavelvet: A whole genome de novo assembler that uses De bruijn graph based method designed specifically for metagenomic data.
Microbialite: Microbially derived organosedimentary carbonate structures.
Oncolite: A type of microbialite that is stromatolithic, meaning it has a laminated internal carbonate layer that is spherical in nature.
Oxic or Oxygenic: Containing or using molecular oxygen. Used to describe photosynthetic processes.
Progressive Mauve: A whole genome alignment program that is useful for genome ordering, editing and detecting rearrangements.
Pavilion Lake: A microbialite-forming ecosystem in British Columbia, Canada.
Ray: A whole genome de novo assembler that uses De bruijn graph based method. It has a built-in metagenomic data detector known as Ray Meta but does not require different working commands.
Recruitment analysis: The process of taking metagenomic reads and aligning them to a known genome (commonly a bacterial genome), in order to assess abundances of similar genotype in a community.
RefSeq: A taxonomic protein database.
Scaffold: A collection of non-contiguous contigs with gaps.
SOAP2: A whole genome de novo assembler that uses De bruijn graph based method.
**Stromatolite:** A layered microbially derived organosedimentary carbonate structure.

**Thrombolite:** A clotted microbially derived organosedimentary carbonate structure.

**Velvet:** A whole genome *de novo* assembler that uses De brujin graph based method.

**Welch's t-test:** An alteration of student’s t-test with assumptions that two samples are independent and that both are drawn from a normal population.
Acknowledgments

I thank my supervisor Curtis A Suttle for guidance, help and funding. I particularly appreciate that he let me follow my own ideas and interests and make important decisions on my own. I thank the members of my committee, J. Thomas Beatty, Steven J Hallam, and Sue Baldwin for guidance, advice, and occasional discussions.

I thank all my collaborators, and particularly Christopher J Grassa, Ian M Power, Greg Dipple, Greg Slater, Sarah Soles, Gordon Southam, Jennifer Abel, and Greg Gavelis for their excellent and significant contributions to the publications included in this thesis. Also to thank Allyson Brady, Darlene Lim, Kyung-Mee (Jenny) Moon, and Leonard Foster for significant contributions on my projects not included within the thesis and occasional discussions about the thesis work. I thank all lab mates and friends for helping in sampling and various things needed to survive a PhD.
Dedication

This thesis is dedicated to my loving wife Rachelle Rene’ White and our beautiful daughter Elizabeth Baylee White. We did it girls! I couldn’t have done it without either of you. I know that the man I am today is because of my love of you girls.

This is also for my Mother, Father, and Uncle Art who are not here to celebrate one of the greatest things a person can ever achieve in their life. This was a promise to my father that I would go “all the way,” with my education and obtain a PhD. However, what I didn't expect is that my thirst for knowledge would be greater now, than when I started.

This is only the beginning.

I love you all wherever you are. Always remember that a part of you lives in me and Elizabeth now and forever.
1.1 What is a microbialite?

1.1.1 Physical classification and brief history

The term microbialite refers to a microbially-derived organosedimentary carbonate structure that is our earliest evidence of life (Grotzinger and Knoll, 1999; Schopf 2006; Dupraz et al., 2009). Microbialites are classified into two supergroups according to their macrofabric carbonate layout, as well as gross morphology (Perry et al., 2007; Riding et al., 2011). The first supergroup contains internal laminated carbonate macrofabric stromatolites (thinly layered piles), oncolites (layered, but spherical), and leiolites (amorphous mesofabric) (Perry et al., 2007; Riding et al., 2011; Figure 1.1). The second supergroup contains un laminated carbonate macrofabric thrombolites (irregular clots), and dendrolites (branched projections that are treelike) (Perry et al., 2007; Riding et al., 2011; Figure 1.1). The classification of microbialites is complicated by their plasticity in nature, with many gradations existing between these basic morphologies. Among living forms, stromatolites and thrombolites are the most common (Dupraz et al., 2009; Figure 1.2).

Microbialites have been intensely studied for over one hundred years (Kalkowsky, 1908). The first mention of the term stromatolite, or “Stromatolith,” was used to describe early Triassic stromatolites (Kalkowsky, 1908; Riding et al., 2011; Figure 1.1). Pia (1933), was the first to classify oncolites as “unattached, regularly or irregularly spheroidal, concentrically to semiconcentrically laminated bodies” (Figure 1.1). Oncolites were reclassified as “a category of stromatolite structure,” thus removing them as a separate category (Pia 1933; Logan et al., 1964). Aitken (1967), was the first to use the term ‘thrombolite,’ from the Greek meaning “thrombos or bloodclot,” from their irregular, clot-like arrangement (Figure 1.1). Burne and Moore (1987),
effectively provided the modern definition we use today, as structures “that are accreted as a result of a benthic microbial community trapping and binding detrital sediment and/or forming the locus of mineral precipitation.” They broadened the definition from one that included biotic factors to that of a “benthic microbial community”, took into account abiotic factors such as “detrital sediment” and added geologic context (“forming the locus of mineral precipitation”).

1.2 Microbialites as the oldest persistent ecosystems

Microbialites represent the oldest known persistent ecosystems and potentially the earliest evidence of life (Schopf 2006; Grotzinger and Knoll, 1999). The geologic record indicates that microbialite diversification occurred concurrently with the emergence of life ~3.50 billion years ago (Gya), followed by a period of dominance extending until ~542 million years ago (Mya) following the Cambrian explosion (Allwood et al., 2006; Dupraz and Visscher 2005; Schopf et al., 2007; Rothschild and Mancinelli, 1990; Figure 1.3). With the age of the Earth estimated to be 4.54 billion years (Gyr), the appearance of microbialites in the fossil record at ~3.85 to 3.5 Gyr suggests that microbialites have persisted for ~85% of geological history (Patterson 1956; Dupraz and Visscher 2005; Figure 1.3). Modern microbialites first appear when cyanobacteria emerge (~3.0 Gyr), persisting for roughly 66% of earth's geologic history (Dupraz and Visscher 2005; Figure 1.3). Thus, modern microbialites are the best available representation of Earth's earliest life forms, and their biology may grant insight into the oldest aspects of the biosphere (Grotzinger and Knoll, 1999; Dupraz and Visscher 2005).

1.3 Global distribution of microbialites

Microbialites have adapted to radiate into many environmental conditions, including extremes of temperature, pH and salinity (Dupraz and Visscher 2005). Microbialites are globally distributed, and can be found in oceans (Reid et al., 2000; Khodadad and Foster, 2012; Mobberley
et al. 2013), freshwaters (Ferris et al., 1997; Laval et al., 2000; Gischler et al., 2008; Breitbart et al., 2009), hypersaline lagoons (Goh et al., 2009; Allen et al., 2009), tropical lagoons (Sprachta et al., 2001), hot springs (Bosak et al., 2012), and remnant mining sites (Power et al., 2011a). To a lesser extent, they have also colonized terrestrial environments, such as soils (Maliva et al., 2000) and caves (Lundberg and Mcfarlane, 2011).

Modern microbialite-forming ecosystems have been found globally suggesting that they are not as rare as previously thought. North America and Antarctica harbor many examples of freshwater microbialites, with Cuatro Cienegás and Yellowstone microbialites being the most studied (Breitbart et al., 2009; Bosak et al., 2012). In North America, microbialites are found in the Yukon (Power et al., 2011a), Southwestern British Columbia (Laval et al., 2000; Ferris et al., 1997), Yellowstone National Park in Wyoming (Bosak et al., 2012), Pyramid Lake Nevada (Arp et al., 1999), and in Mexico at Cuatro Cienegás (Breitbart et al., 2009), Lake Pátzcuaro (Bridgwater et al., 1999; Bischoff et al., 2004) and Lake Alchichica (Couradeau et al., 2011). Antarctica is a haven for microbialites, with many found in pristine condition with little human influence. Antarctica has four ecosystems containing conical stromatolites or thrombolites: Lake Untersee (Andersen et al., 2011), Lake Joyce (Hawes et al., 2011), Lake Hoare (Squyres et al., 1991) and Lake Vanda (Andersen et al., 2011).

The microbialites of Central and South America are found in some of the most extreme environments, including lakes containing toxic levels of heavy metals. Microbialites are found in Central and Mesoamerica, including Laguna Bacalar, Mexico, which borders with Belize (Gischler et al., 2008). In South America, microbialites are found in Brazil in the hypersaline lakes of Lagoa Vermelha and Lagoa Salgada (Vasconcelos et al., 2006; Spadafora et al., 2010), in northern Chile in La Brava (Farías et al., 2014) and at the southern tip of Chile (Graham et al., 2014). Laguna
Socompa on the border of Chile and Argentina is an example of an extreme ecosystem that contains microbialites, as it contains high levels of arsenic (18.5 mg L\(^{-1}\)) and iron (1 mg L\(^{-1}\)) (Farías et al., 2013).

Diverse microbialites, including rare hydromagnesite forms and the largest microbialite ever recorded, are found in Europe. The Ruidera pools in Spain hold a great diversity of freshwater stromatolites (Santos et al., 2010), and Lake Van, Turkey, contains the largest microbialites on record (Kempe et al., 1991; López-García et al., 2005). Salda Gölü Lake, in southern Turkey, also holds the only other example of hydromagnesite microbialites outside of Lake Alchichica, Mexico (Braithwaite and Zedefa, 1994; Couradeau et al., 2011).

Microbialites in Oceania are rare; however, they include modern examples of coral replacement. Hypersaline lakes found in Lake Stonada, Indonesia (Arp et al., 2004), Kiritimati Atoll in the central Pacific (Schneider et al., 2013) and the caldera lakes of Niuafo’ou Island, Tonga (Kazmierczak and Kempe, 2006) contain microbialites. Microbialites in the tropical lagoons of French Polynesia and New Caledonia exist side by side with dead and live corals, illustrating direct competition (Sprachta et al., 2001).

Western Australia has the oldest fossilized examples of microbialites at roughly 3.45 Gyr old (Van Kranendonk et al., 2008), and Shark Bay has one of the few examples of marine stromatolites (Goh et al., 2009; Allen et al., 2009). Two lakes in Australia have documented microbialites, including Lake Thetis (Grey et al., 1990) and Lake Clifton (Konishi et al., 2001). Other lakes in Australia that contain thrombolithic microbialites include Lake Richmond, Lake Walyungup and Lake Coolongup, which are protected and endangered. However, no publications exist describing them.
While China and Russia, the largest combined land mass on the planet, have many examples of microbialite fossils, no studies describe modern versions; or, if studies are available, they have not been translated into English (Wu et al., 2014). In terms of African microbialites, a recent study found some in southern Africa (Smith et al., 2011), and fossils have been found throughout the continent. However, Africa remains largely unstudied when it comes to modern microbialites.

1.4 Proposed microbialite formation hypotheses

There is much debate on whether the formation of ancient microbialites dating back ~3.50 Gya in Western Australia was purely abiotic (Lowe, 1994; Grotzinger and Rothman, 1996) or a combined biotic-abiotic event (Dupraz and Visscher, 2005; Schopf, 2006; Mcloughlin et al., 2008; Dupraz et al., 2009).

1.4.1 Abiotic (spectator hypothesis)

The abiotic model stresses that ancient microbialites and possibly even modern versions are formed completely by abiotic processes (Lowe, 1994; Grotzinger and Rothman, 1996; Mcloughlin et al., 2008). Lowe suggested that the Western Australian stromatolites within the Warrawonna group (3.5 to 3.2 Gyr) lack the structural evidence of biotic processes and therefore are of abiotic origin (1994). Moreover, a laboratory spray deposition experiment showed that stromatolites could be generated abiotically using enamel spray paint with metal-oxide pigment particles (Mcloughlin et al., 2008). A purely abiotic model would suggest that the microbes and/or pathways would be the same between the microbialite and their surrounding environment, as the chemistry alone is driving the formation of the microbialite: microbes would merely be “spectators.” Thus, the abiotic model would suggest that microbes have no role in the formation of microbialites.
Abiotic factors are certainly critical to microbialite formation. These include hard water (≥ 121 mg CaCO₃ L⁻¹) with free divalent cations (mainly Ca²⁺ and/or Mg²⁺), high alkalinity (≥ 121 mg CaCO₃ L⁻¹), stable alkaline pH water (pH ≥ 8), clear water with visible light penetration, and dissolved inorganic carbon (DIC) (Dupraz et al., 2009). No carbonate-derived microbialite-forming ecosystem has been found without at least hard water, high alkalinity and a stable alkaline pH (pH ≥ 8); carbonates cannot form without these factors.

1.4.2 **Abiotic combined biotic hypothesis**

The biotic link to the origin of microbialites has been debated for over one hundred years since it was first suggested by Kalkowsky (1908). Lowe (1995) rejected any biotic role in the formation of the Australian Warrawonna group stromatolites (3.5 to 3.2 Gya). Buick et al. (1995), responded that Lowe violated Descartes’ fourth rule of science, “omettre rien,” directly translating to “omit nothing,” suggesting that Lowe (1995) omitted critical examples relating to a combined hypothesis. The examples not included by Lowe (1995) were the post-Isua δ¹³C isotopic data, and various microfossil examples illustrating that biotic-abiotic processes are involved in the formation of microbialites (Buick et al., 1995). Since that time, sulfur isotopic measurements in the Australian stromatolites from ~3.45 Gya suggest negative δ³⁴S isotopic fractionation which can only be explained by biotic processes (Bontognali et al., 2012, Figure 1.3). Ancient microbialites spanning ~3.5 to 3.0 Gya formed under anoxygenic processes, as sufficient oxygen was not present to drive oxygenic photosynthesis, whether cyanobacteria were present or not (Crowe et al., 2013; Planavsky et al., 2014; Figure 1.3).

Biotic metabolism in modern microbialites is an active and highly complex interaction with abiotic factors. Both oxygenic and anoxygenic metabolism within the surface microbial mat on microbialites is highly productive (5000 to 6200 mg biomass cm⁻² day⁻¹), and similar to the primary
productivity of tropical forests (6000 mg biomass cm$^{-2}$ day$^{-1}$) (Stal, 2000; Centeno et al., 2012). Carbonates and free calcium are initially trapped by adhesive biofilms and mats that are often cyanobacterial in origin (Dupraz and Visscher 2005; Dupraz et al., 2009). The cyanobacterial mat is then degraded by heterotrophic bacteria that liberate HCO$_3^-$ and Ca$^{2+}$, thereby increasing the saturation index and producing nucleation points for carbonate precipitation (Dupraz and Visscher 2005; Dupraz et al., 2009). Cyanobacterial mats contain exopolysaccarides (EPS) that serve as a location of mineral nucleation, while providing a heterotrophic microenvironment favorable for organomineralization via dissimilatory sulfate reduction (Dupraz and Visscher 2005; Dupraz et al., 2009). Light is another critical factor in the biotic-abiotic relationship of microbialite formation. Many microbialite-forming ecosystems are either in shallow locations or have clear light visibility with high light penetration allowing for photoautotrophic microbial mat growth.

The concept that microbes have no role in microbialite formation does not account for evidence that specific mineral precipitation is a function of microbial presence. For the abiotic hypothesis, the question is, why would only certain minerals form, and rates of precipitation differ, with the presence of microbes under similar abiotic conditions? The Atlin wet-land in British Columbia, Canada, contains carbonate precipitating microbial mats. A microcosm experiment showed that abiotic-biotic microcosms promoted carbonate formation at a faster rate and resulted in a completely different mineral (dytingite) than the abiotic microcosm (nesquehonite) (Power et al., 2007). This suggests that both biotic and abiotic forces are required to form particular minerals associated with lithifying microbial mats, including microbialites (Power et al., 2007). Microbialite formation is likely a consequence of complex interactions between both biotic and abiotic factors (Dupraz et al., 2009). However, much is still unknown regarding the effect of microbial community interactions on the formation of microbialites.
1.5 Major microbial taxa and their metabolic pathways that contribute to microbialite formation

Microbialite-forming ecosystems, which include microbial mats, contain highly diverse microbial consortia, including both cellular (bacterial, archaea and eukaryotes) and viral (capsid based) organisms.

1.5.1 *Alphaproteobacteria*

*Alphaproteobacteria* are abundant in marine, freshwater and artificial marine-derived microbialite-forming systems (Havemann and Foster, 2008; Breitbart *et al.*, 2009; Goh *et al.*, 2009; Khodadad and Foster, 2012) and cyanobacterial non-lithifying mats (Varin *et al.*, 2012). Pozas Azules II within Cuarto Cienegás and the high altitude, hypersaline arsenic-rich stromatolites of Lake Scompa within the Argentinean Andes have a high percentage of *Alphaproteobacteria* sequences (~12-15%), mainly belonging to the *Rhodobacterales* (Breitbart *et al.*, 2009; Fariás *et al.*, 2012). The rare hydromagnesite microbialites of Lake Alchichica and the oncolites of Rios Mesquites had cyanobacterial sequences that outnumber the *Alphaproteobacteria* sequences (Breitbart *et al.*, 2009; Couradeau *et al.*, 2011). However, prior amplification of the DNA was used before sequencing in both Lake Alchichica (PCR) and the Rios Mesquites (phi29-MDA), which may have biased the results.

*Alphaproteobacteria* contain members (e.g. *Rhodobacterales*) that are photoheterotrophic, which could influence the formation of microbialites by anoxygenic photosynthesis and microbialite dissolution through fermentation (Dupraz and Visscher, 2005; Figure 1.4). Anoxygenic photosynthesis would been the only type of photosynthesis within the Precambrian stromatolites prior to the origin of cyanobacteria, and alphaproteobacterial anoxygenic phototrophs could have filled that metabolic niche (Bosak *et al.*, 2007). *Alphaproteobacteria*
appear to be important for microbialite nitrogen fixation, even in the presence of cyanobacterial nitrogen fixation, likely because of diel cycles (Havemann and Foster 2008; Figure 1.4).

1.5.2    

**Cyanobacteria**

1.5.2.1    Planktonic cyanobacteria

*Cyanobacteria* in microbialites exist as planktonic cells and mat builders. The planktonic unicellular cyanobacteria are found in the surrounding water. Cyanobacterial calcification is well documented (Dittricha *et al.*, 2003; Obst *et al.*, 2009). *Synechococcus* sp., a model planktonic cyanobacterium, has been shown in multiple studies to precipitate carbonates using oxygenic photosynthesis (Thompson *et al.*, 1990; Dittricha *et al.*, 2003; Obst *et al.*, 2009; Figure 1.4). *Synechococcus* sp. is known to cause “whiting events,” in which calcium carbonate rains down through the water column (Thompson *et al.*, 1997). These “whiting events” are a photosynthetically driven alkalization process that precipitates calcium carbonate principally around the *Synechococcus* cells, which serve as nucleation points of precipitation (Dittricha *et al.*, 2003; Obst *et al.*, 2009). These planktonic cyanobacteria provide the “bricks”, in the form of either magnesium or calcium based carbonates, which are the building blocks of microbialites.

1.5.2.2    Filamentous cyanobacteria, mat-builders

If the planktonic cyanobacteria provide the “bricks” to make microbialites, then mat-building filamentous cyanobacteria provide the “mortar” that stabilizes the structure. Filamentous cyanobacteria are fundamental in modern microbialites, as they provide nutrients, structure, and directly act in the formation processes. Filamentous cyanobacteria, like their planktonic counterparts, influence carbonate precipitation through photosynthesis-induced alkalization (Dupraz and Visscher, 2005; Figure 1.4). Mat-building cyanobacteria produce extracellular polymeric substances (EPS) which trap Ca$^{2+}$ ions near dissolved CO$_3^{2-}$, and provide nucleation
points for precipitation (Dupraz and Visscher, 2005; Dupraz et al., 2009). Commonly found genera of filamentous mat building *Cyanobacteria* in microbialite-forming sites include *Nostoc, Pseudanabaena, Schizothrix, Dichothrix*, and *Lyngbya* (Laval et al., 2000; Dupraz et al., 2009).

Mat-building filamentous cyanobacteria also provide an essential energy source for heterotrophic bacteria in an oligotrophic water column (Lim et al., 2009). These heterotrophic bacteria in turn mediate further lithification and build-up of the microbialite structure (Omelon et al., 2013). Heterotrophs, release calcium and/or magnesium ions that are bound to cyanobacterial EPS mat by degradation (Dupraz and Visscher, 2005). This allows for more free interaction between cations and CO$_3^{2-}$ anions and leads to carbonate precipitation (Dupraz and Visscher, 2005). Phosphorus, nitrogen and various vitamins and co-factors are critical nutrients required for primary production by filamentous cyanobacteria; these nutrients can be sequestered in the mats and then released by heterotrophic degradation, which feeds further primary production. These cycles of autotrophic and heterotrophic mat breakdown are balanced; otherwise, the structures would likely dissolve and the mats would die. Through primary production, the microbialite cyanobacterial mat is an oasis in a desert-like oligotrophic water column.

### 1.5.3 Sulfur-oxidizing bacteria (SOBs)

Sulfur-oxidizing bacteria (SOBs) are a double-edged sword for microbialite formation, as they directly cause dissolution of microbialites, but byproducts of this dissolution provide intermediate substrates that allow for colonization by sulfate-reducing bacteria (SRBs; Figure 1.4). Microbialites are often found in highly oxygenic water (Dupraz et al., 2009), and SOBs remove high amounts of oxygen, allowing for further lithification via sulfate reduction by anaerobic sulfate-reducing bacteria (Dupraz and Visscher, 2005). SOBs are chemolithotrophs that promote
dissolution of microbialites through the oxidation of hydrogen sulfide, which in turn releases calcium ions from the microbialite calcium carbonate matrix (Figure 1.4).

The sulfur oxidation pathway encodes the Sox gene cluster that is responsible for aerobic sulfur oxidization (Friedrich et al., 2005). Sox genes are conserved across Proteobacteria, including Gamma-, Beta- and Alpha-proteobacteria (Friedrich et al., 2005). Beggiatoa sp., a ubiquitous filamentous sulfide-oxidizing Gammaproteobacterium, would be an excellent candidate for the study of the removal of oxygen from the highly oxygenated surface mat of microbialites (Dupraz and Visscher, 2005). The Sox pathway comprises four protein complexes SoxYZ, SoxAX, SoxB and SoxCD, which oxidize reduced sulfur species, including sulfide (HS\(^-\)), elemental sulfur (S), and sulfite (HSO\(_3\)-) (Friedrich et al., 2005). The diversity and abundance of SOBs, and whether Sox-related pathways are found in active microbialite-forming ecosystems, are still unclear. It is also unclear whether sulfur oxidation pathways are a signature of healthy (actively forming) or failing (actively dissolving) microbialites.

1.5.4 Sulfate-reducing bacteria (SRBs)

Sulfate-reducing bacteria (SRBs) are a crucial part of the heterotrophic microbial consortia responsible for lithification in microbialites (Visscher et al., 2000; Dupraz and Visscher, 2005; Gallagher et al., 2012; Figure 1.4). Sulfate reduction is an ancient metabolism, predating oxygenic photosynthesis (Canfield et al., 2000). Sulfate reduction was occurring in microbialites prior to the great oxidation event and the appearance of cyanobacterial mats (Canfield et al., 2000; Wacey et al., 2011; Bontognali et al., 2012; Figure 1.3).

SRBs reduce sulfate under anaerobic conditions, but also exist in active layers of highly oxygenated mats, which presents a metabolic paradox (Dupraz and Visscher, 2005; Dupraz et al., 2009). SRBs in microbialites must be either oxygen tolerant, find anoxic microenvironments, or
form tight consortia with oxygen-depleting anoxic organisms, such as SOBs (Dupraz and Visscher, 2005). Recent rate measurements suggest that SRBs within microbialites must be oxygen tolerant, as measurable amounts of dissimilatory sulfate reduction occurs in the highly oxygenated surface layers of the cyanobacterial mat (Baumgartner et al., 2006; Gallagher et al., 2012).

SRBs may mediate carbonate precipitation in microbialites by altering the saturation index and increasing alkalinity though dissimilatory sulfate reduction (Baumgartner et al., 2006; Gallagher et al., 2012; Figure 1.4). Active microbialite growth is thus suggested to be related to the heterotrophic breakdown of cyanobacterial-derived mats, which provide a rich and complex carbon source (Dupraz and Visscher, 2005). SRBs need electron donors for functional dissimilatory sulfate reduction: these include organic carbon, as acetate, lactate, ethanol, or formate, or molecular hydrogen (Baumgartner et al., 2006; Gallagher et al., 2012). The concept of lithifying mats and non-lithifying mats could be explained by the rates of dissimilatory sulfate reduction and related to the abundance of metabolically active SRBs (Dupraz and Visscher, 2005; Gallagher et al., 2012). However, this process and the underlaying mechanism are not well understood.

SRBs that use the highly conserved dissimilatory sulfate reduction (or dsr) pathway are mainly found among *Proteobacteria*, but are found also in non-proteobacterial lineages. Specifically, the majority of SRB-related taxa are within the class *Deltaproteobacteria*, and fall within the orders *Desulfovibrionales*, *Desulfobacterales*, and *Syntrophobacterales*. Non-proteobacterial lineages that have SRBs include the taxa *Nitrospira*, *Thermodesulfobacteria*, *Thermodesulfobioaceae*, and *Firmicutes* (mainly *Clostridia*). The dsr pathway consists of sulfate adenyltransferase (*Sat*), adenylylsulfate reductase with the A and B subunit (*apsAB*), and dissimilatory sulfite reductase with alpha and beta subunits (*dsrAB*) (Zhou et al., 2011). The
pathway transports sulfate (reactant) with an electron donor (organic carbon, acetate, lactate, formate, hydrogen and/or ethanol), and through various enzymatic steps yields sulfide (Zhou et al., 2011). The diversity and occurrence of the dsr pathway in microbialites is not known, but it is critical to our understanding of microbialite formation (Baumgartner et al., 2006; Gallagher et al., 2012).

1.5.5  **Firmicutes**

*Firmicutes* are Gram-positive bacteria that are abundant in a multitude of environments, including microbialites (Breitbart et al., 2009). *Bacillus* spp. were shown over forty years ago to produce carbonates as a “general phenomenon” (Boquet et al., 1973). *Firmicutes* are known to precipitate calcium carbonate with or without the presence of cyanobacteria, with at least one operon (*lcfA*) geared towards carbonate precipitation (Barabesi et al., 2007).

*Firmicutes* have many mechanisms for facilitating carbonate precipitation. Seven different *Bacillus* species are known to precipitate carbonates, including *B. subtilis* (Barabesi et al., 2007), *B. sphaericus* and *B. lentus* (Dick et al., 2006), *B. megaterium*, *B. cereus*, and *B. thuringiensis* (Dhami et al., 2013), and *B. amyloliquefaciens* (Lee, 2003). *Bacillus* spp., are able to induce carbonate formation through the metabolism of urea (Hammes et al., 2003), heterotrophic ammonification of amino acids (Castanier et al., 1999), carbonic anhydrase (Dhami et al., 2014), EPS formation (Ercole et al., 2007) and the *lcfA* operon (Barabesi et al., 2007). Even at low abundance, members of the *Firmicutes* have metabolic versatility geared towards carbonate precipitation, which may have a role in the formation of microbialites (Paerl et al., 2001; Breitbart et al., 2009; Omelon et al., 2013).

Microbial mats including microbialites have colored pigmented layers (orange to yellow) (Bottos et al., 2008). Little is known about non-photosynthetic pigmented heterotrophic bacteria
in microbialites, although many studies have focused on the oxygenic phototrophs in these layers, and suggested the coloration is related to carotenoid pathways (Nübel et al., 1999; Lionard et al., 2012). Diversity studies (based on 16S rDNA) have found firmicutes in these layers (Bottos et al., 2008; Lionard et al., 2012), and firmicutes synthesize carotenoids that give rise to many different colors of bacteria, including orange (Köcher et al., 2009; Klassen, 2010). To the best of our knowledge, no study has screened for cultivatable isolates of non-photosynthetic, pigmented, heterotrophic members of the Firmicutes that could be responsible for the colored layers in microbial mats and microbialites.

1.5.6 Actinobacteria

Representatives of the phylum Actinobacteria appear to be rare within microbialite-forming ecosystems, and little is known about their abundance, diversity and role. Actinobacteria have been found in smooth and pustular pre-stromatolithic mats in Shark Bay and the dry freshwater stromatolites of Ruidera Pools, Spain (Allen et al., 2009; Santos et al., 2010). Their presence in microbialites does not entail a functional role, but the knowledge gap points to a need for both culture-dependent and independent studies.

Actinobacteria have also been identified in the pigmented layers in microbial mats (Bottos et al., 2008; Lionard et al., 2012), and it has been suggested they may be responsible for the coloration (Klassen, 2010). Although carotenoid biosynthesis also occurs in actinobacteria, to best of our knowledge no study has examined isolates of non-photosynthetic pigmented heterotrophic members of the Actinobacteria to determine if they may contribute to the colored layers found in microbial mats and microbialites.
1.5.7 **Eukaryotes**

Eukaryotes in microbialites are present in low abundance (<1%), by physical counts (Power *et al.*, 2011a) and by sequence abundance (Couradeau *et al.*, 2011; Khodadad and Foster, 2012). Even at low abundance, metazoans have been associated with the global decline of microbialites (Awramik, 1971; Garrett 1970, Figure 1.3). Metazoan grazing following the Cambrian explosion has been correlated to a drop in stromatolite diversity about ~500 to 800 Mya (Awramik, 1971; Garrett 1970, Figure 1.3). However, the metazoan grazing hypothesis is hindered by a lack of fossil evidence; as well, protist grazers presumably predated the Cambrian microbialites far earlier (Bernhard *et al.*, 2013). Another major problem is that unicellular or soft body grazers would leave no obvious fossils, possibly contributing to a virtually nonexistent fossil record (Bernhard *et al.*, 2013).

The question remains, why do modern microbialites exist if grazing is the cause of their decline? This is puzzling, as metazoan grazing occurs even under the oligotrophic conditions in which many microbialites are found (Elser *et al.*, 2005). Possibly, metazoan grazing is controlled by abiotic factors such as low nutrients, high temperature, high salinity, and high pH. Cyanobacteria make cytotoxic compounds that can inhibit and kill metazoans, thus cyanobacterial mats in microbialites could also biotically control metazoan grazing (Neilan *et al.*, 2013; Wu *et al.*, 2014). The type of microbial mat that exists on the surface of a marine microbialite is suggested to alter both the microbialite fabric (whether stromatolithic or thrombolithic) and to control the metazoan community structure (Tarhan *et al.*, 2013). Thus the microbial community also likely plays a role in controlling metazoan grazing (Tarhan *et al.*, 2013).

