Meta-omics analyses of the diversity and metabolism of the uncultivated CL500-3 clade of Planctomycetes in seasonally ice-covered northern lakes

Stephanie Messina-Pacheco

A Thesis
in
The Department
of
Biology

Presented in Partial Fulfillment of the Requirements
For the Degree of Master of Science (Biology) at
Concordia University
Montréal, Québec, Canada

March 2019

CONCORDIA UNIVERSITY

School of Graduate Studies

This is to certify that the thesis prepared

By: **Stephanie Messina-Pacheco**

Entitled: **Meta-omics analyses of the diversity and metabolism of the uncultivated CL500-3 clade of Planctomycetes in seasonally ice-covered northern lakes**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science in Biology

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

_____ External Examiner
Dr. Laurent Potvin-Trottier

_____ Examiner, Chair
Dr. Steve Shih

_____ Examiner
Dr. Brandon Findlay

_____ Thesis Co-supervisor
Dr. Jesse Shapiro (Université de Montréal)

_____ Thesis Supervisor
Dr. David Walsh

Approved by _____
Robert Weladji, Graduate Program Director

_____ Date April 16, 2019
, Dean of Faculty

ABSTRACT

Meta-omics analyses of the diversity and metabolism of the uncultivated CL500-3 clade of Planctomycetes in seasonally ice-covered northern lakes

Stephanie Messina-Pacheco

Owing to cold water temperature and low light penetration through ice, microbial communities were thought to be dormant during winter in seasonally ice-covered lakes. Recent observations of under-ice blooms have challenged this view and raised questions about how microbes contribute to nutrient cycling throughout the year. A previous study demonstrated that Planctomycetes bacteria were commonly abundant under the ice. Therefore, in this thesis, I investigated the diversity and metabolic traits of Planctomycetes in three distinct Canadian lakes over a three-year time series using a combination of 16S rRNA gene sequencing, metagenomic and metatranscriptomic analyses. Winter Planctomycetes assemblages were comprised in large part of the uncultivated CL500-3 clade, which consisted of five subclades (CL500-3a to CL500-3e), each exhibiting a seasonal and biogeographical distribution. Six CL500-3 genomes were reconstructed from Quebec lake metagenomic data and a large fraction of genes were detected in the metatranscriptomic datasets, supporting the notion that CL500-3 subclades were metabolically active throughout the year. CL500-3 genomes were enriched in amino acid degradation pathways and depleted in carbohydrate degrading enzymes compared to other Planctomycetes. The abundance and expression of amino acid degradation pathways suggests that they serve as a source of nitrogen, as well as carbon substrates that feed into the citric acid cycle. Additionally, expression of phosphate-selective porins and chemotactic genes indicates the ability to adapt and respond to cellular and environmental changes. In total, these findings shed light on the metabolism of the widely distributed but poorly characterized CL500-3 clade of Planctomycetes and increase our understanding of the diversity and metabolic processes occurring under-ice in northern lakes ecosystems.

Acknowledgements

I would like to express gratitude toward my supervisor Dr. David Walsh for his continuous guidance, support and for the many opportunities and skill sets allowing me to excel in my scientific career. The door to Dr. Walsh's office was always open whenever I ran into a trouble spot or had a question about my research or writing. He consistently allowed this paper to be my own work, but steered me in the right direction whenever he thought I needed it. I would also like to thank my co-supervisor Jesse Shapiro at Université de Montréal, and my committee members Steven Shih and Brandon Findlay at Concordia University for their support and expertise. Thank you to past and present members of the Walsh lab and various other colleagues for lending a helping hand and for sharing the graduate school experience with me.

For the work presented in this thesis, I would like to specifically thank Morgan Botrel and Stéphanie Massé for coordinating the Lake Sentinels Project, a program funded by the Groupe de Recherche Interuniversitaire en Limnologie et Environnement Aquatique (GRIL), and for sampling Lake Croche and Lake Montjoie. Thank you to Guillaume Grosbois, Tobias Schneider and Maxime Wauthy for sampling Lake Simoncouche. Funding from the Canadian Natural Sciences and Engineering Research Council (NSERC) Discovery Grants (Drs. David Walsh, Milla Rautio, Yannick Huot, Beatrix Beisner) and the Canada Research Chair Program (Drs. David Walsh, Milla Rautio, Yannick Huot) are also acknowledged and appreciated. The metagenomic and metatranscriptomic sequencing and annotation provided by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is recognized and supported under Contract No. DE-AC02-05CH11231. In addition, I would like to acknowledge ÉcoLac, a NSERC-CREATE program, who provided funds during my Master's degree.

Finally, I must express my very profound gratitude to my parents and to my sister for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Contributions of authors

This thesis written by Stephanie Messina-Pacheco entitled "Meta-omics analysis of the diversity and metabolism of the uncultivated CL500-3 clade of Planctomycetes in seasonally ice-covered Northern lakes", includes the work and collaboration of Ola Khawasik, Arthi Ramachandran, Milla Rautio, Beatrix Beisner, Yannick Huot and David Walsh. Stephanie Messina-Pacheco and David Walsh developed the research objectives and methodologies implemented in this thesis. Stephanie Messina-Pacheco performed the metagenomic and metatranscriptomic analyses, as well as the metabolic reconstruction. Ola Khawasik performed the preliminary 16S rRNA analyses, which are included in her Master of Science thesis at Concordia University entitled "Distinct under-ice bacterial communities in seasonally ice-covered northern lakes". The DNA and RNA for the metagenomic and metatranscriptomic datasets were extracted by Arthi Ramachandran. The large-scale "Lake Sentinels" monitoring project on northern seasonally ice-covered lakes in Quebec, Canada took place from 2012 to 2015. This project, associated with the Groupe de Recherche Interuniversitaire en Limnologie et Environnement Aquatique (GRIL), was led by Dr. Milla Rautio at Université du Québec à Chicoutimi, Dr. Yannick Huot at Université de Sherbrooke, Dr. David Walsh at Concordia University, and Dr. Beatrix Beisner, at Université du Québec à Montréal.

Table of contents

Abbreviations

OTU	Operational taxonomic unit
rRNA	Ribosomal RNA
JTT model	Jones-Taylor-Thornton model
ML	Maximum-likelihood
NNI	Nearest-Neighbor Interchange
NJ	Neighbor joining
DoE JGI	Department of Energy Joint Genome Institute
JGI	Joint Genome Institute
IMG/M	Integrated microbial genomes and metagenomes database
MAG	Metagenome-assembled genome
PCA	Principal component analysis
HMM	Hidden Markov models
ANI	Average nucleotide identity
AAI	Average amino acid identity
KEGG	Kyoto Encyclopedia of Genes and Genomes
COG	Cluster of orthologous groups
Pfam	Protein families database
EC number	Enzyme Commission number
GH	Glycoside hydrolase
CAZyme	Carbohydrate-active enzyme
PGDB	Pathway tools genome database
WGS	Whole-genome shotgun
RPKB	Reads recruited per kilobase of transcript per billion reads per sample
RPKM	Reads recruited per kilobase of transcript per million mapped reads
ABC transporters	ATP-binding cassette transporters
PTS transporters	Phosphotransferase system transporters
GS-GOGAT	Glutamine synthetase-glutamine oxoglutarate aminotransferase
GS	Glutamine synthetase
GDH	Glutamate dehydrogenase
TBDR	TonB-dependent receptor
DON	Dissolved organic nitrogen
DIN	Dissolved inorganic nitrogen
LMW	Low molecular weight
HMW	High molecular weight
DFAA	Dissolved free amino acids
C:N:P ratio	Carbon : Nitrogen : Phosphorus ratio
GlcNAc	N-acetylglucosamine
MurNAc	N-acetylmuramic acid
DAP	Diaminopimelic acid
LPS	Lipopolysaccharides
ROS	Reactive oxygen species

List of Figures

List of Tables

1. Introduction

1.1. Winter limnology and seasonally ice-covered lakes

More than 50 million lakes, which encompasses half of the world's lakes, periodically freeze for over 150 days each year (Weyhenmeyer *et al.*, 2011; Verpoorter *et al.*, 2014). Despite this, surprisingly little is known about the ecology, nutrient cycling, and biogeochemical processes occurring under the ice in such lakes (Salonen *et al.,* 2009). Atmospheric warming and the subsequent reduction of ice cover on lakes and rivers is occurring worldwide and highlights an urgent need for research focused on under-ice ecosystem dynamics and their contributions to whole-ecosystem processes (Magnuson *et al.,* 2000; Benson *et al.,* 2012; Hampton *et al.*, 2017). As ice coverage duration decreases, the dynamics of under ice processes are important for understanding and predicting the structure and functioning of seasonally frozen lakes in the near future (Bertilsson *et al.*, 2013; Hampton *et al.*, 2017). Additionally, the number of seasonally ice-covered lakes is expected to grow, as warming temperatures result in melting and prevent ice from forming in lakes that are normally ice-covered year-round, thus forcing open-water conditions for at least some periods of the year (Solomon *et al.,* 2007, Weyhenmeyer *et al.,* 2011, Hampton *et al.*, 2017).

The behavior of lakes and rivers, particularly their biogeochemical processes and food web interactions, is strongly seasonally dependent and recent observations indicate that the seasons are shifting in time in response to climate change (Sommer *et al.*, 1986; Stottlemyer and Toczydlowski, 1999; Thomson, 2009). A global synthesis of summer lake temperature trends revealed that winter ice cover is a major force in determining the characteristics of summer warming trends, demonstrating the cascading effects between seasons (O'Reilly *et al.*, 2015). Cross-seasonal cascades can involve both abiotic and biotic variables, such as when winter ice characteristics influence spring and summer algal growth (Gerten and Adrian, 2000; Straile, 2002; Adrian *et al.,* 2006; Blenckner *et al.,* 2007). Consequently, in water bodies that freeze, the timing and physical characteristics of ice cover are likely to drive some of the most important biological changes for lakes overall (Moore *et al.,* 2009; Salonen *et al.,* 2009; Benson *et al.,* 2012).

1.2 Microbial ecology under ice

Compared to the open water period of the year, under ice conditions in boreal and temperate lakes include low and stable temperatures, slow water movements, limited light availability, and reduced exchange with the surrounding landscape, which can influence community and ecosystem processes (Bertilsson *et al.*, 2013). It was traditionally thought that microbial communities and processes lay dormant under the ice due to a low-light environment, low temperature or nutrient limitation (O'Brien *et al.,* 1992; Cáceres and Schwalbach, 2001; Jewson *et al.,* 2009; Özkundakci *et al.,* 2016). However, recent observations of under ice blooms and the rapid succession of photoautotrophic and heterotrophic organisms have since challenged this traditional view and suggest light penetrating through ice and snow can fuel substantial primary production under the ice (Bertilsson *et al.*, 2013). Phytoplankton blooms may be possible as a result of gradually increasing inputs of solar radiation that melt the snow cover, penetrate the ice, and warm the upper part of the water column, creating conditions that are favorable for photosynthesis (Farmer, 1975; Mironov *et al.,* 2002).

In inland waters, as in other ecosystems, microorganisms are at the center of most biogeochemical processes and largely control ecosystem functioning via their metabolic activities (Rousk and Bengston, 2014). However, attempts to describe the taxonomy and ecology of the freshwater microbiome mainly involve samples collected during the summer months, largely neglecting microbial communities and their activities during the ice-covered period (Bertilsson *et al.*, 2013). Nevertheless, concerns about the consequences of climate change have recently stimulated research on microbial community structure and function during the ice-covered period as well (Salonen *et al.,* 2009). Knowledge on year-round metabolic traits and activities of freshwater microorganisms is required to understand how they will influence other organisms and biogeochemical processes in the future.

1.3. Planctomycetes: an abundant but poorly characterized lineage in freshwaters

Planctomycetes, along with Verrucomicrobia and Chlamydiae form one of the major divisions of the domain Bacteria called the PVC superphylum, which is of increasing importance in understanding microbial biodiversity (Schlesner and Stackebrandt, 1986; Fuerst, 1995; Lindsay *et al.,* 2001; Newton *et al.*, 2011). Planctomycetes were previously described as peptidoglycan-less organisms with proteinaceous cell walls (Fuerst, 1995; Ward *et al.*, 2000).

These Gram-negative microbes, characterized by internal cell compartmentilization and budding reproduction, exist in either spherical or rod-shaped motile bodies (Fuerst, 1995; Ward *et al.*, 2000). The life cycle of many Planctomycetes involves stagnant cells budding to form flagellated motile cells, which swim for a few days before maturing and settling down to reproduce (Lindsay *et al.,* 2001). Maturation results in the loss of their single flagellum, as they synthesize pilli along with a non-cellular stalk, enabling them to attach to surfaces or to each other, forming distinct rosettes (Tekniepe *et al.*, 1981; Newton *et al.*, 2011).

The Planctomycetes phylum has two validated classes: Planctomycetacia, which contains 11 genera, and Phycisphaerae, which contains the single classified species *Phycisphaera mikurensis* (Euzéby, 1997; Fukunaga *et al.*, 2009; Jenkins *et al.*, 2013). Of these 12 genera, ten have representative strains that have been successfully cultured: *Planctomyces, Pirellula, Blastopirellula, Rhodopirellula, Gemmata, Isophaera, Schlesneria, Singulisphaera, Zavarzinella and Phycisphaera* (Schlesner *et al.*, 2004; Fukunaga *et al.*, 2009). Members of this phylum occupy diverse ecological niches, including soils, freshwaters, brackish water, seawater, hot springs, wastewater and waste-treatment bioreactors, marine sediments, organic aggregates, and even invertebrate animals (DeLong *et al.*, 1993; Gray and Herwig, 1996; DeLong *et al.*, 1998; Neef *et al.*, 1998; Strous *et al.*, 1999; Schleifer *et al.*, 2000; Wang *et al.*, 2002; Newton *et al.,* 2011). Some "*Candidatus* Planctomycetes" species are anaerobic ammonium-oxidizing (anammox) bacteria consisting of lithoautotrophs that reduce nitrite and oxidize ammonium to produce dinitrogen gas under anoxic conditions (Jenkins *et al.*, 2013). Aside from the anammox bacteria, most cultured Planctomycetes, including *Phycisphaera mikurensis*, are aerobic organoheterotrophs (Wang *et al.*, 2015). Members of *Phycisphaerae* have been found to be associated with polymeric substrates, such as decaying wood, and can grow on cellobiose and several monosaccharides (Fukunaga *et al.*, 2009; Bienhold *et al.*, 2013). Many members of the better-known *Planctomycetacia* class are able to grow on heteropolysaccharides (Wang *et al.*, 2015). *Rhodopirellula baltica* is one of the best-studied Planctomycetes bacteria. Based on its genome sequence, it can degrade complex polysaccharides such as fucoidan and is hypothesized to live primarily via the decomposition of complex sulphated heteropolysaccharides produced by algae (Glöckner *et al.*, 2003; Wegner *et al.*, 2013). *Planctomycetacia* isolated from *Sphagnum* peat, including species of *Schlesneria*, *Singulisphaera*, and *Zavarzinella*, are also able to degrade various heteropolysaccharides, excluding cellulose and chitin (Dedysh, 2011).

The above description of Planctomycetes is based on only those species that have been cultivated and isolated in the laboratory. Therefore, knowledge of the overall diversity of this phylum based on cultivation approaches is limited (Wang *et al.*, 2015). Traditionally, it has been difficult to observe and cultivate microbes in the laboratory, given the unique conditions in which they grow in the natural environment (Tsementzi *et al.*, 2014). PCR-based surveys of bacteria using 16S rRNA gene sequencing have revealed a vast diversity of uncultivated Planctomycetes in different environments. Many members have mismatches to commonly used universal primers, resulting in the underrepresentation of Planctomycetes in 16S rRNA gene clone libraries (Newton *et al.,* 2011). As a result, little is known about the diversity, distributions and temporal variations of Planctomycetes within different environments (Newton *et al.,* 2011). Nonetheless, despite the fact that Planctomycetes are understudied, they are common in freshwater lake environments and are likely providing important ecosystem functions (Newton et al., 2011).

The majority of Planctomycetes species are uncultured, including the CL500-3 lineage. CL500-3, a poorly characterized clade within the family Phycisphaerae, was discovered in oligotrophic Crater Lake in Oregon (Urbach *et al.*, 2001). The most closely related relatives of the CL500-3 clade are *Phycisphaera mikurensis* and *Phycisphaera algisphaera*, both isolated from marine algae in Japan, and *Tepidisphaera mucosa* discovered in terrestrial hot springs in the Baikal Lake region of Russia (Fukunaga *et al.*, 2009; Yoon *et al*., 2013; Kovaleva *et al*., 2015). Other than the discovery of CL500-3 16S rRNA sequences from various lakes worldwide, a systematic analysis of the diversity and metabolic capacity of this clade has not yet been performed.

1.4. Thesis objectives

In this thesis, the distribution, metabolic gene content and gene expression patterns of CL500-3 across a three-year time-series was investigated using a combination of 16S rRNA gene sequencing, metagenomics and metatranscriptomics. These analyses were performed with the objective of providing insights into the diversity and metabolic potential of CL500-3 associated with open-water and ice-covered conditions in northern lakes. Overall, our analysis advances the understanding of the function, ecophysiology, and distribution of freshwater CL500-3, while highlighting their potential role in freshwater carbon and nitrogen cycling.

2. Materials and Methods

2.1 Sample collection

Sampling was performed at three established lake observatories, located in Quebec, Canada. Two lakes are located in temperate regions, Lake Croche (Mauricie, QC - 45°59N, 74°01W), and Lake Montjoie (Sherbrooke, QC - 45°24N, 72°14W), while the third, Lake Simoncouche (Chicoutimi, QC - 48°14N, 71°15W) is located in a boreal region. Microbial biomass samples were collected throughout the year, for a period of three years. During the open-water period, samples were taken bi-weekly from the epilimnion (the upper layer) and metalimnion (taken at the thermocline, often pooled samples from multiple depths) layers of the water column. In winter, samples representing a mixed water column were collected monthly. A total of 143 samples were collected in pre-acid-washed bottles and subsequently pre-filtered through a 53 µm mesh. The samples were then size-fractionated to distinguish between particle-associated and free-living microorganisms by sequential filtration on a 3 µm polycarbonate filter followed by a 0.22 µm Sterivex filter. After filtration, 1.8 ml of sucrose-based lysis buffer was added to samples collected for DNA extraction, while 1.8 ml of RNAlater (Invitrogen) was added to samples collected for RNA extraction. These samples were then stored at -80 °C until processing.

2.2 Nucleic acid extraction

High quality environmental DNA was extracted from 0.22 µm Sterivex filters using a phenol/chloroform-method modified from Zhou (1996). Sterivex filters were thawed on ice and the storage buffer was removed. The storage buffer was concentrated into Amicon 30 kD filters (500 µl at a time) followed by a centrifugation of 20 minutes at 10,000 g. 500 µl of storage buffer was repeatedly added and re-centrifuged until the buffer was concentrated down to a final volume of 100 µl. Buffer exchange was conducted twice by washing with 500 µl of TENP buffer (600 mg Tris, 740 mg EDTA, 580 mg NaCl, 2 g Polyvinylpyrrolidon and 100 ml milliQ, pH 8). The filters were then removed from the Sterivex, split into halves and placed inside a 2 ml Eppendorf tube. In order to conduct the cell lysis and digestion, 0.37 g of 0.7 mm pre-sterilized Zirconium beads, 60 µl of 20% SDS, 100 µl concentrated buffer exchanged filtrate, 500 µl TENP buffer and 500 µl phenol-chloroform-isoamylalcohol (PCI) 25:24:1 were added to the 2 ml Eppendorf tube and vortexed for 10 minutes. Samples were incubated for 10 minutes in a

60°C water bath, followed by a 1-minute incubation on ice, and a 6-minute centrifugation at 10,000 rpm and 4°C. The resulting supernatant was transferred to a clean 1.5 ml Eppendorf tube where 500 µl phenol-chloroform-isoamylalcohol (PCI) 25:24:1 was added and vortexed briefly. This solution was then centrifuged for 6 min at 10,000 rpm and 4°C and the supernatant was transferred to a new 1.5 ml Eppendorf tube. The PCI step was repeated twice or until the white precipitate was removed. The DNA was precipitated by adding 120 µl of 3 M sodium acetate and 1 ml of 96% ethanol and incubated at -20°C for at least 1.5 hours. The incubation was followed by a centrifugation of 60 minutes at 13,000 rpm and 4°C. The supernatant was decanted, and the pellet was washed with 850 µl of 80% ethanol. The samples were then incubated for 10 minutes on ice, vortexed briefly and centrifuged for 15 minutes at 13,000 rpm and 4°C. The supernatant was removed, and the pelleted DNA was resuspended in 50 µl of TE or Tris-HCl (pH 7.5-8) buffer.

RNA was extracted from 0.22 µm Sterivex filters with a modified protocol which employs both the mirVana miRNA isolation kit (Invitrogen) and the RNeasy RNA cleanup kit (Qiagen) (Shi *et al.*, 2009; Stewart *et al.*, 2010). Samples were thawed and the RNAlater (Invitrogen) surrounding the Sterivex filter was removed (approximately 1700 µl) and discarded. 1700 µl of mirVana lysis buffer was added to the Sterivex filter and vortexed to lyse bacterial cells attached to the filter. Total RNA was then extracted from the lysate according to the mirVana protocol. Purified sample (100 µl) was treated with 2 µl DNase (New England Biotech) incubated at 65ºC for 1-2 hours to remove genomic DNA and concentrated using the RNeasy RNA cleanup kit (Qiagen).

2.3 16S rRNA amplicon sequencing and data analysis

A 16S rRNA gene library was then generated for 143 samples by PCR amplification and sequencing of 16S rRNA genes using next-generation technology. The V3 region of the 16S rRNA gene was amplified using the universal primers 341F (5'CCTACGGGRSGCAGCAG 3') and 515R (5' TTACCGCGGCKGCTGVCAC 3') (Klindworth *et al.*, 2013). Two-step PCR reactions with a final volume of 25 µl contained 0.5 µM $MgCl_2$, 0.2 mM deoxynucleotide, 0.2 µM of each primer and 1U of Phire Hot Start II DNA Polymerase (modified from Berry *et al.*, 2011). The template was amplified using non-barcoded PCR primers for 20 cycles, followed by 1:50 dilution of the PCR product and 10 additional cycles with barcoded PCR primers. The PCR

conditions consisted of an initial 95°C denaturation step for 4 min, a cycling program of 95°C for 30 s, 52°C for 30 s, 72°C for 60 s, and a final elongation step at 72°C for 7 min. Reverse primers were barcoded with a specific IonXpress sequence to identify samples. PCR products were purified using QIAquick Gel Extraction Kit (Qiagen), quantified using Quantifluor dsDNA System (Promega), pooled at an equimolar concentration and sequenced using an Ion Torrent PGM system on a 316 chip with the ION Sequencing 200 kit as described in Sanschagrin and Yergeau (2014).

The V3 region of the 16S rRNA sequences obtained by the Ion Torrent PGM system were analyzed using an open-source bioinformatics data processing pipeline called MOTHUR (Schloss *et al.*, 2009). Sequences were clustered into operational taxonomic units (OTUs) at 97%, generating 6,974 amplicon sequences, which were then assigned to taxonomic groups using the structure proposed in Newton *et al.* 2011. Sequences that did not match the IonXpress barcode, along with sequences with an average quality of < 17, and both the PCR forward and reverse primer sequences were discarded. Potential chimeric sequences were identified using UCHIME and discarded as well (Edger *et al.*, 2011). Additionally, a sequence cut-off of 100 bp was used.

In order to investigate Planctomycetes diversity, a maximum likelihood phylogenetic tree using non-redundant reference 16S rRNA gene sequences obtained from the SILVA database (clustered at 97% identity), along with OTUs taxonomically classified using MOTHUR, was generated. The maximum likelihood phylogenetic tree was generated using MEGA6.06, with a bootstrap of 100 iterations. The substitution model was the Jones-Taylor-Thornton (JTT) model for amino acids, the rates among site were Gamma distribution with 4 gamma categories, the maximum-likelihood (ML) heuristic method was Nearest-Neighbor Interchange (NNI), and the initial tree was neighbor joining (NJ).

2.4 Global biogeography of CL500-3 based on 16S rRNA gene distributions

CL500-3 diversity in lakes across the globe was assessed by searching the SILVA database and NCBI for 16S rRNA sequences belonging to Phycisphaerae. From the maximum likelihood phylogenetic tree described in the previous section, a representative 16S rRNA gene sequence was obtained for each subclade (CL500-3a to CL500-3e). A comprehensive global

distribution profile of each subclade was generated by obtaining the location and coordinates for each representative sequence, along with the respective non-redundant reference 16S rRNA gene sequences (SILVA database) clustered at 97% identity. From this information, each CL500-3 16S rRNA sequence found globally was classified to its respective subclade (CL500-3a to CL500-3e).

2.5 Metagenome sequencing, assembly, annotation and binning

From 73 free-living samples, extracted DNA from a subset of 24 samples were sent to the Department of Energy Joint Genome Institute (JGI) (Walnut Creek, CA, USA) for DNA sequencing on the HiSeq 2500-1TB (Illumina) platform. Paired-end raw reads of 2 × 150bp were generated for all libraries and assembled using MEGAHIT (https://github.com/JGI-Bioinformatics/ megahit) with kmer sizes of 23, 43, 63, 83, 103, 123. Gene prediction and annotation was performed using the DoE JGI IMG/M functional annotation pipeline (Markowitz *et al.*, 2014).

16S rRNA gene data extracted from the assembled metagenomes were taxonomically assigned using MOTHUR. Sequences belonging to Planctomycetes were selected and clustered at 100% identity using CD-HIT version 4.6 (Li and Godzik, 2006). The longest sequence representative of each cluster was added to the CL500-3 phylogenetic tree in order to assign the metagenome-derived 16S rRNA sequences to CL500-3 subclades. The depth of coverage information provided by the JGI allowed for the determination of the relative abundances of each subclade throughout the year.

Metagenomic binning, using MetaWatt version 3.5.2, was performed on individual samples in order to generate metagenome-assembled genomes (MAGs) (Strous *et al.*, 2012). Binning was based solely on tetranucleotide frequencies; therefore, the relative weight of binning coverage setting was set to 0. The identity of the resulting bins was assessed using a concatenated phylogeny of single copy core genes implemented in MetaWatt, MAFFT aligner and FastTreeMP for tree inference (Price *et al.*, 2010). Subsequent manual curation of MAGs was conducted through differential coverage and by implementing a scaffold length cut-off of ≥ 5000 base pairs for further analysis. Differential coverage was determined by recruiting unassembled raw reads from each metagenomic sample to each MAG using bbmap (cutoff of

97%). Principal component analyses (PCAs) were generated for each MAG, showing the differential coverage profiles of each scaffold. A 95% confidence measure was used to ascertain which scaffolds should be binned together based on similar profiles. The outlier scaffolds were then further investigated on IMG to determine which proteins were present and how they were taxonomically assigned. Scaffolds containing ribosomal proteins were kept because it is known that they exhibit a higher coverage due to conserved regions (Wang, *et al.* 2013). Outlier scaffolds determined to be contaminating or heterogeneous were removed.

CheckM, which relies on pplacer, prodigal and HMM, was used to assess binning statistics such as completeness and contamination (Matsen *et al.*, 2010; Eddy, 2011; Hyatt *et al.*, 2012; Parks *et al.*, 2015). Six MAGs of 40 - 73% completeness, with no contamination were obtained. To assess MAG similarity, ANI values were calculated with ANICalculator (https://ani.jgi-psf.org/html/anicalculator.php), while AAI values were calculated using CompareM (https://github.com/dsparks1134/CompareM).

2.6 Comparative genomics and phylogenetic relationships

The distribution of protein-encoding gene content between genomes was determined using proteinortho (Lechner *et al.*, 2011). The inputs included the 6 MAGs, either alone or with 9 reference genomes belonging to Planctomycetes (*Phycisphaera mikurensis, Gemmata obscuriglobus, Isosphaera pallida, Pirellula staleyi, Gimesia maris, Planctopirus limnophila, Rhodopirellula baltica, Rhodopirellula europaea,* and *Rhodopirellula sallentina*). The 15 genomes shared 8 core proteins, 4 of which were selected to build a concatenated gene phylogeny. These proteins included phosphoribosylformyl-glycinamidine synthase I, phosphoribosylformylglycinamidine synthase II, phosphoribosylamine - glycine ligase and ABC-2 type transport system ATP-binding protein. The remaining 4 core proteins were not selected based on poor confidence levels in the orthology, and duplication of single copy genes. A maximum-likelihood tree was inferred for each of the 4 genes selected to ensure that these proteins were conserved and that the phylogenetic relationships between genomes are well represented. In order to create the concatenated phylogenetic tree, sequences for each protein were aligned using MUSCLE and concatenated using Mesquite (Edgar, 2004; Maddison and Maddison, 2018). MEGA6.06 was used to generate a maximum likelihood phylogeny of 15 taxa, using a bootstrap of 100 iterations (Tamura *et al.*, 2013). The substitution model was the Jones-Taylor-Thornton (JTT) model for amino acids, the rates among site were Gamma distribution

with 4 gamma categories, the maximum-likelihood (ML) heuristic method was Nearest-Neighbor Interchange (NNI), and the initial tree was neighbor joining (NJ). To further investigate the relationship between the 6 Quebec MAGs and the 9 Planctomycetes reference genomes, an automated concatenated gene phylogeny was created using PhyloPhlan (Segata *et al.*, 2013).

2.7 Metabolic reconstruction and functional annotation

Inference of protein function and metabolic reconstruction was based on the IMG annotations provided by the JGI, including KEGG, COG, Pfam, EC numbers, and Metacyc annotations. KEGG proteins were classified into their respective functional categories using a mapping file created from the publicly available information found on the KEGG database (https://www.genome.jp/kegg/).

Furthermore, carbohydrate-active enzymes were identified using the dbCAN annotation tool (http://csbl.bmb.uga.edu/dbCAN/annotate.php), which uses hmmsearch against hidden Markov models (HMMs), with an *E*-value cutoff of 1e-10 (Yin *et al.*, 2012). Owing to the nonspecific HMM associated with GH109, this GH was removed from our reported results. In order to compare carbohydrate-active enzyme (CAZyme) abundances between Quebec MAGs and known Planctomycetes genome diversity, the number of CAZymes present in each MAG needed to be corrected for MAG completeness. This was achieved by dividing the number of CAZymes found by the MAG size, and then multiplying by the estimated genome size (MAG size x completeness). Originally, 21 glycoside hydrolases (GH) were found in SC-feb-99, which has a MAG completeness of 73.24% and a bin size of 2473776bp. Using these characteristics, an estimated genome size of 3377630bp was calculated. Finally, 21 GH was divided by the MAG size (2473776bp) and then multiplied by the estimated genome size (3377630bp), which leaves us with 29 estimated glycoside hydrolases. Therefore, if SC-feb-99 were 100% complete, this MAG would comprise of 29 GH genes.

Metabolic reconstruction was facilitated using the PathoLogic algorithm to map functional annotations onto the MetaCyc collection of reactions and pathways. Pathway tools genome databases (PGDBs) for each MAG were constructed, which were then navigated using Pathway Tools (Karp *et al.*, 2009). Additionally, to investigate the significant enrichment of amino acid utilization in CL500-3, carbohydrate and amino acid degradation pathways between

CL500-3 and other heterotrophic freshwater bacteria were compared. PGDBs were generated for the 9 Planctomycetes reference genomes used in the concatenated protein phylogeny, 15 Verrucomicrobia MAGs from Quebec lakes spanning subdivisions V1 through V4 (Tran *et al.*, 2018), and 8 known free-living freshwater bacteria. These PGDBs were created using data from IMG so that they would be comparable to the Quebec MAGs described in this paper. A one-way ANOVA statistical test was performed to evaluate the difference in abundance of amino acid and carbohydrate degrading enzymes between CL500-3 MAGs and Planctomycetes reference genomes. Ratios of amino acid to carbohydrate degradation enzymes were calculated for each genome and differences were considered significant when $P < 0.05$.

2.8 Metatranscriptomic analysis of CL500-3 gene expression patterns

cDNA library preparation and sequencing were performed at the Department of Energy Joint Genome Institute (JGI) (Walnut Creek, CA, USA) on the HiSeq 2500-1TB (Illumina) platform in order to assemble and annotate metatranscriptomic data. Paired-end sequences of 2 × 150bp were generated for all libraries. The metatranscriptome dataset comprises 24 samples (6 Croche, 8 Montjoie and 10 Simoncouche), some overlapping in date with the metagenome dataset. Fragment recruitment was achieved using BBMAP (https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbmap-guide/) to map the reads to the CDS in each of the 6 CL500-3 MAGs, using a cutoff of 97% identity. Gene expression is further investigated using an RPKB value. RPKB (reads recruited per kilobase of transcript per billion reads per sample) was calculated instead of the RPKM (reads recruited per kilobase of transcript per million mapped reads) to control for differences in raw reads between samples.

2.9 Data accessibility

The metadata for the 24 metagenomes and 24 metatranscriptomes can be downloaded at the JGI website, https://jgi.doe.gov/ (Metagenomes - IMG Genome ID: 3300009684, 3300009182, 3300010158, 3300010334, 3300009151, 3300009154, 3300010157, 3300009186, 3300009161, 3300010160, 3300009155, 3300009181, 3300009163, 3300009185, 3300009187, 3300009068, 3300009183, 3300009164, 3300009184, 3300009158, 3300009159, 3300009180, 3300009152, 3300009160; Metatranscriptomes - IMG Genome ID:3300012778, 3300012771, 3300012780, 3300012756, 3300012760, 3300012765, 3300012776, 3300012768, 3300012751, 3300012754, 3300012769, 3300012775, 3300012755, 3300012766, 3300012774, 3300012779,

3300012758, 3300012781, 3300012772, 3300012761, 3300012763, 3300012773, 3300012770, 3300012777).

3. Results

3.1 CL500-3 clade diversity and distribution in Quebec lakes

The abundance and diversity of CL500-3 was assessed in three seasonally ice-covered Quebec lakes (Lake Croche, Lake Montjoie, and Lake Simoncouche) (**Figure S1**) using 16S rRNA sequences from a combination of amplicon and whole-genome shotgun (WGS) metagenomic datasets covering a period of 3 years. In total, 143 samples were collected from the epilimnion and metalimnion of these three seasonally ice-covered lakes during ice-covered and open-water periods of the year, corresponding to 6 winter time-points (January and February 2013, 2014, and 2015) and 6 summer time-points (June, July, August, 2013 and 2014) **(Table S1)**.

Phylogenetic analysis of 16S rRNA gene sequences recovered from Quebec lakes identified five well-resolved CL500-3 subclades (CL500-3a to CL500-3e), with an average of 89% identity between near full-length reference sequences (**Figure 1a, 1b**). The most closely related relatives of CL500-3 were *Phycisphaera mikurensis* and *Phycisphaera algisphaera*, both isolated from marine algae in Japan, and *Tepidisphaera mucosa* discovered in a terrestrial hot spring in the Baikal Lake region of Russia (**Figure 1b**). A global synthesis of CL500-3 16S rRNA sequences lead to the discovery that these subclades were not specific to Quebec lakes, but were present in a wide range of global freshwater ecosystems (**Figure S2**).

According to the 16S rRNA amplicon data, CL500-3 comprised a significant fraction of the bacterial community in Montjoie and Simoncouche, reaching a maximum of 18% of 16S rRNA sequences in Montjoie (**Figure 1c**). Alternatively, CL500-3 represented less than 1.5% of the amplicons from Croche. Overall, the relative abundance of 16S rRNA gene sequences were generally higher during the ice-covered period, particularly in Montjoie. Focusing on the relative abundances of each subclade individually, it was observed that CL500-3a, -3b, and -3c were more abundant in Montjoie, while CL500-3d and -3e were present in both Montjoie and Simoncouche. Additionally, CL500-3a, -3c and -3e generally preferred a winter environment,

while CL500-3b had a tendency to associate with the summer season. CL500-3d was dynamic and had an affinity for either an ice-covered environment in Simoncouche, or an open water environment in Montjoie (**Figure 1c, S3**). These results demonstrated that each subclade exhibited a distinct seasonal and geographical distribution, suggesting substantial ecological diversity exists within CL500-3. Moreover, CL500-3 was common in both free-living and particle-attached fractions of the community. However, a higher abundance overall was observed for the particle-attached fraction, suggesting a preference for a particle-associated lifestyle. Although relative abundance differed between lakes, seasons and lifestyles, these results demonstrate that CL500-3 subclades are often associated with ice-covered conditions, making them an intriguing group with which to investigate genomic and metabolic adaptations to life under ice.

To further explore CL500-3 diversity, we retrieved 16S rRNA sequences from an overlapping metagenomic dataset and used scaffold coverage as a measure of relative abundance (**Figure S4**). As expected, CL500-3 was found in very low abundance in Croche, with a relatively high depth of coverage in the Lake Montjoie samples, especially during the ice-covered periods. Patterns of seasonal and environmental preferences mimicked those obtained from the amplicon data, in that subclades are found in either one or both lakes, and in either the open-water season, under the ice, or both.