In contrast, photoautotrophic eukaryotes (e.g. algae and other protists) could aid in the organosedimentary formation process via photosynthesis-induced carbonate precipitation (Figure
1.4). Even at low sequence abundances of eukaryotes (<1%) in marine microbialite ecosystems (Highbourne Cay, Bahamas), it is suggested that foraminifera influence the microfabric by stabilizing carbonate grains, and their reticulopods, once encased, act as sheaths that effect the lamination of microbialites (Bernhard et al., 2013; Edgcomb et al., 2013). Thus, foraminifera may help to explain the changes in the microbialite microfabric observed in the late Precambrian (Bernhard et al., 2013). In modern microbialites, foraminifera are suggested to move carbonate grains during diel vertical migration, disrupting lamination critical for stromatolite formation and promote clot formation or thrombolithic structure (Bernhard et al., 2013). Foraminifera were found in microbialites in the Highbourne Cay, Bahamas, including thrombolites and stromatolites; however, the highest diversity was observed in thrombolite mats or mats that were biologically turbated (Bernhard et al., 2013; Edgcomb et al., 2013). Metazoans, Stramenopiles, Alveolata, Amoeboza and Rhizaria were found to be active metabolically in various microbialites across Highbourne Cay and Shark Bay (Edgcomb et al., 2013).

1.5.8 Viruses

The role of viruses in microbialite communities has remained elusive. Viruses are the most prevalent “organisms” on Earth, with an estimated $10^{30}$ viruses in the ocean, and around $10^6$ - $10^7$ viruses ml$^{-1}$ in most freshwater systems (Suttle, 2005). Viruses are agents of cell lysis and thus play a role in carbon cycling on a global scale (Suttle, 2005). Additionally, new research has found that many viruses harbor auxiliary metabolic genes, including genes that encode the core proteins involved in photosynthesis ($psbA, psbD$) and phosphate starvation ($phoH$) (Clokie et al., 2010; Thompson et al., 2011). Viral auxiliary metabolic genes have also been shown to redirect their host’s metabolic potential to fuel viral replication (Clokie et al., 2010; Thompson et al., 2011). Some marine cyanophages encode a Calvin cycle inhibitor ($cpl2$) that inhibits the carbon-fixation
actions of Rubisco (Ribulose-1,5-bisphosphate carboxylase/oxygenase), and a virally-encoded transaldolase \((talC)\) redirects the host metabolism to the pentose phosphate pathway, benefiting viral replication (Clokie et al., 2010; Thompson et al., 2011). This evidence suggests that viruses affect the carbon metabolism of an ecosystem, and may influence the chemistry of microbialite formation or dissolution.

1.6 **Previously studied microbialite-forming ecosystems**

Shark Bay in Western Australia, Highborne Cay in Bahamas, and the Cuatro Cienegás basin of Mexico, including Pozas Azules II and Rios Mesquites, are the most studied microbialite-forming assemblages. Metagenomic surveys of the microbial fraction and the viral fraction has been completed for Highborne Cay, and Cuatro Cienegás microbialites. These sites will provide a valuable reference for comparison with the microbialites from Pavilion Lake and Clinton Creek.

1.6.1 **Shark Bay microbialites**

The oldest fossil microbialites are found in Western Australia, dating back ~3.5 Gya (Van Kranendonk et al., 2008). Near these fossil microbialites are active, modern versions within Shark Bay at Hamelin Pool in Western Australia. These are the main example of hypersaline marine microbialites (Goh et al., 2009; Allen et al., 2009). Shark Bay microbialites range in morphology from columnar, club, spheroidal, domal, nodular, to irregular-shaped morphotypes (Burns et al., 2004).

Shark Bay microbialite diversity has been assessed by three 16S rDNA amplicon studies (Burns et al., 2004; Goh et al., 2009; Allen et al., 2009). Currently, only 412 rDNA sequences are available on Genbank from all bacterial surveys of the Shark Bay microbialites. These sequences reveal the distribution of the various bacterial taxa, indicating that *Alphaproteobacteria* (20%), *Actinobacteria* (16%), and *Cyanobacteria* (15%) dominate, with the rest of the sequences
belonging to Planctomycetes (11%), Gammaproteobacteria (10%), and the Firmicutes (9%) (Burns et al. 2004; Papineau et al., 2005; Allen et al., 2009; Goh et al. 2009). Papineau et al. (2005), concluded that Shark Bay microbialites are highly diverse, with 90% of sequences belonging to bacteria and that Shark Bay microbialites have novel Proteobacteria representing 28% of the sequences, followed by Planctomycetes (17%), and Actinobacteria (14%), respectively.

Shark Bay is novel compared to other microbialite-forming ecosystems as it has a relatively high abundance of archaea (~10% of 16S rDNA sequence), which are usually <1% of the microbial community in other microbialites (Burns et al. 2004; Papineau et al., 2005; Allen et al., 2009; Goh et al. 2009). The archaea in Shark Bay microbialites are principally halophiles (Papineau et al., 2005; Allen et al., 2009; Goh et al. 2009). Many unclassified haloarchaea have been detected only in Shark Bay microbialites and not in the surrounding seawater, indicating that Shark Bay microbialites are a potential reservoir for these archaea (Papineau et al., 2005; Allen et al., 2009; Goh et al. 2009). Currently, no metagenomic data are available for Shark Bay microbialites, and only 16S rDNA based studies on prokaryotic diversity (Burns et al. 2004; Papineau et al., 2005; Goh et al. 2009), and a RNA-based study on active eukaryotes (Edgcomb et al., 2013), are available for comparison.

1.6.2 Highbourne Cay microbialites

Like Shark Bay, Highbourne Cay microbialites are found in shallow depths. They have a high diversity of colored mats covering the microbialites, ranging from pink, green, brown and black (Reid et al., 2000; Myshrall et al., 2010; Mobberley et al., 2012). Highbourne Cay microbialite mats have been classified into three types (I - III) based on morphology (Reid et al., 2000). Type I mats are considered young and are dominated by filamentous cyanobacteria with few heterotrophic bacteria. They have low diversity (~128 OTUs) and low sulfate reduction,
photosynthesis, and aerobic respiration rates, and have been described as non-lithifying mats (Reid et al., 2000; Myshral et al., 2010; Khodadad and Foster, 2012; Mobberley et al., 2012). Type II and III mats are considered older lithifying mats with the beginning of carbonate lamination; both have higher rates of sulfate reduction, photosynthesis, aerobic respiration rates than type I mats (Reid et al., 2000; Myshral et al., 2010; Khodadad and Foster, 2012; Mobberley et al., 2012). Type II and III mats have higher bacterial diversity (~181 OTUs) than type I mats (Reid et al., 2000; Myshral et al., 2010; Khodadad and Foster, 2012; Mobberley et al., 2012).

Previous bacterial diversity studies based on 16S rDNA amplicons of Highbourne Cay microbialites reveal that bacteria assigned to the *Proteobacteria* dominate all three mat types (52.0%), followed by *Planctomycetes* (12.1%), *Cyanobacteria* (11.6%) and *Bacteroides* (10.4%) (Myshral et al., 2010; Mobberley et al., 2012). Within the *Proteobacteria*, *Alphaproteobacteria* are dominant, making up 42% of the proteobacterial sequences, followed by 4.9% for both the *Delta*- and *Gamma*-proteobacterial lineages (Myshral et al., 2010; Mobberley et al., 2012). The surrounding water communities differ from those in the stromatolitic mats I-III, with more taxa assigned to the *Alphaproteobacteria* and *Gammaproteobacteria* in the water, and no *Deltaproteobacteria* detected (Myshral et al., 2010; Mobberley et al., 2012). Like Shark Bay, *Proteobacteria* are the dominant group in Highbourne Cay, with *Alphaproteobacteria* being the most abundant OTU in the pink thrombolithic mats (Myshral et al., 2010; Mobberley et al., 2012).

Metagenomes for both the microbial fraction and the viral fraction have been published for Highbourne Cay (Desnues et al., 2008). However, because the surrounding environment was not sampled, including the water, it is hard to know which taxa are specific to the microbialite. Also, multiple displacement amplification (MDA) was used to obtain DNA for sequencing, which is
known to affect quantitative results due to amplification bias (Abulencia et al., 2006; Yilmaz et al., 2010; Abbai et al., 2012).

The Highbourne Cay viral metagenome was dominated by ssDNA microphages, likely due to MDA bias towards circular templates, which enriches for these viral genotypes and skews quantitative results (Kim and Bae, 2011). Moreover, these phage are present in the surrounding seawater (Desnues et al., 2008). Nonetheless, some ssDNA viral genotypes in Highbourne Cay were found to be highly specific, suggesting that microbialites act as islands of viral diversity (Desnues et al., 2008). Highbourne viral metagenomes had contigs that were similar to those in the Sargasso Sea or Cuatro Cienegás. No Caudovirales phage sequences were detected, which is unusual, and may be the results of bias from MDA (Desnues et al., 2008).

1.6.3 **Cuatro Cienegás microbialites: Pozas Azules II and Rios Mesquites**

Cuatro Cienegás, or the “four marshes,” is found in the northern Mexican state of Coahuila. Cuatro Cienegás is home to Pozas Azules II or the “blue pools,” which have thrombolites, and Rios Mesquites (“mesquite tree rivers”), which have oncolites (Breitbart et al., 2009). Cuatro Cienegás has metagenomic data sets for both the viral and microbial fractions, but not samples from the sediment or water (Desnues et al., 2008; Breitbart et al., 2009).

Comparing the viral metagenomic data in Highbourne Cay and Cuarto Cienegás suggests that microbialites act as islands of novel viral genotypes and reflect an ancient marine origin that has persisted in a freshwater ecosystem (Desnues et al., 2008). However, high abundances of ssDNA viruses within both systems are an artifact of the bias of MDA towards circular templates, which enriches for ssDNA sequences and skews quantitative results (Abulenica et al., 2006; Yilmaz et al., 2010; Kim and Bae, 2011; Abbai et al., 2012). Pozas Azules II is similar to
Highbourne Cay, but has more dsDNA phage sequences, including sequences that are similar to common marine phages, which could be due to cross-contamination from marine samples.

1.7 Microbialite-forming ecosystems described in this thesis

Pavilion Lake and Clinton Creek are freshwater microbialite sites that have not been analyzed for metabolic potential or microbial community structure. Both sites are northern latitude freshwater microbialite-forming ecosystems that are very different from previously studied tropical marine microbialite systems. Described below is what is known about microbialites from these sites, including their morphology, geological classification, abiotic chemistry, and morphologically classified organisms.

1.7.1 Pavilion Lake

Pavilion Lake (50.8°N, 121.7°W) lies in Mable Canyon between Squamish and Lillooet, BC, and contains active thrombolithic microbialites (Laval et al., 2000). Pavilion Lake thrombolithic microbialites range from ~2,000 to 12,000 years in age, with the corrected age maximum at ~4,600 years and the uncorrected age maximum at ~12,300, based on Uranium-Thorium dating (Laval et al., 2000). These age estimates put the initial formation of thrombolites in Pavilion Lake after the end of the last ice age (Laval et al., 2000).

Pavilion Lake is a dimictic, alkaline (pH 8.4), oligotrophic (mean total phosphorus - 3.3 μg L⁻¹), freshwater lake with hard water (mean calcium carbonate - 182 mg L⁻¹) (Laval et al., 2000; Lim et al., 2009; Brady et al., 2013; Omelon et al., 2013). Pavilion Lake microbialites consist primarily of calcite thrombolites, which change in morphotype as a function of depth, as follows: shallow domes resembling “open lettuce-beds” at ~10 m, intermediate domes resembling “cabbage heads” at ~20 m, intermediate-deep domes with hollow conical “asparagus-like” outcroppings at
25 m, and deep domes that possess “artichoke leaflets” at ≥25 m (Laval et al., 2000; Brady et al., 2013; Omelon et al., 2013; Figure 1.5).

Morphological characteristics of the cyanobacterial community in Pavilion Lake microbialites have been described (Laval et al., 2000). The microbialites have thin (>5 mm) cyanobacterial-derived mats, comprised of coccoids (Gloeocapsa spp.), rods (Synechococcus spp.), and filamentous cyanobacteria (Fischerella spp. and Pseudoanabaena spp.) (Laval et al., 2000). The water column of Pavilion Lake is dominated by Synechococcus spp. (8.3 x 10^4 cells ml⁻¹), the picoplankton known for causing carbonate precipitation “whiting events,” (Thompson et al., 1997).

Pavilion Lake microbialites have round filamentous balls of cyanobacteria that are >2 mm at 15 to 20 m, which decrease in size with depth (Figure 1.6, Brady et al., 2010). These filamentous balls of cyanobacteria are called “nodules,” which range in pigmentation from purple to green to black. The purple cyanobacterial nodules were morphologically classified as either Oscillatoria sp. or Calothrix sp. (Laval et al., 2000). Recent clone library analysis suggests that the green nodules are Leptolyngbya sp. and the purple nodules are Tolypothrix sp. (Russell et al., 2014). Tolypothrix spp. and Calothrix spp. are morphologically similar and their phylogenetic relationships are not well understood (Berrendero et al., 2011). Green and purple nodules create a microenvironment that has a higher pH and higher dissolved oxygen, and has active lithification of calcite within the filaments (Brady et al., 2010). The nodules are only found at 15 to 25 m depths, which makes it unlikely that they are major players in the formation of the overall microbialite structure, otherwise they would be found at all depths.

The diversity and abundance of microbial taxa and the metabolic potential of Pavilion Lake microbial communities are unknown. Only a single metagenomic study is available for freshwater
microbialites, which did not account for the surrounding environment (e.g. sediments and water column) (Breitbart et al., 2009). Is the microbialite microbial community and metabolic potential the same or different from that in the surrounding environment? This knowledge will either support previous studies that microbial community and metabolic potential are microbialite-specific or will provide evidence of microbes from the surrounding environment (Breitbart et al., 2009; Khodadad and Foster, 2012; Mobberley et al. 2013).

1.7.2 Clinton Creek

Clinton Creek has a subarctic climate and is located near the Arctic Circle, ~77 km northwest of Dawson City, Yukon, Canada (64°26'42"N and 140°43'25"W). It is a flooded open pit from an abandoned asbestos mine, which was active from 1969 to 1978 (Power et al., 2011a). Clinton Creek has served as a model research site for carbon sequestration studies within industrial mining sites (Power et al., 2011a). Sediment mineralogy is highly variable, containing quartz, muscovite, kaolinite, and chrysotile, as well as minor aragonite and trace calcite (Power et al., 2011a). Aragonite will preferentially lithify abiotically over calcite due the inhibiting effect of the high magnesium to calcium ratio (Power et al., 2011a). Clinton Creek microbialites have a columnar thombolithic morphology that is primarily composed of aragonite with spherulitic texture (Power et al., 2011a). The water in the open pit is subsaline (>3.0 g L\(^{-1}\)), alkaline (average pH 8.4), and supersaturated with aragonite (Power et al., 2011a). It is oligotrophic (undetectable phosphorus), has low-iron (~0.04 mg L\(^{-1}\)) and light penetrates to the bottom (Power et al., 2011a). The water is very hard (>200 mg L\(^{-1}\) and >75 mg Ca L\(^{-1}\)) and has high alkalinity (>190 mg HCO\(_3^-\) L\(^{-1}\)) (Power et al., 2011a). There is minimal nutrient input into the open pit due to the lack of surrounding soil and human activity over the last 35 years (Power et al., 2011a).
Power et al. (2011a) were the first to microscopically and isotopically characterize the microbial community of the Clinton Creek microbialite: it was morphologically described to have benthic diatoms (e.g. *Brachysira* spp.), filamentous algae (e.g. *Oedogonium* spp.), dinoflagellates and cyanobacteria. Power et al. (2011a) noted the low microbial growth rates and suggested that the low sediment input had little influence geologically, but did not examine microbe-sediment interactions.

The microbial community structure and metabolic potential of the Clinton Creek microbialites and their surrounding environment are uncharacterized, as are microbialite systems from cold regions in general. As well, the metabolic potential of freshwater microbialites, especially from cold environments, is mostly uncharacterized; only one freshwater microbialite metagenomic dataset exists, which is from a tropical environment and which was biased by using MDA for DNA amplification (Breitbart et al., 2009). Prior metagenomic studies, whether marine or freshwater, did not examine the composition or metabolic potential of the adjacent sediments or surrounding waters; such sampling would provide a context for identifying the microbes and processes that are confined to the microbialites. In Clinton Creek, the microbialite and sediment communities were compared to determine their similarities and differences and which biotic carbonate precipitating pathways, whether photosynthetic or heterotrophic, are enriched in the microbialites over the sediments.

### 1.8 Thesis hypotheses

Outlined below are the hypotheses considered in this thesis and the chapters where they are addressed.

Hypothesis [2] Modern freshwater microbialites in Clinton Creek have a core functional potential and different pathways relating to carbonate precipitation than in the nearby sediments. Presented in Chapter 3.

Hypothesis [3] Agrococcus sp. strain RW1 represents a new species of Actinobacteria that has genomic features associated with carotenoid biosynthesis leading to potential for yellow colony pigmentation. Presented in Chapter 4.

Objective [4] Exiguobacterium sp. strain RW2 represents a new species of Firmicutes that has genomic features associated with carotenoid biosynthesis to potential for orange colony pigmentation. Presented in Chapter 5.
Figure 1.1: Microbialite morphology and carbonate macrofabric layout
Adapted from Riding et al., 2011.
Figure 1.2: Microbialite morphology examples
A) A fossilized stromatolite (~505 Mya). Burgess shale in British Columbia. Photo courtesy of Dr. David Turner, Beaty Biodiversity with permission. Gray scale bar 7 cm.
B) A close up of the fossilized stromatolite (10x from scale bar).
C) Clinton Creek thrombolite from Clinton Creek, Yukon Canada. Photo courtesy of Dr. Ian Malcolm Power (University of British Columbia) with permission. Gray scale bar 3 cms.
D) A close up of the Clinton Creek thrombolite (10x from scale bar).
Figure 1.3: Timeline of microbialite and global atmospheric oxygen concentration over geological time
The figure was adapted from Canfield et al., 2005 and Dupraz and Visscher, 2005. The oxygen data was provided by Dr. Sean Crowe (University of British Columbia) with permission, including data from Lyons et al., 2014 and Crowe et al., 2013. Oxygen data is expressed as %PAL or present atmospheric levels.
Figure 1.4: Generalized equations for microbialite formation and dissolution
Includes examples of key taxa involved or substrates (ethanol only) during light (day) and night (dark) cycles. Adapted from Dupraz and Visscher, 2005 and Dupraz et al., 2009.
Figure 1.5: Pavilion microbialite morphologies
Red scale bar is ~10 cm across for both.
Photo courtesy of Donnie Reid (Nuytco reseach) and Tyler Mckay (UC Davis) with permission.
Figure 1.6: Pavilion Lake nodular cyanobacterial morphologies
Scale of the photo on the right at ~10 cm across.
Photo courtesy of Tyler Mckay (UC Davis) with permission.
Chapter 2:
Modern freshwater microbialite microbial community structure and metabolic potential is distinct from the surrounding environment

2.1 Introduction

Microbialites are thought to form through carbonate precipitation by a benthic microbial community and through trapping of detrital sediment (Burne and Moore, 1987; Dupraz and Visscher, 2005). This suggests that microbialites would comprise microbial communities from detrital sediments and surrounding water. Previous studies have not compared the microbial communities in microbialites with those of the sediments or water (Breitbart et al., 2009; Khodadad and Foster, 2012; Mobberley et al. 2013). Determining the composition and metabolic potential of these microbial communities will allow the identification of microbialite-specific components.

Pavilion Lake is a temperate freshwater system that houses large diversity of microbialite structures (Laval et al., 2000; Lim et al., 2009). Isotopic data suggest that photosynthetic induced alkanilization is a major contributor to carbonate precipitation in Pavilion Lake microbialites. However, the metabolic potential, whether primarily photosynthetic or a mixture of heterotrophic and photosynthetic processes, is unknown (Brady et al., 2009). Radiocarbon analysis shows that the microbialite-associated microbial community is using atmosphere-derived dissolved inorganic carbon (DIC) (Brady et al., 2009). Moreover, elevated $\delta^{13}C$ carbonate values within surface microbial mats and nodules from microbialites at < 20 m depth indicate photosynthetic carbonate precipitation (Brady et al., 2010; Brady et al., 2013). Although photosynthetic processes appear to be important in microbialite formation, the overall metabolic potential of microbialite communities in Pavilion Lake has not been investigated. In this contribution a metagenomic approach was used
to uncover the metabolic potential that is specifically associated with microbialite formation, and to distinguish it from that in microbial communities in the adjacent sediments and water.

As mentioned in Chapter 1, the only metagenomic data sets available for freshwater microbialites are for a tropical system; other metagenomic studies on microbialites are marine (Breitbart et al., 2009; Khodadad and Foster, 2012; Mobberley et al., 2013). The freshwater microbialite metagenomic study of Cuatro Cienegas basin in Mexico (e.g. Pozas Azules II and Rios Mesquites) used multiple displacement amplification to obtain DNA, which can negatively affect quantitative results due to amplification bias (Abulencia et al., 2006; Breitbart et al., 2009; Yilmaz et al., 2010; Abbai et al., 2012). In this study, unamplified total genomic DNA from microbialites and their surrounding environments was sequenced, and used a metagenomic approach to identify constituent taxa and infer their metabolic potential. By examining taxonomic differences and representative genes across diverse metabolic pathways, we can explore the complex interface between microbialite communities and their environment.

2.2 Materials and methods

2.2.1 Sample collection

Triplicate representative microbialites (~10 kg), adjacent water (~100 L, at 10 m, 20 m and 45 m depths), and triplicate sediment samples (~20 g, at 10 m, 20 m, 25 m depths) were collected by divers and by pumping water from Pavilion Lake (50.86°N, 121.74°W) in 2010-11. The water was filtered in series through 120 µm pore-size Nitex screening to remove large plankton, and 1.2 µm pore-size glass-fiber, and 0.45 µm and 0.22 µm pore-size Durapore polyvinylidene difluoride (PVDF) filters (Millipore).
2.2.2 DNA extraction

A sterile razor blade was used to scrape off 3-10 mm (~5 g) from the surface of three morphologically similar microbialites collected at each depth, from which DNA was extracted on-site using a Powersoil Kit (Mobio, Carlsbad, CA). Samples were frozen on-site in liquid nitrogen and placed at -80 °C when returned to the laboratory. Frozen samples were extracted with CTAB (Untergasser, 2008). For the filters, half of each glass-fiber and 0.22 µm pore-size filter was extracted using a PowerWater Kit (Mobio, Carlsbad, CA) and CTAB (Untergasser, 2008). Sediment samples were pelleted by centrifugation (5000 x g) and the DNA extracted from a ~5 g subsample using a PowerSoil Kit (Mobio, Carlsbad, CA) and CTAB (Untergasser, 2008). DNA concentrations were determined on-site using a Nanodrop-3300 (ThermoFisher, Wilmington, DE) with PicoGreen® reagent according to the manufacturer’s instructions (Invitrogen, Carlsbad, CA). Purity was determined by absorbance (260/280 and 260/230) using a Nanodrop-1000 (ThermoFisher, Wilmington, DE). CTAB and MoBio DNA extractions were pooled (50:50) by equimolar DNA to reduce extraction bias and then used for library construction. Surface and 10 m water filter DNA was pooled for metagenomic library construction.

2.2.3 Metagenomic library preparation (454 FLX Titanium and Illumina)

Libraries for 454 FLX Titanium were constructed using random DNA shearing with a Bioruptor (Diagenode Denville, NJ). Fragments were polished and blunt-end ligated (NEBNext, DNA prep kit, Ipswich, MA) to in-house Multiplex Identifier barcode oligos (IDT, Coralville, Iowa), with small fragments removed by magnetic beads (Beckman Coulter, Danvers, MA). The libraries were quantified using a digital PCR quantified standard curve (White 3rd et al., 2009), diluted, and pooled for 454 pyrosequencing with Titanium chemistry (Sick Kids Hospital, Toronto, Ontario).
For Illumina library construction, DNA was sheared by ultrasonication (Covaris M220 series, Woburn, MA), and the fragments end-paired, A-tailed (Lucigen NxSeq DNA prep kit, Middleton, WI) and ligated to TruSeq adapters (IDT, Coralville, Iowa); small fragments were removed twice using magnetic beads (Beckman Coulter, Danvers, MA). The resulting libraries were pooled, and both 250 bp and 100 bp paired-end sequenced on MiSeq (GenoSeq UCLA Los Angeles, CA) and HiSeq (McGill University/Génome Québec, CAN) platforms, respectively. Replicate samples were barcoded separately, then pooled for data analysis. The replicates were pooled based on data from tRFLP analysis that showed samples from a given depth had very similar taxonomic profiles. Pooling provides similar numbers of contigs across depths.

2.2.4 Metagenomic data assembly and analysis

The raw sequencing data were processed as follows: for the 454 data, the raw SFF files were converted to FASTQ format and binned by molecular barcode (MID) using a custom Perl script. Barcodes were removed by Tagcleaner (Schmieder et al., 2010) and sequences cleaned for low quality and homopolymers using PRINSEQ (Schmider and Edwards, 2011). The Illumina data were extracted and demultiplexed using the CASAVA pipeline v1.8 (Illumina, San Diego, CA), and the PhiX spike-in used for sequencing quality control was screened using Bowtie2 (version 2.1.0) then removed using Picard tools (version 1.90) (White 3rd and Suttle, 2013c).

The resulting 454 FLX titanium reads, Illumina overlapping merged reads, and Illumina non-overlapping reads from replicate libraries were combined and assembled (kmer size: 39) using the Ray DeNovo assembler (Boisvert et al., 2010; 2012). Metagenomic rapid annotations using a subsystems technology (MG-RAST) server was used for contig annotation (Meyer et al., 2008). MG-RAST annotation of the contigs used BLAT (BLAST-like alignment tool) annotations based on hierarchical classification against SEED subsystems and RefSeq databases with a minimum E-
value cutoff of $10^{-5}$, a minimum percent identity cutoff of 60%, and a minimum alignment length
cutoff of 15 base pairs.

Statistical analysis was completed using statistical analysis of metagenomic profiles
(STAMP) and R (Parks et al., 2010). STAMP and R were used to parse MG-RAST data for RefSeq
(class level) and SEED subsystems (level I to function) results. The STAMP ANOVAs (including
PCA) were completed using multiple groups, post-hoc tests (Tukey-Kramer at 0.95), an effect size
(Eta-squared) and multiple test correction using Benjamini-Hochberg FDR (false discovery rate)
procedure. SEED subsystems level III scatter plots for samples from microbialites, sediments and
filters were constructed in STAMP using a one-sided Welch's t-test with Welch's inverted
confidence interval (0.95), with Benjamini-Hochberg FDR. The top 25 RefSeq classifications were
normalized for each sample using count-relative abundances. PCA analysis for the normalized top
25 RefSeq classifications used R libraries Ecodist (Dissimilarity-based functions for ecological
analysis), and pvclust (Hierarchical Clustering with P-Values via Multiscale Bootstrap
Resampling) using ward clustering and the Bray-Curtis distance metric at a thousand replicates
(Suzuki and Shimodaira, 2006). The PCA for the normalized top 25 RefSeq classifications was
plotted using R library ggplot2 and a dotplot was done using R libraries Reshape2 using the melt
function, then plotted using ggplot2 (Wickham, 2009).

FragGeneScan was used to translate contigs into protein ORFs (Rho et al., 2010) and
ProPas was used to predict protein isoelectric points (pI) (Wu and Zhu 2012). These data were
further analyzed with R (R development Core Team 2013) then visualized using the R library
ggplot2 (Wickham, 2009).
2.2.5 **Recruitment analysis to draft genomes**

Metagenomic read recruitment from the Pavilion Lake 20 m microbialite metagenome to determine the abundance of *Agrococcus* sp. strain RW1 and *Exiguobacterium* sp. strain RW2 from their isolated environment (White 3rd *et al.*, 2013a-b). To determine genotypic-specific abundance we also recruited metagenomic reads to closely related genomes, which included *Agrococcus lahaulensis* strain DSM17612 (unpublished NCBI download), *Exiguobacterium* sp. strain S17 (Belfiore *et al.*, 2013; Ordoneza *et al.*, 2013), and *Exiguobacterium* sp. strain AT1b (Vishnivetkaya *et al.*, 2011). Metagenomic recruitment analysis used ~7.5 M reads pooled from HiSeq (~100 bp), MiSeq (~250 bp) and 454 FLX Titanium (~300 bp) sequencing platforms from 20 m Pavilion Lake microbialites only. EMBOSS union script was used to find overlapping contigs and merge contigs for draft genomes into a single scaffold that may not be contiguous; however, it yields a single genomic contig, which is required for metagenomic read recruitment (Rice *et al.*, 2000; Kunin *et al.*, 2008a). FR-hit program was used for metagenomic recruitment for the *Exiguobacterium* and *Agrococcus* genomes using default parameters with a minimum identity >70% and an E-value >1e^-5 (Nsiu *et al.*, 2011), the recruitments were then visualized with the R library ggplot2 (Wickham, 2009).

2.2.6 **Metagenomic data depositing**

All the data used in this study are publicly accessible from MG-RAST. From MG-RAST, it is listed in the project name Pavilion Lake surrounding environment under names PLsfcFil (ID 4532785.3), PL20Fil (ID 4532783.3), PL45Fil (ID 4532784.3), PLMB10 (ID 4532771.3), PLMB20 (ID 4532772.3), PLMB25 (ID 4532774.3), PLMB45 (ID 4532775.3), PL10Sed (ID 4526738.3), PL20Sed (ID 4526739.3) and PL25Sed (ID 4526740.3).
2.3 Results and discussion

2.3.1 Metagenomic assembly statistics

This study combined data from 454 FLX titanium pyrosequencing with Illumina MiSeq and HiSeq data, as Illumina sequencing compensates for the error-prone homopolymers of 454, while 454 compensates for Illumina's GC bias and substitution errors (Bentley et al., 2008; Ross et al., 2013). A total of 446 Mbp of assembled contigs from the microbialites and 17 Mbp of assembled contigs from the filters (Table 2.1) allowed differences in microbial structure and metabolic potential to be compared between the microbialites and surrounding water. The sediment metagenomes failed to assemble using Ray (Boisvert et al., 2010; 2012), Velvet (Zerbino and Birney, 2008), MetaVelvet (Namiki et al., 2012) or AbySS (Simpson et al., 2009) (data not shown). Paired-end reads from the sediments only were extended for overlap and pooled with unextended reads for further analysis (Table 2.1). In the future, longer reads will allow for better assemblies of metagenomic data (Wommack et al., 2008), aided by new sequencing technologies such as Pacific Biosciences (Rasko et al., 2011) or the Oxford nanopore, which promise reads as high as 50 kb (Loman et al., 2012).