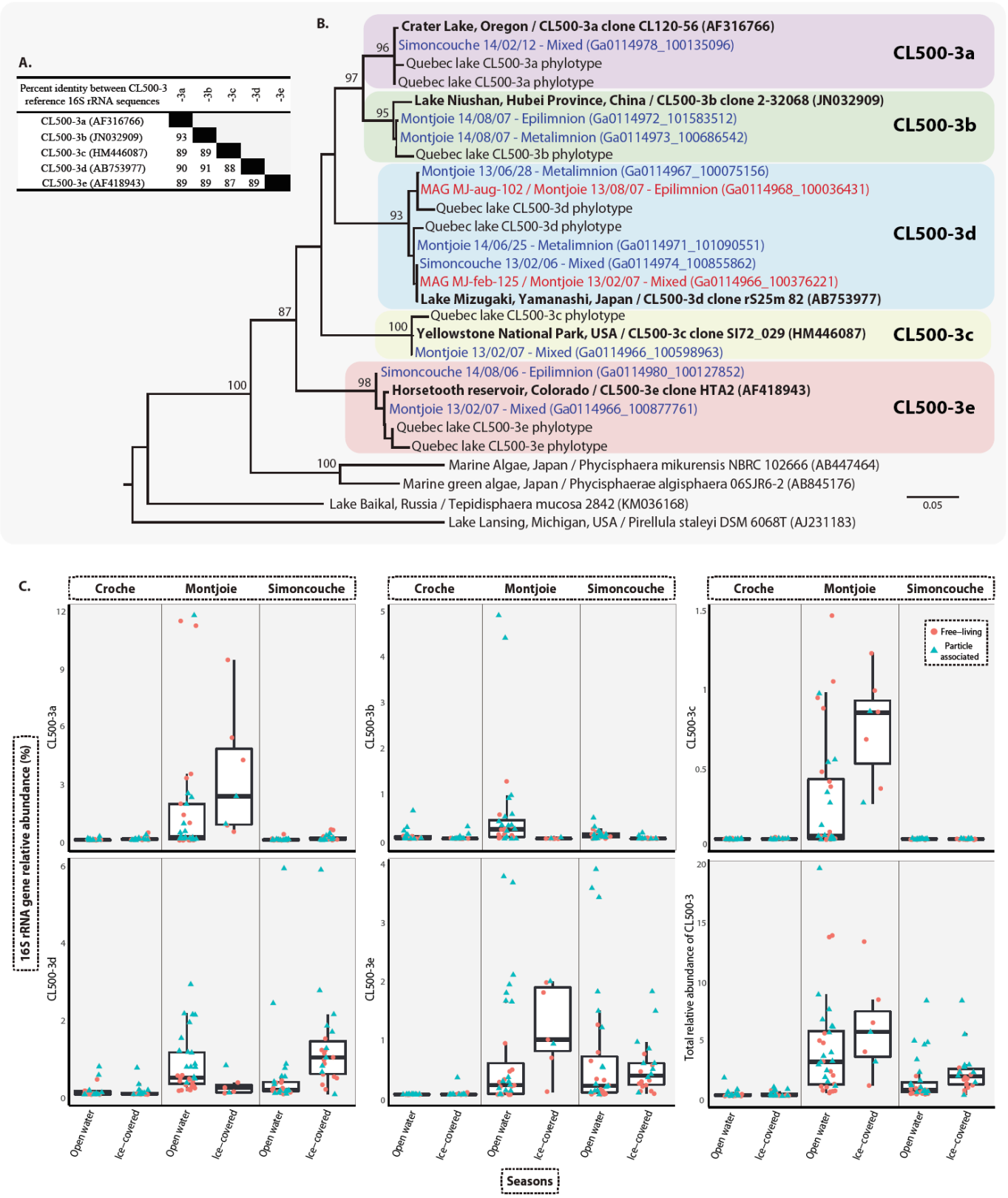

Figure 1.
A. Percent identity between CL500-3 reference 16S rRNA sequences representing each subclade. Sequences were obtained from the NCBI database. **B.** Maximum likelihood phylogenetic tree of 16S rRNA gene sequences representing the five lineages of the uncultivated CL500-3 clade of Phycisphaerae. Sequences in blue represent the metagenomic 16S rRNA gene sequences, while sequences in red represent the 16S rRNA gene sequences found in two of the six MAGs, and the bolded sequences represent the reference sequence for each lineage. **C**. Box plots demonstrating the percent relative abundance of CL500-3, represented by the 16S rRNA amplicon data, organized by lake and season. Mid-lines represent the median, upper and lower boundaries are 25% quartiles, and whiskers represents scores outside the middle 50%.

3.2 Diversity of CL500-3 MAGs from Quebec lakes

CL500-3 genome diversity was investigated using metagenomes generated from 16 samples collected over a nearly two-year period (February 2013- August 2014) in Montjoie and Simoncouche, where CL500-3 populations were detected. Metagenomic binning was performed on each sample individually to facilitate the comparison between seasons and lakes. Six MAGs were selected based on seasonality and high N50 value. The six MAGs originated from samples collected during three different time periods (February, June, August) in Montjoie (MAGs MJ-feb-125, MJ-june-47, MJ-aug-102) and Simoncouche (MAGs SC-feb-99, SC-june-56, SC-aug-92). After manual curation, MAG sizes were between 0.99-2.47 Mb, GC content was 59-70%, completeness ranged from 40-73% and contamination values were 0% (**Table 1)**. The average nucleotide identity (ANI) between the MAGs ranged from 66-92%, while the average amino acid identity (AAI) was 54-93% (**Table S2**).

MAG	MAG size (bp)[a]	Completeness (%)[a]	Estimated genome size (bp)[c]	# of Scaffolds[a]	Cov (x)[b]	GC content[a]	N50 (bp)[a]	Longest scaffold (bp)[a]	# of Proteins[a]	Contamination[a]
MJ-feb-125	1233551	45.19	2729699	139	7.27	59.6	9559	19863	1144	0
SC-feb-99	2473776	73.24	3377630	243	10.58	63	11313	30616	2237	0
MJ-june-47	2002500	69.81	2868500	175	8.72	69.4	12674	49213	1809	0
SC-june-56	1593273	54.55	2920757	140	10.10	69.7	13243	38646	1442	0
MJ-aug-102	993108	40.65	2443070	103	12.75	67.9	10950	25395	943	0
SC-aug-92	2071391	45.76	4526641	269	6.90	67.3	7574	25091	1934	0

Table 1.
Genome characteristics of the six CL500-3 MAGs, recovered from Quebec lake metagenomic samples. ([a] Calculated using CheckM, [b] Calculated using BBMAP by mapping raw reads to scaffolds using 97% identity, [c] Calculated as (MAG size (bp)/completeness) *(100)).

16S rRNA genes were present in two MAGs (MJ-aug-102 and MJ-feb-125) and based on 16S rRNA phylogeny, both were members of CL500-3d (**Figure 1b**). Given that 16S rRNA genes were not identified for each MAG, phylogenetic relationships between CL500-3 MAGs and Planctomycetes genomes were investigated using a concatenated protein phylogeny. The 6 MAGs formed a monophyletic group consisting of two clades, with marine *P. mikurensis* as the most closely related genome (**Figure 2a**). A clade of 4 MAGs (MJ-aug-102, MJ-june-47, SC-june-56, MJ-feb-125) encompassed the two that contained the 16S rRNA sequences assigned to the CL500-3d clade (MJ-aug-102 and MJ-feb-125), suggesting that all 4 MAGs belong to the Cl500-3d subclade. A second clade containing two MAGs (SC-feb-99 and SC-aug-92) was more distantly related. Since the concatenated phylogeny was based on only four proteins,

phylogenetic relationships were further investigated using Phylophlan, an automated pipeline based on hundreds of conserved proteins. In general, the Phylophlan tree agreed with the concatenated protein phylogeny, in that there were two well-resolved monophyletic clades within the CL500-3 clade (**Figure S5**).

3.3 Biogeography of CL500-3 MAGs

CL500-3 MAG biogeography across Quebec lakes and seasons was investigated through fragment recruitment analysis. For the most part, each MAG exhibited the highest coverage in its sample of origin and the recruitment patterns across metagenomes revealed seasonal patterns (**Figure 2b**). For example, all four MAGs originating from summer metagenomes (MJ-aug-102, MJ-june-47, SC-june-56, SC-aug-92) were common in summer metagenomes but rare in winter metagenomes, adding evidence that they represent summer-associated CL500-3 populations. Biogeography patterns revealed that MJ-aug-102, SC-jun-56, MJ-jun-47 are associated with the epilimnion, and SC-aug-92 with the metalimnion. The two MAGs reconstructed from winter metagenomes (MJ-feb-125 and SC-feb-99) consistently exhibited highest coverage in winter metagenomes, although they were also detected within summer metagenomes. These results suggested that the CL500-3 MAGs represent dynamic populations with differences in seasonal associations.

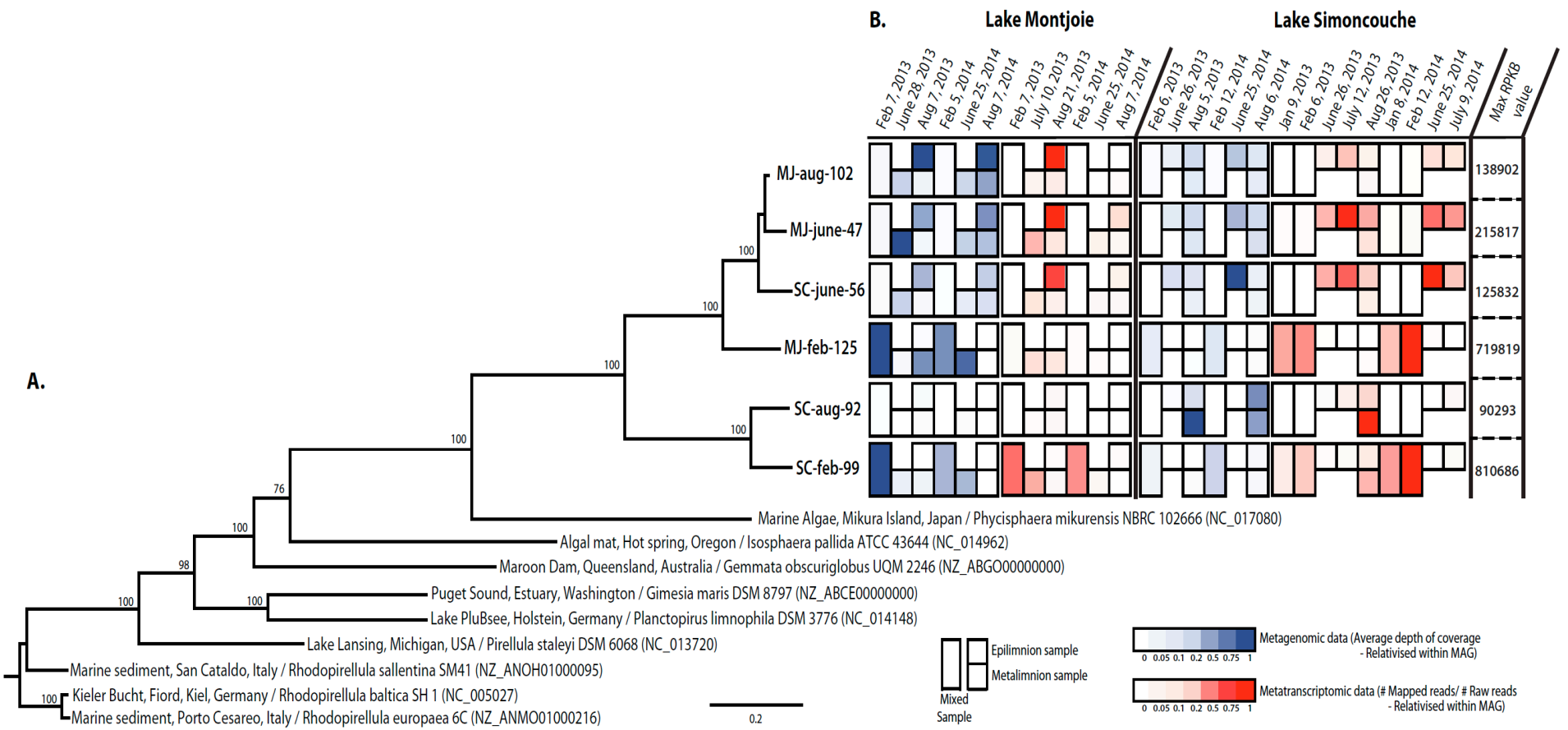

Figure 2.
A. Concatenated protein phylogeny demonstrating the relationship between the six CL500-3 MAGs and 9 Planctomycetes reference genomes. **B.** Heatmap of the relative MAG abundance based on metagenomic coverage (blue) and relative gene expression based on the number of transcripts recruited to each MAG (red), across lakes and seasons. Rectangular boxes represent winter ice-covered samples, with a mixed water column. Square boxes represent summer samples taken from the epilimnion (top squares) or the metalimnion (bottom squares) layers of the water column. Heatmap is comparable across a MAG, but not between different MAGs. Therefore, to give an idea of the level of expression found in each MAG, the Max RPKB value represents the highest gene expression value found in any given sample within a MAG. MAGs: MJ-Feb-125, MJ-june-47, MJ-aug-102, SC-feb-99, SC-june-56, SC-aug-92.

Quebec CL500-3 MAG biogeography and diversity was assessed in metagenomes originating from additional freshwater environments (**Table S3, Figure S6**). These included metagenomes from two Spanish reservoirs (Tous and Amadorio), two Wisconsin Lakes (Trout and Mendota), as well as from under the ice in Lake Baikal. Overall, fragment recruitment from these locations was very low. Results demonstrated that winter CL500-3 MAGs are not detected outside of Quebec lakes, including an absence from the under-ice metagenome from Lake Baikal. Although CL500-3 populations related to SC-june-56, MJ-june-47, and MJ-aug-102 were present in the Tous reservoir during the summer, they were not detected during the winter, providing further evidence that these are in fact summer-associated MAGs. Also, CL500-3 related populations were identified in the Amadorio reservoir particle-associated fraction, providing evidence that CL500-3 can be associated with larger size fractions in other systems as well.

3.4 Metabolic overview of the CL500-3 clade

To investigate the general metabolic potential of CL500-3, protein-encoding genes within the MAGs were assigned to KEGG categories representing various metabolisms. The relative abundance of the genes found in different KEGG categories was similar in all CL500-3 MAGs, with slight variations in a few categories (**Table S4**). According to the KEGG annotations, 46-53% of proteins were involved in metabolism, 37-44% of which were involved in carbohydrate and amino acid metabolism, 27-38% were involved in genetic information processing, 10-14% were involved in environmental information processing and 5-10 % were involved in cellular processes such as cell motility and transport.

To provide a more detailed metabolic overview of CL500-3, a composite metabolic map combining the genomic features encoded in the six MAGs was generated using the Metacyc metabolic pathway database as a reference (**Figure 3**). The metabolic reconstruction suggested that the CL500-3 clade is comprised of aerobic heterotrophic bacteria. Evidence for central metabolic pathways such as glycolysis, the TCA cycle, aerobic respiration, the pentose phosphate pathway (non-oxidative branch) and the non-mevalonate pathway (isoprenoid biosynthesis) was observed in all MAGs. Evidence for autotrophic CO_2 fixation, lithotrophic energy conservation, or anaerobic respiration was not found.

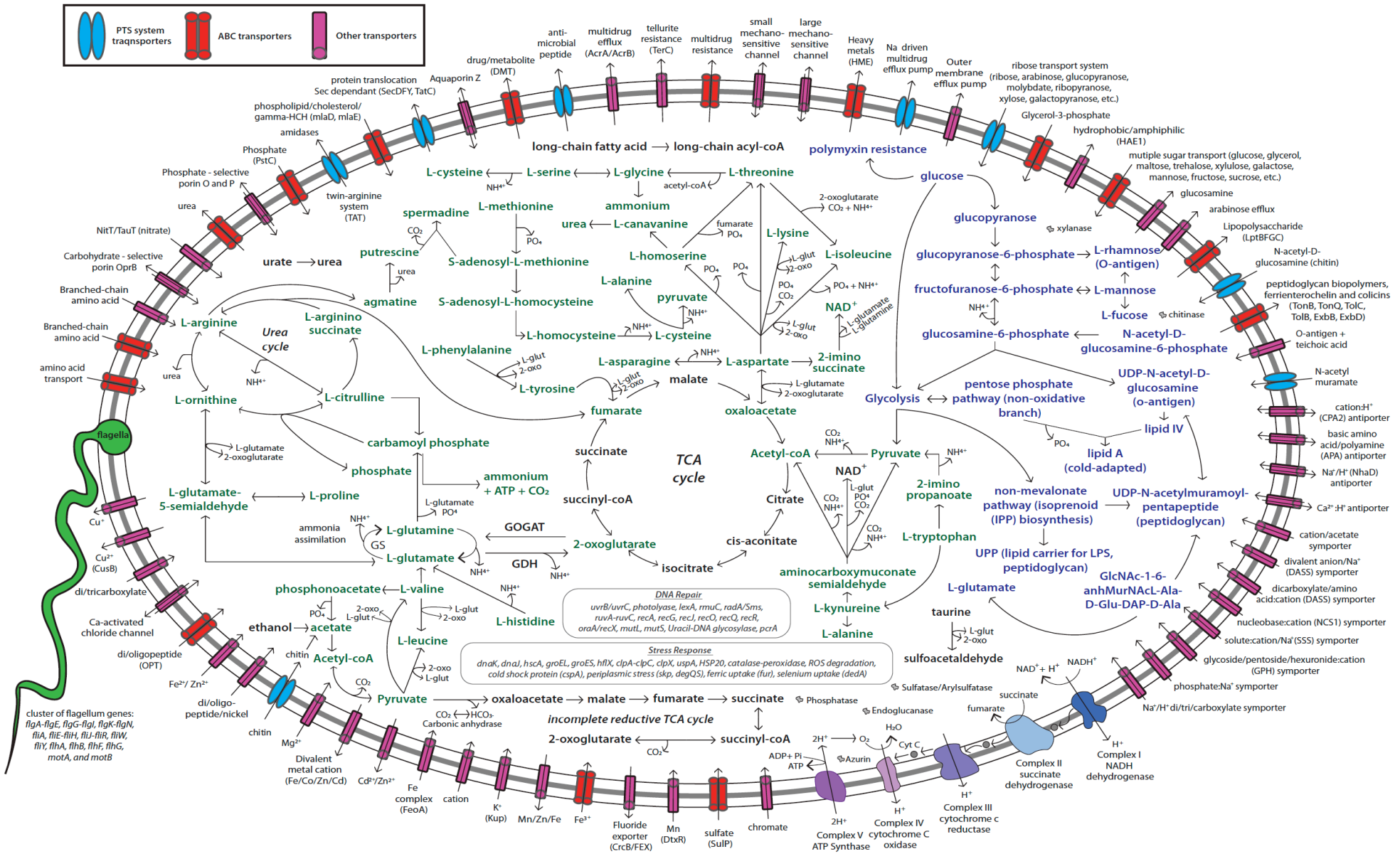

Figure 3.
Metabolic cellular overview representing all lineages of CL500-3. Enzymes, substrates, compounds or products involved in amino acid degradation and/or biosynthesis pathways are highlighted in green, while those involved in peptidoglycan biosynthesis/ degradation, central carbon metabolisms and biosynthesis of several membrane components are highlighted in blue. PTS system, ABC and other transporters are shown around the membrane and are depicted by several shapes (shown in legend). Arrows around the cellular membrane (attached to transporters) show whether the compound is being imported into the cell or exported out of the cel.

3.4.1 Saccharolytic Metabolism

Aquatic Planctomycetes are often associated with the degradation of biopolymers, particularly polysaccharides, which is apparent due to an abundance of carbohydrate-activate enzymes (CAZymes) found within their genomes (Kim *et al.,* 2016). To investigate whether this is a trait also found in CL500-3, CAZyme abundance and diversity in the 6 MAGs was investigated. In summary, between 9-21 glycoside hydrolases (GH), 11-27 glycosyl transferases (GT), 0-2 polysaccharide lyases (PL), 4-19 carbohydrate esterases (CE), 4-12 carbohydrate binding modules (CBM), 0-3 auxiliary activities (AA) and 2-35 dockerin-encoding genes were identified in the MAGs (**Table S5**). When scaled based on genome completeness, the estimated CAZyme abundances increased slightly (**Table 2**), but values were still well below those reported from currently available Planctomycetes genomes, as well as genomes from other well-described saccharolytic bacteria. Specifically, CL500-3 MAGs encode an estimated 11-18% of the GHs found in *P. mikurensis*, 9-31% of GHs found in other aquatic Planctomycetes genomes, 12-15% in two Verrucomicrobia genomes (*Verrucomicrobia spinosum*, *Chthoniobacter flavus*) and lastly, 9-10% found in the genomes of other saccharolytic bacteria including *Lentisphaera araneosa*, *Singulisphaera acidiphilus*, and *Zobellia galactanivorans* **(Table 2)**. Similarly, the GT, PL, CE and CBM categories in CL500-3 were on average 19, 42, 48 and 15% respectively, of the number of these enzyme classes found in other marine and freshwater genomes presented in **Table 2.**

Genomes	Organism	Genome size (bp)	Glycoside hydrolases	Glycosyl transferases	Poly-saccharide lyases	Carbohydrate esterases	Carbohydrate binding modules
MAGs	MJ-feb-125	2729699	15	24	4	27	11
	SC-feb-99	3377630	29	22	0	26	14
	MJ-june-47	2868500	19	29	3	21	20
	SC-june-56	2920757	26	37	4	13	22
	MJ-aug-102	2443070	22	30	5	10	10
	SC-aug-92	4526641	24	59	4	35	11
Plactomycetes genomes	Phycisphaera mikurensis	3803225	144	171	2	22	45
	Isosphaera palida	5472964	86	185	5	53	175
	Gemmata obscuriglobus	9161841	119	248	5	100	293
	Planctomyces maris	7777997	171	182	16	84	297
	Planctomyces limnophilus	5423075	89	139	2	55	162
	Pirellula staleyi	6196199	104	143	3	77	232
	Rhodopirellula sallentina	8186686	371	202	5	68	324
	Rhodopirellula baltica	7145576	191	155	9	82	301
	Rhodopirellula europaea	7446194	202	158	11	78	325
	Rhodopirellula maiorica	8874084	319	255	7	95	401
	Rhodopirellula rubra	8777069	394	226	10	98	368
	Blastopirellula marina	6653746	167	195	4	60	190
	Planctomyces brasiliensis	6006602	157	175	3	67	207
	Schlesneria paludicola	8674627	168	211	7	91	278
	Kuenenia stuttgartiensis	4218325	78	247	0	42	49
Verrucomicrobia genomes	Verrucomicrobia spinosum	8220857	176	264	16	67	287
	Chthoniobacter flavus	7848700	246	309	18	107	315
Other bacterial genomes	Lentisphaera araneosa	6023180	386	116	2	68	465
	Singulisphaera acidiphilus	9755686	213	320	7	104	313
	Zobellia galactanivorans	5521712	278	176	14	55	170

Table 2.
Table showing CAZyme abundances between CL500-3 MAGs and other marine and freshwater genomes. CAZyme abundances for CL500-3 MAGs are estimated based on an estimated genome size. Genomes sizes for MAGs are estimated based on completeness.

Although CL500-3 do not share the abundance of CAZymes observed in related Planctomycetes, they still possess a repertoire of GHs that would allow access to certain polysaccharides (**Table 3**). Twenty-one GH families were represented in CL500-3, with GH13 (glycosidase) being the only one found in all 6 MAGs (**Table 3**). Multiple copies of GH families 13, 29 (alpha-L-fucosidase), and 95 (alpha-L-fucosidase) were present in any single CL500-3 MAG (**Table 3**). GH5 (chitinase), GH35 (beta-galactosidase), GH37 (trehalase) and GH119 (alpha-amylase) were only found within the summer derived MAGs (MJ-june-47, MJ-aug-102,

SC-june-56, SC-aug-92), while GH63 (alpha-glucosidase) and GH133 (alpha-1,6-glucosidase) were only found in the winter derived MAGs (MJ-feb-125, SC-feb-99).

GH family	GH annotation (substrates)	MJ-feb-125	SC-feb-99	MJ-june-47	SC-june-56	MJ-aug-102	SC-aug-92
GH2	beta-mannosidase, beta-galactosidase	-	1	2	1	-	-
GH3	beta-glucosidase	-	1	-	-	-	1
GH5	endoglucanase, chitinase	-	-	-	2	1	-
GH9	endoglucanase	1	-	1	1	1	-
GH10	endo-1,4-beta-xylanase	1	-	-	1	1	1
GH13	Glycosidase, pullulanase,	1	4	2	2	2	1
GH15	Glucoamylase (glucan-1,4-alpha-glucosidase)	-	1	-	-	1	1
GH28	alpha-L-rhamnosidase	1	-	1	1	1	-
GH29	alpha-L-fucosidase, alpha-1,3/1,4-L-fucosidase	-	3	2	2	1	1
GH33	sialidase-1	-	1	1	-	-	-
GH35	beta-galactosidase, exo-beta-glucosaminidase	-	-	-	-	1	1
GH37	Neutral trehalase	-	-	-	-	-	1
GH38	alpha-mannosidase	1	1	1	1	-	1
GH57	alpha-amylase	-	1	-	1	-	-
GH63	alpha-glucosidase	-	1	-	-	-	-
GH74	endoglucanase, xyloglucanase	-	1	-	-	-	1
GH77	4-alpha-glucanotransferase	1	1	1	1	-	-
GH95	alpha-L-fucosidase, alpha-1,2-L-fucosidase	1	2	1	1	-	1
GH102	membrane-bound peptidoglycan lytic murein transglycosylase A	-	2	-	-	-	1
GH119	alpha-amylase	-	-	1	-	-	-
GH133	Glycogen debranching enzyme (alpha-1,6-glucosidase)	-	1	-	-	-	-
Total	-	**7**	**21**	**13**	**14**	**9**	**11**

Table 3.
Actual number of glycoside hydrolases found for various GH families in each CL500-3 MAG (A dash (-) means that a GH belonging to that family was not found in the MAG).

3.4.2 Metabolism of low-molecular weight labile compounds

Ecological differences in aquatic bacteria are often associated with variability in the use of labile compounds such as sugars and amino acids as growth substrates (Salcher, 2013). Various genes encoding multi-sugar ATP-binding cassette (ABC) transporters, a phosphotransferase system (PTS) ribose transport system and N-acetyl-D-glucosamine (chitin monomer) transporters were identified in all of the MAGs, with the majority observed in SC-feb-99. Glycerol-3-phosphate transporters (MJ-feb-125, SC-feb-99, MJ-june-47, SC-june-56, SC-aug-92) and glycerol-3-phosphate dehydrogenases (SC-feb-99, SC-june-56, MJ-aug-102) were also identified. Enzymes involved in sugar metabolism such as trehalose degradation, stachyose degradation and xylose degradation were represented in all 6 MAGs. Additionally, a number of amino acid transporters, as well as many amino acid degradation pathways (18/20) were identified in all MAGs. The observation of such an abundance of amino acid degradation

pathways suggests that CL500-3 may be especially adapted to the use of amino acids as growth substrates.

To investigate the significance of amino acid utilization in CL500-3, the diversity of carbohydrate and amino acid degradation pathways observed here was compared to other heterotrophic freshwater bacteria. The metabolic profiles of CL500-3 were compared to profiles from Planctomycetes reference genomes used in the concatenated protein phylogeny, a diversity of Verrucomicrobia MAGs previously described from Quebec lakes, and a selection of typical free-living freshwater bacteria. This analysis demonstrated that there was on average 1.6-fold more enzymes involved in amino acid degradation compared to carbohydrate degradation in the CL500-3 MAGs (**Table 4**). In contrast, the reference Planctomycetes genomes have on average 2-fold the number of enzymes involved in carbohydrate degradation compared to amino acid degradation. Additionally, a one-way ANOVA was performed to compare abundances of amino acid to carbohydrate degrading enzymes between CL500-3 and other Planctomycetes. A resulting p-value of < 0.05 (3.8×10^{-8}) was obtained, suggesting that amino acid utilization within CL500-3 is significant. The distribution was more diverse within the Verrucomicrobia MAGs which were enriched in either amino acid or carbohydrate degradation genes. Other free-living freshwater bacteria including species of Actinobacteria, Limnohabitans, and Polynucleobacter exhibit a higher number of amino acid degrading enzymes than carbohydrate degrading enzymes (**Table 4**).

	Genome	Genome size (bp)	Completeness (%)	# of Amino acid degradation enzymes	# of Carbohydrate degradation enzymes	Amino acid/ Carbohydrate degradation	# of Amino acid biosynthesis enzymes	# of Carbohydrate biosynthesis enzymes	Amino acid/ Carbohydrate biosynthesis
CL500-3 MAGs	MJ-feb-125	1233551	45.19	**26**	15	**1.7**	34	17	2.0
	SC-feb-99	2473776	73.24	**42**	27	**1.6**	51	42	1.2
	MJ-june-47	2002500	69.81	**41**	22	**1.9**	46	27	1.7
	SC-june-56	1593273	54.55	**35**	19	**1.8**	42	29	1.4
	MJ-aug-102	993108	40.65	**10**	7	**1.4**	17	12	1.4
	SC-aug-92	2071391	45.76	**36**	21	**1.7**	41	28	1.5
Reference Planctomycetes	Phycisphaera mikurensis	3884382	100	29	**75**	**0.4**	97	59	1.6
	Isosphaera pallida	5472964	100	46	**55**	**0.8**	102	66	1.5
	Gemmata obscuriglobus	9104366	100	52	**64**	**0.8**	117	76	1.5
	Gimesia maris	7776747	100	54	**88**	**0.6**	128	78	1.6
	Planctopirus limnophila	5423075	100	38	**52**	**0.7**	104	63	1.7
	Pirellula staleyi	6196199	100	41	**55**	**0.7**	105	56	1.9
	Rhodopirellula sallentina	8186686	100	42	**121**	**0.3**	110	71	1.5
	Rhodopirellula baltica	7145576	100	40	**99**	**0.4**	101	58	1.7
	Rhodopirellula europaea	7191307	100	43	**104**	**0.4**	104	60	1.7
Verrucomicrobia MAGs	CHTH-244	3240000	83.28	34	**35**	**1.0**	90	60	1.5
	CHTH-196	2790000	70.16	15	**24**	**0.6**	56	50	1.1
	opi-128	1850000	77.74	25	**35**	**0.7**	66	34	1.9
	opi-242	1190000	57.77	18	**19**	**0.9**	49	22	2.2
	opi-474	2810000	92.17	41	**43**	**1.0**	96	54	1.8
	Pedos-303	4470000	89.7	27	**40**	**0.7**	82	72	1.1
	Pedos-510	2510000	82.43	27	**30**	**0.9**	57	30	1.9
	Pedos-1123	2730000	95.05	**26**	22	**1.2**	82	45	1.8
	v1-33	4080000	78.24	32	**45**	**0.7**	64	32	2.0
	v1-115	4140000	82.07	**34**	23	**1.5**	95	43	2.2
	v1-159	2630000	95.92	23	**31**	**0.7**	**90**	58	1.6
	v1-690	2240000	70.61	12	**39**	**0.3**	**60**	27	2.2
	v1-1361	2250000	72.93	**23**	22	**1.0**	**76**	40	1.9
	xiphi-315	1420000	87.73	**24**	20	**1.2**	**53**	20	2.7
	xiphi 554	1660000	95.95	**26**	23	**1.1**	**53**	35	1.5
Freshwater bacteria	Actinobacterium sp. acAcidi	2690658	91.56	**60**	23	**2.6**	**95**	39	2.4
	Actinobacteria bacterium IMCC26077	1551612	100	**25**	12	**2.1**	**84**	25	3.4
	Limnohabitans sp. Rim47	2937972	100	**52**	10	**5.2**	**91**	50	1.8
	Limnohabitans sp. 63ED37-2	3372992	100	**66**	11	**6.0**	**104**	45	2.3
	Polynucleobacter necessarius STIR1	1560469	100	**16**	3	**5.3**	**71**	33	2.2
	Polynucleobacter sinensis MWH-HuW1	2317790	100	**30**	8	**3.8**	**96**	42	2.3
	Polynucleobacter yangtzensis MWH-JaK3	2044982	100	**23**	5	**4.6**	**82**	37	2.2
	Rhodoluna lacicola MWH-Ta8	1430433	100	28	**32**	**0.9**	**75**	31	2.4

Table 4.
Number of amino acid/carbohydrate degradation enzymes and amino acid/carbohydrate biosynthesis enzymes found within the CL500-3 MAGs, reference Planctomycetes genomes, Verrucomicrobia MAGs from the same Quebec lakes, and some other freshwater bacteria. The amino acid/carbohydrate degradation column depicts the ratio of amino acid to carbohydrate degradation enzymes for each genome. The amino acid/carbohydrate biosynthesis column depicts the ratio of amino acid to carbohydrate biosynthesis enzymes. Genome sizes for CL500-3 MAGs are not estimated, but rather actual MAG sizes.

3.4.3 Nitrogen, phosphorus and other DOM metabolisms

The success of aquatic bacteria often depends on their ability to compete for nitrogen and/or phosphorous that may be in limiting supply (Cotner, 2010). Although the evidence for use of amino acids by CL500-3 is strong, there is only weak genomic evidence for inorganic nitrogen acquisition. Typical ammonia transporters were not found, and although genes for a NitT/TauT transporter system for nitrate were present in both SC-aug-92 and SC-feb-99, an assimilatory

nitrate reductase was not identified. However, high substrate affinity nitrogen assimilation enzymes encoding the glutamine synthetase-glutamine oxoglutarate aminotransferase (GS-GOGAT) system were identified including an ATP dependent glutamine synthetase (GS), a NADH/NADPH dependent glutamate synthase, as well as a glutamate dehydrogenase (GDH). GS was found in MJ-aug-102, SC-aug-92, and in SC-feb-99, while a glutamate synthase was present in SC-aug-92 and SC-feb-99. GDH was found in all MAGs, except for MJ-aug-102. Additionally, various exoenzymes such as phosphatases and proteases, implicated in the hydrolysis of dissolved organic forms of nutrients, were found in all MAGs throughout the year. Moreover, a urea cycle and a urea transporter, potentially used to assimilate and recycle nitrogen, were found in MJ-june-47, MJ-feb-125, SC-feb-99, SC-june-56 and SC-aug-92.

Genes involved in phosphate transport and metabolism, such as an inorganic phosphate transport system, phosphate-selective porins O and P, alkaline phosphatase (*phoA*), phosphate starvation proteins (*phoH*) and a component of the high-affinity phosphate-specific transport system (*pstABC*) were identified within the MAGs. An alkaline phosphatase (*phoA*) was present in SC-june-56 and SC-aug-92, phosphate starvation proteins (*phoH*) were found everywhere except for in MJ-aug-102, and a component of the high-affinity phosphate-specific transporter system (*pstABC*) was discovered in the SC-feb-99 MAG. Phosphate-selective porins O and P were found in all MAGs except for SC-june-56. In fact, four phosphate-selective porins were found in SC-feb-99, and three were found in MJ-feb-125, MJ-june-47, MJ-aug-102 and SC-aug-92. Detection of an osmotic shock response regulator (*ompR*), which forms a two-component system with a phosphate regulon sensor histidine kinase (*phoR*), was also observed within all MAGs except for MJ-aug-102.

Additionally, various transporters and receptors implicated in the incorporation and recycling of dissolved matter such as exb, Ton, and Tol biopolymer and iron transporters, especially the TonB-dependent receptors (TBDR), were also present in each MAG.

3.4.4 Cell motility

Genes involved in bacterial motility including flagellar biosynthesis and chemotaxis genes were identified in all MAGs, mainly in the MAGs derived from summer metagenomes (MJ-june-47, MJ-aug-102, SC-june-56, SC-aug-92). Genes required for a functional flagellum

were observed, including *flgA-flgE*, *flgG-flgI*, *flgK-flgN*, *fliA*, *fliE-fliH*, *fliJ-fliR*, *fliW, fliY*, *flhA*, *flhB*, *flhF*, *flhG*, *motA*, and *motB*. These genes maintained a scattered presence throughout the MAGs. Chemotaxis genes including *cheA-cheD*, *cheR*, *cheW-cheY*, sensor histidine kinase and response regulators *wspA-wspF*, type II secretory pathway genes, pilus assembly proteins and methyl-accepting chemotaxis proteins, were also dispersed throughout the 6 MAGs.

3.5 CL500-3 gene expression patterns across Quebec lakes and seasons

To investigate gene expression patterns of the CL500-3 MAGs during ice-covered and open-water periods, a temporally overlapping metatranscriptomic time-series from Montjoie and Simoncouche was mapped to the MAGs, providing a view of CL500-3 transcriptional activity across lakes and seasons. Gene expression, reported as RPKB value, was observed for all MAGs and for the most part paralleled MAG distributions in the lakes (**Figure 2b**). MJ-jun-47, SC-jun-56, MJ-aug-102, and SC-aug-92 exhibited the highest transcript abundance during the summer time points. MJ-feb-125 exhibited the highest expression in Simoncouche during the winter, while the highest expression for SC-feb-99 was observed in Simoncouche during the winter and in Montjoie during the summer. In general, gene expression of SC-feb-99 was the highest, with 92% of genes being expressed, while expression of MJ-feb-125 was the lowest, with 18% of genes expressed. As for MJ-june-47, SC-june-56, MJ-aug-102, and SC-aug-92, expression was observed for 28, 47, 51 and 86% of genes respectively.

3.5.1 Gene expression profiling

Previously, the diversity of protein-encoding genes found within CL500-3 was described. In this following section, potential metabolisms occurring under the ice and during the open water season was determined by investigating gene expression profiles throughout the year. SC-feb-99 exhibited high gene expression during the winter in Lake Simoncouche, and during the summer in Lake Montjoie for protein-encoding genes involved in a myriad of KEGG metabolisms (**Figure 4)**, as well as enzymes involved in amino acid degradation, amino acid biosynthesis, and cell structure pathways. Compared to other metabolisms, increased expression was observed for transcripts involved in lipid metabolism, folding, sorting and degradation, replication and repair, and translation for MJ-feb-125, mainly during the winter in Simoncouche (**Figure 4)**. As for the MAGs derived from summer metagenomes (MJ-june-47, MJ-aug-102, SC-june-56, SC-aug-102), gene expression was observed mainly during the summer in both lakes. SC-aug-92 exhibited the highest expression for KEGG metabolisms in general compared

to the other summer-derived MAGs, particularly for protein encoding genes involved in cell motility, and several environmental and genetic information processes (**Figure 4**). In MJ-june-47 and SC-june-56, expression was the highest for genes involved in cell motility (**Figure 4**), cell structure and amino acid degradation. MJ-aug-102 presents a low expression profile in general, with the highest expression observed for translation during the summer in Simoncouche (**Figure 4**). Genes involved in cell motility, amino acid biosynthesis, and several environmental and genetic information processes (**Figure 4**), exhibited increased levels of expression in SC-aug-92.