2.3.2 Microbialites communities differ from the surrounding environment

The microbial community structure and metabolic potential of Pavilion Lake microbialites were distinct from the surrounding environment, based on GC content, RefSeq taxonomic classifications, including SEED and MetaCyc functional gene assignments. Across depths, the GC content in microbialites differed from that in adjacent waters and sediments; in contrast, predicted protein isoelectric content (pI) was not as useful a distinguishing metric. (Figure 2.1). The GC content across microbialites as a function of depth was very similar, implying that the microbial communities were similar.
Supporting the results from GC content analysis, the RefSeq classification also showed that the microbialite communities were collectively different from those in the water and sediments. STAMP ANOVA on the RefSeq classifications identified thirteen bacterial and seven archaeal classes that were significantly enriched in microbialites over the surrounding environment (Table 2.2, p < 0.01). PCA based on RefSeq classifications also indicated that the metagenomes from microbialite was distinct from those in the water and sediment (Figure 2.2A). These analyses demonstrated that the microbialite communities are distinct from those of neighboring sediments and water.

The predicted metabolic potential based on SEED and MetaCyc functional gene assignments are also consistent with a microbialite community that is distinct from those in the sediments and water. ANOVA using STAMP on the highest level classification (Level I) in the SEED database indicates that membrane transport, aromatic metabolism, motility, potassium metabolism, cell signaling and virulence genes are significantly enriched in microbialites over the neighboring environment (Table 2.3, p < 0.05). At the next level of metabolic organization within the SEED subsystems (Level II), ANOVA predicted more than 90 metabolic features that were overrepresented in the microbialite metagenomes; PCA using these data also indicates that the metabolic potential of the microbialites is distinct from that of the surrounding environment (Figure 2.3A). Welch's one sided t-test cluster analysis using >300 level III SEED subsystems indicates a weak correlation between the functional genes in the microbialites and sediments ($R^2 = 0.546$), but a stronger correlation between the microbialite and water as indicated by the data from the filters ($R^2 = 0.871$) (Figure 2.3C). PCA and NMDS analysis of MetaCyc features supports the SEED PCA, revealing that microbialites are distinct from the sediments and water (Figure 2.4).
Overall, the analyses of metagenomic data indicate that the microbialite communities and their metabolic potential are distinct from those in the surrounding environment.

The microbial communities and the metabolic potential of the surrounding environments differed significantly from those of the microbialites; the sediment metagenomes were predicted to have more sequences associated with Betaproteobacteria and Spirochaetia (Figure 2.5). The predicted metabolic potential of sediment metagenomes was associated with nitrogen metabolism (e.g. ammonification) and anaerobic fermentation (Figure 2.6). The water metagenomes had more phototrophic eukaryotes (e.g. Chlorophyceae) with a photosynthetic metabolic potential (e.g. carotenoid biosynthesis) (Figure 2.5 and 2.6). Bacteroidetes and Chlorophyta are low GC containing taxa which could explain the distinction between the water and microbialite metagenomes based on GC content (Figure 2.1). Thus, the surrounding environments had a microbial community and metabolic potential that was divergent from microbialites.

2.3.3 Pavilion Lake microbialite core microbial community and related pathways

The microbialites of Pavilion Lake were dominated by Proteobacteria, Acidobacteria, and Cyanobacteria (Figure 2.2B). Based on RefSeq taxonomic classification ANOVA using STAMP, the classes Alphaproteobacteria, Deltaproteobacteria, Acidobacteria and Gloeobacteria were significantly more abundant in the microbialite metagenomes (Table 2.2; p <0.01). The dominance of sequences associated with the members of the phyla Proteobacteria (mainly Alphaproteobacteria and Deltaproteobacteria classes), and Cyanobacteria is consistent with results from other marine and freshwater microbialite communities (Havemann and Foster 2008; Breitbart et al., 2009; Goh et al., 2009; Mobberley et al., 2010; Khodadad and Foster 2012).
There were significantly more microbialite sequences classified as *Alphaproteobacteria* relative to the neighboring environment, which consisted of functional groups that were photoheterotrophic and heterotrophic (Figure 2.7; Table 2.2; \( p < 0.05 \)). Microbialite alphaproteobacterial specific contigs were assigned to genera of photoheterotrophic (e.g. *Rhodomicrobium*, *Rhodopseudomonas*, and *Rhodospirillum*), heterotrophic (e.g. *Sphingomonas*) and nitrogen-fixing (e.g. *Argobacterium*, *Bradyrhizobium*, and *Rhizobium*) bacteria (Figure 2.7; \( p < 0.05 \)). Alphaproteobacterial-based nitrogen fixation complements cyanobacterial nitrogen fixation in microbialites likely because of cyanobacterial diel cycles (Havemann and Foster 2008).

Deltaproteobacterial-associated sequences within the microbialite metagenomes were assigned to genera of dissimilatory sulfate reducing (e.g. *Desulfobacterium* and *Desulfovibrio*) and heterotrophic carbonate precipitating (e.g. *Myxococcus*) bacteria (Jimenez-Lopez *et al.*, 2011), which were significantly more abundant in microbialites than in the neighboring environments (Figure 2.8). Sulfate-reducing deltaproteobacteria are often found where carbonates precipitate, and are important drivers of the “alkalinity engine,” by pushing the saturation index higher via increasing alkalinity (Gallagher *et al.*, 2012). *Myxococcus* spp. are abundant in a variety of microbialite-forming systems and can directly precipitate carbonate through the release of ammonium (Ben Chekroun *et al.*, 2004; Jimenez-Lopez *et al.*, 2011). Analyses of the Pavilion Lake microbialite metagenomes confirmed prior investigations showing that members of the *Alphaproteobacteria* include photoheterotrophs and nitrogen-fixers, and the *Deltaproteobacteria* include sulfate-reducers (Havemann and Foster 2008; Breitbart *et al.*, 2009; Goh *et al.*, 2009; Mobberley *et al.*, 2010; Khodadad and Foster 2012).

Sequences associated with filamentous cyanobacterial mat-builders from the genera *Anabaena*, *Lyngbya*, *Microcoleus*, *Nostoc*, *Oscillatoria* and the planktonic *Cyanothece* and
Acryochoris were enriched significantly in microbialites (Figure 2.9, p < 0.05). Pathways for synthesis of cyanoglobin and cyanophycin, as well as copper metabolism, which are associated with cyanobacterial mat-builders, were highly enriched in the microbialites over the surrounding environment. (Figure 2.10, p < 0.05). Cyanoglobin is a peripheral membrane protein that binds oxygen with high affinity, is highly expressed under low oxygen and could be restricted to some strains of Nostoc spp. and Anabaena spp. (Hill et al., 1996). Cyanophycin is formed in filamentous cyanobacteria in response to low or changing DIC to O2 ratios (Liang et al., 2014). Copper homeostasis genes were abundant in microbialites, which is common for cyanobacterial derived mats, as copper is essential for growth (Varin et al., 2012) but also toxic at levels ≥10 mM (Burnat et al., 2009). The microbialite metagenome indicates that the metabolic potential of filamentous cyanobacterial mats is adaptive to metal homeostasis (e.g. copper), as well as carbon and oxygen limitation (e.g. cyanoglobin and cyanophycin).

Sequences associated with planktonic cyanobacteria (Synechococcus spp.) represented 40 to 60% of the cyanobacterial sequences in the water but <27% in microbialites based on RefSeq classification. This confirms the high abundances (8.3 x 10⁴ cells ml⁻¹) of Synechococcus spp. reported by Laval et al. (2000). The low abundance of sequences associated with Synechococcus spp. in cyanobacterial mats is likely a reflection of their planktonic lifestyle. However, planktonic cyanobacteria are involved in “whiting events,” during which photosynthetic alkalization induces carbonate precipitation from the water (Thompson and Ferris, 1990; Thompson et al., 1997). Measured sedimentation rates in Pavilion Lake are 0.07 g y⁻¹ with 4.3% being carbonates (Lim et al., 2009), indicating a minor contribution of whiting events to microbialite formation. This is consistent with phototrophic and heterotrophic processes in the cyanobacterial mats dominating microbialite formation (Omelon et al., 2013).
More than 60% of the microbialite acidobacterial specific contigs were assigned to *Candidatus Solibacter* spp., and were more abundant in the microbialites than in the sediments or water (Table 2.2, p-value < 0.01). Representatives of the genus have been reported from other permanently cold environments, including Arctic tundra, and have cold-adapted glycoside hydrolases and transferases that breakdown diverse carbon sources and allow growth in cold oligotrophic conditions (Rawat *et al.*, 2012), such as found in Pavilion Lake (Lim *et al.*, 2009). These conditions are analogous to those in cyanobacterial mats on Pavilion Lake microbialites (Lim *et al.*, 2009). Solibacterial contigs in the microbialites contain both glycoside hydrolases and transferases that could degrade filamentous cyanobacterial mats, and reverse EPS-related calcium binding inhibiton (Dupraz and Visscher, 2005).

Recent studies have suggested that *Candidatus Chloracidobacterium thermophilum*, a novel photoheterotroph, could be present in microbialites (Couradeau *et al.*, 2011; Russell *et al.*, 2014). *Candidatus Chloracidobacterium thermophilum* was found at low abundance (<1% of contigs) in Pavilion Lake microbialites based on RefSeq classification. Despite low abundance of sequences for *Candidatus C. thermophilum*, evidence of functional genes in the microbialites included some for housekeeping (RecA, gyrB, tRNA synthetase), carotenoid biosynthesis, glycoside processing (hydrolases, transferases) and CRISPRs. While, no evidence for photosystems was found among the *Candidatus C. thermophilum* like sequences, there were many hypothetical genes that were unclassified.

Urealytic metabolism has been hypothesized to be involved in microbialite formation due to its carbonate precipitating effects, but its detection in microbialites has remained elusive (Castanier *et al.*, 1999). *Gamma* and *Deltaproteobacteria* specific urease beta subunits and urease accessory proteins (UreD/F) were found only amongst the Pavilion Lake microbialite
metagenomes (Figure 2.10). The linkage of urease related genes to *Proteobacteria* was unexpected due to the strong experimental evidence supporting *Firmicutes* (mainly *Bacillus* spp.) as the dominant taxa contributing urease related genes (Boquet *et al*., 1973; Hammes *et al*., 2003; Lee, 2003; Dick *et al*., 2006; Dhami *et al*., 2013).

Two potential sources of urea in the microbialite are metazoans and bacteria. Crustaceans excrete urea as a by-product of heterotrophic metabolism (Weihrauch *et al*., 2009). As well, some bacteria produce urea by the degradation of purines, arginine and allantoin (Jørgensen, 2006). Notably, gammaproteobacteria (e.g. *Pseudomonas* spp.) can produce urea without an external supply of organic nitrogen (Therkildsen *et al*., 1997). Urea is likely produced locally in the microbialite mat and is degraded by urease as rapidly as it is produced, supporting the model that microbialites become progressively lithified as the photosynthetically derived carbonate becomes in-filled through subsequent carbonate precipitation by heterotrophic activity (Omelon *et al*., 2013).

Antibiotic and heavy-metal resistance pathways were associated with *Proteobacteria* and by RefSeq classification were over-represented in the microbialites (Figure 2.11). These included antibiotic resistance pathways assigned to the *Alpha*, *Beta* and *Gamma* classes of *Proteobacteria* that include multidrug efflux transporter families (e.g. MexF and Bcr/CflA) and class A beta-lactamases (Figure 2.11). Genes related to antibiotic resistance in the Pavilion Lake microbialites could be in response to toxic organic molecules produced by cyanobacterial mats (Neilan *et al*., 2013). Cobalt-zinc-cadmium resistance genes were also enriched in microbialites and assigned to *Alpha*, *Beta* and *Gamma* classes of *Proteobacteria* (Figure 2.11). Pavilion Lake has low levels of zinc (0.01 to 0.03 mg L⁻¹) and undetectable levels of cobalt, iron, arsenic and cadmium (Lim *et al*., 2009). The presence of heavy metal resistance genes may be involved in resistance, homeostasis
or sequestration of metals. Conceivably, historical mining in the area (Stevenson, 1940) could have contaminated Pavilion Lake, and selected for resistance to heavy metals. Antibiotic resistance has also been linked to heavy-metal stress, suggesting that selection for resistance to one can lead to resistance to the other in complex bacterial communities (Nisanian et al., 2014).

Primary alcohol fermentation related to genes associated with acetone, butanol and ethanol biosynthesis were assigned to Alpha- and Beta-proteobacteria, and were enriched in the microbialite metagenomes relative to the surrounding environment (Figure 2.11). Primary alcohol fermentation has been linked to microbialite dissolution; however, fermentation also provides substrates that fuel dissimilatory sulfate reduction, which can precipitate carbonate and thus could offset carbonate lost by fermentation (Dupraz and Visscher, 2005; Gallagher et al., 2012). Members of the Proteobacteria appear to be major constituents of Pavilion Lake microbialites, and potentially provide important metabolic roles, such as resistance to antibiotics and heavy-metals, and primary alcohol fermentation, which have not previously been attributed to microbialites, and which suggest an important role for Proteobacteria in these systems.

2.3.4 Low abundance members of Pavilion Lake microbialite microbial community

Bacterial phyla such as Gemmatimonadetes, Poribacteria and Chloroflexi were in low abundance (<2% of the contigs) in the microbialite (Figure 2.2B). The ANOVA in STAMP using RefSeq classification found that Chloroflexi and Gemmatimonadetes were significantly more abundant in the microbialites than in the sediments or the water (Table 2.2, p-value <0.01). Gemmatimonadetes and Chloroflexi are likely part of the heterotrophic consortia in Pavilion Lake microbialites. Members of the Gemmatimonadetes synthesize carotenoids to protect against reactive oxygen species and possibly photo-oxidation, which would be useful to avoid photo-oxidation in the microbialite cyanobacterial mats (Takaichi et al., 2010). Porbacteria are
associated with marine sponges (Fieseler et al., 2004), and freshwater sponges are found throughout Pavilion Lake, even growing on microbialites.

Sequences assigned to *Archaea* in the microbialites were <1% of the contigs (Figure 2.2B). Little is known about the role of archaea in microbialites, but prior metagenomic evidence also suggests they are of minor importance (Mobberley et al., 2010; Khodadad and Foster 2012). For example, in both nonlithifying and lithifying stromatolitic mats at Highbourne Cay there were no taxonomic or metabolic differences attributed to archaea (Mobberley et al., 2010; Khodadad and Foster 2012). While the abundance of archaea is low, there are statistically significant differences in their occurrence between microbialites and the surrounding environments. RefSeq ANOVA assigned reads to the taxa *Halobacteria, Thermococci, Thermoplasmata, Methanopyri, Methanomicrobia*, and unclassified *Korarchaeota* (Table 2.2, p < 0.01), supporting the presence of archaea. Moreover, the SEED subsystem annotations found archaeal specific pathways in the microbialites for DNA replication, DNA recombination and tRNA modification, as well as tRNAs/rRNA methyltransferases and predicted archaeal chaperones associated with the Archease 2 pathway (Figure 2.12). However, given their low abundance and likely anaerobic lifestyle, it is unlikely that archaea play a major functional role within the microbialites, although more study is warranted.

*Exiguobacterium* sp. strain RW2 was isolated from a microbialite collected from 20 m in Pavilion Lake; however, its abundance is unknown (White 3rd et al., 2013b). Based on the taxonomic assignment of contigs from the metagenomic data, sequences most similar to those from *Exiguobacterium* spp. were more abundant in the microbialites than in the water or sediments (Figure 2.13a). Sequenced genomes from isolates of *Exiguobacterium* spp. (strains RW2, S17 and AT1b) were used to recruit against ~7.5M metagenomic reads from a 20 m microbialite to
determine which genotype recruited the most sequences. More reads (0.46% of total reads) recruited to the draft genome *Exiguobacterium* sp. strain RW2 than to the genomes of its closest relatives, including *Exiguobacterium* sp. strain AT1b (0.25% of total reads), and *Exiguobacterium* sp. strain S17 (0.33% of total reads) (Figure 2.13B). As well, *Exiguobacterium* sp. strain RW2 had the highest proportion of identical hits to metagenomic data, consistent with its presence in the system (Figure 2.13B). Members of the genus *Exiguobacterium* are associated with hard carbonate structures, calcium-rich springs, and alkaline environments (Vishnivetskaya et al., 2011; Ordoñez a et al., 2013; White et al., 2013b). As well, isolates from modern microbialites, including *Exiguobacterium* sp. strain RW2 and *Exiguobacterium* sp. strain S17, appear to code for pathways involved in resistance to heavy metals (Belfiore et al., 2013; Ordoñez a et al., 2013; White et al., 2013b). This is supported by analysis using tBLASTx (1e⁻³) that showed reads recruited to *Exiguobacterium* sp. strain RW2 at 95 to 100% identity were heavy-metal resistance genes, which are relatively more abundant in the microbialites (Figure 2.11).

*Agrococcus* sp. strain RW1 was isolated from the same 20 m Pavilion Lake microbialite as *Exiguobacterium* sp. strain RW2 (White 3rd et al., 2013a-b). However, like *Exiguobacterium* sp. strain RW2, the abundance of *Agrococcus* sp. strain RW1 is unknown (White et al., 2013a). In order to estimate the contribution of members of this genus to the microbialite community, the genomes of *Agrococcus* sp. strain RW1 and *A. lahaulensis* strain K22-21 were recruited against the ~7.5 M metagenomic reads from the 20 m microbialites. Both genomes recruited about 1.6% of the reads, indicating that *Agrococcus* spp. are significant members of this community. Although *A. lahaulensis* recruited a few hundred more reads, *Agrococcus* sp. strain RW1 had more reads that recruited at 95 to 100% identity (Figure 2.14). Similar to the *Exiguobacterium* sp. strain RW2
genome, these reads were confirmed by tBLASTx \(10^{-3}\) to be heavy-metal resistance genes (Figure 2.11).

### 2.3.5 Metabolic potential relating to photosynthetic or heterotrophic processes

The metabolic potential of the microbialites is dominated by heterotrophic processes. Sequences related to photosynthesis including those encoding photosystems and electron transport proteins were ranked 28th out of 29 based on SEED subsystems (Figure 2.15). In contrast, pathways related to carbohydrate metabolism (carbon-related pathways) were ranked 2nd and accounted for \(~9\%\) of the contigs. Among the carbon-related pathways \(~45\%\) were related to central (TCA cycle) and one-carbon metabolism (e.g. serine-glyoxlate cycle), while another \(45\%\) were related to degradation (e.g. fermentation, glycoside hydrolases and other hydrolytic enzymes) (Figure 2.16). Only \(10\%\) of the microbialite-specific contigs were annotated as carbon-fixation related (e.g Calvin-Benson cycle) (Figure 2.16). Stable-isotope studies suggest that photosynthetic processes are linked to carbonate precipitation in the microbialites (Brady et al., 2010, Omelon et al., 2013), even though the metabolic potential is dominated by heterotrophic processes. Microbialite formation in Pavilion Lake appears to be localized carbonate precipitation associated with cyanobacterial photosynthesis and heterotrophic processes such as urealytic metabolism, dissimilatory sulfate reduction and heterotrophic mat degradation, which could reduce EPS related carbonate inhibition (Dupraz et al., 2009). \(^{13}\)C enrichment in microbialite carbonates may not only be a product of cyanobacterial photosynthesis, photoheterotrophs such as acidobacteria and proteobacteria may also contribute; however, this remains to be demonstrated.

### 2.3.6 Viral community and viral defense

Even though only the cellular fraction from the water was analyzed, viral sequences represented \(>1\%\) of reads, whereas in total DNA extracted from microbialites they comprised
ANOVA in STAMP based on RefSeq classification further confirms that viruses from the water were significantly more abundant than in microbialites or sediments than water (Figure 2.17). Specifically, T4-like and large algal viruses (e.g. Phycodnaviridae) dominated the viral sequences in the water and were more abundant than in the microbialites and sediments (Figures 2.18). The low proportion of viral reads in the microbialite data may reflect the fact that in the RefSeq database there are many sequenced dsDNA viral genomes from water samples, but not from microbialites. Despite the limited data, viruses in the water appeared to have higher abundances of proteins related to phage structure (tail fibers), phage replication and DNA replication (Figure 2.18).

An active role for phages in the microbialites is suggested by the higher relative abundances of predicted genes involved with CRISPRs and phage excision (Figure 2.17 and 2.18). CRISPR cas genes were associated with the following taxonomic groups: Chloroflexi (e.g. Dehalococcoides), Deltaproteobacteria (e.g. Myxococcus and Desulfuromonadales), filamentous cyanobacteria (e.g. Anabaena, Nostoc, Rivularia) and Firmicutes (e.g. Clostrida) based on tBLASTx (1e⁻³) analysis. As well, more putative genes involved in phage integration and excision occurred in the microbialites (Figure 2.17 and 2.18). Also, CRISPRs were predicted to be associated with key members involved in microbialite formation, such as filamentous cyanobacteria and Myxococcus spp., implying that the microbialite community is under continuous selective pressure from viral infection.

It is important to emphasize that the viral DNA was from the >0.22 μM cellular fraction captured on filters. Most free viral particles would have passed through the filters; hence, most viral sequences were likely from infected cells or from viruses attached to particles. It is not uncommon for filters with pore sizes much larger than viruses to contain many viral sequences
(Zeigler Allen et al., 2012). The most abundant viral contigs in the water were for T4-like cyanophages and phycodnaviruses (Figure 2.18), although gene-specific primers failed to amplify DNA from either group from any of the samples, suggesting the viruses were divergent from those targeted by the primers (Chen and Suttle, 1995; Filée et al., 2005).

Consistent with reports for other microbialites (Desnues et al., 2008), relatively few viral sequences were recovered in this study. Yet, the occurrence of phage integration and CRISPR-cas sequences implies that the Pavilion Lake microbialites are under selection from viral infection. It is also likely that the lack of viral sequences from freshwater microbialites in databases compromised their identification.

**2.4 Conclusion**

Prior to this study, it was unknown if the microbial community profile and metabolic potential of microbialites was distinct from those in neighboring environments. Hypotheses on the formation of microbialites have suggested that material from the neighboring sediments is the source of the microbialite-associated microbial community (Burne and Moore, 1987). The data presented here clearly show that the two microbial communities differ markedly. Russell et al. (2014) also found taxonomically distinct microbial communities in non-lithifying soft-mat biofilms and microbialites in Pavilion Lake. As well, metagenomic analysis of marine microbialites in Highbourne Cay showed distinctly different communities associated with lithifying and non-lithifying microbial mats (Khodadad and Foster, 2012).

While differences among the Pavilion Lake metagenomes can be attributed to differing selection pressures among environments, the microbialite community is likely composed of taxa essential for microbialite growth, as well as opportunists taking advantage of nutrients and the matrix supplied by filamentous cyanobacterial mats. Isotopic data suggests that Pavilion Lake
microbialites are formed by photosynthetic influences (Brady et al., 2010, 2013), however, our data suggest that photosynthetic gene presence is quite low. The microbialite community is dominated by filamentous cyanobacteria, proteobacteria and acidobacteria, all of which include phototrophs that may contribute to precipitation of carbonate.

Unlike other microbialites, those in Pavilion Lake are enriched for pathways that include heavy-metal and antibiotic resistance, as well as primary alcohol fermentation. These pathways are associated with members of the Proteobacteria, which are numerically dominant, likely convey resistance to toxins and heavy metals, and may influence carbonate formation through photosynthesis and urea metabolism. Omelon et al. (2013) suggested that heterotrophs contribute to the lithification of thrombolites in Pavilion Lake by triggering additional carbonate precipitation, a hypothesis supported here by evidence for urealytic metabolism and dissimilatory sulfate reduction, which has not been reported previously in microbialites.

The prevalence of CRISPR-cas systems and phage excision genes imply that the microbialites are under continuous selective pressure from viral infection. In particular, the presence of CRISPRs assigned to taxa that precipitate carbonates (Cyanobacteria, Deltaproteobacteria and Firmicutes) suggest that viruses play an important role in the microbialite communities in Pavilion Lake.
2.5 Tables and figures

Table 2.1: Pavilion Lake metagenome assembly statistics

<table>
<thead>
<tr>
<th>Depth</th>
<th>Microbialites</th>
<th>Water</th>
<th>Sediments*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Depth</td>
<td>10 m</td>
<td>20 m</td>
<td>25 m</td>
</tr>
<tr>
<td># Contigs or Reads</td>
<td>510,415</td>
<td>881,449</td>
<td>499,134</td>
</tr>
<tr>
<td>Total Length (Bp)</td>
<td>115,196,707</td>
<td>178,186,326</td>
<td>111,159,198</td>
</tr>
<tr>
<td>N50 &gt;500bp</td>
<td>647</td>
<td>752</td>
<td>717</td>
</tr>
<tr>
<td>Avg &gt;500bp</td>
<td>721</td>
<td>783</td>
<td>760</td>
</tr>
<tr>
<td>Med &gt;500bp</td>
<td>571</td>
<td>633</td>
<td>616</td>
</tr>
<tr>
<td>Largest (Bp)</td>
<td>3882</td>
<td>6079</td>
<td>2965</td>
</tr>
<tr>
<td>G+C%</td>
<td>60.8</td>
<td>57.9</td>
<td>60.1</td>
</tr>
</tbody>
</table>

*Not assembled reads only
**Sfc: surface water filter (1 to 10 m) (-) not available.
Table 2.2: Taxonomic classes (RefSeq) that are overrepresented in microbialites relative to the surrounding environment as determined by ANOVA using STAMP.

<table>
<thead>
<tr>
<th>Bacteria</th>
<th>p-val</th>
<th>Effect size</th>
<th>Water: Avg Rfreq (%)</th>
<th>Water: SD (%)</th>
<th>MB: Avg Rfreq (%)</th>
<th>MB: SD (%)</th>
<th>Sed: Avg Rfreq (%)</th>
<th>Sed: SD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acidobacteriia</td>
<td>1.96E-03</td>
<td>0.889</td>
<td>0.314</td>
<td>0.164</td>
<td>0.839</td>
<td>0.139</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Acidobacteria (Unclass)</td>
<td>8.34E-05</td>
<td>0.972</td>
<td>0.096</td>
<td>0.053</td>
<td>0.849</td>
<td>0.094</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Alphaproteobacteria</td>
<td>4.52E-04</td>
<td>0.947</td>
<td>7.078</td>
<td>0.266</td>
<td>19.768</td>
<td>2.196</td>
<td>2.679</td>
<td>2.008</td>
</tr>
<tr>
<td>Chloroflexi (class)</td>
<td>9.30E-05</td>
<td>0.973</td>
<td>0.449</td>
<td>0.142</td>
<td>1.481</td>
<td>0.116</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Deferrribacters (class)</td>
<td>1.13E-03</td>
<td>0.917</td>
<td>0.059</td>
<td>0.014</td>
<td>0.117</td>
<td>0.020</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Deinococci</td>
<td>1.82E-03</td>
<td>0.892</td>
<td>0.335</td>
<td>0.131</td>
<td>0.613</td>
<td>0.081</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Deltaproteobacteria</td>
<td>1.81E-04</td>
<td>0.961</td>
<td>2.151</td>
<td>0.933</td>
<td>7.850</td>
<td>0.448</td>
<td>0.398</td>
<td>0.563</td>
</tr>
<tr>
<td>Gemmatimonadetes</td>
<td>1.39E-03</td>
<td>0.904</td>
<td>0.208</td>
<td>0.160</td>
<td>0.797</td>
<td>0.114</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Gloeobacteria</td>
<td>7.68E-05</td>
<td>0.972</td>
<td>0.056</td>
<td>0.038</td>
<td>0.360</td>
<td>0.030</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Korarchaeota (Unclass)</td>
<td>1.49E-03</td>
<td>0.900</td>
<td>0.076</td>
<td>0.042</td>
<td>0.310</td>
<td>0.062</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Solibacteres</td>
<td>1.16E-04</td>
<td>0.973</td>
<td>0.386</td>
<td>0.279</td>
<td>3.346</td>
<td>0.331</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Thermomicrobia (class)</td>
<td>1.44E-03</td>
<td>0.905</td>
<td>0.058</td>
<td>0.030</td>
<td>0.481</td>
<td>0.111</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Thermotogae (class)</td>
<td>5.03E-03</td>
<td>0.845</td>
<td>0.044</td>
<td>0.023</td>
<td>0.138</td>
<td>0.006</td>
<td>0.023</td>
<td>0.033</td>
</tr>
<tr>
<td>Euryarchaeota (Unclass)</td>
<td>8.67E-04</td>
<td>0.928</td>
<td>0.010</td>
<td>0.007</td>
<td>0.039</td>
<td>0.004</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Halobacteria</td>
<td>8.11E-04</td>
<td>0.931</td>
<td>0.043</td>
<td>0.017</td>
<td>0.172</td>
<td>0.029</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Korarchaeota (Unclass)</td>
<td>4.49E-04</td>
<td>0.945</td>
<td>0.000</td>
<td>0.000</td>
<td>0.009</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Methanomicrobia</td>
<td>4.79E-04</td>
<td>0.943</td>
<td>0.062</td>
<td>0.036</td>
<td>0.312</td>
<td>0.044</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Methanopyri</td>
<td>3.85E-03</td>
<td>0.860</td>
<td>0.000</td>
<td>0.000</td>
<td>0.005</td>
<td>0.002</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Thermococci</td>
<td>2.64E-03</td>
<td>0.877</td>
<td>0.014</td>
<td>0.013</td>
<td>0.051</td>
<td>0.007</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>Thermoplasmata</td>
<td>3.70E-03</td>
<td>0.863</td>
<td>0.002</td>
<td>0.003</td>
<td>0.012</td>
<td>0.003</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 2.3: Functional annotations (SEED subsystem level I) that are overrepresented in microbialites relative to the surrounding environment as determined by ANOVA using STAMP

<table>
<thead>
<tr>
<th>Function</th>
<th>P-val (corr)</th>
<th>Effect Size</th>
<th>Water: Avg Rfreq (%)</th>
<th>Water: SD (%)</th>
<th>MB: Avg Rfreq (%)</th>
<th>MB: SD (%)</th>
<th>Sed: Avg Rfreq (%)</th>
<th>Sed: SD (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Membrane Transport</td>
<td>8.12E-003</td>
<td>0.793</td>
<td>1.998</td>
<td>0.157</td>
<td>2.514</td>
<td>0.112</td>
<td>0.941</td>
<td>0.577</td>
</tr>
<tr>
<td>Metabolism of Aromatics</td>
<td>3.00E-004</td>
<td>0.934</td>
<td>1.270</td>
<td>0.074</td>
<td>1.775</td>
<td>0.111</td>
<td>0.872</td>
<td>0.109</td>
</tr>
<tr>
<td>Motility and Chemotaxis</td>
<td>3.60E-002</td>
<td>0.649</td>
<td>0.606</td>
<td>0.025</td>
<td>1.063</td>
<td>0.039</td>
<td>0.454</td>
<td>0.356</td>
</tr>
<tr>
<td>Potassium metabolism</td>
<td>6.40E-003</td>
<td>0.810</td>
<td>0.182</td>
<td>0.035</td>
<td>0.296</td>
<td>0.013</td>
<td>0.084</td>
<td>0.068</td>
</tr>
<tr>
<td>Regulation and Cell signaling</td>
<td>1.82E-003</td>
<td>0.874</td>
<td>1.090</td>
<td>0.028</td>
<td>1.506</td>
<td>0.079</td>
<td>0.978</td>
<td>0.133</td>
</tr>
<tr>
<td>Virulence, Disease and Defense</td>
<td>3.45E-004</td>
<td>0.928</td>
<td>2.069</td>
<td>0.195</td>
<td>2.997</td>
<td>0.057</td>
<td>1.167</td>
<td>0.327</td>
</tr>
</tbody>
</table>

Figure 2.1: Pavilion Lake microbialite morphology with GC content and protein isoelectric points (pI)
Data for GC content and pI for water at 25 m and the sediment at 45 m were not available for comparison.
Figure 2.2: Pavilion Lake microbial community composition (A) PCA plot of the normalized top 25 RefSeq classifications. Clustering was done using a ward matrix and Bray-Curtis distance cut-offs, bootstrapped with one thousand replicates. (B) Dotplot of the normalized RefSeq classifications (top 25) in log relative abundances.
Figure 2.3: PCA and scatter plots using SEED subsystems based on ANOVA in STAMP for Pavilion Lake metagenomes
(A) PCA plot (multiple groups, subsystem level II)
(B) Scatter plot microbialites vs sediments (SEED subsystems level III).
(C) Scatter plot microbialites vs water filter (SEED subsystems level III).
For scatter plots only, each dot represents a single functional gene.
Figure 2.4: MetaCyc pathway annotations for Pavilion Lake metagenomes
(Top) PCA by ward clustering followed by bootstrapping of thousand replicates using Bray-Curtis distance metric.
(Bottom) NMDS by ward clustering followed by bootstrapping of thousand replicates using Bray-Curtis distance metric.
Figure 2.5: Box and whisker plots for significant taxa represented in water filters and sediments by RefSeq classification using ANOVA in STAMP based on multiple groups
Figure 2.6: Post-havoc confidence interval plots functional classification for represented metabolic potential in surrounding environment (water filters and sediments). SEED subsystems classifications are based on ANOVA in STAMP based on multiple groups. Microbialites are labeled in green, followed by sediments in orange and water in blue.
Figure 2.7: Post-havoc confidence interval plots for *Alphaproteobacteria* using RefSeq classification based on ANOVA in STAMP for significant genus-level members using multiple groups. Microbialites are labeled in green, followed by sediments in orange and water in blue.
Figure 2.8: Post-havoc confidence interval plots for *Deltaproteobacteria* using RefSeq classification. Based on ANOVA in STAMP for significant genus-level members represented in Pavilion Lake metagenomes (microbialite, water filter and sediment) using multiple groups. Microbialites are labeled in green, followed by sediments in orange and water in blue.
Figure 2.9: Post-havoc confidence interval plots for Cyanobacteria using RefSeq classification. Based on ANOVA in STAMP for significant genus-level members represented in Pavilion Lake metagenomes (microbialite, water filter and sediment) using multiple groups. Microbialites are labeled in green, followed by sediments in orange and water in blue.
Figure 2.10: Post-havoc confidence interval plots for cyanobacterial based functional genes using SEED subsystem classification and urealytic pathways
Based on ANOVA in STAMP for significant functional classifications represented in Pavilion Lake metagenomes (microbialite, water filter and sediment) using multiple groups.
The urealytic genes are for general presence in the metagenomes and are not cyanobacterial specific. Microbialites are labeled in green, followed by sediments in orange and water in blue.
Figure 2.11: Box and whisker plots for novel functional classifications (SEED subsystem) represented in microbialites based on ANOVA in STAMP using multiple groups
Figure 2.12: Box and whisker plots for novel Archaeal based functional classifications (SEED subsystem) represented in microbialites based ANOVA in STAMP using multiple groups
Figure 2.13: *Exiguobacterium* presence in Pavilion Lake microbialite metagenome

A) *Exiguobacterium* post-havoc confidence interval plots (0.95) based on ANOVA parameters for novel metabolic potential differences for microbialites and surrounding environment using multiple groups in STAMP.

B) Metagenomic recruitment plot of Pavilion lake 20 m microbialite reads (~7.5 Million) to three *Exiguobacterium* genetically related to *Exiguobacterium* sp. strain RW2 (minimum identity >70% and an E-value >1e⁻⁵).

*Exiguobacterium* sp. strain RW2 was isolated from a microbialite at 20 m (White 3rd *et al*., 2013b).
Figure 2.14: Metagenomic recruitment plot for Pavilion Lake microbialite against *Agrococcus*

FR-Hit recruitment plot of Pavilion lake 20 m microbialite reads (~7.5 Million) to *Agrococcus* sp. strain RW1 and to *Agrococcus lahaulensis* strain DSM17612 (Minimum identity >70% and an E-value >1e⁻⁵).

* *Agrococcus* sp. strain RW1 was isolated from a microbialite at 20 m (White 3rd *et al.*, 2013a).
Figure 2.15: Microbialite metagenome functional annotation ranking using SEED subsystem (Level I)
Microbialite metagenomes are listed as a function of depth (m).
Figure 2.16: Microbialite metagenome carbohydrate related SEED subsystem (Level II) functional annotations
Microbialite metagenomes are listed as a function of depth (m).
Figure 2.17: Box and whisker plots for viruses and CRISPRs in Pavilion Lake RefSeq and SEED subsystem represented in microbialites based on ANOVA in STAMP using multiple groups
Figure 2.18: Post-havoc confidence interval plots for viral taxa using RefSeq and viral specific functional genes in SEED
Based on ANOVA in STAMP for genus-level and functional gene-level classifications that were statistically significant by analysis using multiple groups.
Chapter 3:

Comparative analysis of microbiome of rapidly forming modern microbialites from the Canadian Yukon

3.1 Introduction

The extent to which microbialite communities are similar or different from surrounding sediments and waters remains unresolved due to a limited number of comparative sampling efforts. Sampling surrounding sediments and waters provide the environmental context needed to better constrain the study of common and unique aspects of microbialite community structure and function. This information may be used to uncover conserved patterns of microbial community assembly and the metabolic pathways mediating microbialate formation under different environmental conditions.

Freshwater microbialites from the Clinton Creek mine tailings pond located in Yukon Territory manifest a rapid growth rate allowing study of the interplay between the biotic and abiotic factors driving microbialite formation on a human time-scale. Clinton Creek microbialites have an estimated maximum accretion rate of ~5 mm per year (Power et al., 2011a), more than ten times faster than other modern microbialite-forming ecosystems, including Highbourne Cay (~0.33 mm per year) (Planavsky and Ginsburg, 2009) and Pavilion Lake (0.05 mm per year) (Brady et al., 2009).

Here we use a metagenomic approach to explore the community structure and metabolic potential of Clinton Creek microbialites in relation to those in the adjacent sediments in an effort to better understand the metabolic drivers of rapid growth in this freshwater ecosystem. We focus
on describing metabolic pathways mediating photosynthetic or heterotrophic carbonate precipitation and the taxonomic distribution of these pathways. We then compare the metagenomic sequence information from Clinton Creek microbialites and sediments to those in diverse non-lithifying mats, sediments, microbialites and corals, to explore whether a core community and associated functional potential can be defined for microbialites.

3.2 Materials and methods

3.2.1 Site description and water chemistry

Clinton Creek (64°26′42″N and 140°43′25″W) is a flooded and abandoned open-pit asbestos mine located in the subarctic ~77 km northwest of Dawson City, Yukon, Canada (Power et al., 2011a; Figure 3.1A to B). The mine was active for eleven years until closing in 1978 (Power et al., 2011a). Microbialites presumably began to form soon after the mine closed and the open-pit flooded. The photic zone likely occupies the full depth of the open-pit pond and there is minimal nutrient input, due to the lack of surrounding soil (Power et al., 2011a). Sediments in the open pit are composed of quartz, muscovite, kaolinite, chrysotile, and minor amounts of aragonite and trace calcite (Power et al., 2011a). The microbialites are columnar, up to 15 cm in height, and are primarily composed of aragonite with spherulitic fabric (Power et al., 2011a; Figure 3.1D). The open pit water is subsaline (>3.0 g L⁻¹), oligotrophic (undetectable phosphate), and alkaline in pH (8.4); the distribution of cation concentrations are Mg²⁺ >> Ca²⁺ >> Na⁺ > K⁺ > Si⁴⁺, while anions concentrations are SO₄²⁻ >> dissolved inorganic carbon (DIC) > Cl⁻ (Power et al., 2011a). The iron concentration in the water is very low, and phosphate is undetectable, indicating that the pond is oligotrophic, which is common in microbialite forming systems (Dupraz et al., 2009; Lim
The water is supersaturated with aragonite (saturation index = 0.6), the dominant mineral in the microbialites, as well as calcite \([\text{CaCO}_3]\) (Power et al., 2011). Aragonite will preferentially precipitate, as calcite crystal growth is inhibited at high magnesium to calcium ratios, as is the case in Clinton Creek (Power et al., 2011a).

### 3.2.2 Sampling, DNA extraction, purity and concentration measurements

Microbialites and sediment samples were obtained in July 2011 and were stored at 4ºC prior to DNA extraction. Community genomic DNA was extracted from ~10 g of each of the triplicate microbialites and sediment samples using the PowerMax soil DNA isolation kit (MoBio Laboratories, Inc., Carlsbad, CA, USA), following the manufacturer's instructions. DNA concentrations were determined using a Nanodrop-3300 (ThermoFisher, Nandrop Wilmington, DE) with PicoGreen® reagent used according to the manufacturer’s instructions (Invitrogen, Carlsbad, CA). The purity of extracted DNA and samples was determined by absorbance (260/280 and 260/230 ratios) using a Nanodrop-1000 (ThermoFisher, Nandrop Wilmington, DE). Genomic DNA from each replicate was pooled prior to Illumina library construction.

### 3.2.3 Illumina HiSeq/MiSeq library construction quality control and quantification

For Illumina library construction, DNA was sheared by ultrasonication (Covaris M220 series, Woburn, MA), and the fragments end-paired, A-tailed (Lucigen NxSeq DNA prep kit, Middleton, WI) and ligated to TruSeq adapters (IDT, Coralville, Iowa); small fragments were removed twice using magnetic beads (Beckman Coulter, Danvers, MA) (White 3rd et al., 2013a-c). No PCR enrichment was used to amplify libraries to avoid PCR duplication bias.
Libraries were checked for size and adapter-dimers using a Bioanalyzer HighSens DNAchip (Agilent). Libraries were quantified using Qubit (Invitrogen, Carlsbad, CA), according to the manufacturer’s instructions, by qPCR using a microfluidic digital PCR quantified standard curve (White 3\textsuperscript{rd} et al., 2009). The resulting libraries were pooled, and sequenced using both 250 bp and 100 bp paired-end sequencing on the Illumina MiSeq at GenoSeq UCLA Los Angeles, CA and HiSeq at McGill University/Génomique Québec, CAN platforms, respectively.

3.2.4 **Analysis of Illumina sequencing data**

Raw Illumina sequencing data were screened for PhiX spike-in contaminants using Bowtie2 (version 2.1.0) which were removed using Picard tools (version 1.90) (White 3\textsuperscript{rd} and Suttle, 2013c). Reads were quality checked using FastQC, then paired-end reads were merged by FLASH (version 1.2.5) and assembled with the Ray assembler (kmer size: 39 and 55) (Boisvert et al., 2010; 2012, White 3\textsuperscript{rd} and Suttle, 2013c), FragGeneScan was used to predict and translate open reading frames (ORFs) of the contigs (Rho et al., 2010) and ProPas (Wu et al., 2012) was used to calculate predicted protein isoelectric points (pI).

The assemblies were annotated using Metagenomic Rapid Annotations using Subsystems Technology (MG-RAST) (Meyer et al., 2008). MG-RAST analysis of the contigs, used BLAT (BLAST-like Alignment Tool) annotations based on hierarchical classification against M5RNA (MG-RAST ribosomal specific database), SEED subsystems and RefSeq databases with a minimum e-value cutoff of $10^{-5}$, a minimum percent identity cutoff of 60\%, and a minimum alignment length cutoff of 15 base pairs. The SEED, RefSeq and M5RNA (MG-RAST rRNA database) classifications were normalized using relative count abundances for each sample. PCA
analysis for the normalized RefSeq classifications (top 25) used R libraries Ecodist (dissimilarity-based functions for ecological analysis), and pvclust (hierarchical clustering with p-values via multiscale bootstrap resampling) using ward clustering and the Bray-Curtis distance metric, with a thousand replicates (Suzuki and Shimodaira, 2006). Principal component analysis (PCA) for the normalized RefSeq classifications was plotted using R library ggplot2 (Wickham, 2009). A dotplot of the top 25 normalized RefSeq classifications was completed using R libraries Reshape2, using the melt function, and then plotted using ggplot2 (Wickham, 2009). The annotations were parsed analyzed using statistical analysis of metagenomic profiles (STAMP) (Parks and Beiko, 2010). MG-RAST annotations using SEED subsystems for Clinton Creek microbialite and sediment contigs were loaded into STAMP and compared for metabolic potential using the one-sided G-test (with Yates' + Fishers') alternative to the chi-square test, with asymptomatic confidence intervals (0.95), using the Benjamini-Hochberg FDR procedure (Parks et al., 2010).

In addition to MG-RAST, metabolic pathways were predicted using MetaPathways, a modular pipeline for gene prediction and annotation that uses pathway tools and the MetaCyc database to construct environmental pathway/genome databases (ePGBDs) (Paley and Karp, 2006; Konwar et al., 2013; Caspi et al., 2014). Metapathways uses the seed-and-extend homology search algorithm LAST (local alignment search tool) for annotations of ORFs with a minimum of 180 bp and minimum alignment length cutoff of 50 bp (Kielbasa et al., 2011). Venn diagrams were constructed from predicted MetaCyc pathways based on normalized pathway size, and the number of ORFs associated with each pathway were examined using R, and then plotted using ggplot2 (Wickham, 2009).
3.2.5 Comparative metagenomics

MG-RAST annotations for Clinton Creek microbialite (ID 4532705.3) and sediment (ID 4532704.3) metagenomes were loaded into STAMPS and compared against arctic mats and sediments (Ward Hunt Ice Shelf mat (ID 4532782.3), MarkHam Ice Shelf (ID 4532781.3), and Lost Hammer Sediments (ID 4532786.3), mats from Yellowstone (Octopus Springs Mat (ID 4443749.3), and Mushroom Springs Mat (ID 4443762.3) and Antarctica mats, marine-derived lakes, freshwater lakes and marine (McMurdo ice shelf mat (ID 4532780.3), Marine Lake 1 (ID 4443683.3), Marine Lake 5 (ID 4443682.3), Marine 8 (ID 4443686.3), Marine 9 (ID 4443687.3), and Ace Lake (ID 4443684.3) using a multiple group ANOVA in STAMP with SEED subsystems (Level III) annotations by post-hoc tests (Tukey-Kramer at 0.95), an effect size (Eta-squared) and multiple test correction using the Benjamini-Hochberg FDR procedure. Clinton Creek microbialites were further examined against polar mats, Cuarto Ciénegas microbialites (IDs 4440060.4, 4440067.3) and marine stromatolites from Highbourne Cay (ID 4440061.3) in STAMP by the one-sided G-test (with Yates’ + Fisher’s) alternative to chi-squared, with asymptomatic confidence intervals (0.95), using Benjamini-Hochberg FDR procedure. Corals (IDs 4445755.3, 4445756.3) and Whale fall (IDs 4441619.3, 4441620.3 44416564.4) metagenomes were comparing with SEED subsystem III in STAMP using a one-sided Welch's t-test with Welch's confidence interval (0.95), and Storey’s FDR multiple correction test.
3.2.6 **Metagenomic data depositing**

The data used in this study are freely available from the MG-RAST metagenomics analysis server under the project name; Yukon microbialites’ under the names Clinton Creek microbialite (ID 4532705.3) and Clinton Creek sediment (ID 4532704.3).

3.3 **Results and discussion**

3.3.1 **Metagenomic assembly and contig properties**

The assemblies were selected based on the number of contigs (>100bp), N50/N90 values, longest contig, and total length (bp) of the assembly (Table 3.1). Based on these analyses, a kmer size of 39 was used for all further analysis (Table 3.1). A kmer size of 55 generally yielded longer but fewer contigs, which would not allow for a comparable differential analysis between microbialites and sediments (Table 3.1). Nevertheless, only 0.64% and 1.74% of the reads from the sediment and microbialite metagenomes, respectively, assembled into contigs, indicating that both environments had complex microbial communities.

The GC content differed in the microbialite compared to the sediment, whereas protein isoelectric points (pI) were similar (Figure 3.2). The lower GC content in the sediments was likely due to more representatives from low GC containing phyla such as *Bacteroides* and *Firmicutes*. The higher GC content in the microbialite data is likely due to the high relative abundance of sequences assigned to anoxic photoheterotrophic *Alphaproteobacteria*, including *Rhodobacterales* (59 to 65% GC content) and *Rhodomicrobium* (62.2% GC content). GC content across bacterial genomes can be highly variable amongst microbial taxa, although amino-acid usage is typically...
similar (Lightfield et al., 2011), which is consistent with our metagenomic data that GC is more variable than pI content (Figure 3.2).

3.3.2 Community composition

The microbialite and sediment microbial communities were dominated by bacteria with low abundances of eukaryotes and archaeal sequences (Table 3.2). From RefSeq annotations, <1% of the microbialite and sediment contigs were of archaeal origin (Table 3.2, RefSeq), and no archaeal ribosomal genes were detected in either the sediment or microbialite contigs (Table 3.2, M5RNA). Clinton Creek microbialites, similar to Highbourne Cay marine microbialites and the freshwater microbialites from Cuarto Cienegás, had low abundances (<1%) of archaea and eukaryotes (Breitbart et al., 2009; Khodadad et al., 2012; Mobberley et al., 2013). Eukaryotes were rare, as they made up <1% of the sediment and microbialite contigs of Clinton Creek (Table 3.2), although common taxa such as diatoms, dinoflagellates, cryptomonas, chlamydomonades and fungi were detected, which support microscopy data (Power et al., 2011a). Diatoms and other protists have been observed by microscopy and detected in the metagenomic data from Clinton Creek, but their contribution to the formation of microbialite structures requires further study (Power et al., 2011a). Diatoms may influence carbonate precipitation through photosynthetic alkalinization, akin to processes found in cyanobacteria, or through the ammonification of amino acids and the metabolism of urea (Tesson et al., 2008; Allen et al., 2011).

Sequences from metazoans represented a small proportion of those from the eukaryotes, and were <0.1% of the total assigned reads in the Clinton Creek microbialites. The sequences included representatives from the taxa Nematoda, Platyhelminthes, Microsporidia, Cnidaria (e.g.
Hydra spp.) and Arthropoda (e.g. insects). Our data supports microscopic observations showing low abundances of metazoans (Power et al., 2011a). Metazoan grazing has been a prime suspect in the global decline of microbialites as it removes cyanobacterial mats, thereby reducing microbialite formation by removing the main carbon source and structural components (Awramik, 1971; Garrett 1970). The low abundance of metazoans in Clinton Creek suggests that there grazing impact on the microbialites is quite low.

The Proteobacteria dominated the taxonomic groups in the Clinton Creek sediments and microbialites, with >50% of the predicted ORFs and >35% of the ribosomal sequences being assigned to this taxon (Figure 3.3A). The microbialite contigs were dominated by anoxic photoheterotrophic Alphaproteobacteria (e.g. Rhodobacterales), while sediment contigs had more nitrogen-fixing Gammaproteobacteria (e.g. Pseudomonas spp.) (Figure 3.3A). Alphaproteobacteria are commonly found among microbialite-forming microbial consortia and likely play an important role as nitrogen fixers during darkness, when heterocystous cyanobacteria are not active (Havemann and Foster, 2008).

Sequences assigned to the Deltaproteobacteria represented ~10% of the predicted Proteobacterial ORFs (based on RefSeq) from the sediments and microbialites (Figure 3.3A). The microbialite Deltaproteobacterial contigs consisted mainly of Myxococcus spp.; whereas, members of the Desulfuromonadales dominated in the sediments. Myxococcus spp. are abundant in a variety of microbialite-forming systems and can directly precipitate carbonate through the release of ammonium (Ben Chekroun et al., 2004; Jimenez-Lopez et al., 2011). The microbialites and sediments had similar abundances of Desulfurovibrionales, Desulfobacterales and
Syntrophobacterales, the major orders of dissimilatory sulfate reducers. The dissimilatory sulfate-reducers in the Deltaproteobacteria may be critical drivers of the alkalinity engine and thereby induce carbonate precipitation (Gallagher et al., 2012). The occurrence of the major groups of dissimilatory sulfate-reducing bacteria (Desulfovibrionales, Desulfobacterales and Syntrophobacterales) in Clinton Creek microbialites is not surprising; however, the proportion of sequences assigned to these taxa, including genes involved in dissimilatory sulfate-reduction, were similar in the sediments. Thus, the sediments and ground water could be sources of these dissimilatory sulfate-reducing bacteria in microbialites, or the bacteria may be dispersed as spores from other environments.

Sequences assigned to Cyanobacteria were the fourth most abundant (Figure 3.3A). The microbialites had 4-fold more protein-coding ORFs that were classified as Cyanobacteria than the sediments (Figure 3.3A). The cyanobacterial ribosomal sequences (based on the M5RNA database) recovered from the microbialites were assigned to filamentous cyanobacteria from the genera Tolypothrix and Leptolyngbya, and unclassified Antarctic Cyanobacteria. In contrast, no ribosomal sequences were recovered from the sediments that were assigned to the Cyanobacteria (Figure 3.3A; M5RNA). Based on RefSeq classification, in the microbialites there were more contigs assigned to filamentous genera including Microcoleus, Lyngbya, Nodularia, and Anabaena, and more unicellular calcifying Synechococcus, than in the sediments. The sediments had fewer sequences assigned to genera of filamentous and unicellular Cyanobacteria (e.g. Synechococcus). Cyanobacterial carbon fixation produces the nutrient-rich EPS that likely supports the growth of the heterotrophic microbial consortium, and fuels microbialite formation by
increasing the saturation index to favor carbonate precipitation (Dupraz and Visscher 2005; Dupraz et al., 2009).

Contigs assigned to the phylum Gemmatimonadetes comprised 7 to 8% of the protein-coding ORFs in the sediments and microbialites (Figure 3.3A). To our knowledge, this is the first report of protein-coding sequences in metagenomic data from microbialites that have been assigned to the Gemmatimonadetes, although they were also found in other environments. The contigs annotated mainly as hypothetical proteins, although there were annotations to ATPases, Zn-dependent peptidases and glucose/sorbosone dehydrogenase-like genes. Glucose/sorbosone dehydrogenases transform various sugar moieties into vitamins, including L-ascorbic acid (vitamin C). Gemmatimonadetes can also make D-glucono-1,5-lactone from D-glucose, which can acidify the extracellular environment and lead to dissolution of carbonate (Dupraz and Visscher, 2005; Miyazaki et al., 2006; Fender et al., 2012). Whether they are involved in microbialite formation or dissolution, or are just opportunists, needs to be elucidated.

3.3.3 Metabolic potential

The metagenomic data predicts that the metabolic potential of the Clinton Creek microbialites relating to carbonate precipitation is dominated by photosynthesis; whereas, the sediment metagenomes indicated more heterotrophic metabolism (e.g. respiration) (Figure 3.4A). SEED subsystem level I (i.e. the highest functional classification) annotations indicated that carbohydrate metabolism relating to carbon fixation, DNA metabolism and photosynthesis pathways were more abundant in the microbialites than sediments (Figure 3.4A, FDR p-value < 0.01). Lower level SEED subsystem predictions (level III) also revealed more photosynthetic
pathways (e.g. photosystem II reaction centers, and carotenoid and chlorophyll biosynthesis) in microbialites than sediments (Figure 3.4B, FDR p-value < 0.01). These photosynthetic pathways in microbialites were assigned to genera of filamentous cyanobacteria such as Microcoleus, Lyngbya, Nodularia, and Anabaena, which were not found in the sediments.

Power et al. (2011a) found that aragonite-associated biomass within the Clinton Creek microbialites was modestly enriched in $^{13}$C by 0.8‰, relative to aragonite not associated with biomass, consistent with carbonate precipitation by phototrophs. Electron microscopy of the microbialites confirmed that phototrophs were associated with carbonate enriched in $^{13}$C (Power et al., 2011a). The isotopic data and high abundance of sequences associated with photosynthetic pathways support the model that carbonate precipitation is mediated through photosynthetic-induced alkalization. Heterotrophic processes may also contribute to carbonate precipitation, as it is hypothesized that microbialites become progressively lithified as the photosynthetically derived carbonate becomes in-filled through subsequent carbonate precipitation by heterotrophic activity (Omelon et al., 2013).

The metapathway pipeline was used for MetaCyc pathway annotations to complement the SEED functional gene annotations. MetaCyc predicted that pathways revealed that most pathways were shared between microbialites and sediments (Figure 3.5A). Only 13 pathways were restricted to the sediments, while 240 were unique to the microbialites; 358 pathways were shared (Figure 3.5A). The forty most abundant shared pathways were housekeeping genes with functions such as protein, nucleic-acid, lipid and carbohydrate biosynthesis and degradation (Figure 3.5B). In the microbialites, MetaCyc annotations predicted higher levels of glutamine degradation I, which
results in the donation of nitrogen in the form of ammonium, while glutamine biosynthesis appeared to be higher in the sediments (Figure 3.5B). Both the sediments and the microbialites were predicted to recycle ammonium through the ammonium assimilation cycle I-II (Figure 3.5B), which would supply nitrogen, and support primary production of the filamentous cyanobacterial mats in microbialites, thus leading to further carbonate precipitation.

### 3.3.4 Comparative metagenomic analysis

The metagenomic data are consistent with the Clinton Creek microbialites and sediments being functionally more related to polar mats than marine or tropical freshwater microbialites. Using SEED subsystem level III, PCA indicates better clustering to polar mats and sediments from the Arctic and Antarctica (Figure 3.6). The Clinton creek data cluster closest to data from Markham and Ward Hunt ice-shelf mats from the Canadian High Arctic (Figure 3.6; Vanin et al., 2012). Functionally, the Markham mats and Clinton Creek microbialites are most strongly correlated based on SEED subsystem level III annotations (Figure 3.7, R$^2 = 0.952$). Overall, based on functional pathways in SEED, the polar mats (e.g. Markham, Ward Hunt, and Mcmurdo) were strongly correlated with the Clinton Creek microbialites. Markham mats, like the microbialites of Clinton Creek, are dominated by members of the *Proteobacteria*, and *Gemmatimonadetes* also comprise 3% of the 16S rDNA sequences (Bottos et al., 2008). Polar mats, whether on microbialites or ice shelves, appear to have functional similarities; this likely relates to shifts in temperature, including temperatures well below freezing (Varin et al., 2012).

SEED-based functional annotations for metagenomic data from the Clinton Creek microbialites were also compared with those from tropical freshwater microbialites from Cuarto
Cienegás and Highbourne Cay marine stromatolites (Figure 3.7). The Clinton Creek microbialite metagenome had weak correlations with microbialites from two locations in Cuarto Cienegás, including Pozas Azules II (Figure 3.7, \( r^2 = 0.702 \)), Rio Mesquites (Figure 3.7, \( r^2 = 0.633 \)), and Highbourne Cay marine stromatolites (Figure 3.7, \( r^2 = 0.018 \)). Arctic polar mats had stronger correlations for SEED pathways than Cuarto Cienegás microbialites and Highbourne Cay stromatolite metagenomes (Figure 3.7, \( r^2 > 0.900 \)). Cuarto Cienegás microbialites and Highbourne Cay are tropical; hence, many pathways related to cold-adaptation that are present in the Clinton Creek microbialites and polar mats (Varin et al., 2012) are absent. The SEED functional gene classifications for Highbourne Cay stromatolite were the least correlated with those from the Clinton Creek microbialites, consistent with marine and freshwater microbialite communities being quite different.

The Clinton Creek microbialite metagenomic data were distinct from non-lithifying mats or biofilms from Octopus and Mushroom Springs in Yellowstone (Figure 3.6), whale fall and corals (Figure 3.8 to 3.9). These data reveal that Clinton Creek microbialites are closely related to polar mats, due possibly to cold-adaptation, and differ greatly from tropical microbialites. This reveals that under the correct chemistry (e.g. alkaline pH, high DIC, and dissolved Ca\(^{2+}\) or Mg\(^{2+}\)), polar mats on ice shelves could have at least the metabolic potential to make microbialites.

### 3.4 Conclusion

The northernmost and fastest growing microbialites known are located at subarctic Clinton Creek (Yukon, Canada). DNA from representative microbialites was extracted and directly sequenced, without bias from DNA amplification, and used to produce the largest set of assembled
metagenomic data for a freshwater microbialite-forming ecosystem. The microbialite metagenomes revealed a high proportion of photosynthetic genes, implying that microbialite formation is driven by photosynthesis-induced alkalinization, which is consistent with data from $^{13}$C isotopic enrichment (Power et al., 2011a). Predicted metabolic pathways overlapped extensively between microbialite and sediment communities, particularly with respect to housekeeping genes; however, they have distinct core communities with microbialites dominated by *Alphaproteobacteria* (mainly anoxic phototrophs like *Rhodobacterales*) and sediments dominated by *Gammaproteobacteria* (mainly heterotrophic nitrogen-fixing *Pseudomonas* spp.).