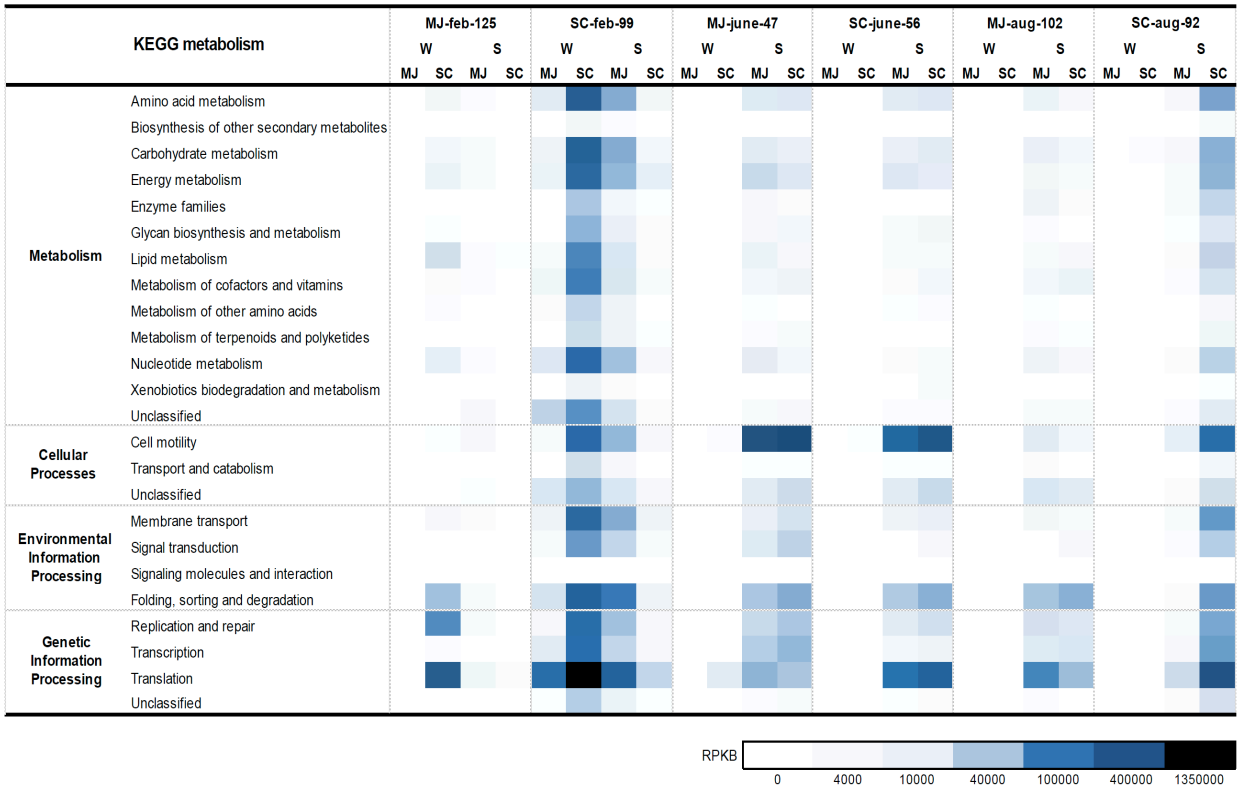

Figure 4.
Heatmap demonstrating the relative gene expression of various KEGG metabolisms found within the six CL500-3 MAGs. Relative gene expression was calculated by compiling the RPKB values of all genes involved in a specific metabolism. The resulting RPKB values for each metabolism were then organized by lake and season. W: winter samples, S: summer samples, MJ: Lake Montjoie, SC: Lake Simoncouche, MAGs: MJ-Feb-125, MJ-june-47, MJ-aug-102, SC-feb-99, SC-june-56, SC-aug-92.

3.5.2 Gene expression of saccharolytic metabolism

Expression of various GH families was detected in all MAGs, with a higher expression in general being observed for SC-feb-99, especially during the winter in Simoncouche (**Figure 5**). As expected, the MAGs derived from summer metagenomes (MJ-june-47, MJ-aug-102, SC-

june-56, SC-aug-92) contained GH genes that were transcriptionally active during the summer in both lakes, and the MAGs derived from winter metagenomes (MJ-feb-125, SC-feb-99) contain GH genes that were expressed during both the winter in Simoncouche and the summer in Montjoie. Expression of GH genes in MJ-feb-125 was low. GH families 9 and 13 were the most highly expressed in MJ-june-47, MJ-aug-102, SC-june-56, during the summer in both lakes, while GH15 (glucan-1,4-alpha- glucosidase), GH35, GH37 and GH102 (membrane-bound peptidoglycan lytic murein transglycosylase A) had a higher expression in SC-aug-92 during the summer in Lake Simoncouche. GH5 was also observed with high expression during the summer in SC-june-56, while GH5 and GH35 had a higher expression in MJ-aug-102. GH13 was the only one to be expressed in all MAGs. Although GH29 and GH95 were present in 5 of the 6 MAGs, they were mainly active in SC-feb-99. Expression of GH families 5, 10, 28, 35, 37, 119 was only observed for the summer derived MAGs (MJ-june-47, MJ-aug-102, SC-june-56, SC-aug-92), while expression of GH2, GH33, GH63, and GH133 was observed solely in the winter derived MAGs (MJ-feb-125, SC-feb-99) (**Figure 5**).

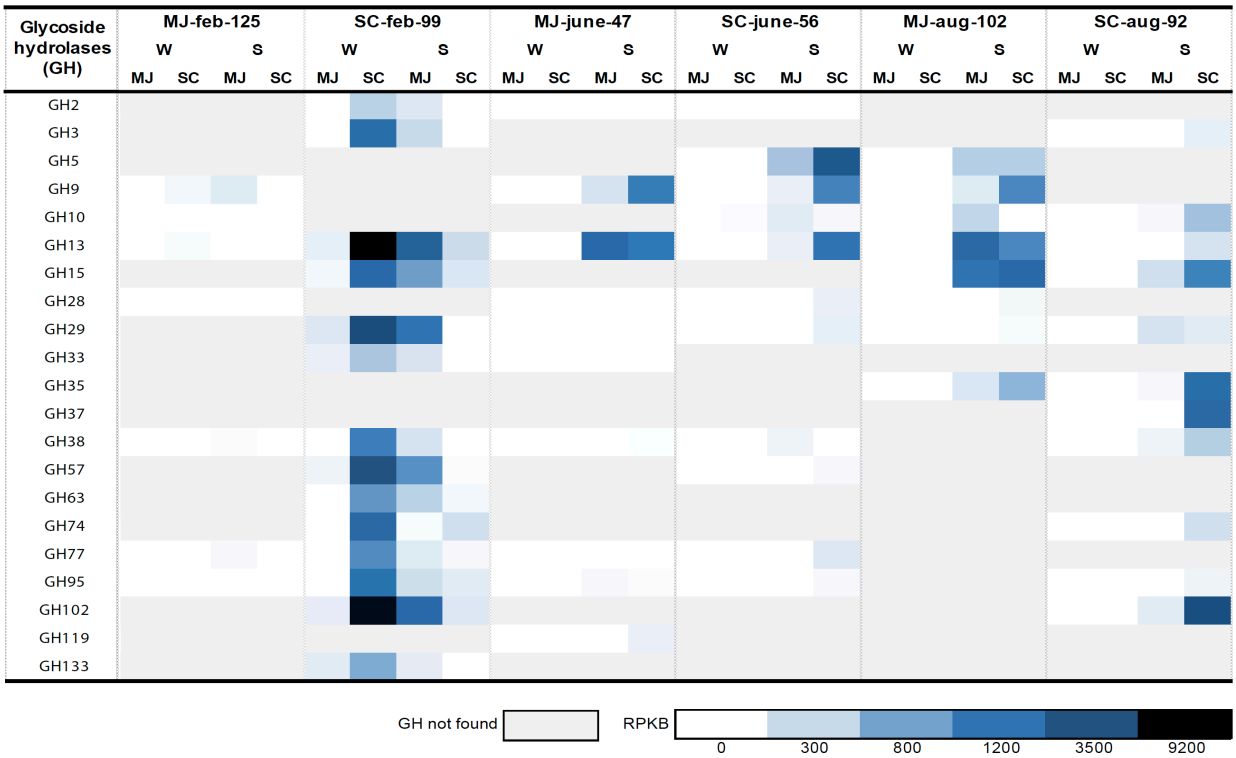

Figure 5.
Heatmap demonstrating the relative gene expression of each glycoside hydrolase family found within the six CL500-3 MAGs. Relative gene expression was calculated by compiling the RPKB values of all GHs affiliated to the same family. The resulting RPKB values for GH family were then organized by lake and season. W: winter samples, S: summer samples, MJ: Lake Montjoie, SC: Lake Simoncouche, MAGs: MJ-Feb-125, MJ-june-47, MJ-aug-102, SC-feb-99, SC-june-56, SC-aug-92.

3.5.3 Metabolism of low-molecular weight labile compounds

Expression was observed for enzymes involved in the degradation of specific sugars in SC-aug-92 and SC-feb-99 (≤ 5,630 RPKB). Genes encoding multi-sugar ATP-binding cassette (ABC) transporters were the most highly expressed in SC-feb-99, during the summer in Montjoie and the winter in Simoncouche, and in SC-aug-92 during the summer (≤ 10,555 RPKB). A phosphotransferase system (PTS) ribose transport system was transcriptionally active in MJ-june-47, SC-feb-99 and SC-aug-92, with the highest expression observed in SC-feb-99 during the summer in Montjoie and the winter in Simoncouche (≤ 6,604 RPKB). A transporter for N-acetyl-D-glucosamine (chitin monomer) was active in all Simoncouche MAGs (SC-feb-99, SC-june-56, SC-aug-92) and in MJ-june-47 (≤ 1,779 RPKB). Glycerol-3-phosphate transporters and glycerol-3-phosphate dehydrogenases were active in SC-aug-92 during the summer and SC-feb-99, during the summer in Montjoie and winter in Simoncouche (≤ 1,018 RPKB).

Additionally, the expression of enzymes involved in amino acid biosynthesis and amino acid degradation was observed throughout the 6 MAGs, with a higher expression on average in SC-aug-92 during the summer, and in SC-feb-99 during the summer in Montjoie and the winter in Simoncouche (≤ 14,437 RPKB). Furthermore, a number of amino acid transporters were active mainly in SC-feb-99 during the summer in Montjoie and winter in Simoncouche (≤ 2,424 RPKB). The expression of enzymes involved in amino acid degradation pathways provided further evidence that CL500-3 may be especially adapted to utilize amino acids as growth substrates.

3.5.4 Expression patterns of nitrogen, phosphorus and other DOM metabolisms

The majority of the enzymes and transporters involved in nitrogen metabolism were expressed throughout the year. Minimal relative expression of a nitrate transporter (NitT/TauT) was observed in SC-feb-99, during the summer in Montjoie and the winter in Simoncouche, and in SC-aug-92, during the summer in Simoncouche (≤ 2,147 RPKB). High expression of enzymes involved in the GS-GOGAT system, with the exception of glutamate dehydrogenases, was observed mostly in SC-feb-99 and SC-aug-92 (≤ 12,468 RPKB). Glutamate dehydrogenases were expressed in all MAGs except for MJ-aug-102 (≤ 5,476 RPKB). Additionally, various phosphatases and proteases, implicated in the hydrolysis of dissolved organic forms of nutrients,

were expressed in all MAGs throughout the year. Moreover, a urea cycle and a urea transporter, potentially used to assimilate and recycle nitrogen, were active in all MAGs except for MJ-aug-102, with expression being the highest in SC-feb-99 and SC-aug-92 (≤ 6,561 RPKB).

Expression of genes involved in phosphate transport and metabolism, such as alkaline phosphatase (*phoA*) was observed in the MJ-aug-102 and SC-aug-92 during the summer, while phosphate starvation proteins (*phoH*) were highly expressed in SC-feb-99, during the summer in Montjoie and the winter in Simoncouche (≤ 952 RPKB, ≤ 15,596 RPKB, respectively). A high-affinity inorganic phosphate transporter system (*pst*) was expressed in all Simoncouche MAGs (SC-feb-99, SC-june-56, SC-aug-92), MJ-june-47 and in MJ-aug-102 (≤ 11,324 RPKB). The most transcriptionally active protein in SC-feb-99 was a phosphate-selective porin O and P, with a value of 329,601 RPKB during the winter in Simoncouche and 1,002,896 RPKB during the winter in Montjoie. Three other phosphate selective porins were identified in SC-feb-99 during the winter in Simoncouche (≤ 14,631 RPKB, ≤ 17,097 RPKB, ≤ 30,298 RPKB). Out of three phosphate porins identified for MJ-feb-125, two of them showed minimal expression, while the other was highly expressed during the winter in Simoncouche (250,256 RPKB). MJ-june-47, MJ-aug-102 and SC-aug-92 contain three phosphate porins that were highly active in Montjoie and Simoncouche (≤ 8,987 RPKB, ≤ 50,793 RPKB, ≤ 79,892 RPKB, respectively). Additionally, the expression of an osmotic shock response regulator (*ompR)* was observed in all MAGs except for MJ-aug-102, with a higher expression in general being observed in SC-feb-99 during the winter in Simoncouche (≤ 3,614 RPKB).

The expression of various exb, Ton, and Tol biopolymer and iron transporters was detected, suggesting an alternative route for the incorporation and recycling of dissolved organic matter in these freshwater ecosystems (Tang *et al.*, 2012). Expression of five TolC outer membrane proteins was observed during the summer for the 4 summer derived MAGs (MJ-june-47, MJ-aug-102, SC-june-56, SC-aug-92), and during the summer in Montjoie and winter in Simoncouche for the SC-feb-99 MAG (≤ 4,028 RPKB). Expression of TolB and another component of the Tol biopolymer transport system was observed during the winter in Simoncouche in SC-feb-99 (2,587 RPKB, 1,863 RPKB, respectively). Four TonB proteins, which link inner and outer membranes, and TonB-dependent receptors (TBDRs) were expressed in SC-feb-99 throughout the year and in SC-aug-102 during the summer in Simoncouche (8,768

RPKB). Various biopolymer transport proteins exbB and exbD were expressed throughout the year, mainly during the summer in SC-aug-92 (≤ 6,997 RPKB). A few iron transport systems were also expressed in all MAGs, mainly in SC-feb-99 during the winter in Simoncouche and summer in Montjoie (≤ 3,197 RPKB, ≤ 759 RPKB, respectively).

3.5.5 Cell motility

Most flagellar biosynthesis and chemotaxis genes were expressed in M-june-47, SC-feb-99, SC-june-47, and SC-aug-92 with the highest expression observed during the winter in Simoncouche and the summer in Montjoie for SC-feb-99, and during the summer in Simoncouche for SC-aug-92. Expression of only a few motility genes was observed in MJ-feb-125 and MJ-aug-102. Flagellin, with an RPKB value of 46,478 exhibited the highest expression for SC-aug-92 during the summer in Simoncouche. The same gene exhibited an RPKB value of 79,105 during the winter in Simoncouche for SC-feb-99. Type II secretory pathway, pseudopilin PulG involved in bacterial secretion and cell motility was the most active gene found in SC-aug-92 during the summer in Lake Simoncouche, with an RPKB value of 111,226, and the second most active gene in SC-feb-99, with an RPKB value of 161,085 during the winter in Simoncouche. Type IV pilus assembly protein PilA was the most active gene found during the summer in both June MAGs (MJ-june-47, SC-june-56) with RPKB values of 444,921 and 215,287 respectively in Montjoie and 476,348 and 346,955 respectively in Simoncouche.

3.5.6 Expression of hypothetical proteins

Hypothetical proteins were among the most highly expressed in the CL500-3 MAGs. The genes with the highest expression in MJ-feb-125 and MJ-aug-102 were hypothetical proteins, with RPKB values of 1,220,089 during the winter in Simoncouche and 149,688 during the summer in Montjoie respectively. Furthermore, in SC-feb-99, there were 25 hypothetical proteins with expression values greater than 10,000 RPKB, with the highest expression during the winter in Simoncouche being 102,986 RPKB, and the highest expression during the summer in Montjoie being 10,426 RPKB. MJ-june-47 and SC-june-56 both encode hypothetical proteins as one of their most highly expressed proteins. Lastly, out of the top 44 most highly expressed proteins in SC-aug-92, 10 of them were annotated as hypothetical proteins.

4. Discussion

Fundamental questions remain to be answered concerning the diversity and functionality of communities of microorganisms recently discovered under the ice, as well as deciphering the metabolic traits and activities that differentiate the microorganisms present during winter and summer in seasonally ice-covered lakes. Previous analysis of 16S rRNA gene data from Quebec lakes revealed that bacterial community composition was distinct during ice-covered and open-water periods, suggesting differences in the metabolic strategies of the bacterial populations living under the ice (Khawasik, 2017). Moreover, a wide diversity of Verrucomicrobia strongly associated with ice-covered periods, along with a range of metabolic genes expressed during winter, were previously described in the same Quebec lakes (Tran *et al.*, 2018). These findings contribute to the emerging view that microorganisms are important but widely overlooked components of lake ecosystems under the ice. To better understand the contribution of winter communities to lake metabolism and nutrient cycling, year-round investigation of the metabolic traits and activities of freshwater microorganisms is warranted.

Based on this 16S rRNA data analysis, groups of bacteria found at higher relative abundances under the ice, such as Planctomycetes, were investigated. Planctomycetes, which are normally found in low abundance in freshwater lakes, demonstrated a high relative abundance in both open water and ice-covered periods of the year in Lake Simoncouche and Lake Montjoie in Quebec, Canada. As part of a larger initiative to better understand under-ice microbial metabolic diversity and activity in seasonally ice-covered lakes, Planctomycetes became the focus of this thesis. Investigations of microbial metabolism and gene expression patterns were conducted in seasonally ice-covered lakes using 16S rRNA sequence, metagenomics, metatranscriptomics analysis.

4.1 CL500-3 clade diversity and distribution in aquatic ecosystems

In depth phylogenetic analysis of Planctomycetes 16S rRNA sequences revealed a dominance of the uncultivated CL500-3 clade, which comprises five subclades (CL500-3a to CL500-3e). According to amplicon data, each subclade exhibited a seasonal and geographical distribution, suggesting that substantial diversity exists within this poorly described freshwater lineage. CL500-3a, -3c and -3e were generally associated with ice-cover, while CL500-3b had a tendency to inhabit the summer season. CL500-3d was dynamic and had an affinity for either an

ice-covered environment in Simoncouche, or an open water environment in Montjoie. Based on these findings, we can hypothesize that the subclades associated with ice-cover are cold-adapted. Although CL500-3 is uncultivated and a genome has yet to be sequenced, my global synthesis of available CL500-3 16S rRNA sequences lead to the discovery that these subclades were not specific to Quebec lakes. Alternatively, they were present in a wide range of global freshwater ecosystems, including Lake Baikal in Russia, Songhua River in China, Crater lake in Oregon, the Amazon river in Brazil, and various lakes in Germany, Spain and the Netherlands.