The microbiome of the Clinton Creek microbialites was distinct from that of mats and biofilms from corals, whale fall, Yellowstone hot springs, and marine stromatolites. While Clinton Creek microbialites shared some functional potential with microbialites from Cuarto Cienegás, the Arctic mats (e.g. Markham and Ward Hunt) shared greater functional potential, possibly because of cold-adaptation. The shared metabolic potential between Clinton Creek microbialites and the polar ice-shelf Markham, Ward Hunt and McMurdo mats suggests that under favorable chemical conditions (e.g. alkaline pH, high DIC, and dissolved Ca$^{2+}$ or Mg$^{2+}$), Arctic mats have the metabolic potential to make microbialites.

At Clinton Creek, precipitation of carbonate minerals within the mine tailings has been shown to be a significant sink for atmospheric CO$_2$ (Wilson et al., 2009). This study supports the view that cyanobacteria generate alkalinity and sustain heterotrophic communities, which together drive the fast formation of microbialites at Clinton Creek.
### Table 3.1: Assembly statistics for Clinton Creek metagenomes

<table>
<thead>
<tr>
<th></th>
<th>Microbialite (k = 39)</th>
<th>Microbialite (k = 55)</th>
<th>Sediment (k = 39)</th>
<th>Sediment (k = 55)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. Contigs &gt;100bp</td>
<td>109722</td>
<td>689</td>
<td>59928</td>
<td>171</td>
</tr>
<tr>
<td>Total length (Bp) &gt;100bp</td>
<td>27180461</td>
<td>219687</td>
<td>12690391</td>
<td>151415</td>
</tr>
<tr>
<td>Mean &gt;100bp</td>
<td>247</td>
<td>318</td>
<td>211</td>
<td>885</td>
</tr>
<tr>
<td>N50 &gt;100bp</td>
<td>248</td>
<td>289</td>
<td>200</td>
<td>1307</td>
</tr>
<tr>
<td>N90 &gt;100bp</td>
<td>184</td>
<td>231</td>
<td>160</td>
<td>383</td>
</tr>
<tr>
<td>Med Length &gt;100bp</td>
<td>240</td>
<td>269</td>
<td>186</td>
<td>582</td>
</tr>
<tr>
<td>No. Contigs &gt;500bp</td>
<td>371</td>
<td>46</td>
<td>557</td>
<td>99</td>
</tr>
<tr>
<td>Total length (bp) &gt;500bp</td>
<td>230702</td>
<td>37861</td>
<td>508559</td>
<td>127068</td>
</tr>
<tr>
<td>Mean &gt;500bp</td>
<td>621</td>
<td>823</td>
<td>913</td>
<td>1283</td>
</tr>
<tr>
<td>N50 &gt;500bp</td>
<td>588</td>
<td>611</td>
<td>849</td>
<td>1591</td>
</tr>
<tr>
<td>Med &gt;500bp</td>
<td>574</td>
<td>566</td>
<td>608</td>
<td>916</td>
</tr>
<tr>
<td>Largest</td>
<td>9491</td>
<td>9752</td>
<td>33503</td>
<td>5896</td>
</tr>
<tr>
<td>Total GC count</td>
<td>16801673</td>
<td>135459</td>
<td>7329632</td>
<td>82490</td>
</tr>
<tr>
<td>GC%</td>
<td>61.82</td>
<td>61.66</td>
<td>57.76</td>
<td>54.48</td>
</tr>
</tbody>
</table>

k = kmer size (bp)
Table 3.2: Domain classification of the microbial community in Clinton Creek (%)

<table>
<thead>
<tr>
<th>Domains*</th>
<th>Microbialite (RefSeq)</th>
<th>Microbialite (M5RNA)</th>
<th>Sediment (RefSeq)</th>
<th>Sediment (M5RNA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bacteria</td>
<td>99.05</td>
<td>73.91</td>
<td>98.22</td>
<td>92.50</td>
</tr>
<tr>
<td>Eukaryotes</td>
<td>0.60</td>
<td>25.00</td>
<td>1.16</td>
<td>7.50</td>
</tr>
<tr>
<td>Archaea</td>
<td>0.32</td>
<td>0.00</td>
<td>0.45</td>
<td>0.00</td>
</tr>
<tr>
<td>Viruses</td>
<td>0.02</td>
<td>0.00</td>
<td>0.13</td>
<td>0.00</td>
</tr>
<tr>
<td>Unassigned</td>
<td>0.01</td>
<td>1.09</td>
<td>0.04</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Based on MG-RAST annotations. RefSeq: Protein coding ORFs. M5RNA: Ribosomal rRNA genes.
Figure 3.1: Clinton Creek sample site and examples of microbialite morphology

A) Map of northwestern North America illustrating the location of Clinton Creek Yukon, Canada.

B) Photograph of the Clinton Creek open pit-pond; the red dot indicates the sampling location.

C) Photograph of the microbialites along the periphery of the open pit pond
   Scale bar equals ~100cm.

D) Complete microbialite and cross-section of a microbialite. Scale bar represents ~10cm.
Figure 3.2: Molecular properties of the Clinton Creek microbialite and sediment contigs
A) GC content (%) B) Predicted protein isoelectric (pI) content (%)
Figure 3.3: Microbial community structure of Clinton Creek metagenomes

A) Dot pot of representative taxonomic groups from Clinton Creek sediments and microbialites using RefSeq (protein coding ORFs) and M5RNA (rRNA, MG-RAST rRNA database) in log relative abundances. “Other” denotes low abundance taxa that were <1% of the total ORF or rRNA, individually, but were all combined here into one point.

B) PCAs of top 25 taxonomic groups from Clinton Creek sediments and microbialites by RefSeq (ORFs) and M5RNA (rRNA, MG-RAST rRNA database) classification.
Figure 3.4: Extended error plots for functional gene annotations for Clinton Creek metagenomes in STAMP using SEED subsystems
A) SEED subsystem level I (highest level classification in SEED).
B) SEED subsystem level III (3rd lowest classification in SEED out of four levels).
Extended error plots used a one sided G-test (w/Yates' + Fisher's) with asymptomatic confidence intervals (0.95) using Benjamini-Hochberg FDR procedure.
*Red asterisks are significant photosynthetic pathways.
Figure 3.5: MetaCyc pathway annotations for Clinton Creek metagenomes
A) Venn diagram of MetaCyc pathways
B) The top 40 shared MetaCyc pathways from Venn diagram.
Figure 3.6: Functional gene comparative PCA plot for Clinton Creek metagenomes
Based on ANOVA for multiple groups using SEED subsystem level III in STAMP.
Figure 3.7: Scatter plots of functional gene annotations using SEED subsystem level III One sided G-test (w/Yates’ + Fisher's) with asymptomatic confidence intervals (0.95) using Benjamini-Hochberg FDR procedure in STAMP. Each dot represents a unique functional classification gene.
Figure 3.8: Comparative functional gene PCA plot for Clinton Creek against whale fall metagenomes
In STAMP, a one-sided Welch's t-test with Welch's confidence interval (0.95), and Storey’s FDR multiple correction test.
Figure 3.9: Comparative functional gene PCA plot for Clinton Creek against coral metagenomes. In STAMP, a one-sided Welch’s t-test with Welch’s confidence interval (0.95), and Storey’s FDR multiple correction test.
Chapter 4:

The genome of *Agrococcus* sp. strain RW1 isolated from modern microbialites reveals genetic promiscuity and elements responsive to environmental stresses

4.1 Introduction

Little is known about non-photosynthetic pigmented bacteria in modern microbialite mats. We screened over one hundred isolates from Pavilion Lake for bacteria that were pigmented but non-photosynthetic using a culture-dependent method, to supplement our culture-independent metagenomic approach. As part of this screen we isolated yellow *Agrococcus* sp. strain RW1 described here and orange *Exiguobacterium* sp. strain RW2 described in Chapter 5 pigmented bacteria from a microbialite collected at 20 m depth in Pavilion Lake (White 3rd et al., 2013a-b). To date, there are no completed genomes for the genus *Agrococcus*, making inferences on the abundance, distribution, and metabolic potential of this genus intractable. Additionally, the pathway responsible for the yellow pigmentation in the genus *Agrococcus* has not been described.

The genus *Agrococcus* was described based on two strains of *Agrococcus jenensis* isolated from soil and the surface of sandstone (Groth et al. 1996). The genus is classified within the family *Microbacteriaceae*, members of which have diaminobutyric acid within the cell wall, which could impart the distinctive lemon-yellow color, although whether diaminobutyric acid is involved in the pigmentation is unknown (Groth et al., 1996). *Agrococcus* spp. have been isolated from a wide range of environments, including air (Zlamala et al., 2002), a coal mine (Dhanjal et al., 2011),
cheese (Bora et al., 2007), cold-desert soil (Mayilraj et al., 2006), forest soil (Zhang et al., 2010),
a medieval wall painting (Wieser et al., 1999), dried seaweed (Lee, 2008), and the phyllosphere of
potato plants (Behrendt et al., 2008). There are eight described species of Agrococcus as follows:
A. jenensis (Groth et al., 1996), A. baldri (Ziamala et al., 2002), A. carbonis (Dhanjal et al., 2011),
A. casei (Bora et al., 2007), A. citreus (Wieser et al., 1999), A. jejuensis (Lee 2008), A. lahaulensis
(Mayilraj et al., 2006), A. terreus (Zhang et al., 2010), and A. versicolor (Behrendt et al., 2008).
However, little is known about the metabolism, genetic richness and diversity of this genus.

Here, we present the polyphasic characterization of a new species of Agrococcus isolated
from modern microbialites, including the first complete reference genome. Genomic analysis
revealed signatures of environmental adaption to the cold oligotrophic conditions of Pavilion Lake.
Promiscuous mobile elements were found in two plasmids involved in heavy metal resistance and
DNA transposition. The genomes of Agrococcus sp. RW1 and A. lahaulensis K22-21 both encode
a carotenogenic gene cluster that could be responsible for producing the characteristic lemon-
yellow pigmentation found in Agrococcus.

4.2 Materials and methods

4.2.1 Isolation, growth conditions, phage induction, biochemical and antibiotic
susceptibility tests

Agrococcus sp. strain RW1 was isolated by plating 0.5 g of a thrombolic microbrialite,
collected from 20 m in Pavilion Lake, British Columbia (50.86677 °N, 121.74191 °W), on Luria
- Bertani medium [1% (w/v) tryptone, 0.5% (w/v) yeast extract, 1% (w/v) NaCl, pH 7] solidified
with 1.5% (w/v) agar, and incubated at 30 °C for 3 days. Solidified LB was also used for assessing
growth at various temperatures (4, 11, 16, 20, 25, 30, 35, 42, 50 °C), at pH 7 and 1% (w/v) NaCl. To assess growth at various pH (5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 10, 10.5, 11), standard solidified LB was used at 30 °C and 1% (w/v) NaCl. Growth at different salinities was tested using solidified LB (pH 7) spiked with various amounts of NaCl (0-6%, 9, 12, 13, 16%) at 30 °C. Cultures were maintained in LB broth or on agar at 1% NaCl, pH 7 and at 30 °C. M-agar medium [0.5% (w/v) tryptone, 0.25% (w/v) yeast extract, 1% (w/v) NaCl, 1.5% (w/v) agar, pH 7], a diluted version of LB-agar medium, was used to determine filamentous growth under carbon-limitation. Strain characteristics, including colony and cell morphologies were determined by standard methods (Murry et al., 1994). Oxidase tests, biochemical enzyme assays, and carbohydrate utilization tests were completed using API20E (BioMérieux) test strips according to the manufacturer's instructions. Antibiotic susceptibility of strain RW1 was determined by the Kirby-Bauer method using antibiotic discs on solidified LB (Collee et al., 1996). Prophage induction assays were done either by adding 0.2 μg ml⁻¹ of mitomycin C, or by heating cultures to 45 °C, incubating at 30 °C for 3-10 h and monitoring changes in turbidity until a decrease of OD₆₀₀ to 0.1 or less was observed (starting OD₆₀₀ = 0.6). At several time points, cells were pelleted at 3,250 x g and the supernatant filter-sterilized twice through a 0.22 μm pore-size Millex Durapore PVDF membrane (Millipore) filters before plating using a double-agar overlay plaque assay (Kropinski et al., 2009).

4.2.2 Phospholipid fatty acid analysis

Phospholipid fatty acids (PLFAs) were extracted from cultures grown in 25 ml LB broth for 4 d at 22 °C. Cultures were transferred into pre-combusted vials for an overnight solvent extraction in a 1:2:0.8 ratio solution of dichloromethane (DMC), methanol (MeOH), phosphate-
buffered saline (PBS) [137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄ 2H₂O, 2 mM KH₂PO₄, pH 7.4] (Bligh and Dyer, 1959). The extract was filtered through a separatory funnel where DCM and water were added to achieve a mixture of MeOH:DCM:water of 1:1:0.9 (Bligh and Dyer, 1959). The lower organic phase was removed and purified into polar, neutral, and non-polar fractions using liquid chromatography through silica-gel. Phospholipids in the polar fraction were subjected to mild alkaline methanolysis to produce fatty-acid methyl esters (FAMEs) (Guckert et al., 1985). FAMEs were separated, identified, and quantified using gas chromatography mass spectrometry (GC/MS) (Agilent Technologies Inc., Santa Clara, CA, USA) with a DB-5MS capillary column (30 m x 0.32 mm I.D. x 0.25 µm film thickness) at a temperature regime of 50 °C (1 min), 20 °C min⁻¹ to 130 °C, 4 °C min⁻¹ to 160 °C, and 8 °C min⁻¹ to 300 °C (5 min). PLFAs were identified by retention time and mass spectra relative to those of reference standards (Bacterial Acid Methyl Ester Mix, Matreya Inc., Pleasant Gap, PA, USA; and Supelco 37 Component FAME Mix, Sigma-Aldrich Co., Bellefonte, PA, USA). A modified picolinyl ester derivatization was used to determine the branching point in unknown compounds (Destaillats and Angers, 2002; Dowd, 2012). Dimethyl disulphide adducts derivatives were prepared to determine the double bond position in unsaturated fatty acids (Nichols et al., 1986).

4.2.3 **Light and scanning electron microscopy (SEM)**

Exponentially growing cells were harvested after 48 h of growth in LB-medium and then viewed by light microscopy under oil immersion at 100x. For SEM, a culture in stationary phase (~72 h), was pelleted at 3,250 x g for 10 min, and the LB medium was then exchanged and cells were fixed in 2.5% glutaraldehyde spiked phosphate-buffered saline (PBS) [137 mM NaCl,
2.7 mM KCL, 10 mM Na$_2$HPO$_4$ 2H$_2$O, 2 mM KH$_2$PO$_4$, pH 7.4] solution for 30 min on ice. The fixed culture was filtered onto a 0.2 µm pore-size Supor polycarbonate membrane (Pall Port Washington, NY). The filter was washed with PBS and post-fixed in 1% OsO$_4$ for 1 h. The stained filter was passed through a graded ethanol series (25, 50, 70, 95, 100%) at 10 min intervals, and at 100% EtOH, were critical-point dried. A sputter coater applied 5 nm of gold/palladium alloy onto the cells prior to imaging by SEM using a Hitachi S4700 microscope.

4.2.4 DNA extraction, PCR and Illumina library construction

DNA was extracted using QIAamp followed by MinElute cleanup columns (Qiagen Germantown, MD). Full length 16S rDNA was amplified using the universal primers 27f and 1492r (Lane et al., 1991). To ensure coverage of the full length 16S rDNA sequence, another PCR amplification was completed using primers 341f and 907r (Muyzer et al., 1993; 1998). PCR products were sequenced using the standard Sanger method on an ABI3730 (Applied Biosystems Foster City, Ca). The Illumina MiSeq library was constructed using the NxSeq library prep kit (Lucigen Middleton, WI) without the final 14-cycle PCR enrichment to avoid PCR bias, and quality control of the resulting library was completed using Agilent high-sensitivity DNA chips and digital droplet PCR (Hindson et al., 2011; White 3rd et al., 2013 a). The library was sequenced using 250 bp paired-end sequencing on the Illumina MiSeq at GenoSeq UCLA Los Angeles, CA.

4.2.5 Phylogenetic analysis

Sanger sequences obtained from the 27f-1492r and 341f-907r PCR products were merged into full-length 16S rDNA sequence using Consed (Gordon et al., 1998) with manual editing. The phylogenetic position of Agrococcus sp. strain RW1 was assessed using the error-corrected whole-
genome assembled 16S rDNA (~99% similar to PCR amplified) rather than the PCR amplified sequence, due to a lower chance of PCR based substitution errors and greater sequence length (1409 bp vs. 1354 bp). MLST (multiple locus sequencing typing) marker analysis was completed by extracting inferred amino-acid sequences for rpoB (β subunit of bacterial RNA polymerase, ~1156 aa), RecA (recombination protein A, ~352 aa), gyrB (DNA gyrase subunit B, ~679 aa), and ppK (Polyphosphatkinase, ~752 aa) from draft and completed genomes by BLASTP analysis or from prior MLST analysis (A. jenensis strain DSM9580 only, Stackebrandt et al., 2007), and then concatenated them into a ~2939 aa sequence. All phylogenetic analyses were aligned and trimmed using MUSCLE (parameters: -400 gap open with zero gap extended) and clustered using UPGMB. Trees were constructed in MEGA using Maximum likelihood followed by bootstrapping (1000 replicates) and the Jukes-Cantor substitution model (Edgar 2004; version 5.10, Tamura et al., 2011).

4.2.6 Whole genome assembly and genome finishing

Raw Illumina sequencing data were screened for PhiX spike-in contaminants using Bowtie2 (version 2.1.0), removed using Picard tools (version 1.90) (White 3rd and Suttle, 2013c). For the draft genome assembly, read-error correction using AllPaths-LG (version 44837), Celera assembly (version 7.0, including plasmid pHC-CG425: listed in the paper as a 1,427 bp single contig), and read partitioning using Jellyfish (version 1.1.10), were done as described (White 3rd et al., 2013a). Ray genome assembly for Agrococcus sp. RW1 was completed using the error-corrected reads (from AllPath-LG, version 44837) with a k-mer size of 55 bp (Boisvert et al., 2010).
Progressive Mauve was used to find the best representative assembly and contig order, and to complete the genome (Darling et al., 2010). Contigs from Celera and Ray assemblies were pooled, and the remaining gaps were closed by recursive alignments in Mauve. The draft *A. lahaulensis* K22-21 genome from NCBI (version ASM42510v1) was used for genome ordering. The ordered and aligned overlapping contigs were merged using the EMBOSS union script, yielding three circular contigs (Rice et al., 2000).

To confirm the three circular contigs as separate circular genomes, read mapping was used. Error-corrected, phiX-removed reads were mapped back to the genome and plasmids using Bowtie2 with the very sensitive local option (Langmead and Salzberg, 2012). The Bowtie2 readmapping output file (Sam file) was visually inspected using the Tablet program (Milne et al., 2013).

Annotation was completed on RAST using SEED (Aziz et al., 2008). RAST server parameters used SEED subsystems with FigFam under the Glimmer 3 option (Meyer et al., 2009). In addition to RAST, metabolic pathways were predicted using MetaPathways, a modular pipeline for gene prediction and annotation that uses pathway tools and the MetaCyc database to construct environmental pathway/genome databases (ePGBDs) (Paley and Karp, 2006; Konwar et al., 2013; Caspi et al., 2014). Metapathways uses the seed-and-extend homology search algorithm LAST (local alignment search tool) for annotations of ORFs with a minimum of 180 bp and minimum alignment length cutoff of 50 bp (Kielbasa et al., 2011).

Annotations were further analyzed for comparison to the genome of *A. lahaulensis* strain K22-21, for genome synteny, average amino acid identity, and phage lifestyle prediction.
The genome circular plot was constructed using CGviewer (Grant and Stothard, 2008); Celera (k0-k1250) and Ray (k0-k1250v2) assemblies of the A. lahaulensis strain K22-21 genome were mapped to the Agrococcus sp. strain RW1 genome using tBLASTx at an e-value of 1e^{-3} with 50% identity and 25 bp overlap, and then displayed in the CGviewer genome plot. Synteny plots were completed in the RAST server module using a BLAST based dot-plot format (Aziz et al., 2008). Average amino-acid identity (AAIr) analysis and functional gene similarities were calculated on the RAST server module and then parsed with a web-based tool called AAIr (Aziz et al., 2008; Krebs et al., 2013). RAST server annotation predicted a prophage element in the genome, which was analyzed for lytic or lysogenic lifestyle using the phage classification tool set (PHACTS) (McNair et al., 2012).

4.3 Results and discussion

4.3.1 Morphology and growth characteristics

The morphology of Agrococcus sp. strain RW1 shared features with other members of the genus but had a novel phenotype of mycelial-like growth. Agrococcus sp. strain RW1 cells were coccoid during logarithmic growth, and were irregular rod-like or coccoid in stationary phase (Figure 4.1/Table 4.1). Cells of strain RW1 were 0.5 to 0.7 μm in diameter, similar to other described members of the genus (Zhang et al., 2010; Table 4.1). On solidified LB, the colonies were bright lemon-yellow, smooth and circular, and were 0.5 to 2 mm in diameter after ~72 h of growth at 30°C. After one week of growth at 25°C on diluted LB or M-agar, colonies were pale-yellow to white with mycelial-like branched irregular filaments. Mycelial-like growth morphology is common among actinobacteria, including isolates of Streptomyces spp.; however, this phenotype
has not been reported for other members of the *Microbacteriaceae* (Doroghazi and Metcalf, 2013). While branched mycelial-like growth has not been reported for members of the genus, they are known to live on surfaces (e.g. sandstone or seaweed) (Groth *et al.*, 1996). Possibly, mycelial-like growth may favor nutrient acquisition during growth on solid substrates.

*Agrococcus* sp. strain RW1 grew under a wide range of conditions. Growth occurred from pH 6 to 10, at 0 to 6% added NaCl, and over a temperature range of 11 to 42°C (Table 4.1). A close relative of strain RW1, *A. lahaulensis* strain K22-21, has a narrower temperature range of growth between 25 and 37°C, but can grow at salt concentrations as high as 7% (Mayilraj *et al.*, 2006; Table 4.1). Strain RW1 had the highest reported growth temperature for the genus at 42°C, but had no observed growth at <4°C (Zhang *et al.*, 2010; Table 4.1). Although growth of strain RW1 was not observed at 4°C, the temperature of the permanently cold hypolimnion of Pavilion Lake suggests that strain RW1 may find warmer regions of the microbialite mat or that warmer excursions in temperature occasionally occur (Lim *et al.*, 2009).

### 4.3.2 PLFA characterization and comparative analysis

The PLFA composition of *Agrococcus* sp. RW1 was distinct from that of other characterized strains of *Agrococcus* spp., including its close relative *A. lahaulensis*. *Agrococcus* sp. RW1 had half the amount of iC16:0 but three times as much C16:0, compared to *A. lahaulensis* (Mayilraj *et al.*, 2006, Table 4.2). The branched unsaturated PLFA iC15:1Δ4 was 3.5% of the total PLFAs found in *Agrococcus* sp. RW1; it has only been found in trace amounts (<1%) in *A. versicolor* strain K114/01T (Behrendt *et al.*, 2008; Table 4.2). Branched monoenoic PLFAs such as iC15:1Δ4 are typically used as biomarkers for anaerobic sulfate reducing bacteria (Kohring *et
al., 1994); however, Agrococcus sp. RW1 grows aerobically and does not reduce sulfate. Although the PLFA profiles between Agrococcus sp. RW1 and A. lahaulensis K22-21 are quite similar, these differences suggest that the two isolates are from different taxonomic groups.

### 4.3.3 Evolutionary placement of Agrococcus sp. RW1 within the genus

Phylogenetic analysis of the 16S rDNA gene indicates that Agrococcus sp. RW1 is a close relative of A. lahaulensis, in the same clade but evolutionarily distinct (Figure 4.2). In support of this, Agrococcus lahaulensis K22-21 is more closely related to an isolate from human skin than to Agrococcus sp. RW1 (Figure 4.2). However, full-length 16S rDNA sequence alone was unable to resolve whether Agrococcus sp. RW1 and A. lahaulensis are different species.

MLST analysis also supports the argument that Agrococcus sp. strain RW1 and A. lahaulensis are in the same clade, but was unable to resolve whether they are separate species (Figure 4.3). MLST needs a minimum of seven loci to assign species-level classification of closely related bacterial taxa and only four loci are available for the genus Agrococcus; hence, species classification using MLST is currently not tractable (Maiden et al., 2013). Due to low amounts of protein-coding sequences from related isolates, MLST alone was unable to determine whether Agrococcus sp. RW1 and A. lahaulensis are different species.

Agrococcus sp. RW1 and A. lahaulensis are distinct species based on whole genome analysis based on gaps in genomic synteny and functional gene annotations. Agrococcus sp. strain RW1 was the first published draft genome (White 3rd et al., 2013a); subsequently, the genome of A. lahaulensis strain K22-21 was released on NCBI. We mapped the assemblies of A. lahaulensis and Agrococcus sp. RW1 (both Ray and Celera) against the final circular chromosome of
Agrococcus sp. RW1 using tBLASTx, and only the A. lahaulensis assembly showed gaps (Figure 4.4/Table 4.3). Synteny plots revealed 12 large gaps between the genomes of Agrococcus sp. RW1 and A. lahaulensis, along with 1752 non-conserved intergenic regions in A. lahaulensis (Figure 4.5). A comparison of functional gene annotations for Agrococcus sp. RW1 and A. lahaulensis K22-21, using both SEED (RAST-based) and MetaCyc (MetaPathways-based), revealed >200 conserved genes, suggesting high genomic plasticity, and that few genes are shared between the isolates (Figure 4.5). The whole genomic comparison between Agrococcus sp. RW1 and A. lahaulensis K22-21 also indicates the isolates are close relatives, but the differences between the genomes suggests high genomic plasticity within the group.

Average amino-acid identity score between the two genomes supports classifying Agrococcus sp. RW1 and A. lahaulensis K22-21 as different species (Figure 4.5). Average amino-acid identity is a robust measure for bacterial species classification that is based on whole-genome sequences, and that is comparable to DNA-DNA hybridization (Konstantinidis and Tiedje, 2005). The standard cut-off to distinguish bacterial isolates as different species is <70% similarity by DNA-DNA hybridization; this corresponds to <95% average amino-acid identity (Konstantinidis and Tiedje, 2005). The average amino-acid identity score for Agrococcus sp. RW1 and A. lahaulensis K22-21 was only 86.2% based on bidirectional whole-genome best-hit protein analysis using RAST annotation, supporting the classification of the isolates as different species (Konstantinidis and Tiedje, 2005; Krebs et al., 2013; Figure 4.5)
4.3.4 Biochemical properties and antibiotic susceptibility

*Agrococcus* sp. strain RW1 shared many biochemical properties with other members of the genus, except for β-galactosidase activity. *Agrococcus* spp. strain RW1, like other members of the genus, is Gram positive, negative for oxidase, arginine dihydrolase, lysine decarboxylase, ornithine decarboxylase, citrate utilization, hydrogen sulfide production, urease, indole production, inositol/myo-inositol utilization and acetoin production, but positive for catalase (Behrendt et al., 2008). However, strain RW1 was positive for β-galactosidase activity; whereas, other isolates have only weak or no activity (Behrendt et al., 2008; Table 4.4). In *A. lahaulensis* there is no reported β-galactosidase activity or predicted gene in the genome (Mayilraj et al., 2006, Table 4.4). β-galactosidase activity in strain RW1 may be a response to living in microbialite mats, and allow for the degradation of filamentous cyanobacterial mats containing complex exopolysaccharides with β-galactoside moieties.

Tests for antibiotic sensitivity in strain RW1 indicate that sensitivity to penicillin, tetracycline and streptomycin, as well as rifampin is shared with other *Agrococcus* spp. (Weiser et al., 1999; Table 4.5). *Agrococcus* sp. RW1 was sensitive to tobramycin, vancomycin and clindamycin, but resistant to cefixime, sulfisoxazole, oxacillin, trimethoprim and a mixture of sulfamethoxazole/trimethoprim. *Agrococcus citreus* DSM 12453T and *A. jenensis* strains DSM9580T and DSM9996 are sensitive to oxacillin at 5 μg and weakly sensitive to polymyxin (Weiser et al., 1999); whereas, strain RW1 was resistant to oxacillin at 1 μg and sensitive to polymyxin (Table 4.5). The resistance of *Agrococcus* sp. RW1 to 1 μg of oxacillin, given that other *Agrococcus* spp. are sensitive to doses of 5 μg (Wieser et al., 1999), suggests that *Agrococcus* sp.
RW1 could be sensitive to higher oxacillin concentrations. β-lactamase is commonly involved in oxacillin resistance, but evidence for its occurrence was not found in the genomes of either *Agrococcus* sp. RW1 or *A. lahaulensis* K22-21 (Hou et al., 2007). Although there were no putative antibiotic resistance genes predicted within the genome of *Agrococcus* sp. RW1, there were predicted pathways for aromatic compound degradation, including salicylate and gentisate catabolism, which may be involved in resistance. The antibiotic resistance in strain RW1 may have evolved from exposure to toxic organic molecules produced by cyanobacterial mats (Neilan et al., 2013).

4.3.5 **Mobile DNA and viral elements in *Agrococcus* sp. RW1**

Mobile DNA elements, which are plasmid encoded in *Agrococcus* sp. RW1, are involved in integration, transposition, and heavy-metal resistance. Two plasmids were discovered in strain RW1, named pHCCG425 and pLC-RRW783. The 1,427-bp plasmid pHCCG425 had a GC content of 67.8% (4.7% lower than the main chromosome) and encoded a putative integrase and a hypothetical protein of unknown function. However, pHCCG425 shares strong similarities to gene clusters in other *Actinobacteria* including *Brevibacterium linens* and *Mycobacterium* spp. Plasmid pLC-RRW783 was 31,975 bp with a GC content of 70.6% (2% lower than the main chromosome) and 36 ORFs (Figure 4.6). Plasmid pLC-RRW783 contained ORFs annotated for mercuric reduction, arsenic resistance, various metal-dependent proteases, peptidases and ATPases, cadmium and unknown transporters, and an unclassified oxidoreductase. There were also three putative transposases, as well as DNA processing/binding proteins with a suggested role
in DNA transposition and/or integration. Seven ORFs in pLC-RRW783 had no predicted function and are annotated as hypothetical proteins.

Plasmids such as pLC-RRW783 that carry genes associated with heavy-metal resistance could be common in Gram-positive bacteria from cold alkali environments. The plasmid pGIAK1 was found in Bacillus sp. isolated from the Antarctic and is similar to pLC-RRW783 (Guo and Mahillon, 2013). Both plasmids pGIAK1 and pLC-RRW783 encode arsenic (including arsR) and cadmium resistance genes (Guo and Mahillon, 2013). pGIAK1 has been transferred from an isolate of Bacillus sp. to isolates of Staphylococcus spp., so it is possible that pLC-RRW783 could have been transferred by conjugation, phage transduction or transposition (Guo and Mahillon, 2013). The copy number of pLC-RRW783 is unknown, as is whether it can be transferred, or if its genes are expressed.

Annotation of the Agrococcus sp. RW1 genome revealed a putative prophage-like element. The 34,174 bp region on the genome resembled Siphoviridae prophages found in Actinobacteria, and had 43 predicted ORFs and a GC content of 70.1%, which was ~2% less than the GC content of the genome. The phage classification tool set (PHACTS) predicted that the lifestyle of this prophage would be temperate (52% of the trees voting for temperate lifestyle at standard deviation of 0.040) (McNair et al., 2012). Temperate phages are not previously known to be found in the genus Agrococcus. The addition of heat and mitomycin C did not result in induction; hence it is unknown if the prophage can enter the lytic cycle. No prophage or phage genes are predicted in the genome of A. lahaulensis. The prophage in Agrococcus sp. RW1, has conserved genes for a phage tape-measure protein and protease that are related to other Streptomyces phages including
VWB, phi-c31, phi-BT1, as well as the *Mycobacterium* phage Brujita. These phages belong to the *Siphoviridae* and have a temperate lifestyle, which is consistent with predictions using PHACTS (Gregory *et al.*, 2003; Van Dessel *et al.*, 2005).

### 4.3.6 *Agrococcus* sp. RW1 life in a cold microbialite mat

Pavilion Lake microbialites exist in permanently cold (4 ° to 8 °C) water (Lim *et al.*, 2009), and the genome of *Agrococcus* sp. RW1 has signatures of cold adaptation. The *Agrococcus* sp. RW1 genome encoded a single copy of *cspA* gene (cold shock protein A), which is consistent with growth under the *in situ* cold conditions of Pavilion Lake. CspA is induced and essential for growth at <10 °C, and functions as a molecular chaperone that binds mRNA, preventing secondary structure formation and ensures translation at low temperatures (Yamanka and Inouye, 1997). CspA is essential and expressed during subzero temperature growth in the genus *Exiguobacterium* (Rodrigues *et al.*, 2008). In *Agrococcus* sp. RW1, CspA function under cold stress has not been confirmed, nor has growth at <10°C, although analysis of the genome predicts the ability to grow under colder conditions.

Heavy-metal resistance gene clusters were also located on the main chromosome of *Agrococcus* sp. RW1. A mercury resistance operon spans positions 501 to 8136 bp, and includes a transposase, mercuric ion reductase, FAD dependent oxidoreductase, thiol oxidoreductase, and arsenic resistance protein are gene duplications presumably from a prior pLC-RRW783 plasmid transposition. Pavilion Lake has very low levels of zinc (0.01 to 0.03 mg L⁻¹) and undetectable levels of cobalt, copper, chromium, arsenic and cadmium (Lim *et al.*, 2009); hence, it is unclear why an organism from Pavilion Lake would carry gene clusters for heavy-metal resistance.
However, mercury was actively mined near Pavilion Lake in the 1940s (Stevenson, 1940), indicating that mercury could have been at a higher concentration in the lake, making these heavy-metal resistance genes a possible remnant of this event. *Agrococcus* sp. RW1 is a limited component of the Pavilion Lake microbialite community (<1% metagenomic recruitment) (unpublished data, Thesis Chapter 2). Sequences in the metagenome that recruited with high similarity (>95%) by confirmatory tBLASTx ($10^{-3}$) to the *Agrococcus* sp. RW1 genome were predicted heavy-metal resistance genes (unpublished data, Thesis Chapter 2). This suggests that *Agrococcus* sp. RW1 contributes heavy-metal resistance genes, thereby adding new capabilities for metal metabolism to the Pavilion Lake microbialite community.

The *Agrococcus* sp. RW1 genome encodes a phosphate (Pho) regulon for the high-affinity uptake of phosphate, which could be an adaptation to the low concentration of total phosphorus (3.3 µg L$^{-1}$) present in Pavilion Lake (Lim et al., 2009). Phosphate (Pho) regulon encoded by *Agrococcus* sp. RW1 included the phosphate permease protein (PstA), phosphate regulon sensor protein (PhoR) and the phosphate regulon transcriptional regulatory protein (PhoB). The phosphate regulon proteins (PhoR/PhoB) and PstA are not predicted within the *A. lahaulensis* genome. Genes associated with phosphorus scavenging have been found in other freshwater microbialites as well (Breitbart et al., 2009). The response of the Pho regulon in *Agrococcus* sp. RW1 under phosphorus limitation remains to be tested; however, it may be a footprint of living in an oligotrophic environment.

*Agrococcus* sp. RW1 is able to metabolize a wide range of carbon compounds, including amygdalin (Table 4.3). The ability of *Agrococcus* sp. RW1 to use amygdalin was surprising.
because has not been reported for other Agrococcus spp., and amylodalin-specific glycosylases were not predicted in the genome. The ability to use amylodalin is found in the distant relative, Rhodococcus kunmingensis, an actinobacterium isolated from soil (Wang et al., 2008). Analysis of the Agrococcus sp. RW1 genome predicts carbohydrate utilization pathways for mannose, fructose, D-gluconate, trehalose, D-ribose, and glycogen, as well as for chitin, lactate, glycerate, deoxyribose and deoxynucleoside catabolism. Carbohydrate utilization tests of D-glucose, D-mannitol and D-sucrose validated the putative metabolic potential observed in the Agrococcus sp. RW1 genome (Table 4.3). Agrococcus sp. RW1 was able to grow on many more single carbon sources compared to other members of the genus, possibly allowing access to carbon provided by cyanobacterial mats (Breitbart et al., 2009; Table 4.3).

4.3.7 Carotenoid biosynthesis

The pathway responsible for the yellow pigmentation in the genus Agrococcus has not been described, although it has been suggested that diaminobutyric acid within the cell wall could impart the distinctive lemon-yellow colony color (Groth et al., 1996). However, it is unlikely that diaminobutyric acid is completely responsible for the yellow pigmentation in Agrococcus spp., as diaminobutyric acid is a quite rare. And, many other bacteria (e.g. Cronobacter sakazakii) have yellow colonies in the absence of diaminobutyric acid (Zhang et al., 2014). As well, actinobacteria, which form yellow colonies, produce C_{40} carotenoids (e.g. canthaxanthin and echinenone) and C_{50} carotenoids (e.g. flavuxanthin) (Tao et al., 2007; Klassen, 2010).

Cronobacter sakazakii strain BAA894 is a Gammaproteobacteria that also produces yellow pigmented colonies (Zhang et al., 2014). The carotenoid biosynthetic pathway from C.
sakazakii strain BAA894, when reconstructed in E. coli, produced lycopene, β-carotene, cryptoxanthin or zeaxanthin (Zhang et al., 2014). The production of zeaxanthin or zeaxanthin glycoside from C. sakazakii strain BAA894 pathways in E. coli changed the colony pigmentation from white to yellow (Zhang et al., 2014).

We therefore screened Agrococcus sp. RW1 and A. lahaulensis K22-21 genomes for genes involved in the production of carotenoids, and based on these results, propose a C\textsubscript{40}/C\textsubscript{50} carotenoid biosynthetic pathway (Figure 4.7). Both Agrococcus sp. RW1 and A. lahaulensis K22-21 have the genetic potential to produce lycopene, β-carotene, canthaxanthin, echinenone, and/or zeaxanthin or astaxanthin (Figure 4.7). In both Agrococcus sp. RW1 and A. lahaulensis K22-21 the carotenoid biosynthetic pathway includes predicted genes for phytoene synthase, phytoene dehydrogenase, beta-carotene ketolase, and a second beta-carotene-like ketolase that likely converts echinenone to canthaxanthin, which are characteristic of C\textsubscript{40}/C\textsubscript{50} carotenoid pathway (Tao et al., 2007; Figure 4.7). Genes within the carotenogenic gene cluster of Agrococcus sp. RW1 and A. lahaulensis K22-21 shared similarity to Brevibacterium linens, a C\textsubscript{50} carotenoid producer (Krubasik et al., 2000).

The proposed carotenoid biosynthetic pathways in both Agrococcus sp. RW1 and A. lahaulensis K22-21 are similar to the pathway described in Cronobacter sakazakii, which produces yellow pigmented colonies (Zhang et al., 2014). While we predict that zeaxanthin is responsible for the yellow pigmentation in Agrococcus, astaxanthin, another yellow-pigmented carotenoid, could alternatively be responsible (Figure 4.7). No hydrolases are predicted by either the Agrococcus sp. RW1 or A. lahaulensis genome, which is the only enzyme known to convert β-carotene to zeaxanthin or canthaxanthin to astaxanthin (Klassen, 2010; Zhang et al., 2014). It is
also possible that a hypothetical gene within the genome is functioning as a hydrolase to convert β-carotene to zeaxanthin or canthaxanthin to astaxanthin. Preliminary HPLC data suggests that the yellow pigment is carotenoid like however, we were unable to determine its structure.

The function of yellow pigmentation within *Agrococcus* sp. RW1 remains unclear. Low dissolved organic carbon (DOC) in Pavilion Lake allows for high penetration of solar UV radiation, which benthic microbial mats partially protect against with carotenoids (Lim *et al.*, 2009; Lionard *et al.*, 2012). Similarly, the yellow pigment may help protect *Agrococcus* sp. RW1 from photo-damage in Pavilion Lake.

### 4.4 Conclusions

*Agrococcus* sp. strain RW1 was isolated from a modern microbialite, and has a number of features that distinguish it from previously characterized members of this genus. These include the presence of PLFA iC15:1Δ^4_, β-galactosidase activity, the amygdalin utilization, and filamentous growth in diluted medium. Phylogenetic analysis using either 16S rDNA or MLST could not resolve *Agrococcus* sp. RW1 and *A. lahaulensis* as different species, but placed them consistently in the same clade. However, whole genome analysis indicates that *Agrococcus* sp. RW1 and *A. lahaulensis* K22-21 are separate species, based on relatively low functional gene conservation, less than 95% amino-acid identity, and the presence of many non-conserved intergenic regions within the *A. lahaulensis* genome. We suggest the renaming of *Agrococcus* sp. RW1 to *Agrococcus pavilionensis* RW1 based on our polyphasic analysis.

The genome of *Agrococcus* sp. RW1 suggests genetic promiscuity due to mobile elements, plasmids and predicted prophage. The plasmids could act as mobile elements through replicative
transposition leading to high genomic plasticity between *Agrococcus* sp. RW1 and *A. lahaulensis*. The LC-RRW783 plasmid and chromosome encodes genes related to heavy-metal resistance, which could contribute to the heavy metal-resistance genes found in the metagenome. Although the predicted prophage in *Agrococcus* sp. RW1 could not be induced to enter the lytic phase, such a lysogen may also act as a mobile element in the genome.

The presence of cold-shock proteins, a low phosphorous response regulon (Pho), heavy-metal resistance, and carotenogenic gene clusters that are likely responsible for the yellow pigmentation, are features that are likely highly conducive to living in Pavilion Lake. These results will provide a blueprint for future studies of members of this genus including their distribution, stress response, pigment synthesis, and phage interactions.
4.5 **Tables and figures**

Table 4.1: Comparison of physiological properties of *Agrococcus* selected strains

<table>
<thead>
<tr>
<th></th>
<th>Strain RW1</th>
<th><em>A. lahaulensis</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th><em>A. baldri</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th><em>A. citreus</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th><em>A. jenensis</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th><em>A. terreus</em>&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Habitat of isolation</td>
<td>microbialite</td>
<td>desert soil</td>
<td>K22-21</td>
<td>V-108</td>
<td>D-1/1a</td>
<td>2002-39/1</td>
</tr>
<tr>
<td>Cell size (µm)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Length</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>width</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.5 - 0.7</td>
<td>1.0 - 1.5</td>
<td>1.1 - 1.7</td>
<td>1.1 - 1.7</td>
<td>0.7 - 1.7</td>
<td>0.8 - 1.0</td>
</tr>
<tr>
<td></td>
<td>0.3 - 0.5</td>
<td>0.6 - 1.0</td>
<td>0.7 - 1.0</td>
<td>0.7 - 1.0</td>
<td>0.7 - 1.0</td>
<td>0.4 - 0.5</td>
</tr>
<tr>
<td>Growth temp range(°C)</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>11</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>30</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>37</td>
<td>+</td>
<td>+</td>
<td>W</td>
<td>+</td>
<td>V</td>
<td>+</td>
</tr>
<tr>
<td>42</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>NA</td>
<td>-</td>
<td>NA</td>
</tr>
<tr>
<td>pH growth range</td>
<td>+</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>10</td>
<td>+</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>-</td>
</tr>
<tr>
<td>NaCl tolerance (%)</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td>0</td>
<td>+</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>+</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>NA</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>-</td>
<td>+</td>
<td>NA</td>
<td>+</td>
<td>NA</td>
<td>-</td>
</tr>
</tbody>
</table>

- Negative, + Positive, W: Weak growth, V: Variable growth, NA: not available, Temp: Temperature
All growth measurements for strain RW1 were taken after 3 d.
<table>
<thead>
<tr>
<th></th>
<th>Strain RW1</th>
<th>A. lahaulensis&lt;sup&gt;1&lt;/sup&gt; K22-21</th>
<th>A. baldri&lt;sup&gt;1&lt;/sup&gt; V-108</th>
<th>A. citreus&lt;sup&gt;1&lt;/sup&gt; D-1/1a</th>
<th>A. jenensis&lt;sup&gt;1&lt;/sup&gt; 2002-39/1</th>
</tr>
</thead>
<tbody>
<tr>
<td>C&lt;sub&gt;14:0&lt;/sub&gt;</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>tr</td>
<td>tr</td>
</tr>
<tr>
<td>iC&lt;sub&gt;14:0&lt;/sub&gt;</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>tr</td>
<td>tr</td>
</tr>
<tr>
<td>C&lt;sub&gt;15:0&lt;/sub&gt;</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>tr</td>
<td>ND</td>
</tr>
<tr>
<td>iC&lt;sub&gt;15:1Δ4&lt;/sub&gt;</td>
<td>3.5</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>iC&lt;sub&gt;15:1&lt;/sub&gt;</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>1.9</td>
</tr>
<tr>
<td>aiC&lt;sub&gt;15:1&lt;/sub&gt;</td>
<td>ND</td>
<td>tr</td>
<td>ND</td>
<td>ND</td>
<td>0</td>
</tr>
<tr>
<td>iC&lt;sub&gt;15:0&lt;/sub&gt;</td>
<td>8.8</td>
<td>9.9</td>
<td>5.7</td>
<td>10</td>
<td>12.2</td>
</tr>
<tr>
<td>aiC&lt;sub&gt;15:0&lt;/sub&gt;</td>
<td>46.6</td>
<td>48.4</td>
<td>44.9</td>
<td>53.1</td>
<td>57.8</td>
</tr>
<tr>
<td>C&lt;sub&gt;16:0&lt;/sub&gt;</td>
<td>5.5</td>
<td>1.8</td>
<td>3.0</td>
<td>1.7</td>
<td>2</td>
</tr>
<tr>
<td>iC&lt;sub&gt;16:0&lt;/sub&gt;</td>
<td>2.8</td>
<td>5.8</td>
<td>7.5</td>
<td>12.0</td>
<td>12.6</td>
</tr>
<tr>
<td>iC&lt;sub&gt;17:0&lt;/sub&gt;</td>
<td>3.1</td>
<td>4.8</td>
<td>1.5</td>
<td>1.7</td>
<td>1.9</td>
</tr>
<tr>
<td>aiC&lt;sub&gt;17:0&lt;/sub&gt;</td>
<td>29.7</td>
<td>27.6</td>
<td>24.3</td>
<td>13.2</td>
<td>9.3</td>
</tr>
<tr>
<td>aiC&lt;sub&gt;17:1&lt;/sub&gt;</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>C&lt;sub&gt;18:0&lt;/sub&gt;</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>tr</td>
<td>tr</td>
</tr>
</tbody>
</table>

Values are percentages of total phospholipid fatty acids. ND: not detected, tr: traces (<1%). Strain RW1: *Agrococcus* sp.<sup>1</sup> Mayilraj *et al.*, 2006.
Table 4.3: *Agrococcus* spp. final assembly statistics

<table>
<thead>
<tr>
<th>Strain RW1</th>
<th><em>A. lahaulensis</em> K22-21</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Assembler</strong></td>
<td>Celera*</td>
</tr>
<tr>
<td><strong>Assembly Name</strong></td>
<td>k0-k1250</td>
</tr>
<tr>
<td><strong>No. Contigs</strong></td>
<td>50</td>
</tr>
<tr>
<td><strong>No. Scaffold</strong></td>
<td>NA</td>
</tr>
<tr>
<td><strong>Genome Size (bp)</strong></td>
<td>2,878,403</td>
</tr>
<tr>
<td><strong>N50</strong></td>
<td>133,224</td>
</tr>
<tr>
<td><strong>N90</strong></td>
<td>31,609</td>
</tr>
<tr>
<td><strong>Largest Contig (bp)</strong></td>
<td>308,429</td>
</tr>
<tr>
<td><strong>G+C%</strong></td>
<td>72.53</td>
</tr>
</tbody>
</table>

*Published in White 3rd *et al.*, 2013a.*
Table 4.4: Biochemical properties of *Agrococcus* spp. selected strains

<table>
<thead>
<tr>
<th>Activity of:</th>
<th>Strain RW1</th>
<th>A. <em>lahaulensis</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th>A. <em>baldri</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th>A. <em>citreus</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th>A. <em>jenensis</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th>A. <em>versicolor</em>&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Strain RW1</td>
<td>K22-21</td>
<td>V-108</td>
<td>D-1/1α</td>
<td>2002-39/1</td>
<td>K114/01(T)</td>
</tr>
<tr>
<td>Gelatinase</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>ND</td>
</tr>
<tr>
<td>β-galactosidase</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>W</td>
</tr>
<tr>
<td>Assimilation of:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-glucose</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D-mannitol</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>W</td>
</tr>
<tr>
<td>D-sorbitol</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L-Rhamnose</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-sucrose</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Amygdalin</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L-arabinose</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

ND: Not detected, W: Weakly positive, Strain RW1: *Agrococcus* sp.

<sup>1</sup>Behrendt *et al.*, 2008.
Table 4.5: Antibiotic susceptibility of *Agrococcus* spp. selected strains

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Disc content</th>
<th>Strain RW1</th>
<th>A. citreus&lt;sup&gt;1&lt;/sup&gt;</th>
<th>A. jenensis&lt;sup&gt;2&lt;/sup&gt;</th>
<th>A. jenensis&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulfamethoxazole +</td>
<td>23.75 +</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Trimethoprim</td>
<td>1.25 µg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Penicillin</td>
<td>10 IU</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Clindamycin</td>
<td>2 µg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Rifampin*</td>
<td>5 / 30 µg</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Polymyxin</td>
<td>300 IU</td>
<td>-</td>
<td>W</td>
<td>W</td>
<td>W</td>
</tr>
<tr>
<td>Cefixime</td>
<td>5 µg</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Sulfisoxazole</td>
<td>300 µg</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Oxacillin**</td>
<td>1 / 5 µg</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tetracycline</td>
<td>30 µg</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Trimethoprim</td>
<td>5 µg</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Tobramycin</td>
<td>10 µg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Vancomycin</td>
<td>30 µg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Streptomycin</td>
<td>10 µg</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

+ Resistant, - Sensitive, W: Weakly Sensitive, ND: No Data

*Strain RW1 was sensitive to 5 µg, whereas the other strains were sensitive at 30 µg.

**Strain RW1 was resistant to 1 µg, whereas the other strains were sensitive at 5 µg.

Strain RW1 was not tested at 5 µg Oxacillin, only at 1 µg.

<sup>1</sup>Wieser *et al.*, 1999.
Figure 4.1: *Agrococcus* sp. strain RW1 microscopy
A) Light microscopy image during logarithmic phase of growth (48 h). Scale bar: ~100 μm.
B) SEM image during stationary phase of growth (72 h). Scale bar: ~10 μm.
Observed here at standard growth conditions in LB medium at 30°C.
Figure 4.2: Maximum-likelihood tree of 16S rRNA gene sequences for Agrococcus spp. and related taxa (~1409 bp) Bootstrap values greater than 50% at given branches.
Figure 4.3: MLST (multiple locus sequence typing) maximum-likelihood tree of Microbacteriaceae (~2939 amino acids) Bootstrap values greater than 50% at given branches.
Figure 4.4: Genome plot (~2.6Mb) of *Agrococcus* sp. strain RW1
Genome key (left corner): starts with inner most ring which is a genome ruler followed by GC skew (purple/green) and ends with two outer rings which contain protein-coding ORFs, tRNAs and rRNAs.
Figure 4.5: Genome synten and Venn diagrams of *Agrococcus* sp. RW1 vs. *A. lahaulensis* K22-21
A) RAST bidirection blast genome synten dot-plot with Progressive Mauve alignments as axes. Red dots are positive blast hits based on RAST genome comparision module. Red line regions at the axes are regions of synteny based on Progressive Mauve. Average amino-acid identity was calculated by RAST functional module with web based tool AAIr (Krebs et al., 2013).
B) Venn Diagrams based on RAST SEED/Figfams and Metapathways MetaCyc annotations.
Figure 4.6: *Agrococcus* sp. RW1 LC-RRW783 plasmid genomic plot
Figure 4.7: Proposed carotenoid biosynthetic pathway for Agrococcus

*Phytoene synthase, **lycopene cyclase, ***beta - carotene ketolase, ****beta - carotene like ketolase are predicted in the genomes of Agrococcus sp. RW1 or A. lahaulensis K22-21.
Chapter 5:

Illumination of an *Exiguobacterium* through the comparative lenses of classical characterization and modern genomic analysis

### 5.1 Introduction

Here we describe a second non-photosynthetic pigmented isolate from a 20 m Pavilion Lake microbialite, *Exiguobacterium* sp. strain RW2. The pathway responsible for the orange colony pigmentation in the genus *Exiguobacterium* has not been described.

The genus *Exiguobacterium* is found in many environments due to its tolerance of a wide range of conditions, including pH (5 to 11), salinity (NaCl: 0 to 16%), heavy metals, and temperature (-12°C to 55°C), which is rooted in its diverse metabolic capabilities (Kim et al., 2005; Vishnivetskaya et al., 2009; Belfiore et al., 2013; White 3rd et al., 2013b). The first described member of the genus, *E. aurantiacum* DSM6208^T^, was isolated from a potato processing plant (Collins et al., 1983). The genus *Exiguobacterium* has fifteen named species, which were isolated from diverse environments, including permafrost (Rodrigues et al., 2006), deep-sea vents (Crapart et al., 2007) and hot springs (Vishnivetskaya et al., 2009, 2011; Table 5.1 and Table 5.2).

The distribution of *Exiguobacterium* spp. across many different environments is related to their growth under a wide range of conditions. Consequently, isolates of *Exiguobacterium* have been used as model organisms for understanding thermal adaptation (Rodrigues et al., 2008), heavy metal stress (Belfiore et al., 2013), and the function of non-marine proteorhodopsin (Gushchin et al., 2013). Their metabolic capabilities have industrial and
biotechnological applications, including production of biofuels (Vijayalaxmi et al., 2013), cheese (Lusk et al., 2012) and L-lactate (Jiang et al., 2013a-b), removal of chemicals associated with the textile industry, including chloronitrophenols (CNPs) and N-methylated diaminotriphenylmethane (Arora et al., 2012; Wang et al., 2012), use as biodetergents (Anbu et al., 2013), and for degrading shrimp biowaste (Anil and Suresh, 2013). Further insights into the genetic potential of these microbes may yield a greater understanding of their metabolic capabilities in nature and possibly reveal other industrial applications.

Seven genomes from the genus *Exiguobacterium* have been completed as either drafts (many contigs) or complete genomes (as a single chromosome with plasmids). These include draft genomes of *Exiguobacterium* sp. strain S17 from a stromatolitic microbialite (Ordoñez et al., 2013), strain 8-11-1 from lake sediments (Jiang et al., 2013), strain OS77 from a Japanese hot-spring (Nonaka et al., 2014), and complete genomes for *E. sibiricum* strain 255-15 from permafrost (Rodrigues et al., 2008), *E. antarcticum* strain B7 from a microbial mat in Antarctica (Azevedo et al., 2012), *Exiguobacterium* sp. strain MH3 from the rhizosphere of duckweeds (complete genome, Tang et al., 2013), and *Exiguobacterium* sp. strain AT1b from Mammoth Terrace hot spring, Yellowstone National Park (Vishnivetskaya et al., 2011). Eleven other ongoing genome sequencing projects through the Joint Genome Institute include *E. acetylicum* strain DSM20416 (Farrow et al., 1994), *E. aurantiacum* strain DSM6208 (Collins et al., 1983), *E. marinum* strain DSM16307 (Kim et al., 2005), *E. oxidotolerans* strain JCM12280 (Yumoto et al., 2004) and *E. sibiricum* strain 7-3 (Vishnivetskaya et al., 2009).
In this contribution, classical and modern genomic analyses are used to characterize the pigmented bacterium *Exiguobacterium* sp. strain RW2, an isolate from a modern freshwater microbialite. The genome encodes many putative proteins, potentially involved in the adaptive physiological response to ranges of salinity, pH, temperatures and heavy-metal stress. As well, a putative carotenoid biosynthetic pathway is likely related to the characteristic orange pigmentation found in the genus. These results provide further insight into the potential of this genus to exploit such a wide range of environmental conditions.

5.2 Materials and methods

5.2.1 Isolation, growth conditions, biochemical and antibiotic susceptibility tests

*Exiguobacterium* sp. strain RW2 was isolated by plating 0.5 g of a thrombolitic microbialite, collected from 20 m depth in Pavilion Lake, British Columbia (50.86 °N, 121.74 °W), onto M-agar medium [0.5% (w/v) tryptone, 0.25% (w/v) yeast extract, 1% (w/v) NaCl, 1.5% (w/v) agar, pH 7], followed by incubation at 30 °C for 3 d (White 3rd et al., 2013b). M-agar was also used for maintenance (at 30 °C) and for assessing growth with respect to temperature (4, 5, 11, 16, 20, 25, 30, 37, 42, 45, 50, 55 °C at pH 7 and 1% NaCl), pH (4, 5, 6, 6.5, 7, 7.5, 8, 10, 10.5, 11, 12 at 30 °C) and salinity (0, 1, 3, 6, 9, 12, 13, 16 % NaCl at 30 °C) for 3 d. Strain characteristics, including colony and cell morphologies were determined by standard methods (Murry et al., 1994). Oxidase tests and biochemical enzyme assays and carbohydrate utilization were done using API20E (BioMérieux) test strips on cultures re-suspended in sterile water. Antibiotic susceptibility was determined by the Kirby-Bauer method using antibiotic discs on M agar (Collee et al., 1996). Prophage induction assays were done either by adding 0.2 µg ml⁻¹ of
mitomycin C, or heating cultures to 45 °C, incubating at 30 °C for 3-10 h and monitoring changes in turbidity until a decrease of OD<sub>600</sub> to 0.1 or less was observed (starting OD<sub>600</sub> 0.6). At several time points cells were pelleted at 3,250 x g for 10 mins and the supernatent filter-sterilized twice through a 0.22 µm pore-size Millex Durapore PVDF membrane (Millipore) filter before plating using a double agar overlay plaque assay (Kropinski et al., 2009).

### 5.2.2 Phospholipid fatty acid analysis (PLFA)

Phospholipid fatty acids were extracted from cultures grown in 25 ml M broth for 4 d at 22 °C. Cultures were transferred into pre-combusted vials for an overnight solvent extraction in a 1:2:0.8 ratio of dichloromethane (DMC), methanol (MeOH) and phosphate-buffered saline (PBS) [137 mM NaCl, 2.7 mM KCl, 10 mM Na<sub>2</sub>HPO<sub>4</sub> 2H<sub>2</sub>O, 2 mM KH<sub>2</sub>PO<sub>4</sub>, pH 7.4] (Bligh and Dyer, 1959). The extract was filtered through a separatory funnel where DCM and water were added to achieve a mixture of MeOH:DCM:water of 1:1:0.9 (Bligh and Dyer, 1959). The lower organic phase was removed and purified into polar, neutral, and non-polar fractions using liquid chromatography through silica-gel. Phospholipids present in the polar fraction were subjected to mild alkaline methanolysis to produce fatty-acid methyl esters (FAMEs) (Guckert et al., 1985). FAMEs were separated, identified, and quantified using gas chromatography mass spectrometry (GC/MS) (Agilent Technologies Inc., Santa Clara, CA, USA) with a DB-5MS capillary column (30 m x 0.32 mm I.D. x 0.25 µm film thickness) under a temperature regime of 50 °C (1 min), 20 °C min<sup>-1</sup> to 130 °C, 4 °C min<sup>-1</sup> to 160 °C, and 8 °C min<sup>-1</sup> to 300 °C (5 min). PLFAs were identified by retention time and mass spectra relative to those of reference standards (Bacterial Acid Methyl Ester Mix, Matreya Inc., Pleasant Gap, PA, USA; and Supelco 37
Component FAME Mix, Sigma-Aldrich Co., Bellefonte, PA, USA). A modified picolinyl ester derivatization was used to determine the branching point in unknown compounds (Destaillets and Angers, 2002; Dowd, 2012). Dimethyl disulphide adducts derivatives were prepared to determine the double bond position in unsaturated fatty acids (Nichols et al., 1986).

5.2.3 **Light and scanning electron microscopy (SEM)**

Exponentially growing cells harvested 26 h after being transferred to liquid M-medium were viewed by light microscopy under oil immersion at 100x. For SEM, a culture in mid-stationary phase 72 h after being transferred to liquid M-medium was pelleted at 3,250 x g for 10 min, then the M medium was exchanged and cells were fixed in 2.5% glutaraldehyde spiked phosphate-buffered saline (PBS) [137 mM NaCl, 2.7 mM KCl, 10 mM Na₂HPO₄ 2H₂O, 2 mM KH₂PO₄, pH 7.4] solution for 30 min on ice. The fixed culture was filtered onto a 0.2 µm pore-size Supor polycarbonate membrane (Pall Port Washington, NY). Cells on the filter were washed with PBS and post-fixed in 1% OsO₄ for 1 h. The fixed culture was filtered onto a 0.2 µm pore-size Supor polycarbonate membrane (Pall Port Washington, NY). The filter was washed with PBS and post-fixed in 1% OsO₄ for 1 h. The stained filter was passed through a graded ethanol series (25, 50, 70, 95, 100%) at 10 min intervals, and at 100% EtOH, was critical-point dried. A sputter coater applied 5 nm of gold/palladium alloy onto the cells prior to imaging by SEM using a Hitachi S4700 microscope.