In addition, an association of CL500-3 MAGs with both open water and ice-covered conditions was revealed. Consequently, it appears that CL500-3 populations represented by our MAGs are resident members of the microbial community throughout the year. Investigations of metagenomes from Spanish reservoirs, Wisconsin Lakes, and under the ice in Lake Baikal revealed that winter CL500-3 MAGs are not detected outside of Quebec lakes. SC-june-56, MJ-june-47, and MJ-aug-102 are in fact summer populations and CL500-3 is detected in particle-associated size fractions in other systems as well. Based on the fact that a global synthesis of CL500-3 16S rRNA sequences revealed the presence of all subclades worldwide, and the fact that our CL500-3 MAGs are associated with other systems, our findings contribute to the emerging view that Planctomycetes, specifically CL500-3, is an important component of freshwater ecosystems. Knowledge of their genomic content and metabolic traits can contribute to our understanding of lake metabolism and nutrient cycling in aquatic ecosystems worldwide. Additionally, given their widespread distribution across freshwater ecosystems and their prominence under ice, CL500-3 may serve as a suitable model for investigating metabolic adaptations and lifestyle strategies associated with seasonally ice-covered lakes.

4.2 Lifestyle strategies

According to the amplicon data, the highest abundances of CL500-3 subclades were often observed in the particle-associated fraction compared to the free-living fraction in Lake Montjoie and Lake Simoncouche, suggesting that CL500-3 prefers to be attached to particles or that they form colonies with each other, as opposed to living freely in the water column. Although it seems that CL500-3 prefers a particle-associated lifestyle, they were also common in the free-living fractions of the community, indicating an ability to switch between these lifestyles. This is also supported by the fact that Planctomycetes lifestyles involve stagnant cells budding to form

flagellated motile cells, which swim for a few days before maturing and settling down to reproduce (Lindsay *et al.*, 2001). Maturation results in the loss of their single flagellum, as they synthesize pili along with a non-cellular stalk, enabling them to attach to surfaces or to each other, forming distinct rosettes (Tamura *et al.*, 2013; Newton *et al.*, 2011). This is consistent with their detection in particle-associated size fractions in previous studies (DeLong *et al.*, 1993; Crump *et al.*, 1999). During winter, cells may persist in a free-living state but also associate with phytoplankton during bloom onset and progression. Future studies comparing gene expression patterns between particle-attached and free-living cells would be informative in understanding the metabolic shifts that occur during these transient bloom events.

4.3 Year-round metabolic overview and gene expression of CL500-3

To investigate CL500-3 metabolism further, a metabolic overview depicting the major metabolisms and transporters discovered within all subclades of CL500-3 year-round was generated. Due to the lack of 16S rRNA gene data within the MAGs, and the absence of cultured representatives for CL500-3 subclades, it is extremely difficult to characterize the MAGs as specific subclades. Therefore, a composite map of all five subclades was used, which revealed a range of metabolic activities potentially occurring throughout the year.

Genes encoding proteins involved in metabolic pathways common for heterotrophic bacteria, such as glycolysis, the citric acid cycle, the pentose-phosphate pathway, and oxidative phosphorylation, were present and expressed in the CL500-3 complex. Additionally, genes involved in amino acid transport and metabolism, nitrogen metabolism, phosphorus metabolism, as well as cell motility, were observed.

4.4 Saccharolytic metabolism

Phytoplankton blooms, occurring during the open water or ice-covered period of the year, not only provide organic materials to the higher trophic food web, but also provide dissolved organic matter, and suspended organic matter for bacterial attachment and subsequent degradation by heterotrophic bacterial communities (Twiss *et al.*, 2012). Prior studies employing cultivation, genomics, metagenomics and radiolabeled DOM uptake assays have suggested the importance of Planctomycetes in polysaccharide degradation in aquatic and soil environments (Cottrell and Kirchman, 2000; Glöckner *et al.*, 2003; Weiner *et al.*, 2008; Bertilsson *et al.*, 2007;

Martinez-Garcia, 2012; Wang *et al.*, 2015; Orsi *et al.*, 2016). Additionally, a number of studies have implicated Planctomycetes in the degradation of phytoplankton-derived carbohydrates (Rabus et al., 2002; Zeigler *et al.*, 2012). The degradation of cell wall polysaccharides requires glycoside hydrolases (GH), which catalyze the initial step of converting high molecular weight polysaccharides into oligo- or monosaccharides that are sufficiently small to be transported into the cell for further processing (Martinez-Garcia, 2012).

The GH abundance and diversity within a genome may differ among individual organisms within a phylum and determines the width of the substrate spectrum and the complexity of the carbohydrates used by that organism (He *et al.*, 2017). When comparing the *Phycisphaera mikurensis* genome, which encodes 149 GH genes to any one of the 6 MAGs, which encode on average 32 GH genes, it can be inferred that the marine relative may be a specialist in polysaccharide degradation and may be able to use a wider range of more-complex polysaccharides than the CL500-3 populations. Through the comparison of GH genes found within relatively large Planctomycetes reference genomes (3.8 - 9.16 Mb) and CL500-3 subclades, it is apparent that although CL500-3 genomes are smaller, with estimated sizes of 2.4 – 4.5 Mb, they comprise only a fraction of the genes in question. This suggests that, unlike their closely related lineages, CL500-3 subclades are not as specialized for the degradation of carbohydrates.

A previous study demonstrated variations in the distribution of different GH categories among three different Planctomycetes strains, which raises the possibility that these bacteria have preferred niches defined by substrate availability (Kim *et al.*, 2016). Similarly, the Quebec MAGs exhibit variations in GH categories and abundances, suggesting that CL500-3 subclades also have preferred niches, adding evidence to their biogeographical and possibly seasonal diversity within Quebec lakes. Furthermore, various GH genes were identified and expressed throughout the year, suggesting that CL500-3a - CL500-3e interact with phytoplankton during both the summer and the winter.

Similar to Planctomycetes, a number of studies, including a previous study using our Quebec datasets, have implicated Verrucomicrobia in the degradation of phytoplankton-derived carbohydrates in both open water and ice-covered periods of the year in lakes (Paver and Kent,

2010; Parveen *et al.*, 2013; Bižić-Ionescu *et al.*, 2014; Beall *et al.*, 2016, Tran *et al.,* 2018). This along with numerous traits found in our Quebec lake datasets, suggests a capacity to couple Planctomycetes and Verrucomicrobia growth to phytoplankton, including winter phytoplankton blooms (Tran *et al.,* 2018). Further studies are needed in order to assess relationships between phytoplankton and the bacterial community under ice.

Additionally, pathways for the utilization and degradation of a variety of sugars, as well as various transporters for multiple sugars, were common in the CL500-3 complex and expressed during both open-water and under ice conditions. As many of these sugars are abundant carbohydrate monomers in plankton and plant cell walls, the presence of these pathways, together with that of GH genes, suggests that these CL500-3 populations may use plankton- and plant-derived saccharides. Furthermore, CL500-3 is able to utilize sugar alcohol carbon sources, since they have a glycerol-3-phosphate transporter and a glycerol-3-phosphate dehydrogenase (Cabello-Yeves *et al.*, 2017).

4.5 Metagenomic evidence for DON and DOM utilization

In both marine and freshwaters, many components of the dissolved organic nitrogen (DON) and dissolved inorganic nitrogen (DIN) pools, including ammonium, nitrate, and nitrite can play an active role in supplying nitrogen nutrition directly or indirectly to bacteria and phytoplankton and, in doing so, may affect the species composition of the ambient microbial community (Orsi *et al.,* 2016). DON supports a significant amount of heterotrophic production in aquatic ecosystems (Karl *et al.*, 2001; Bronk, 2002; Aluwihare and Meador, 2008; Orsi *et al.*, 2016). Yet, to date, the identity and diversity of microbial groups that transform DON are not well understood. DON consists of numerous low and high molecular weight (LMW, HMW) compounds including amino acids, *N*-acetyl amino polysaccharides, such as chitin and peptidoglycan, dissolved proteins and uncharacterized proteinaceous matter (Tanoue *et al.*, 1995; McCarthy *et al.*, 1997, McCarthy *et al.*, 1998; Aluwihare *et al.*, 2005; Orsi *et al.*, 2016). Many studies have demonstrated that microbial groups linked to the uptake of both high and low molecular weight DON, are involved in DON cycling (Cottrell and Kirchman, 2000; Kirchman, 2002; Malmstrom *et al.*, 2005; Alderkamp *et al.*, 2006; McCarren *et al*., 2010; Nikrad *et al.*, 2012; Dupont *et al.*, 2012; Rinta-Kanto *et al.*, 2012; Teeling *et al.*, 2012; Liu *et al.*, 2013; Nikrad *et al.*, 2014; Sharma *et al.*, 2014; Mayali *et al.*, 2015; Orsi *et al.*, 2016). LMW-DON such as

dissolved free amino acids (DFAA) released by protein hydrolysis are readily assimilated by bacteria in aquatic systems and can support up to 50% of bacterial production in oceans (Hobbie *et al.*, 1968; Kirchman, 2000; Orsi *et al.*, 2016). The importance of DFAA as sources of carbon, nitrogen and energy for the growth of both marine and freshwater heterotrophic bacteria has long been known (Zehr *et al.,* 1985; Gardner and Lee, 1975; Williams *et al.,* 1976; Wheeler and Kirchman, 1986; Jørgensen 1987; Simon & Rosenstock, 1992; Jørgensen *et al.,* 1993; Middelboe *et al.*, 1995; Gardner *et al.* 1998). Alternatively, a major source of HMW-DON results from phytoplankton cell death due to viral lysis, grazing, and passive diffusion (Fogg, 1971; Billen and Fontigny, 1987; Fuhrman, 1999; Worden *et al.*, 2015; Orsi *et al.*, 2016). The DON pool in natural waters is not inert and can be an important sink and source of nitrogen. There is a need for greater appreciation and understanding of the potential role of microbial communities as dynamic participants in DON cycling within aquatic ecosystems, particularly in freshwater environments.

From the metabolic reconstruction of the CL500-3 complex, amino acid transporters along with an abundance of genes involved in amino acid degradation and biosynthesis pathways were detected and expressed during the summer as well as during the winter suggesting an involvement of CL500-3 in the cycling of LMW and HMW-DON. In a previous marine study, researchers sought to better understand the organisms responsible for transforming HMW-DON in the upper ocean, therefore they isotopically labeled ^{13}C and ^{15}N from phytoplankton and added the protein extract to euphotic zone water from the Pacific Ocean (Orsi *et al.*, 2016). Orsi *et al.* have discovered diverse, uncultivated Planctomycetes assimilating protein secreted by phytoplankton, with the majority of OTUs being affiliated with the CL500-3 clade. Consequently, it can be hypothesized that CL500-3 consume polysaccharides and proteins derived from phytoplankton.

In addition to degrading particulate matter, CL500-3 may also act as a sink of HMW-DOM through the utilization of dissolved proteins. It is hypothesized that amino acids are transported from the water column into the cell, and subsequently degraded. The degradation of these amino acids results in ammonium and many other by-products that can be used for growth and survival of CL500-3 as well as other bacterial communities through nutrient remineralization. Nutrient remineralization, the release of nutrients in biologically available

forms into the water column, is accomplished via excretion or passive leakage of nutrients from bacterial or algal cells, cell lysis, and hydrolysis of dissolved organic forms of nutrients by exoenzymes produced by algae and bacteria such as phosphatases and proteases (Denef *et al.,* 2015). Various exoenzymes were observed and expressed throughout the CL500-3 MAGs, suggesting they take part in the hydrolysis of dissolved organic matter (DOM) and DON. The resulting ammonium produced by amino acid degradation is then released during bacterial remineralization of DOM in the water column (Goldman *et al*., 1987). To conclude, CL500-3 may be playing an important role in DON degradation and subsequent production of inorganic nitrogen to fuel primary production.

In order to further investigate the significance of nitrogen metabolism in CL500-3 throughout the year, carbohydrate and amino acid degradation pathways were compared between CL500-3 and other heterotrophic freshwater bacteria. This analysis demonstrated that there were on average 1.6-fold more enzymes involved in amino acid degradation compared to carbohydrate degradation in the CL500-3 MAGs. As expected, the reference Planctomycetes have on average 2-fold the number of enzymes involved in carbohydrate degradation compared to amino acid degradation, since they are known polysaccharide degraders. These results indicate that unlike other Planctomycetes genomes, CL500-3a to CL500-3e are not carbohydrate degradation specialists, but rather seem to be involved in amino acid degradation and protein cycling. Similarly, other free-living freshwater bacteria including species of Actinobacteria, Limnohabitans, and Polynucleobacter exhibit a higher number of amino acid degrading enzymes than carbohydrate degrading enzymes. This suggests and adds further evidence that like other free-living freshwater heterotrophs, CL500-3 is involved in the cycling of dissolved organic nitrogen in the water column. Additionally, through the assimilation and remineralization of proteins, CL500-3 can act as a nitrogen sink, supporting microbial metabolism throughout the year, especially in the winter.

Bacterial uptake is one of the major sinks of nitrogen in freshwater ecosystems and degradation as a result of bacterial activity probably accounts for the major flux of nitrogen out of the DON pool in most aquatic systems (Berman 2003, Orsi *et al.,* 2016). In the case of streams, rivers, lakes and freshwater reservoirs, much of the DON is often derived from terrestrial leaching and runoff (Berman 2003). Ice-cover results in the disruption of terrestrial

leaching and runoff, which subsequently results in a dramatic decrease of DON input into the ecosystem. Phytoplankton blooms occurring under the ice supplying DON, as well as heterotrophic bacteria, such as CL500-3 implicated in nutrient remineralization, are important in supporting other microbial communities and bacterial growth during the winter.

Generally, a relatively large amount of exb, Ton, and Tol biopolymer and iron transporters was detected. The presence and expression of these transporters suggests an alternative route for incorporation and recycling of dissolved organic matter in freshwater bodies, which was already described in marine ecosystems (Tang *et al.*, 2012). In addition, metabolic annotation revealed the presence of TonB-dependent receptors (TBDR). A previous study investigated the mechanism for which some groups of bacteria acquire HMW proteins, which led to the discovery of a HMW transmembrane transport mechanism. TonB-dependent receptor (TBDR) proteins transport HMW compounds that exceed the typical range of normal porins, by catalyzing high affinity transport of Ni-, Cu-, Fe-chelates, proteins, siderophores and polysaccharides across the outer membrane (Schauer et al., 2008). TBDRs have been implicated in the utilization and cycling of dissolved proteins by marine microbes and are important for microbial competition in the ocean, as cells with TBDR were observed to outcompete other microbes lacking TBDR in the presence of labile HMW-DOM released after a phytoplankton bloom in the North Sea (McCarren *et al.,* 2010; Teeling *et al.,* 2012; Orsi *et al.,* 2016). Evidence provided by Orsi *et al.,* suggests that TBDR receptors are a potential mechanism used by certain marine microbes for protein utilization, a finding that warrants further investigation, especially in freshwater ecosystems. Taken together, the presence and expression of TBDRs, exb, Ton, and Tol biopolymer and iron transporters in the winter and the summer suggests that CL500-3 is implicated in the cycling of DOM and DON throughout the year.

Further evidence that CL500-3 is implicated in the cycling and remineralization of DON as opposed to the utilization of inorganic nitrogen is demonstrated by a lack of transporters for inorganic nitrogen into the cell, particularly ammonia. Although a NitT/TauT transporter for nitrate was present, nitrate reductases were not, which impedes the cell from fully transforming nitrate to its usable form, ammonium. Hence, I postulate that CL500-3 does not rely on inorganic nitrogen and in fact may utilize DON as a sole source of nitrogen.

Additionally, almost every amino acid degradation pathway found within CL500-3 creates glutamine or glutamate as a by-product, which is then converted to 2-oxoglutarate by the glutamine oxoglutarate aminotransferase (GOGAT) system. The expression of genes encoding high substrate affinity nitrogen assimilation enzymes involved in the GS-GOGAT system was identified in all MAGs. This system, important for the incorporation of nitrogen into the cell material, is controlled by an ATP dependent glutamine synthetase (GS), allowing for the interconversion of glutamate and glutamine. GS coupled with an NADH/NADPH dependent glutamate synthase (a.k.a. GOGAT), converts 2-oxoglutarate (a.k.a. α–ketoglutarate) to glutamate. Glutamate dehydrogenase (GDH) can then catalyze the direct incorporation of ammonium from glutamate back into 2-oxoglutarate (a.k.a. α–ketoglutarate). 2-oxoglutarate and several other carbon substrates resulting from the degradation of amino acids such as fumarate, oxaloacetate and acetyl-CoA are then used as a source of energy to fuel the cell through the citric acid cycle.

4.6 Urea and metabolic interactions

Urea, produced by water column bacteria, is often a significant source of nitrogen and carbon for aquatic microbiota and phytoplankton (Glibert *et al.,* 1995; Cho *et al.,* 1996; Bronk *et al.,* 1998; Shaw *et al.,* 1998; Mitamura *et al.,* 2000; Berman 2003). Urea may be produced intracellularly through the urea cycle, the biosynthesis of putrescine and the degradation of urate, and in many bacteria, urea in the cell can be broken down by urease into NH^{4+} and CO_2 (Solomon, 2010). A urea cycle and a urea transporter are found and expressed in all MAGs except for MJ-aug-102, especially in Lake Simoncouche. However, a urease was not found in the CL500-3 MAGs, suggesting that rather than transporting the urea into the cell and using it as a source of nitrogen, CL500-3 is producing urea and excreting it to the water column through urea transporters. This can serve as a way to rid of excess nitrogen when other nutrients are limited. A recent study demonstrated through secretion experiments that nitrogen-containing compounds such as nitrite, ammonium, and urea were secreted in 73.5% of the analyzed cases, suggesting maintenance of an appropriate carbon-to nitrogen ratio in the cell (Pacheco *et al.*, 2019).

Heterotrophic bacteria can be both producers and consumers of urea (Jørgensen, 2006). The expression of urea utilization genes by Verrucomicrobia has been observed in Lake Montjoie and Lake Simoncouche, suggesting that Verrucomicrobia are consumers of urea (Tran

et al., 2018). Potential cross-feeding of urea between Verrucomicrobia and CL500-3 may be occurring, suggesting a mutualistic relationship between these communities. Additionally, mutualistic relationships may also be occurring between CL500-3 and phytoplankton. Although urea is usually present in lakes at ambient concentrations below 1 μM-N, it can contribute 50% or more of the total nitrogen used by planktonic communities (Solomon et al., 2010). Urea produced by CL500-3 and other heterotrophic bacteria, may subsequently be consumed by phytoplankton for growth. Moreover, due to decreased atmospheric and terrestrial nitrogen inputs into the water column during ice-cover, an ecological network between producers and consumers of urea can be beneficial and serve as an important nitrogen source and sink for bacterial species living under the ice. Urea may be a more important source of nitrogen nutrition for aquatic microbiota than has been previously reported. Further investigation of the role played by urea as a source of nitrogen in aquatic ecosystems is certainly warranted.