5.2.4 **DNA extraction, PCR and Illumina library construction**

DNA was extracted using QIAamp followed by MinElute cleanup columns (Qiagen Germantown, MD). Full length 16S rDNA was amplified using the universal primers 27f and
1492r (Lane et al., 1991). To ensure coverage of the full length 16S rDNA sequence, another PCR amplification was completed subsequently using primers 341f and 907r (Muyzer et al., 1993; 1998). PCR products were sequenced using the Sanger method on an ABI3730 (Applied Biosystems Foster City, Ca). The Illumina MiSeq library was constructed using the Lucigen NxSeq library prep kit (Lucigen Middleton, WI) without the final 14-cycle PCR enrichment to avoid PCR bias; quality control of the resulting library was completed using Agilent high-sensitivity DNA chips and digital droplet PCR (Hindson et al., 2011; White 3rd et al., 2013 a-c). The library was sequenced using 250 bp paired-end sequencing on the Illumina MiSeq at GenoSeq UCLA Los Angeles, CA.

5.2.5 Whole genome assembly, assembler comparison, genome finishing and annotation

Raw Illumina sequencing data was screened for PhiX spike-in contaminants using Bowtie2 (version 2.1.0) then removed using Picard tools (version 1.90) (White 3rd and Suttle, 2013c). The draft genome assembly was completed using Ray (version 2.2.0) using the error-corrected reads (from AllPath-LG, version 44837) using a k-mer size of 55 bp (Boisvert et al., 2010; White 3rd et al., 2013b). The genome of Exiguobacterium sp. strain AT1b was downloaded from NCBI for comparison, scaffolding and contig ordering. Progressive Mauve was used to find the best representative assembly, contig order among the various assemblies, and for scaffolding (Darling et al., 2010). The Exiguobacterium sp. strain RW2 contigs were ordered and aligned in Progressive Mauve to Exiguobacterium sp. strain AT1b to visualize genome order (Vishnivetskaya et al., 2011). The remaining gaps between contigs were closed by recursive
alignments using Mauve. The ordered contigs with overlaps were merged into a singular circular contig using the EMBOSS union script (Rice et al., 2000).

Error-corrected, phiX-removed reads used in the *Exiguobacterium* sp. strain RW2 Ray assembly were mapped back to the final singular circular genome using Bowtie2 with the very sensitive local option (Langmead and Salzberg, 2012). The Bowtie2 read mapping output file (Sam file) was visually inspected by the Tablet program (Milne et al., 2013). The contigs that were screened for overlaps and read mapping depth (>10x) were then merged manually.

Annotations were further analyzed by comparison to other genomes, genome synteny, and average amino-acid identity. A genome circular plot of *Exiguobacterium* sp. strain RW2 was constructed using CGviewer (Grant and Stothard, 2008), and compared to the genomes of *E. sibiricum* strain 255-15, *Exiguobacterium* sp. strain AT1b and *Exiguobacterium* sp. strain MH3 using tBLASTx at an e-value of 1e\(^{-3}\) with 50% identity and 25 bp overlap, and then displayed in the CGviewer genome plot (Grant and Stothard, 2008). Synteny plots were completed in the RAST server module using a BLAST based dot plot format (Aziz et al., 2008). Average amino-acid identity (AAIr) analysis and functional gene similarities were calculated on the RAST server module and then parsed with a web-based tool AAIr (Aziz et al., 2008; Krebs et al., 2013).

### 5.2.6 Phylogenetic analysis and multiple locus sequence typing (MLST) analysis

Sanger sequences obtained from 27f-1492r and 341f-907r PCR products were merged into a full length 16S rDNA reference sequence using Consed (Gordon et al., 1998) with manual editing. The phylogenetic position of *Exiguobacterium* sp. strain RW2 was assessed using the error-corrected whole-genome assembled 16S rDNA (~99% similar to PCR amplified) rather
than the PCR amplified sequence, due to a lower chance of PCR based substitution errors and having greater total sequence (1353 bp vs. 1276 bp). All sequences for phylogenetic analysis were aligned, trimmed using MUSCLE parameters (-400 gap open with zero gap extended), and then clustered using UPGMB (Edgar 2004). MLST (multiple locus sequencing typing) marker analysis was completed by extracting amino-acid sequences for RpoB (β subunit of bacterial RNA polymerase, ~1180 aa), RecA (recombination protein A, ~349 aa), GyrB (DNA gyrase subunit B, ~643 aa), DnaK (Heat shock protein 70, ~614 aa), IDH (isocitrate dehydrogenase ~422 aa), and IleS (isoleucyl-tRNA synthetase, ~ 921 aa) from *Exiguobacterium* spp. genomes by BLASTP analysis. The MLST protein marker genes (RpoB, RecA, GyrB, DnaK, IDH and IleS) were then concatenated based on strain into ~3856 aa sequences. Maximum-Likelihood trees were constructed using the trimmed 16S rRNA gene sequences (~1353 bp), 5S-16S-23S combined sequences (~4350 bp), and the concatenated MLST sequences (~3856 aa) with bootstrapping (1000 replicates), using the Jukes-Cantor substitution model in MEGA (version 5.10, Tamura et al., 2011).

### 5.3 Results and discussion

#### 5.3.1 Morphology and growth characteristics

*Exiguobacterium* sp. strain RW2 is a Gram-positive facultative anaerobe. After 48 h growth at 30°C on M-agar, it forms bright orange, smooth, circular colonies that are typically 3 to 4 mm in diameter (Table 5.2). In liquid M-medium at 30 °C, the cells are coccoid in logarithmic growth phase after 24 to 48 h, and irregular shaped after ~72 h in late stationary phase, coincident with the beginning of biofilm formation (Figure 5.1). Biofilm formation could facilitate
association with the microbial mats found on microbialites in Pavilion Lake, and may be a survival mechanism in nutrient-poor or starvation states (Shao et al., 2013). Hence, biofilm formation could be a survival strategy for *Exiguobacterium* sp. strain RW2 in the cold oligotrophic conditions of Pavilion Lake (Lim et al., 2009), and is consistent with the isolation of *Exiguobacterium antarcticum* strain DSM14480 and *E. antarcticum* strain B7 from a cold microbial mat and biofilm (Vishnivetskaya et al., 2009).

*Exiguobacterium* sp. strain RW2 grows under a wide range of temperatures, from 4 to 50 °C, which is the largest temperature range for growth reported for the genus (Table 5.2). Other isolates from Antarctica and permafrost can grow at even lower temperatures, with *E. antarcticum* strain DSM 14480, and *E. sibiricum* strains 255-15, 7-3, and 190-11 growing at -2.5 °C (Vishnivetskaya et al., 2009). Temperatures at which Pavilion Lake microbialites are stably 4°C throughout the year (Lim et al., 2009), (Lim et al., 2009), removing the selective pressure for growth at lower temperatures. At the other end of the temperature range, *Exiguobacterium* sp. strain RW2 grows at 50°C, and is the first member of this genus that has been isolated from a permanently cold environment that has been demonstrated to grow at such a high temperature. Although growth at temperatures as high as 55°C is not uncommon among isolates of *Exiguobacterium* spp., strains that grow at 50°C or above have previously only been isolated from hot springs and deep-sea hydrothermal vents (Crapart et al., 2007; Vishnivetskaya et al., 2009). Growth at such high temperatures is not a requirement for life in Pavilion Lake; however, *Exiguobacterium* sp. RW2 is adapted for higher temperatures, suggesting that its distribution is not restricted to cold oligotrophic environments.
Exiguobacterium sp. strain RW2 grows in acidic (pH 5) and alkaline (pH 11) media; whereas, other isolates of Exiguobacterium spp. only grow between pH 6 and 10 (Vishnivetskaya et al., 2009, Table 5.2). Growth at a pH range of 5 to 11 has not been reported for other typed strains of Exiguobacterium spp. (Table 5.2); however the marine isolates E. profundum (Crapart et al., 2007), E. aestuarii and E. marinum (Kim et al., 2005) grow at pH 5.5 and are unable to grow at pH 10.

The growth of Exiguobacterium sp. strain RW2 at NaCl concentrations between 0 and 7% is different from most strains of Exiguobacterium spp, although some marine strains can grow in the presence of 19% NaCl (Kim et al., 2005; Lopez-Cortes et al., 2006; Table 5.2). Strain RW2 is the first non-marine strain able to grow in the presence of 7% NaCl (Table 5.2). Growth over such wide ranges of temperature, pH, and salinity makes Exiguobacterium sp. strain RW2 unique among isolates of the genus.

5.3.2 Biochemical properties and antibiotic susceptibility

Exiguobacterium sp. strain RW2 differed from other typed strains as it has urease function, no oxidase activity, and cannot produce acetoin (Table 5.3). Even though the genome of Exiguobacterium sp. strain RW2 encodes putative cytochrome oxidases (including types B and C), out of fifteen strains of Exiguobacterium spp., only strain RW2, E. aestuarii, E. marinum and E. aurantiacum lack demonstrated oxidase function (Fruhling et al., 2002). Additionally, within the genus, urease activity has only been found in Exiguobacterium sp. strain RW2 (Table 5.3), although a gene encoding urease was not found in the genome, suggesting that it may be plasmid encoded. Urease can induce carbonate precipitation and is found in carbonate
precipitating members of \textit{Bacillus} spp. (Hammes \textit{et al.}, 2003), which is a distant relative of strain RW2. Finding an active urease in strain RW2 would set the stage for further experimentation relating to urease mediated carbonate precipitation.

Little is known about antibiotic sensitivity in \textit{Exiguobacterium} spp. With the exception of \textit{E. indicum} and \textit{E. acetylicum}, isolates of \textit{Exiguobacterium} spp. are sensitive to clindamycin (Chaturvedi \textit{et al.}, 2008), although not all isolates have been tested and the concentrations have not been uniform (Table 5.4). Homologs of genes such as \textit{ermB}, \textit{ermTR}, and \textit{mefA/E} that are typically associated with clindamycin resistance were not found in genomic data for \textit{Exiguobacterium} spp. (Gygax \textit{et al.}, 2006). Additionally, \textit{Exiguobacterium} sp. strain RW2 was tested against eleven other antibiotics (Table 5.4); resistance was only found for sulfisoxazole (300 μg). Sensitivity was shown for sulfamethoxazole and trimethoprim in combination, consistent with observations that some strains of \textit{Exiguobacterium} spp. are sensitive to sulfamethoxazole (Chaturvedi and Shivaji, 2006). Sulfamethoxazole and sulfisoxazole resistance genes such as \textit{sul1}, \textit{sul2} and \textit{dhfr} are carried on plasmids (Barman \textit{et al.}, 2010; Toleman \textit{et al.}, 2007; Sherley \textit{et al.}, 2004), as are resistance genes for trimethoprim. Plasmid pMC1 in \textit{Exiguobacterium} sp. strain S3-2 carries resistance genes for tetracycline, trimethoprim, florfenicol and erythromycin (Yang \textit{et al.}, 2014). There was no evidence for the presence of \textit{sul1}, \textit{sul2} or \textit{dhfr} genes in strain RW2; hence the sulfisoxazole resistance mechanism is unknown. A vancomycin resistance type B protein was predicted by the strain RW2 genome, although strain RW2 is not resistant to vancomycin (Table 5.4) suggesting that the pathway is incomplete or that the protein serves another function.
5.3.3 PLFA characterization and comparative analysis

The PLFA profiles of strain RW2 were dominated by branched fatty acids, with minor contributions from saturated and monounsaturated PLFA (Table 5.5). The major phospholipids in strain RW2 are Iso-C13:0 (15.7%), anteiso-C13:0 (15.4%), and iso-C17:0 (25.6%), which is similar to the profiles from *E. himgiriensis* K22-26^T^ and *E. marinum* MTCC 7751^T^ (Singh *et al.*, 2012). The percentages of iso-C17:0 and C17:0, at 25.6% and 1.3%, respectively, are the highest reported percentages of these PLFAs for any characterized strain of *Exiguobacterium* spp. (Table 5.5). In strain RW2, the fourth most abundant phospholipid at 7.6% is Iso-C17:1Δ^5^, which is not found in any other *Exiguobacterium* isolate (Table 5.5). Iso-17:1Δ^5^ occurs in *Bacillus alvei*, a distant relative of strain RW2 (Moss and Daneshvar, 1992). The high proportion of PLFA Iso-17:1Δ^5^ is consistent with strain RW2 being assigned as a new species within the genus *Exiguobacterium*.

5.3.4 Evolutionary placement of *Exiguobacterium* sp. RW2 within the genus

Phylogenetic analysis of 16S rDNA places *Exiguobacterium* sp. strain RW2 in a strongly supported clade with other isolates of *Exiguobacterium* spp. (Figure 5.2), with the closest relative being *Exiguobacterium* sp. strain GIC31, an isolate from glacial ice in Greenland (Vishnivetskaya *et al.*, 2009). However, the 16S rDNA sequence similarity between strains RW2 and GIC31 is <97%, suggesting that they are separate species.

Analysis of the combined ribosomal 5S-16S-23S operon also indicates that *Exiguobacterium* sp. strain RW2 is phylogenetically distinct but in a strongly supported clade related to *Exiguobacterium* sp. strain AT1b (Figure 5.3). The combined ribosomal operon of
Exiguobacterium sp. strain RW2 shares the highest sequence similarity to Exiguobacterium sp. strain S17 and Exiguobacterium sp. strain 8-11-1 (Figure 5.3). No strains related to Exiguobacterium sp. strain RW2 based on combined ribosomal operon have been officially named species (Figure 5.3). We suggest that Exiguobacterium sp. strain RW2 is a distinct species based on sequences for the 16S rDNA and combined ribosomal operon. Both strain RW2 and strain S17 were isolated from modern microbialites, and are in a clade with Exiguobacterium sp. strain AT1b (based on ribosomal 5S-16S-23S operon) suggesting that they could share a distant ancestor from a high temperature environment (Vishnivetskaya et al., 2011; Belfiore et al., 2013; Ordoñez et al., 2013; White 3rd et al., 2013b).

MLST analysis is consistent with combined ribosomal operon and 16S rDNA gene analysis that Exiguobacterium sp. strain RW2 is phylogenetically distinct but in a strongly supported clade related to Exiguobacterium sp. strain AT1b (Figure 5.4). MLST analysis needs a minimum of seven loci with regions containing conservation across the genus, but also containing variable regions (Rebollar et al., 2012; Maiden et al., 2013). RecA, RpoB, GyrB, DnaK, IDH, and IleS amino-acid sequences were combined for MLST marker sequence analysis because they are conserved across Exiguobacterium spp., as well as in the order Firmicutes (Rebollar et al., 2012; Maiden et al., 2013). MLST analysis suggests that Exiguobacterium sp. strain RW2 is sister species to E. aurantiacum strain DSM14480 (Figure 5.4). This suggests that Exiguobacterium sp. strain RW2 is a new species based on MLST analysis. Isolates of Exiguobacterium spp. from Cuatro Cienegás form distinct genetic clusters based on whether they were isolated from water or sediment (Rebollar et al., 2012). Our MLST analysis support combined ribosomal operon
analysis that microbialite-isolated strains RW2 and S17 are genetically similar and may share a distant ancestor from a high temperature environment (Vishnivetskaya et al., 2011; Belfiore et al., 2013; Ordoñez et al., 2013; White 3rd et al., 2013b).

5.3.5 Comparative genome analysis

The draft and completed genomes of *Exiguobacterium* spp. were similar in terms of genome size, GC content, and number of protein-coding genes. As a genus, the sizes of sequenced genomes ranges from 2.82 to 3.16 Mb, with 47.5% to 53% GC content, and 2,941 to 3,323 genes that encode a potential of 2,772 to 3,203 predicted proteins (Table 5.6). Some genomes were larger than the initial estimates from pulse-field gel electrophoresis (PFGE), which ranged from 2.4 to 2.7 Mb (Vishnivetskaya et al., 2009, Table 5.6). All completed genomes have a GC skew; whereby, about half of the chromosomes have a positive skew, while the other half have a negative skew (Figure 5.5, Vishnivetskaya et al., 2011). Strain RW2 had a lower number of rRNA gene operons compared to other members of the genus based on the RAST annotation (Table 5.6).

Whole genome analysis supports the phylogenetic analysis and is consistent with strain RW2 being assigned to a new species. tBLASTX analysis of *E. sibiricum* strain 255-15, *Exiguobacterium* sp. strain AT1b, and *Exiguobacterium* sp strain MH3 against the genome of RW2 showed many gaps, indicating low similarity (Figure 5.5). Whole genome based average amino acid analysis is a robust measure for bacterial species classification that is comparable to DNA-DNA hybridization; with this analysis, <95% average amino acid identity cut-off allows for assignment of a new species (Konstantinidis and Tiedje, 2005; Richter et al., 2009). Among
the completed genomes, strain RW2 had the highest genome synten to strain AT1b (Figure 5.6). The high genome synten between strain RW2 and AT1b were consistent with their close phylogenetic placement (section 5.3.4). Strain RW2 has the highest overall amino-acid identity to *Exiguobacterium* sp. strain S17 (87.76%, Figure 5.6), but low gene synten due to the high number of contigs in the assembly of strain S17 (Figure 5.6 and Table 5.6). Since <95% of its genome is shared with other sequenced representatives of the genus and genomic synten is general low (except with its near relative, strain AT1b), strain RW2 can be considered a new species within the genus.

*Exiguobacterium* sp. strain RW2 shared >50% of its protein coding genes with other members of the genus and more distant relatives in the phylum *Firmicutes* (e.g *Planococcus halocryophilus* and *Bacillus halodurans*) based on SEED annotations (Figure 5.6 to 5.8). Among isolates with >80% amino-acid identity scores, strain RW2 shares 1659 to 1743 protein-coding genes (average, 1705 of 3092), while 106 to 135 (average, 115) are not shared with *Exiguobacterium* strains S17, AT1b, and 8-11-1) (Figure 5.6 and Table 5.6). Strain RW2 shares 1455 to 1775 of its protein coding genes (average, 1641 out of 3092) with isolates that have <65% amino acid identity, with 86 to 254 (average, 170) not found in *Exiguobacterium* spp. strains B7, OS-77, MH3, 255-15, *Planococcus halocryophilus* and *Bacillus halodurans* (Figure 5.7 to 5.8 and Table 5.6). *Exiguobacterium* sp. strain RW2 shared 1455 out of 3092, of its protein-coding genes with *Planococcus halocryophilus*; however, 254 protein-coding genes are unique to *Exiguobacterium* sp. strain RW2 (Figure 5.8 and Table 5.6). *Planococcus halocryophilus* is a distant relative of stain RW2 that is adapted for drastic shifts in temperature and salinity.
(Mykytczuk et al., 2013). The high percentage of shared protein-coding genes between *Exiguobacterium* sp. strain RW2 and other members of the genus suggests a large core genome with low genomic plasticity.

Comparisons of MetaCyc pathway annotations among strain RW2 and representative genomes of *Exiguobacterium* spp. with high (>80%) and low (<65%) amino-acid identity revealed many pathways that were shared within the genus and few that were specific to strain RW2. When compared to genomes with high amino-acid identity, there were only six pathways that were specific to strain RW2 (Figure 5.9A), and only eighteen were specific when compared to genomes with low amino-acid identity (Figure 5.9B). Strain RW2 shared 240 MetaCyc pathways with strains that had high amino-acid identity (Figure 5.9A), and 222 with strains that had low amino-acid identity (Figure 5.9B). Even genomes with <65% amino-acid identity shared >200 pathways with strain RW2, indicating that there are many shared metabolic pathways within the genus (Figure 5.9B). The MetaCyc pathway annotations are consistent with the SEED annotations, and suggest low genomic plasticity within the genus.

5.3.6 **Carbohydrate metabolism**

The biochemical tests indicated that strain RW2 uses monosaccharides, polysaccharides, and sugars as carbohydrate sources, unlike other members of the genus (Table 5.3). Strain RW2 has the metabolic potential to use a wide variety of monosaccharides, polysaccharides, sugar alcohols, and amino sugars, but cannot use D-glucose or D-sucrose, unlike other members of the genus (Table 5.3). Although sucrose use was not observed, the genome of strain RW2 predicts a complete sucrose utilization pathway, suggesting that it was not functional or that only
fermentation occurs (Table 5.3). Pathways for lactose uptake and use were present, but only a LacZ (β-galactosidase) for lactose utilization appeared to be present and functional (Table 5.3). Strain RW2 could not use D-mannitol, D-inositol, D-sorbitol, L-rhamnose, D-melibiose, and L-arabinose as sole carbon sources (Table 5.3). D-mannitol utilization was predicted by the genome, but was not used, suggesting that only fermentation occurs (Table 5.3). Amygdalin was used, as is the case for several members of the genus, although is not predicted by the genome (Table 5.3).

The genome of strain RW2 suggests the metabolic potential to perform various forms of fermentation including mixed acid, lactate, and acetyl-CoA fermentation to butyrate. The genome of strain RW2 also suggests that chitin, N-acetylglucosamine, maltose, maltodextrin, trehalose, glycerol, glycerol-3-phosphate, glycogen, fructose, gluconate, ribose, fructooligosaccharides (FOS), raffinose and deoxyribose can be used, although none were tested as growth substrates here. Further growth substrates should be tested to better understand carbohydrate utilization in strain RW2.

5.3.7 Amino acid biosynthesis and catabolism

Strain RW2, like other members of the genus, appears unable to make branched-chain amino acids, but has the metabolic potential for the ammonification of amino acids, which is linked to carbonate precipitation (Castanier et al., 1999; Rodrigues et al., 2008; Vishnivetskaya et al., 2011). Whether Exiguobacterium spp. are auxotrophic for branched-chain amino acids is unknown, as there is no experimental evidence, and >30% of the ORFs are for hypothetical proteins. Pathways for the complete degradation of branched-chain amino acids were predicted by the genome of strain RW2, including full isoleucine, leucine and valine degradation pathways.
Moreover, arginine dihydrolase was functional in strain RW2; whereas, tryptophan deaminase, lysine and ornithine decarboxylases were not (Table 5.3). L-asparagine was not an essential amino acid for strain RW2, but the genome encodes aspartate aminotransferase, aspartate-ammonia ligase and L-asparaginase in an L-asparagine biosynthetic pathway. If L-asparaginase is active in strain RW2, it could generate alkalinity through the conversion of L-asparagine to ammonia by the ammonification of amino acids (Castanier et al., 1999). Further experimentation is needed to confirm the metabolic potential of ammonification of amino acids in strain RW2 and whether it is functionally involved in carbonate precipitation.

5.3.8 Motility and flagellum biosynthesis

While flagella were not observed by SEM or light microscopy (Figure 5.1), strain RW2 has complete flagellum biosynthesis pathways (including the fih, fli, fig operons), as does strain AT1b (Vishnivetskaya et al., 2011) and E. sibiricum strain 255-15 (Rodrigues et al., 2008). Under low temperature, E. sibiricum strain 255-15 loses its flagellum (-2.5 °C), and expression of flagellum synthesis operons are downregulated (Rodrigues et al., 2008). Rodrigues et al. (2008) suggested that at lower temperatures, aqueous media becomes more viscous, rendering flagella impractical. Under these prohibitively cold conditions, considerable energy could be saved by repressing flagellar assembly. While RW2 did not exhibit visible flagella, our proteomic survey data suggests that flagellar proteins are expressed at 10 °C but not at 4 °C (data not shown). Hence, strain RW2 appears to downregulate flagellar pathways at a much warmer temperature than the permafrost-dwelling E. sibiricum strain 255-15 (4 °C vs. -2.5 °C).
5.3.9 Stress response temperature, heavy-metals and salinity

*Exiguobacterium* sp. strain RW2 is unlikely grow at subzero temperatures due to the lack of the *cspA* gene (cold-shock protein A). Cold-shock protein A is induced at temperatures <11 °C, where it functions as a molecular chaperone, binding mRNA to prevent secondary structure formation and ensure efficient translation at low temperatures (Yamanka and Inouye, 1997). *Exiguobacterium sibiricum* strain 255-15 grows at -2.5 °C and has four copies of *cspA* (Rodrigues *et al.*, 2008). Strain AT1b cannot grow below 10 °C; whereas, strain RW2 grows at 4 °C (White 3rd *et al.*, 2013b; Table 5.2). Strain AT1b, RW2 and S17 do not encode *cspA*, and hence would not be expected to grow at subzero temperatures.

Strain AT1b and RW2 were able to grow at temperatures as high as 50 °C (Vishnivetskaya *et al.*, 2011: Table 5.2). Both strains contain copies of genes (*dnaJ, dnaK* and *grpE*) encoding heat-shock related proteins, and a repressor of the heat-shock cascade, *hrcA*. However, unlike strain AT1b, strain RW2 does not encode *hspR*, the transcriptional repressor of the heat-shock protein operon (Schmid *et al.*, 2005). In *Deinococcus radiodurans*, the Δ*hspR* mutant has a lower survival rate at >48 °C; however, this didn’t occur in strain RW2, which survived at over 48 °C (White 3rd *et al.*, 2013b; Table 5.2). Further experimentation is needed to confirm the essential genes required by strain RW2 to grow at 50 °C.

Cold-stress responses may confer protection to high salinity in a cross-protective manner (Tsuzuki *et al* 2011, Mykyczuk *et al.*, 2013). Strain RW2, grows at 0 to 7% NaCl and can grow at temperatures <11 °C (White 3rd *et al.*, 2013b; Table 5.2). Transcriptome analysis of *Rhodobacter sphaeroides* under salt stress showed significant upregulation of genes encoding
cold-shock DNA-binding proteins (Tsuzuki et al., 2011). Genome and transcriptome analysis of the hypercryophile Planococcus halocryophilus strain Or1 suggests that adaptation to osmotic stress (mainly NaCl stress) may allow for synergistic adaptation to cold stress, as strain Or1 is able to grow in both high salt (~18%) and subzero temperatures (-15°C) (Mykyczuk et al., 2013). OpuA and opuD are essential genes found in P. halocryophilus (Mykyczuk et al., 2013) that provide osmoprotection using glycine-betaine uptake systems, and are upregulated under NaCl stress (Hahne et al., 2010). Unlike P. halocryophilus, strain RW2 does not encode opuA or opuD. The lack of the Opu proteins may be why strain RW2 did not grow in media with >7% NaCl. Strain RW2 was able to grow in 0% added NaCl, but little is known about cellular functions contributing to low salt survival.

The occurrence of heavy-metal resistance genes in strain RW2 was unexpected, given the low levels of zinc (0.01 to 0.03 mg L⁻¹) and undetectable levels of cobalt, copper, chromium, arsenic and cadmium in Pavilion Lake (Lim et al., 2009). Agrococcus sp. strain RW1, which was isolated from the same microbialite as strain RW2, also carried genes for heavy-metal resistance (Thesis Chapter 3; White 3rd et al., 2013b). Other members of the genus Exiguobacterium, such as the closely-related strain S17, have similar heavy-metal resistance gene clusters, and display high resistance to arsenic (Belfiore et al., 2013). However, the gene encoding the ACR3 arsenite efflux pump is only found in strain S17 (Ordoñez et al., 2013). Lake Socompa, where strain S17 was isolated, differs from Pavilion Lake in having high levels of arsenic (18.5 mg L⁻¹) and iron (1 mg L⁻¹) (Farias et al., 2013). Historical mining in the Pavilion Lake region (Stevenson, 1940) suggests that higher metal content occurred in the past. Alternatively, heavy metal genes may
have a secondary role, as in *Rhodobacter sphaeroides* in which it upregulates arsenic resistance genes under high salt stress (Tsuzuki *et al.*, 2011).

The Pavilion Lake microbialite metagenomes had a number of heavy metal resistance genes which could be contributed by strain RW2 (unpublished data, Thesis Chapter 2). The abundance of strain RW2 in Pavilion lake microbialites was <0.5% by metagenomic recruitment analysis (unpublished data, Thesis Chapter 2). We suggest that strain RW2, like strain RW1, contributes heavy-metal resistance genes to the Pavilion Lake microbialite metabolic potential (confirmed by tBLASTx 1e⁻³); however, further experimentation is needed to determine whether these genes function under heavy-metal stress (unpublished data, Thesis Chapter 2).

5.3.10 Carotenoid biosynthesis

The characteristic orange pigmentation of colonies in *Exiguobacterium* spp. is distinct from other members of the phylum *Firmicutes*. However, the phylum has carotenoid biosynthesis, including the C₃₀ pathway that is responsible for colony pigmentation (Pelz *et al.*, 2005; Köcher *et al.*, 2009; Klassen, 2010). Genes from the 4-4’ diapocarotenoids C₃₀ carotenoid pathway were present in all the published *Exiguobacterium* spp. genomes, including diapophytoene synthase (*crtM*, annotated as phytoene synthase), diapophytoene desaturase (*crtN*, annotated as phytoene desaturase), and diapophytoene desaturase (*crtNb*, annotated as second copy of phytoene desaturase) (Pelz *et al.*, 2005; Köcher *et al.*, 2009; Klassen, 2010; Figure 5.10; Table 5.7). In *Exiguobacterium* sp. strain RW2 the phytoene-based C₃₀ carotenoid pathway is located on the same genetic locus between 1521530 to 1527203 on the genome. These genomic predictions suggest that diaponeurosporene-4-oic acid through the C₃₀ carotenoid pathway could
be responsible for the orange colony pigmentation in *Exiguobacterium* spp. (Köcher et al., 2009; Klassen, 2010). *Halobacillus halophilus* DSM2266 loss-of-function mutants for diapophytoene desaturase (*ΔcrtNb/Nc*) accumulate diaponeurosporene-4-oic acid which causes orange colony pigmentation (Köcher et al., 2009).

To our knowledge, a hypothesis has not been advanced to explain the function of pigments in *Exiguobacterium* spp. Low concentration of dissolved organic carbon (DOC) in Pavilion Lake allows for high penetration of solar UV radiation, which benthic microbial mats may partially protect against using carotenoids (Lim et al., 2009; Lionard et al., 2012). *Agrococcus* sp. strain RW1, was also isolated from the same Pavilion Lake microbialite, and is also bright yellow (unpublished data, Thesis Chapter 4), suggesting that the pigments may serve in photo-protection.

5.4 Conclusions

Polyphasic analysis, using chemotaxonomic markers, and sequences for 16S rDNA, the combined ribosomal operon, MLST analysis and whole genome amino-acid identity is consistent with *Exiguobacterium* sp. strain RW2 being assigned as a new member of the genus. As well, novel chemotaxonomic markers including urease function, amygdalin utilization, and the presence of 7.6% iC17:1Δ5 (Table 5.3 and Table 5.5) are consistent the designation of a new species, and we suggest the name *Exiguobacterium pavilionensis* strain RW2.