4.7 Phosphorus metabolism in CL500-3 MAGs

Heterotrophic organisms that engulf living and nonliving organic material often acquire nutrients such as nitrogen and phosphorus in excess of their requirements. The excess nutrients are excreted into the aquatic environment (Berman, 2003). However, questions are raised as to what happens when these nutrients are lacking, rather than present in excess. The C:N:P ratio is extremely important for the maintenance and proper growth of bacteria, and when phosphorus becomes the limiting factor, specific mechanisms for the adaptation to these cellular changes are employed (McCarthy *et al.,* 2016). CL500-3 populations represented by these MAGs may be able to grow and proliferate under low-phosphorus conditions within the cell, suggested by the presence and expression of genes responding to phosphate limitation. A few phosphate-selective porins O and P were identified with high levels of expression in all MAGs except for SC-june-56. The outer membrane of gram-negative bacteria functions as a selective permeability barrier between the cell and environment, which is mediated by the phosphate-selective porin O and P protein (Modi *et al.*, 2015). When carbon and nitrogen are abundant and in excess within the cell, phosphate may become the limiting factor. CL500-3 can then acquire phosphate from the water column by increasing porin activity, therefore increasing the amount of phosphate that can enter the cell.

Similar to nitrogen, a major source of phosphorus is from land run-off, which decreases drastically during ice-cover thus hindering phosphorus availability (Berman 2003). The highest expression of these phosphate-selective porins was observed during the winter in Lake Simoncouche, suggesting that CL500-3 was experiencing a high C:N:P ratio, indicating the cell was limited in phosphorus. Therefore, by increasing the expression of the phosphate porins, phosphate was acquired from the surrounding environment. In addition to maintaining the cell's integrity, this mechanism is advantageous to CL500-3 in the face of limited phosphorus inputs into the aquatic environment. Competition for the acquisition of phosphorus may arise between microbial communities, wherein CL500-3 would be prosperous. Moreover, the high expression of these porins would decrease phosphorus availability in the water column during the winter. However, in transitioning from winter ice-cover to spring, ice-melt will result in a large input of phosphorus from terrestrial and atmospheric sources (Seitzinger, 2005). Therefore, winter ice cover is a major force in determining limited phosphorus availability during the start of spring, which will then drastically increase by outside forces, stabilize during the summer, and possibly become depleted again during the winter due to decreased inputs and increased bacterial usage, demonstrating the cascading effects between seasons.

Additionally, the detection of an osmotic shock response regulator (*ompR*), which forms a two-component system with a phosphate regulon sensor histidine kinase (*phoR*), creates a protein phosphorylation cascade that allows for cell adaptation to environmental and intracellular changes (Ivanova, 2017). CL500-3 MAGs also contain phosphate starvation proteins (*phoH*), alkaline phosphatase, and inorganic phosphate transporters, which suggest that CL500-3, like freshwater Actinobacteria, are important in inorganic phosphorus sequestering which could contribute to prevent eutrophication (Ghai *et al.*, 2014).

4.8 Cell motility

Many flagellar biosynthesis and chemotaxis genes were detected, and mainly expressed in all Simoncouche derived MAGs (SC-feb-99, SC-june-56, SC-aug-92), and in the June derived Montjoie MAG (MJ-june-47). Although gene expression is the highest for the summer derived MAGs (MJ-june-47, SC-june-56, SC-aug-92) in both lakes, expression is also relatively high during the winter for SC-feb-99 in Simoncouche. Phytoplankton blooms and resource availability under ice are often patchy, therefore motility and chemotaxis may be useful adaptations to efficiently exploit hotspots of organic and inorganic nutrients (Stocker and

Seymour, 2012; Tran *et al.,* 2018). Since bacterial cells are operating at the micro-scale rather than meters, the ability to detect and respond to chemical gradients and the notion of motility towards hotspots of nutrients is advantageous.

Type II secretory pathway, pseudopilin PulG involved in bacterial secretion and cell motility was the most active gene found in SC-aug-92 during the summer in Lake Simoncouche, and the second most active gene in SC-feb-99 during the winter in Simoncouche. Type IV pilus assembly protein PilA was the most active gene found during the summer in both June MAGs Additionally, flagellin exhibited the highest expression for SC-aug-92 during the summer in Simoncouche. Expression of these genes, especially since they represent some of the highest expressed genes, suggests CL500-3 bacteria may be actively sensing and responding to both summer and winter nutrient pulses, such as phytoplankton blooms. Additionally, since expression was observed during the open water period as well as under the ice, I can hypothesize that CL500-3 is actively scavenging for nutrients throughout the year.

4.9 Concluding remarks

A key finding of CL500-3 metabolism is the unique implications of these bacteria in DON cycling and nutrient remineralization compared to other Planctomycetes genomes. CL500-3 may play an important role in nitrogen dynamics and may be involved in mutualistic relationships under the ice, especially with Verrucomicrobia. Planctomycetes and Verrucomicrobia, both found under the ice in our Quebec lakes, seem to inhabit distinct yet overlapping niches under the ice. While Planctomycetes and Verrucomicrobia are both implicated in saccharolytic metabolisms, CL500-3 are also specialists in amino acid degradation. Additionally, while CL500-3 are implicated in producing urea, Verrucomicrobia may act as a consumer (Tran *et al.,* 2018).

More information is needed to elucidate the eco-physiological characteristics of these widespread and abundant, yet largely overlooked bacterial lineages. Despite their relevance in global nutrient cycles, Planctomycetes are still largely undersampled. It is only within the past decade that access to more efficient and powerful molecular tools has enabled systematic exploration of the diverse and largely uncultivable microbial communities that typically populate lakes (Newton *et al.,* 2011). Within this time, only a handful of studies have paid any attention to

microbial communities during the ice-covered period of the year. There is a growing acknowledgment among aquatic ecologists as to the need to study the full annual cycle of lakes in order to understand lake dynamics, particularly as lake temperature increases worldwide (O'Reilly *et al.*, 2015). In addition to a few other studies, our findings substantiate the idea that understanding the under-ice microbiome is important for predicting the dynamics of seasonally ice-covered lakes in the future (Bertilsson *et al.*, 2013; Tran *et al.*, 2018). Although this study was illuminating regarding summer and winter associated patterns of CL500-3 and shows that CL500-3 is indeed proliferating and contributing to nutrient cycling under the ice, a more complete picture of CL500-3 and other uncultured bacterial lineages, and a much wider range of studies must be executed before generalizable patterns can be reported. Moreover, meta-omics studies such as this are, in essence, hypothesis-generating. Future directions for research on CL500-3 and microbial communities in general include targeted enrichment/cultivation and in situ rate measurement-based approaches, which will validate and quantify microbial metabolic contributions to nutrient cycling in lake environments throughout the year. Together, this information will improve our ability to predict and model freshwater ecosystem and biogeochemical dynamics and to determine their role in the landscapes in the face of environmental change.

5. Supplementary Results

5.1 Additional metabolic features of the CL500-3 clade

5.1.1 Stress response

Every living organism must cope with environmental changes that may represent stress situations, including temperature changes, nutrient availability or oxidative injury (Requena, 2012). Various genes involved in stress response were found within the CL500-3 genomes including, *dnaK*, *dnaJ*, *hscA*, *groEL*, *groES*, *hflX*, *clpA-clpC, clpX*, *uspA,* heat shock protein (*hsp20*), *catalase-peroxidase*, cold shock protein (*cspA*), periplasmic stress (*skp, degS, degQ*), ferric uptake (*fur*), and selenium uptake (*dedA*). Enzymes involved in the reactive oxygen species (ROS) degradation pathway were also present in all MAGs except for MJ-feb-125.

5.1.2 DNA repair

From microbes to multicellular eukaryotic organisms, all cells contain genes and pathways responsible for genome maintenance (Lenhart *et al.,* 2012). Several genes involved in DNA repair including *uvrB/uvrc*, *photolyase*, *lexA*, *rmuC*, *radA/Sms*, *ruvA-ruvC*, *recA*, *recG*, *recJ*, *recO*, *recQ*, *recR*, *oraA/recX*, *mutL*, *mutS*, *uracil-DNA glycosylase*, and *pcrA*, were dispersed throughout the MAGs. Additionally, a large-conductance mechanosensitive channel (MscL) was also identified in MJ-june-47 and SC-feb-99.

5.1.3 Cell structure

Many enzymes involved in the generation of the cellular wall were identified. Peptidoglycan, a biopolymer in the bacterial cell wall, is built from alternating units of N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc). The genes required to synthesize these subunits were found in the MAGs. Diaminopimelic acid (DAP), often found in the peptide linkage chains that make up the cell wall of gram-negative bacteria, was also found in all MAGs. Peptidoglycan biosynthesis and transport enzymes such as *murA-murG*, *murJ*, and *murQ*, *mviN*, *ldcA*, *mpaA*, were found mainly in MJ-feb-125, MJ-june-47, SC-feb-99, and SC-june-56. Enzymes involved in peptidoglycan degradation were mostly found in MJ-feb-125, SC-feb-99. The presence of genes involved in the transport and biosynthesis of lipopolysaccharides (LPS) and lipid A*,* a hydrophobic anchor for outer membrane lipopolysaccharides in gram-negative bacteria, were detected with no apparent patterns in the 6 MAGs. Additionally, membrane proteins involved in the export of O-antigen and teichoic acid were detected in all

MAGs. Enzymes implicated in a polymyxin resistance pathway were identified in MJ-june-47, SC-june-56 and SC-aug-92.

5.2 Gene expression of additional metabolic features

5.2.1 Stress response

Generally, a higher expression was observed for genes involved in stress response in SC-feb-99 during the summer in Montjoie and during the winter in Simoncouche (≤ 51,458 RPKB, ≤ 122,844 RPKB, respectively). High expression was also observed for MJ-june-47 and MJ-aug-102 during the summer, in both lakes (≤ 49,953 RPKB, ≤ 47,629 RPKB, respectively**)**. Enzymes involved in reactive oxygen species (ROS) degradation metabolisms were active in all MAGs except for MJ-feb-125.

5.2.2 DNA repair

Furthermore, high expression was observed for genes involved in DNA repair in SC-feb-99, during the summer in Montjoie and during the winter in Simoncouche (≤ 31,289 RPKB, ≤ 7,828 RPKB, respectively). High expression for DNA repair genes was also observed in SC-june-56 and SC-aug-92 during the summer in Simoncouche (≤ 49,199 RPKB, ≤ 10,306 RPKB, respectively). Expression of a large-conductance mechanosensitive channel (MscL) was observed in MJ-june-47 during the summer and SC-feb-99 during the winter in Simoncouche and the summer in Montjoie (1,153 RPKB, 6,547 RPKB, 1,114 RPKB, respectively).

5.2.3 Cell structure

Peptidoglycan biosynthesis and transport enzymes are observed with higher expression values on average in MJ-june-47, SC-june-56 and SC-aug-92 during the summer in both lakes **(**≤ 3,667 RPKB), and in SC-feb-99 during the winter in Simoncouche and the summer in Montjoie **(**≤ 9,331 RPKB, ≤ 1,410 RPKB, respectively). Enzymes involved in peptidoglycan degradation were mostly expressed in Lake Simoncouche during the winter for the SC-feb-99 MAG **(**≤ 7,906 RPKB). MurNAc, GlcNAc and DAP, components of the bacterial cell wall, all exhibited higher expression values in MJ-june-47, SC-aug-92 and SC-june-56 during the summer, and in SC-feb-99 during the winter in Simoncouche and the summer in Montjoie. The expression of genes involved in the transport and biosynthesis of lipopolysaccharides (LPS) and lipid A, were observed in SC-aug-92 during the summer in both lakes and SC-feb-99 during the winter in Simoncouche **(**≤ 3,227 RPKB, ≤ 3,743 RPKB, respectively). Additionally, genes involved in the

biosynthesis and export of O-antigen and teichoic acid were also observed in SC-aug-92 during the summer in both lakes and SC-feb-99 during the winter in Simoncouche (≤ 2,016 RPKB, ≤ 5,962 RPKB, respectively). Expression of enzymes implicated in a polymyxin resistance pathway were expressed in all MAGs, with higher expression values observed in SC-june-56, SC-aug-92 during the summer in both lakes and SC-feb-99 during the winter in Simoncouche (≤ 1,314 RPKB, ≤ 730 RPKB, respectively).

6. Supplementary Discussion

6.1 Additional metabolic features

The bacterial stress response enables bacteria to survive adverse and fluctuating conditions in their immediate surroundings (Eliora, 2012). Genes potentially involved in the adaptation to environmental stress are detected and expressed throughout the year in the CL500-3 genomes. In addition, CL500-3 genomes harbor many mechanisms to repair DNA and reduce the damaging effects of stress and UV radiation. DNA replication allows for the faithful duplication of the genome, whereas DNA repair pathways preserve DNA integrity in response to damage originating from endogenous and exogenous sources (Lenhart *et al.,* 2012). A large-conductance mechanosensitive channel (MscL) was also identified and expressed in the complex, which forms a channel in the inner membrane that has been shown to open and close its gate in response to tension in the lipid bilayer, thus is believed to protect cells from lysis upon osmotic shock (Sukharev, 2001).

6.2 Cellular characteristics

Peptidoglycan, a structure that controls the shape and integrity of almost all bacterial cells, has been a controversial topic when it comes to Planctomycetes. Many years ago, studies proposed that Planctomycetes are environmental peptidoglycan-less organisms (König *et al.*, 1984; Liesack *et al.*, 1986; Fuerst, 1995; Ward *et al*., 2000; McCoy and Maurelli, 2006; Cayrou *et al.*, 2010). However**,** in 2011, Fuerst decided to test this notion by employing modern bioinformatic approaches. In all bacteria, peptidoglycan consists of alternating residues of b-1,4-linked amino sugars GlcNAc and MurNAc, which are cross-linked via peptide chains by DAP (White, 2007). Fuerst demonstrates that Planctomycetes genomes harbour the genes required for peptidoglycan synthesis. Other studies have also been conducted in order to detect the components of peptidoglycan in Planctomycetes. Jogler *et al.* employed a radioactive kinase

assay to analyze cells of three representative Planctomycetes species, *Planctomyces limnophilus, Gemmata obscuriglobus* and *Rhodopirellula baltica*, and subsequently discovered that all three genomes contained the materials needed to synthesize peptidoglycan (2012). Other researchers have also provided evidence for the existence of peptidoglycan in Planctomycetes (Jeske *et al.,* 2015; van Teeseling *et al.,* 2015). In our study, I confirm that CL500-3 has the metabolic potential to synthesize peptidoglycan, as well as recycle the sugars found within its backbone.

A further topic concerning Planctomycetes has also been under scrutiny in the microbial world. In 2005, Fuerst stated that Planctomycetes are clearly divergent from both Gram-negative and Gram-positive cells and they do not have an outer membrane. However, in 2011, evidence strongly supporting the 'classical' Gram-negative organization of the Planctomycetes cell wall was proposed. Fuerst *et al.* provided evidence supported by cryo-electron tomography (CET), that Planctomycetes possess an outer membrane comparable to that of other Gram-negative bacteria (Fuerst *et al.,* 2011). Furthermore, lipopolysaccharide (LPS) biosynthesis pathways have been identified in Planctomycetes, providing evidence that they are gram-negative (Sutcliffe, 2010). I provide evidence supporting CL500-3 as being true gram-negative bacteria such as, the presence and expression of genes involved in the transport and biosynthesis of LPS, lipid A, and outer membrane proteins. Additionally, a polymyxin resistance pathway is identified providing further evidence that CL500-3 possess gram-negative like features. Polymyxins will attack lipid A molecules and cause them to lyse, therefore since CL500-3 is resistant, the lipid A will stay intact, and the bacterial cell will retain gram-negative-like features.

6.3 Concluding remarks

To conclude, CL500-3 demonstrates signs of possessing a peptidoglycan layer within the cell wall. In addition, integral components of peptidoglycan, along with ample evidence of outer membrane proteins were observed and expressed in both summer and winter. As expected, stress response genes and DNA repair mechanisms are reported throughout the year as well. Overall, our study supports the increasing recognition that the winter period, in addition to the summer period, represents a dynamic and metabolically important period for lakes despite ice cover. The Planctomycetes phylum is diverse, with different species filling various niches and environments. Further investigations and experiments, either *in vitro*, or *in silico*, are needed in order to investigate the cell wall along with cellular maintenance mechanisms in Planctomycetes, and update the information provided about previously described lineages.

References

Adrian, R., Wilhelm, S., Gerten, D. (2006). Life-history traits of lake plankton species may govern their phenological response to climate warming. *Global Change Biology*, 12 (4), 652-661.

Alderkamp, A., Sintes, E., Herndl, G. (2006). Abundance and activity of major groups of prokaryotic plankton in the coastal North Sea during spring and summer. *Aquatic Microbial Ecology*, 45 (3), 237-246.

Aluwihare, L.I., Repeta, D.J., Pantoja, S., Johnson, C.G. (2005). Two chemically distinct pools of organic nitrogen accumulate in the ocean. *Science* 308: 1007–1010. doi:10.1126/science.1108925

Aluwihare, L., Meador, T. (2008). Chemical Composition of Marine Dissolved Organic Nitrogen. *Nitrogen in the Marine Environment*. Academic Press: San Diego, CA, USA, pp 95–133. doi:10.1016/B978-0-12-372522-6.00003-7.

Antia, N.J., Harrison, P.J., Oliveira, L. (1991) Phycological reviews: the role of dissolved organic nitrogen in phytoplankton nutrition, cell biology, and ecology. *Phycologia* 30: 1–89

Beall, B.F.N., Twiss, M.R., Smith, D.E., Oyserman, B.O., Rozmarynowycz, M.J.,Binding, C.E., *et al.* (2016) Ice cover extent drives phytoplankton and bacterial community structure in a large north-temperate lake: implications for a warming climate. *Environmental Microbiology*, 18: 1704–1719.

Benson, B.J., Magnuson, J.J., Jensen, O.P., Card, V.M., Hodgkins, G., Korhonen, J., *et al.* (2012). Extreme events, trends, and variability in northern hemisphere lake-ice phenology (1855-2005). *Climatic Change: An Interdisciplinary, International Journal Devoted to the Description, Causes and Implications of Climatic Change*, 112 (2), 299-323. doi:10.1007/s10584-011-0212-8

Berman, T., Bronk, D.A. (2003). Dissolved organic nitrogen: A dynamic participant in aquatic ecosystems. *Aquatic Microbial Ecology*, 31, 279-305.

Bertilsson, S., Eiler, A., Nordqvist, A., & Jørgensen, N. (2007). Links between bacterial production, amino-acid utilization and community composition in productive lakes. *The ISME Journal*, 1 (6), 532-544.

Bertilsson, S., Burgin, A., Carey, C., Fey, S., Grossart, H., Grubisic, L., *et al.* (2013). The under-

ice microbiome of seasonally frozen lakes. *Limnology and Oceanography*, 58 (6), 1998-2012. doi:10.4319/lo.2013.58.6.1998

Bienhold, C., Pop, R., Wenzhöfer, F., Dittmar, T., Boetius, A. (2013). How deep-sea wood falls sustain chemosynthetic life. *PLoS One*, 8 (1), 53590. doi:10.1371/journal.pone.0053590

Billen, G., & Fontigny, A. (1987). Dynamics of a Phaeocystis-dominated spring bloom in Belgian coastal waters. II. Bacterioplankton dynamics. *Marine Ecology Progress Series*, 37 (2-3), 249-257.