*Exiguobacterium* sp. RW2 is able to grow over wide ranges of temperature, salinity and pH. Phylogenetic trees constructed from MLST, 16S rDNA and combined ribosomal operon analyses, as well as high genome synteny and whole-genome amino-acid identity (>80%) indicate that *Exiguobacterium* sp. strains RW2 and AT1b are close relatives; yet, AT1b was
isolated from Mammoth Terrace hot spring in Yellowstone National Park (Vishnivetskaya et al., 2011). These results are consistent with the wide geographically and environmental range where *Exiguobacterium* spp. can be found.

We suggest that diaponeurosporene-4-oic acid, may be responsible for the orange pigmentation through the C$_{30}$ carotenoid biosynthesis pathway. This pathway is universally conserved in all currently available *Exiguobacterium* genomes. These carotenoids may function in *Exiguobacterium* as photo-protection, and further experimentation is needed to validate the pathway and to demonstrate this functional relationship.

*Exiguobacterium* spp. are highly adaptable and thrive under a variety of conditions, including wide ranges of pH, temperature, salinity and heavy metals. Consequently, *Exiguobacterium* spp. colonize a wide range of systems and are useful models for understanding adaptation to a wide variety of environmental stresses.
### 5.5 Tables and figures

Table 5.1: Names and isolation location of *Exiguobacterium* species

<table>
<thead>
<tr>
<th>Species Name</th>
<th>Environment of isolation</th>
<th>Reference</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>E. acetylicum</em></td>
<td>Creamery waste</td>
<td>Farrow <em>et al.</em>, 1994</td>
<td>Levine and Soppeland, 1926 classified &quot;<em>Flavobacterium acetylicum</em>&quot; renamed &quot;<em>Brevibacterium acetylicum</em>&quot; then reclassified by Farrow <em>et al.</em>, 1994</td>
</tr>
<tr>
<td><em>E. antarcticum</em></td>
<td>Microbial mat/biofilm</td>
<td>Fruhling <em>et al.</em>, 2002</td>
<td>Lake Fryxell mat in Antarctica</td>
</tr>
<tr>
<td><em>E. aestuarii</em></td>
<td>Marine tidal flat</td>
<td>Kim <em>et al.</em>, 2005</td>
<td>Daepo beach in yellow sea near Mokpo City, Korea,</td>
</tr>
<tr>
<td><em>E. alkalophilum</em></td>
<td>Wastewater drained sludge</td>
<td>Kulshreshtha <em>et al.</em>, 2013</td>
<td>Beverage industry facility New Delhi, India</td>
</tr>
<tr>
<td><em>E. aquaticum</em></td>
<td>Freshwater lake</td>
<td>Raichand <em>et al.</em>, 2012</td>
<td>Tikkar Tal Lake, Haryana, India</td>
</tr>
<tr>
<td><em>E. aurantiacum</em></td>
<td>Potato processing plant</td>
<td>Collins <em>et al.</em>, 1983</td>
<td>Type species for the genus</td>
</tr>
<tr>
<td><em>E. himgiriensis</em></td>
<td>Soil</td>
<td>Singh <em>et al.</em>, 2013</td>
<td>Lahaul-Spiti valley, Indian Himalayas</td>
</tr>
<tr>
<td><em>E. indicum</em></td>
<td>Glacial melt water</td>
<td>Chaturvedi and Shivaji 2006</td>
<td>Hamta glacier (4270 m above sea level ), Himalayans</td>
</tr>
<tr>
<td><em>E. martemiae</em></td>
<td>Brine shrimp cysts</td>
<td>Lopez-Cortes <em>et al.</em>, 2006</td>
<td><em>Artemia franciscana</em> cysts</td>
</tr>
<tr>
<td><em>E. mexicanum</em></td>
<td>Brine shrimp cysts</td>
<td>Lopez-Cortes <em>et al.</em>, 2006</td>
<td><em>Artemia franciscana</em> cysts</td>
</tr>
<tr>
<td><em>E. oxidotolerans</em></td>
<td>Fish processing plant</td>
<td>Yumoto <em>et al.</em>, 2004</td>
<td>Hokkaido, Japan</td>
</tr>
<tr>
<td><strong>Species Name</strong></td>
<td><strong>Environment of isolation</strong></td>
<td><strong>Reference</strong></td>
<td><strong>Isolation Notes</strong></td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------</td>
<td>---------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>E. profundum</td>
<td>Hydrothermal Vent</td>
<td>Crapart <em>et al.</em>, 2007</td>
<td>Northeast Pacific rise at ~ 2600m</td>
</tr>
<tr>
<td>E. sibiricum</td>
<td>Permafrost</td>
<td>Rodrigues <em>et al.</em>, 2006</td>
<td>Siberian permafrost (3 Mya old)</td>
</tr>
<tr>
<td>E. soli</td>
<td>Glacial moraine</td>
<td>Chaturvedi <em>et al.</em>, 2008</td>
<td>McMurdo dry valleys, Antarctica</td>
</tr>
<tr>
<td>E. undae</td>
<td>Garden pond water</td>
<td>Fruhling <em>et al.</em>, 2002</td>
<td>Wolfenbüttel, Lower Saxon Germany</td>
</tr>
</tbody>
</table>

**Published strains in NCBI**

<table>
<thead>
<tr>
<th><strong>Species Name</strong></th>
<th><strong>Environment of isolation</strong></th>
<th><strong>Reference</strong></th>
<th><strong>Isolation Notes</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain RW2</td>
<td>Thromatolithic microbialite</td>
<td>White 3rd <em>et al.</em>, 2013b</td>
<td>Pavilion lake 20 m microbialites</td>
</tr>
<tr>
<td>Strain AT1B</td>
<td>Alkaline hot spring</td>
<td>Vishnivetskaya <em>et al.</em>, 2011</td>
<td>Mammoth Terrace hot spring (CaCO₃ rich), Yellowstone National Park Wyoming, USA</td>
</tr>
<tr>
<td>Strain MH3</td>
<td>Rhizosphere of duckweeds</td>
<td>Tang <em>et al.</em>, 2013</td>
<td>Lemna minor duckweed from Shenzhen, China</td>
</tr>
<tr>
<td>Strain OS77</td>
<td>Alkaline hot spring</td>
<td>Nonaka <em>et al.</em>, 2014</td>
<td>Tepid spring (41 °C, pH 7.9) Oguni-tyo, Kumamoto, Japan.</td>
</tr>
<tr>
<td>Strain S17</td>
<td>Stromatolithic microbialite</td>
<td>Ordoñez <em>et al.</em>, 2013</td>
<td>Polyextremophile strain tolerate heavy metals (As)</td>
</tr>
</tbody>
</table>

Officially named species are published species: these names are recognized by IJSEM (international journal of systematic and evolutionary microbiology) or other journals.

Isolates published in NCBI as genome announcements that are unnamed, not classically described, or not characterized by IJSEM or other journal, are listed as an *Exiguobacterium* sp.
Table 5.2: Properties of *Exiguobacterium* selected strains

<table>
<thead>
<tr>
<th>Habitat of isolation</th>
<th>Microbialite</th>
<th>Morine</th>
<th>Microbial Mat</th>
<th>Permafrost</th>
<th>Glacial Water</th>
<th>Garden Pond</th>
<th>Fish processing Plant</th>
<th>Creamery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strain RW2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>E. soli</em>&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2 – 4</td>
<td>2 – 3</td>
<td>2 – 3</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 5</td>
<td>2 – 5</td>
</tr>
<tr>
<td><em>E. antarcticum</em>&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2 – 4</td>
<td>2 – 3</td>
<td>2 – 3</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 5</td>
<td>2 – 5</td>
</tr>
<tr>
<td><em>E. sibiricum</em>&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2 – 4</td>
<td>2 – 3</td>
<td>2 – 3</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 5</td>
<td>2 – 5</td>
</tr>
<tr>
<td><em>E. undae</em>&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2 – 4</td>
<td>2 – 3</td>
<td>2 – 3</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 5</td>
<td>2 – 5</td>
</tr>
<tr>
<td><em>E. indicum</em>&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2 – 4</td>
<td>2 – 3</td>
<td>2 – 3</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 5</td>
<td>2 – 5</td>
</tr>
<tr>
<td><em>E. oxidotolerans</em>&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2 – 4</td>
<td>2 – 3</td>
<td>2 – 3</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 5</td>
<td>2 – 5</td>
</tr>
<tr>
<td><em>E. acetylicum</em>&lt;sup&gt;1&lt;/sup&gt;</td>
<td>2 – 4</td>
<td>2 – 3</td>
<td>2 – 3</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 5</td>
<td>2 – 5</td>
</tr>
<tr>
<td>Colony size (mm)</td>
<td>2 – 4</td>
<td>2 – 3</td>
<td>2 – 3</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 4</td>
<td>2 – 5</td>
<td>2 – 5</td>
</tr>
<tr>
<td>Colony shape</td>
<td>Round</td>
<td>Round</td>
<td>Round</td>
<td>Round</td>
<td>Round</td>
<td>Round</td>
<td>Round</td>
<td>Irregular</td>
</tr>
<tr>
<td>Colony color</td>
<td>Orange</td>
<td>Orange</td>
<td>Orange</td>
<td>Orange</td>
<td>Orange</td>
<td>Orange</td>
<td>Yellow-orange</td>
<td>Yellow-orange</td>
</tr>
<tr>
<td>Growth temp range (ºC)</td>
<td>-2.5</td>
<td>NA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>w</td>
<td>w</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>37</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>40</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>45</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Max growth temp (ºC)</td>
<td>50</td>
<td>30</td>
<td>41</td>
<td>40</td>
<td>41</td>
<td>30</td>
<td>40</td>
<td>37</td>
</tr>
<tr>
<td>pH growth range</td>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NaCl growth range</td>
<td>5.8%</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td></td>
<td>7.0%</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

- Negative, + Positive, W: Weak growth, NA: Not available, Temp: temperature. All growth measurements for strain RW2 were taken after 2 d, except 5 ºC which took 10 d to grow.

<sup>1</sup>Chaturvedi et al., 2008
Table 5.3: Biochemical tests for selected strains of *Exiguobacterium* spp.

<table>
<thead>
<tr>
<th>Activity for:</th>
<th>Strain RW2</th>
<th>E. solit DVS 3Y</th>
<th>E. antarcticum DSM14480</th>
<th>E. sibiricum DSM17290</th>
<th>E. undae DSM14481</th>
<th>E. indicum DSM15368</th>
<th>E. oxidotolerans JCM12280</th>
<th>E. acetylicum DSM20416</th>
</tr>
</thead>
<tbody>
<tr>
<td>Catalase</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Oxidase</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>β-galactosidase</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Arginine dihydrolase</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Lysine decarboxylase</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ornithine decarboxylase</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Urease</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tryptophan deaminase</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Citrate utilization</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Gelatinase</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Production of:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>H₂S</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Indole</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Acetoin</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Assimilation of:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D-mannitol</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Inositol</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>D-sorbitol</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>L-Rhamnose</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-Melibiose</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Amygdalin</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>L-arabinose</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>D-glucose</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>D-sucrose</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

All growth measurements for strain RW2 were taken after 24 h.

¹Chaturvedi et al., 2008.
Table 5.4: Antibiotic susceptibility for selected strains *Exiguobacterium* spp.

<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Disc Content</th>
<th>Strain RW2</th>
<th><em>E. antarcticum</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th><em>E. sibiricum</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th><em>E. undae</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th><em>E. indicum</em>&lt;sup&gt;1&lt;/sup&gt;</th>
<th><em>E. acetylicum</em>&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sulfamethoxazole/Trimethoprim</td>
<td>23.75 μg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Sulfamethoxazole</td>
<td>1/1.25 μg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Clindamycin*</td>
<td>50 μg</td>
<td>ND</td>
<td>-</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Penicillin</td>
<td>10 IU</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Rifampin</td>
<td>5 μg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Polymyxin</td>
<td>300 IU</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Cefixime</td>
<td>5 μg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Sulfisoxazole</td>
<td>300 μg</td>
<td>+</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Oxacillin</td>
<td>1 μg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Tetracycline</td>
<td>30 μg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Trimethoprim</td>
<td>5 μg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Tobramycin</td>
<td>10 μg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Vancomycin</td>
<td>30 μg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>Streptomycin</td>
<td>10 μg</td>
<td>-</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
</tbody>
</table>

+ Resistant, - Sensitive, ND: No data
All growth measurements for strain RW2 were taken after 24 h.
<sup>1</sup>Chaturvedi et al., 2008.
Table 5.5: Selected phospholipid fatty acids from *Exiguobacterium* spp. strains (%)

<table>
<thead>
<tr>
<th></th>
<th>Strain RW2</th>
<th><em>E. antarcticum</em> DSM14481&lt;sup&gt;a&lt;/sup&gt;</th>
<th><em>E. aurantiacum</em> DSM6208&lt;sup&gt;a&lt;/sup&gt;</th>
<th><em>E. sibiricum</em> 7-3&lt;sup&gt;a&lt;/sup&gt;</th>
<th><em>E. himgiriensis</em> K22-24&lt;sup&gt;b&lt;/sup&gt;</th>
<th><em>E. marium</em> MTCC7751&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>iC&lt;sub&gt;13:0&lt;/sub&gt;</td>
<td>15.7</td>
<td>12</td>
<td>18</td>
<td>9</td>
<td>12.2</td>
<td>1</td>
</tr>
<tr>
<td>aC&lt;sub&gt;13:0&lt;/sub&gt;</td>
<td>15.4</td>
<td>11</td>
<td>12</td>
<td>11</td>
<td>16.4</td>
<td>16.8</td>
</tr>
<tr>
<td>iC&lt;sub&gt;15:0&lt;/sub&gt;</td>
<td>5.6</td>
<td>11</td>
<td>4</td>
<td>13</td>
<td>14.2</td>
<td>14.1</td>
</tr>
<tr>
<td>aC&lt;sub&gt;15:0&lt;/sub&gt;</td>
<td>2.1</td>
<td>2</td>
<td>ND</td>
<td>3</td>
<td>5.8</td>
<td>2.2</td>
</tr>
<tr>
<td>C&lt;sub&gt;16:1Δ5&lt;/sub&gt;</td>
<td>2.7</td>
<td>18</td>
<td>10</td>
<td>8</td>
<td>2.4</td>
<td>3.6</td>
</tr>
<tr>
<td>iC&lt;sub&gt;16:0&lt;/sub&gt;</td>
<td>1.8</td>
<td>ND</td>
<td>ND</td>
<td>2</td>
<td>3.8</td>
<td>6.9</td>
</tr>
<tr>
<td>brC&lt;sub&gt;16:1&lt;/sub&gt;</td>
<td>1.2</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>C&lt;sub&gt;16:0&lt;/sub&gt;</td>
<td>3.9</td>
<td>13</td>
<td>27</td>
<td>17</td>
<td>2.8</td>
<td>4.7</td>
</tr>
<tr>
<td>iC&lt;sub&gt;17:1Δ5&lt;/sub&gt;</td>
<td>7.6</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>iC&lt;sub&gt;17:0&lt;/sub&gt;</td>
<td>25.6</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>16.1</td>
<td>16.8</td>
</tr>
<tr>
<td>C&lt;sub&gt;17:0&lt;/sub&gt;</td>
<td>1.3</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>tr</td>
<td>tr</td>
</tr>
<tr>
<td>aC&lt;sub&gt;17:0&lt;/sub&gt;</td>
<td>5.7</td>
<td>ND</td>
<td>ND</td>
<td>3</td>
<td>4.8</td>
<td>3.6</td>
</tr>
<tr>
<td>C&lt;sub&gt;18:1Δ5&lt;/sub&gt;</td>
<td>1.9</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>C&lt;sub&gt;18:0&lt;/sub&gt;</td>
<td>2.3</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>tr</td>
<td>tr</td>
</tr>
<tr>
<td>iC&lt;sub&gt;18:0&lt;/sub&gt;</td>
<td>2.3</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>1.6</td>
<td>tr</td>
</tr>
</tbody>
</table>

ND: Not detected, tr: traces (<1%)

<sup>a</sup>Rodrigues et al., 2006, <sup>b</sup>Singh et al., 2012.
| Table 5.6: *Exiguobacterium* spp. assembly and annotation statistics |
|---|---|---|---|---|---|---|---|---|---|
| **Draft Genomes** | **Completed Genomes** |
| | Strain RW2 | Strain S17 | Strain OS-77 | Strain 8-11-1 | Strain RW2 | Strain AT1b | Strain MH3 | *E. antarcticum* | *E. sibiricum* |
| Sequence Method | Illumina MiSeq | 454 GS FLX | 454 GS FLX | Illumina HiSeq | Illumina MiSeq | 454 Sanger | Illumina HiSeq | SOLiD | 454 Sanger |
| Assembler | Ray | Newbler | Newbler | Velvet | Ray | Phred Phrap | SOAP denovo | Velvet Edena | Phred Phrap |
| Assembly Name | r23 | Exi | ASM | V1 | r23 | ASM | ASM | ASM | ASM |
| Genome coverage | 300x | 63x | 45x | 300x | 100x | 20x | 140x | 400x | - |
| No. Contigs | 23 | 163 | 23 | 31 | 1 | 1 | 1 | 1 | 3 |
| Scaffolds | - | - | - | 5 | - | - | - | - | - |
| Genome Size (bp) | 3,019,504 | 3,127,363 | 3,151,479 | 2,906,962 | 3,019,504 | 2,999,895 | 3,164,195 | 2,815,863 | 3,040,786 |
| N50 | 705,844 | 34,407 | 349,882 | 340,222 | - | - | - | - | - |
| Largest Contig (bp) | 947,149 | 122,784 | 929,170 | 663,994 | - | - | - | - | - |
| G+C% | 52.1 | 53.1 | 47.1 | 52.8 | 52.1 | 49 | 47.2 | 47.5 | 47.7 |
| Annotation Method | RAST | RAST | RAST | RAST | RAST | NCBI | NCBI | NCBI | NCBI |
| Gene No. | - | - | - | - | - | 3,141 | 3,332 | 2,941 | 3,155 |
| Pseudo genes | 0 | - | - | - | - | 23 | 41 | 76 | - |
| rRNAs operons | 2 | 2 | 2 | 2 | 2 | 9 | 9 | 9 | 9 |
| tRNAs | 48 | 47 | 49 | 14 | 46 | 68 | 60 | 66 | 69 |
| Plasmids | - | - | - | - | 0 | 0 | 0 | 0 | 2 |
| SEED subsystems | 370 | 370 | 388 | 373 | 395 | 372 | 398 | - | 430 |

- Not available, Draft genome contains many contigs (>20), whereas a completed genome contains a usually a single chromosome with plasmids.
Table 5.7: *Exiguobacterium* spp. carotenoid biosynthesis genes

<table>
<thead>
<tr>
<th></th>
<th>Strain RW2</th>
<th>Strain S17</th>
<th>Strain 8-11-1</th>
<th>Strain AT1b</th>
<th><em>E. antarcticum</em> B7</th>
<th>Strain MH3</th>
<th>Strain OS-77</th>
<th><em>E. sibiricum</em> 255-15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phytoene synthase</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Phytoene desaturase</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>4</td>
</tr>
</tbody>
</table>
Figure 5.1: *Exiguobacterium* sp. strain RW2 microscopy
A) SEM image of biofilm during late-stationary phase of growth (72 h). Scale bar: ~10 μm.
B) Light microscopy image during logarithmic phase of growth (24 h). Scale bar: ~100 μm.
Observed here at standard growth conditions in liquid M medium at 30°C
Figure 5.2: *Exiguobacterium* maximum-likelihood tree based on 16S rDNA gene sequences (~1353 bp) Bootstrap values greater than 40% at the given branches.
Figure 5.3: *Exiguobacterium* combined ribosomal operon maximum-likelihood tree
Bootstrap values greater than 50% at the given branches.
Figure 5.4: MLST maximum-likelihood phylogenetic tree
MLST based on concatenated rpoB, RecA, gyrB, DnaK, IDH and IARS protein sequence (~3856 amino acids). Bootstrap values greater than 50% at the given branches.
Figure 5.5: Genome plot (~3 Mb) of *Exiguobacterium* sp. strain RW2
Genome key (left corner): starts with inner most ring which is a genome ruler, followed by GC skew (purple/green) and ends with two outer rings, which contain protein-coding ORFs, tRNAs and rRNAs genes.
Figure 5.6: Genome synteny and functional Venn diagrams (SEED) with high amino acid identities (AAIr >80%) to *Exiguobacterium* sp. RW2.
Genomes are listed in megabase pairs (Mb). Red dots are positive BLAST hits based on RAST genome comparison module. Amino acid identity (AAIr) was calculated by the RAST functional module with web based tool AAIr (Krebs et al., 2013). Venn diagrams are listed by strain name based on RAST functional annotations (SEED/FigFams).
Figure 5.7: Genome synteny and functional Venn diagrams (SEED) with low amino acid identities (AAIr <65%) to *Exiguobacterium* sp. RW2
Genomes are listed in megabase pairs (Mb). Red dots are positive BLAST hits based on RAST genome comparison module. Amino acid identity (AAIr) was calculated by the RAST functional module with web based tool AAIr (Krebs et al., 2013). Venn diagrams are listed by strain name based on RAST functional annotations (SEED/FigFams).
Figure 5.8: Genome synteny and functional Venn diagrams (SEED) with low amino acid identities (AAIr <65%) to *Exiguobacterium* sp. RW2
Genomes are listed in megabase pairs (Mb). Red dots are positive BLAST hits based on RAST genome comparison module. Amino acid identity (AAIr) was calculated by the RAST functional module with web based tool AAIr (Krebs et al., 2013). Venn diagrams are listed by strain name based on RAST functional annotations (SEED/FigFams).
Figure 5.9: MetaCyc pathway annotations for Exiguobacterium genomes
A) Venn diagram comparing isolates with >80% amino acid identity (AAI) to strain RW2
B) Venn diagram comparing isolates with <62% amino acid identity (AAI) to strain RW2
Labels on Venn diagrams are listed by strain name with strains of Exiguobacterium spp. (8111, AT1b, MH3, RW2 and S17) and E. sibiricum (255-15) and E. antarcticum (B7) shown.
Figure 5.10: Proposed carotenoid biosynthesis pathway in the genus *Exiguobacterium*

*Diapophytoene synthase (crtM) annotated as phyotene synthase.

**Dipophytoene desaturase (crtN) annotated as phyotene desaturase.

***Dipophytoene desaturase (crtNb) annotated as phyotene desaturase.

All genes are found in all *Exiguobacterium* spp. genomes listed in Table 5.7.
Chapter 6: Conclusion

6.1 Summary

This thesis presents an in-depth metagenomic analysis of two freshwater microbialite-forming ecosystems (Pavilion Lake and Clinton Creek). The microbial community structure and metabolic potential of the microbialites and adjacent environment are described. While there are specific features to both systems, there are also many similarities between them, and with other freshwater and marine microbialite-forming ecosystems. In particular, the microbialite communities were dominated by *Proteobacteria* and filamentous *Cyanobacteria*, as has been found in other investigations (Havemann and Foster, 2008; Papineau et al., 2005; Allen et al., 2009; Baumgartner et al., 2009; Myshrall et al., 2010; Couradeau et al., 2011; Farias et al., 2013; Russell et al., 2014).

Pavilion Lake microbialite communities were distinct from those in the surrounding environment in relation to GC content, community composition and metabolic potential. The microbialites had significantly higher abundances of sequences assigned to metabolic pathways related to heavy metals, antibiotic resistance, urea metabolism, and primary alcohol fermentation that are associated with members of the *Proteobacteria*. These pathways, and hence the proteobacteria, likely play an important role in these microbialite systems by protecting against heavy metals and toxins, influencing microbialite formation through photosynthetic and urealytic pathways, and stimulating primary and heterotrophic production by liberating nutrients through the degradation of the EPS rich cyanobacterial mats, which would also remove EPS-related inhibition of carbonate precipitation. Therefore, although isotopic data support photosynthesis-
related processes as being the drivers of carbonate precipitation (Brady et al., 2009; Omelon et al., 2013), our data imply that heterotrophic processes dominate microbialite metabolism, and also likely play an important role. Moreover, the isotopic data do not distinguish between photosynthesis-related carbonate-induced precipitation by cyanobacteria, or by more abundant microbial taxa, including Acidobacteria and Proteobacteria. We suggest that Acidobacteria and Proteobacteria play a role in photosynthesis-related carbonate-induced precipitation in microbialites. Pavilion Lake microbialite formation is likely a localized event in the cyanobacterial mat, caused by a combination of heterotrophic and photosynthetic processes. This is consistent with observations that sedimentation rates in Pavilion Lake are quite low, suggesting that photosynthetically induced water-based whiting events are not a major factor in Pavilion Lake microbialite formation (Lim et al., 2009; Omelon et al., 2013).

Genetic signatures of viruses were evident in water from Pavilion Lake, and the presence of CRISPR-cas systems in the microbialites indicates that viruses are interacting with the microbialite community. CRISPRs were abundant in genomes associated with carbonate precipitators, including members of the Deltaproteobacteria (e.g. Myxococcus and Desulfuromonadales), Cyanobacteria (e.g. Anabaena, Nostoc, Rivularia) and Firmicutes (e.g. Clostrida). This suggests that viral infection may influence the microbialite-forming community.

Subarctic Clinton Creek (Yukon, Canada) houses the northernmost and fastest growing microbialites known (Power et al., 2011a). Microbialite contigs had a high proportion of photosynthetic genes, implying that microbialite formation is driven by photosynthesis-induced alkalization. Predicted metabolic pathways overlapped extensively between microbialite and
sediment communities, particularly with respect to housekeeping genes; however, they have distinct core communities, with microbialites dominated by Alphaproteobacteria (mainly anoxic phototrophs like Rhodobacterales), and sediments dominated by Gammaproteobacteria (mainly heterotrophic nitrogen-fixing Pseudomonas spp.).

The Clinton Creek microbialite communities were distinct from microbialites from marine and tropical freshwater systems and microbial mats on corals, whale fall and hot springs. While Clinton Creek and Cuarto Ciengás microbialites shared some functional potential, the functional potential of the Arctic mats was the most similar with those from Clinton Creek, likely because of cold-adaptation. This suggests that given the correct abiotic conditions, polar mats on ice shelves could have the metabolic potential to make microbialites. The microbialites in Clinton Creek appear to be primarily formed as the result of cyanobacterial photosynthetic alkalinization in the mat.

To complement the metagenomic study in Pavilion Lake, a culture-based study was completed, resulting in two isolates, Exiguobacterium spp. and an Agrococcus spp. from a microbialite collected at 20m from Pavilion Lake. Polyphasic characterization of the isolates included growth studies, biochemical testing, lipid analysis, microscopy, antibiotic testing, phylogenetic analysis and whole-genome comparative analysis. Based on these analyses, the isolates were proposed as new species, Agrococcus pavilionensis strain RW1 and Exiguobacterium pavilionensis strain RW2. Genome sequencing and comparative genome and phylogenetic analyses studies were performed to find the evolutionary placement and metabolic potential of the isolates in relation to other members of the genera. Although both isolates appear to be at relatively
low abundance in the Pavilion Lake microbialites, they contribute heavy-metal resistance genes that account for stress response metabolism. *Exiguobacterium pavilionensis* strain RW2 had a functional urease gene, which may contribute to carbonate precipitation through urealytic pathways. In addition, hypothetical carotenoid pathways were discovered that may be responsible for the coloration in *Agrococcus* spp. and *Exiguobacterium* spp., and it is hypothesized that the pigments may contribute to photo-protection.

### 6.2 Future directions

The characterization of a Pavilion Lake microbialite has set the stage for further comparative metagenomic studies of microbialites from neighboring Kelly Lake. The microbialites in Kelly Lake have no visible cyanobacterial mats; however, like Pavilion Lake, microbialite formation is suspected to be driven by photosynthetic alkalinization. There are no reports on the microbial communities or metabolic potential of microbialites from Kelly Lake.

While bacterial communities and their metabolic potential were the focus of this thesis, eukaryotic communities and viruses need to be examined in greater detail to determine their roles in freshwater microbialite systems. Preliminary analysis of metagenomic data from Pavilion Lake suggests that the abundance of eukaryotes is low; however, the metagenomic data from the water were significantly enriched with sequences assigned to photosynthetic eukaryotes. The viral communities in Pavilion Lake suggest that interactions occur between viruses in the water and the bacterial community in the microbialite, as evidenced by the presence of CRISPR-cas sequences. A viral metagenomic study of water from both Pavilion and Kelly Lakes could be used to link viral CRISPR spacers to members of the microbialite bacterial community.
The nodule-associated cyanobacteria of Pavilion Lake have elevated $\delta^{13}$C carbonate values, suggesting they influence carbonate precipitation by photosynthetic alkalinization resulting in the formation of microstromatolithic structures (Brady et al., 2010). A metaproteomic analysis of the nodules could provide insight into how the carbonates in these structures are formed and by which metabolic pathways.

Finally, a global study comparing all of the metagenomic data for microbialites and microbial mats in relation to the environments in which they occur could help to solve questions relating to microbialite-specific metabolic potential and microbial community structure. While a comprehensive metagenomic analysis has not been conducted on many microbialites, a small comparison of freshwater microbialites (Clinton Creek, Kelly and Pavilion Lakes) and marine microbialites (Highbourne Cay and Shark Bay) can now be pursued.


Anil Kumar PK, Suresh PV. (2013). Biodegradation of shrimp biowaste by marine Exiguobacterium sp. CFR26M and concomitant production of extracellular protease and


Braithwaite CJR, Zedefa V. (1994). Living hydromagnesite stromatolites from Turkey. Sed. Geology. 92, 1-5


Edgcomb VP, Bernhard JM, Summons RE, Orsi W, Beaudoin D, Visscher PT. (2013). Active eukaryotes in microbialites from HighbourneCay, Bahamas, and Hamelin Pool (Shark Bay), Australia. ISME J. (Epub ahead of print)


Kennard JM and James NP. (1986). Thrombolites and stromatolites; two distinct types of microbial structures. Palaios 1, 492-503.


Olson JM, Pierson BK. (1986). Photosynthesis 3.5 thousand million years ago. Photosynth Res. 9, 251-259.


White RA 3rd, Suttle CA. (2013). (c). The Draft Genome Sequence of Sphingomonas paucimobilis Strain HER1398 (Proteobacteria), Host to the Giant PAU Phage, Indicates That It Is a Member of the Genus Sphingobacterium (Bacteroidetes). Genome Announc. 1, e00598-13