Bižić-Ionescu, M., Amann, R., and Grossart, H.P. (2014) Massive regime shifts and high activity of heterotrophic bacteria in an ice-covered lake. *PLoS One*, 9: 1–17.

Blenckner, T., Adrian, R., Livingstone, D.M., Jennings, E., Weyhenmeyer, G.A., George, D.G. *et al.* (2007). Large-scale climatic signatures in lakes across Europe: A meta-analysis. *Global Change Biology*, 13 (7), 1314-1326.

Bronk, D.A., Glibert, P.M., Malone, T.C., Banahan, S., Sahlsten, E. (1998) Inorganic and organic nitrogen cycling in Chesapeake Bay: autotrophic versus heterotrophic processes and relationships to carbon flux. *Aquatic Microbial Ecology*, 15:177–189.

Bronk, D., (2002). Dynamics of DON in Biogeochemistry of Marine Dissolved Organic Matter. *Academic Press*: San diego, CA, USA, pp 153–247.

Cabello-Yeves, P. J., Ghai, R., Mehrshad, M., Picazo, A., Camacho, A., Rodriguez-Valera, F. (2017). Reconstruction of Diverse Verrucomicrobial Genomes from Metagenome Datasets of Freshwater Reservoirs. *Frontiers in Microbiology*, 8, 2131. doi:10.3389/fmicb.2017.02131

Cáceres, C.E. Schwalbach, M.S. (2001). How well do laboratory experiments explain field patterns of zooplankton emergence? *Freshwater Biology*, 46 (9), 1179-1189. doi:10.1046/j.1365-2427.2001.00737.x

Cayrou, C., Raoult, D., Drancourt, M. (2010). Broad-spectrum antibiotic resistance of Planctomycetes organisms determined by E test. *The Journal of Antimicrobial Chemotherapy,* 65 (10), 2119-22. doi:10.1093/jac/dkq290

Cho, B., Azam, F. (1988). Major role of bacteria in biogeochemical fluxes in the ocean's interior. *Nature*, 332 (6163), 441-443. doi:10.1038/332441a0

Cho, B., Park, M., Shim, J., Azam, F. (1996) Significance of bacteria in urea dynamics in coastal surface waters. *Marine Ecology Progress Series*, 142:19–26.

Cotner, J.B., Hall, E.K., Scott, J.T., Heldal, M. (2010). Freshwater Bacteria are

Stoichiometrically Flexible with a Nutrient Composition Similar to Seston. *Frontiers in Microbiology,* 1: 132. doi:10.3389/fmicb.2010.00132

Cottrell, M., & Kirchman, D. (2000). Natural assemblages of marine proteobacteria and members of the Cytophaga-Flavobacter cluster consuming low- and high-molecular-weight dissolved organic matter. *Applied and Environmental Microbiology*, 66 (4), 1692-1697.

Crump, B.C., Armbrust, E., Baross, J.A. (1999). Phylogenetic analysis of particle-attached and free-living bacterial communities in the Columbia River, its estuary, and the adjacent coastal ocean. *Applied and Environmental Microbiology*, 65 (7), 3192-3204.

Dedysh, S. (2011). Cultivating Uncultured Bacteria from Northern Wetlands: Knowledge Gained and Remaining Gaps. *Frontiers in Microbiology*. 2. 184. doi:10.3389/fmicb.2011.00184.

DeLong, E.F., Franks, D.G., Alldredge, A.L. (1993). Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnology and Oceanography*. 38: 924–934. doi: 10.4319/lo.1993.38.5.0924

DeLong, E., Lanoil, B., & Giovannoni, S. (1998). Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales. *Applied and Environmental Microbiology*, 64 (8), 3075-3078.

Denef, V.J., Mueller, R.S., Chiang, E.Y., Liebig, J.R., Vanderploeg, H.A. (2015). Chloroflexi CL500-11 Populations That Predominate Deep-Lake Hypolimnion Bacterioplankton Rely on Nitrogen-Rich Dissolved Organic Matter Metabolism and C1 Compound Oxidation. *Applied and Environmental Microbiology*, 82 (5), 1423-1432. doi: 10.1128/AEM.03014-15

Devos, D.P., Christian, J., John, A. (2013). The 1st EMBO workshop on PVC bacteria—Planctomycetes-Verrucomicrobia-Chlamydiae superphylum: Exceptions to the bacterial definition? *Antonie Van Leeuwenhoek*, 104 (4), 443-449. doi:10.1007/s10482-013-0026y

Dupont, C., Rusch, D., Yooseph, S., Lombardo, M., Alexander Richter, R., Valas, R., *et al.* (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *The ISME Journal*, 6 (6), 1186-1199. doi:10.1038/ismej.2011.189

Eddy S. R. (2011). Accelerated Profile HMM Searches. *PLoS Computational Biology*, 7 (10), e1002195.

Edgar R.C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32 (5), 1792-1797. doi:10.1093/nar/gkh340

Edgar, R.C., Haas, B., Clemente, J., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, 27 (16), 2194-2200.

Erbilgin, O., McDonald, K., & Kerfeld, C. (2014). Characterization of a Planctomycetal organelle: A novel bacterial microcompartment for the aerobic degradation of plant saccharides. *Applied and Environmental Microbiology*, 80 (7), 2193-2205. doi:10.1128/AEM.03887-13

Euzéby, J.P. (1997). List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet. *International Journal of Systematic and Evolutionary Microbiology*. 47: 590–592. http://dx.doi.org/10.1099/00207713-47-2-590.

Farmer, D.M. (1975). Penetrative convection in the absence of mean shear. *Quarterly Journal of the Royal Meteorological Society*, 101 (430), 869-891. doi:10.1002/qj.49710143011

Fogg, G.E. (1971). The extracellular products of algae in freshwater. *Archives Hydrobiology.* 5: 1–25.

Fuerst, J. A. (1995). The Planctomycetes: Emerging models for microbial ecology, evolution and cell biology. *Microbiology*. (13500872), 141 (7), 1493–1506.

Fuerst, J.A., Sagulenko, E. (2011). Beyond the bacterium: Planctomycetes challenge our concepts of microbial structure and function. *Nature Reviews Microbiology*, 9 (6), 403-413. doi:10.1038/nrmicro2578

Fuhrman, J. A. (1999). Marine viruses and their biogeochemical and ecological effects. *Nature*, 399 (6736), 541-548.

Fukunaga, Y., Kurahashi, M., Sakiyama, Y., Ohuchi, M., Yokota, A., Harayama, S. (2009). Phycisphaera mikurensis gen. nov., sp. nov., isolated from a marine alga, and proposal of Phycisphaeraceae fam. nov., Phycisphaerales ord. nov. and Phycisphaerae classis nov. in the phylum Planctomycetes. *The Journal of General and Applied Microbiology*, 55 (4), 267-75.

Gardner, W.S., Lee, G.F. (1975) The role of amino acids in the nitrogen cycle of Lake Mendota. *Association for the Sciences of Limnology and Oceanography*, 20:379–388.

Gardner, W.S., Cavaletto, J.F., Bootsma, H.A., Lavrentyev, P.J., Tanvone, F. (1998) Nitrogen cycling rates and light effects in tropical Lake Maracaibo, Venezuela. *Association for the Sciences of Limnology and Oceanography,* 43: 1814–1825.

Gerten, D., Adrian, R. (2000). Climate-driven changes in spring plankton dynamics and the

sensitivity of shallow polymictic lakes to the North Atlantic oscillation. *Limnology and Oceanography*, 45(5), 1058-1066. doi:10.4319/lo.2000.45.5.1058

Ghai, R., Mizuno, C., Picazo, A., Camacho, A., Rodriguez-Valera, F. (2014). Key roles for freshwater actinobacteria revealed by deep metagenomic sequencing. *Molecular Ecology*, 23 (24), 6073-6090. doi:10.1111/mec.12985

Glibert, P.M., Conley, D.J., Fisher, T.R., Harding, L.W., Malone, T.C. (1995) Dynamics of the 1990 winter/spring bloom in Chesapeake Bay. *Marine Ecology Progress Series*, 122:27–43.

Glöckner, F.O., Kube, M., Bauer, M., Teeling, H., *et al.* (2003). Complete genome sequence of the marine Planctomycete Pirellula sp. Strain 1. *Proceedings of the National Academy of Sciences of the United States of America*. 100. 8298-8303. doi:10.1073/pnas.1431443100.

Goldman, J. C., Caron, D. A., Dennett, M. R. (1987). Regulation of gross growth efficiency and ammonium regeneration in bacteria by substrate C:N ratio. *Association for the Sciences of Limnology and Oceanography*, 32 1239-1252.

Gray J.P., Herwig R.P. (1996). Phylogenetic analysis of the bacterial communities in marine sediments. *Applied and Environmental Microbiology*, 62 (11), 4049-4059.

Grossart, H., Simon, M. (2007). Interactions of planktonic algae and bacteria: Effects on algal growth and organic matter dynamics. *Aquatic Microbial Ecology*, 47 (2), 163-176.

Hampton, S.E., Galloway, A.W., Powers, S.M., Ozersky, T., Woo, K.H., Batt, R.D., *et al.* (2017) Ecology under lake ice. *Ecology Letters*. 20: 98–111. https://doi.org/10.1111/ele.12699

He, S., Stevens, S., Chan, L. K., Bertilsson, S., Del Rio, T.G., Tringe, S. G., *et al.* (2017). Ecophysiology of Freshwater Verrucomicrobia Inferred from Metagenome-Assembled Genomes. *mSphere*, 2(5), e00277-17. doi:10.1128/mSphere.00277-17

Hobbie, J.E., Crawford, C.C., Webb, K.L. (1968). Amino acid flux in an estuary. *Science*. 159: 1463–1464. doi:10.1126/science.159.3822.1463

Hyatt, D., LoCascio, P., Hauser, L., & Uberbacher, E. (2012). Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* (oxford, England), 28 (17), 2223-2230. doi:10.1093/bioinformatics/bts429

Ivanova, A. A., Naumoff, D. G., Miroshnikov, K. K., Liesack, W., Dedysh, S. N. (2017). Comparative Genomics of Four Isosphaeraceae Planctomycetes: A Common Pool of Plasmids and Glycoside Hydrolase Genes Shared by Paludisphaera borealis PX4T,

Isosphaera pallida IS1BT, Singulisphaera acidiphila DSM 18658T, and Strain SH-PL62. *Frontiers in Microbiology*, 8, 412. doi:10.3389/fmicb.2017.00412

Jenkins, C., Staley, J.T. (2013). Planctomycetes: Cell Structure, Origins and Biology. *History, classification and cultivation of the Planctomycetes*, p 1–38. doi:10.1007/978-1-62703-502-6_1.

Jeske, O., Schüler, M., Schumann, P., Schneider, A., Boedeker, C., Jogler, M. (2015). Planctomycetes do possess a peptidoglycan cell wall. *Nature Communications*, 6, 7116. doi:10.1038/ncomms8116

Jewson, D.H., Granin, N.G., Zhdanov, A.A., Gnatovsky, R.Y. (2009). Effect of snow depth on under-ice irradiance and growth of aulacoseira baicalensis in Lake Baikal. *Aquatic Ecology: A Multidisciplinary Journal Relating to Processes and Structures at Different Organizational Levels*, 43 (3), 673-679. doi:10.1007/s10452-009-9267-2

Jogler, C., Waldmann, J., Huang, X., Jogler, M., Glöckner, F., Mascher, T., & Kolter, R. (2012). Identification of proteins likely to be involved in morphogenesis, cell division, and signal transduction in Planctomycetes by comparative genomics. *Journal of Bacteriology*, 194 (23), 6419-6430. doi:10.1128/JB.01325-12

Jørgensen, N.G. (1987) Free amino acids in lakes: concentrations and assimilation rates in relation to phytoplankton and bacterial production. *Association for the Sciences of Limnology and Oceanography*, 32:97–111.

Jørgensen, N.G., Kroer, N., Coffin, R.B., Yang, X.H., Lee, C. (1993) Dissolved free amino acids, combined amino acids, and DNA as sources of carbon and nitrogen to marine bacteria. *Marine Ecology Progress Series,* 98:135–148.

Jørgensen, N.G. (2006). Uptake of urea by estuarine bacteria. *Aquatic Microbial Ecology*, 42 (3), 227-242.

Karl, D.M., Björkman, K.M., Dore, J.E., Fujieki, L., Hebel, D.V., Houlihan, T., *et al.* (2001). Ecological nitrogen-to-phosphorus stoichiometry at Station ALOHA. *Deep Sea Research Part II: Topical Studies in Oceanography*. 48. 1529-1566. doi:10.1016/S0967-0645(00)00152-1.

Karp, P. D., Paley, S. M., Krummenacker, M., Latendresse, M., Dale, J. M., Lee, T. J., Kaipa, P., Gilham, F., Spaulding, A., *et al.* (2009). Pathway Tools version 13.0: Integrated software for pathway/genome informatics and systems biology. *Briefings in bioinformatics*, 11 (1), 40-79.

Khawasik, O. (2017). Distinct under-ice bacterial communities in seasonally ice-covered northern lakes. *Masters of Science thesis in Biology*. Retrieved from Concordia University.

Kim, J., Brawley, S., Prochnik, S., Chovatia, M., Grimwood, J., Jenkins, J., *et al.* (2016). Genome analysis of Planctomycetes inhabiting blades of the red alga Porphyra umbilicalis. *PLoS One*, 11 (3), 0151883. doi:10.1371/journal.pone.0151883

Kirchman, D.L. (2000). Uptake and regeneration of inorganic nutrients by marine heterotrophic bacteria. In: D.L. Kirchman (ed.). *Microbial ecology of the oceans*. Wiley-Liss, New York. 9. 261-288.

Kirchman, D.L. (2002). The ecology of Cytophag-Flavobacteria in aquatic environments. *FEMS Microbiology Ecology*. 39: 91–100. doi: 10.1111/j.1574-6941.2002.tb00910.x.

Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., Glöckner, F. (2013). Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Research*, 41 (1), 1. doi:10.1093/nar/gks808

König, E., Schlesner, H., & Hirsch, P. (1984). Cell wall studies on budding bacteria of the Planctomyces/Pasteuria group and on a Prosthecomicrobium sp. *Archives of Microbiology*, 138 (3), 200-205. doi:10.1007/BF00402120

Kovaleva, L., Merkel, A., Novikov, A., Baslerov, R., Toshchakov, S., Bonch-Osmolovskaya, E. (2014). Tepidisphaera mucosa gen. nov., sp. nov., a new moderately thermophilic member of Phycisphaerae class in the Planctomycetes, and proposal of a new family Tepidisphaeraceae and a new order Tepidisphaerales. *International journal of Systematic and Evolutionary Microbiology*. doi:10.1099/ijs.0.070151-0.

Lazier, J., Mann, K. (1989). Turbulence and the diffusive layers around small organisms. Deep Sea Research Part A, *Oceanographic Research Papers*, 36 (11), 1721-1733. doi:10.1016/0198-0149(89)90068-X

Lechner, M., Findeiss, S., Steiner, L., Marz, M., Stadler, P., & Prohaska, S. (2011). Proteinortho: Detection of (Co-)orthologs in large-scale analysis. *BMC Bioinformatics*, 12, 124. doi:10.1186/1471-2105-12-124

Li, W., Godzik, A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22:1658-9.

Liesack, W., König, H., Schlesner, H., Hirsch, P. (1986). Chemical composition of the

peptidoglycan-free cell envelopes of budding bacteria of the Pirella/Planctomyces group. the Pirella/Planctomyces group. *Archives of Microbiology*, 145 (4), 361-366. doi:10.1007/BF00470872

Lindsay, M.R., *et al.* (2001). Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. Archives of microbiology. 175. 413-29. doi:10.1007/s002030100280.

Liu, Z., Liu, S., Liu, J., Gardner, W. (2013). Differences in peptide decomposition rates and pathways between hypoxic and oxic coastal environments. *Marine Chemistry*, 157, 67-77. doi:10.1016/j.marchem.2013.08.003

Maddison, W., Maddison, D. (2009). MESQUITE: a modular system for evolutionary analysis. 11. Version 3.40 http://mesquiteproject.org

Magnuson, J.J., Robertson, D.M., Benson, B.J. *et al.* (2000). Historical trends in lake and river ice cover in the northern hemisphere. *Science*, 289 (5485), 1743-1746. doi:10.1126/science.289.5485.1743

Malmstrom, R., Cottrell, M., Elifantz, H., & Kirchman, D. (2005). Biomass production and assimilation of dissolved organic matter by SAR11 bacteria in the Northwest Atlantic Ocean. *Applied and Environmental Microbiology*, 71 (6), 2979-2986. doi:10.1128/AEM.71.6.2979-2986.2005

Markowitz, V., Chen, I., Palaniappan, K., Chu, K., Szeto, E., Pillay, M., *et al.* (2014). IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Research*, 42 (D1), 567. doi:10.1093/nar/gkt963

Martinez-Garcia, M., Brazel, D.M., Swan, B.K., Arnosti, C., Chain, P.S.G., Reitenga,K.G., *et al.* (2012). Capturing single cell genomes of active polysaccharide degraders: An unexpected contribution of Verrucomicrobia. *PLoS One*, 7(4). doi:10.1371/journal.pone.0035314

Matsen, F., Kodner, R., Virginia Armbrust, E. (2010). pplacer: Linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics*, 11.

Mayali, X., Stewart, B., Mabery, S., Weber, P.K. (2015) Temporal succession in carbon incorporation from macromolecules by particle-attached bacteria in marine microcosms. *Environmental Microbiology reports*. doi:10.1111/1758-2229.12352.

McCarren, J., Becker, J.W., Repeta, D.J., Shi, Y., Young, C.R., Malmstrom, R.R., *et al.* (2010).

Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proceedings of the National Academy of Sciences of the United States of America*. 107. 16420-16427. doi:10.1073/pnas.1010732107.

McCarthy, M., Pratum, T., Hedges, J.I., Benner, R. (1997). Chemical composition of dissolved organic nitrogen in the ocean. *Nature*. 390. 150-154. doi:10.1038/36535.

McCarthy, M.D., Hedges, J.I., Benner, R. (1998). Major bacterial contribution to marine dissolved organic nitrogen. *Science* (New York, N.Y.). 281. 231-4.

McCarthy, M.J., Gardner, W., Lehmann, M., Guindon, A., Bird, D. (2016). Benthic nitrogen regeneration, fixation, and denitrification in a temperate, eutrophic lake: Effects on the nitrogen budget and cyanobacteria blooms. *Limnology and Oceanography*. doi:10.1002/lno.10306

McCoy, A.J., Maurelli, A.T. (2006). Building the invisible wall: updating the chlamydial peptidoglycan anomaly. *Trends in Microbiology*. 14, 70–77. doi: 10.1016/j. tim.2005.12.004

Mével G., Vernet M., Goutx M., Ghiglione J.F. (2008). Seasonal to hour variation scales in abundance and production of total and particle-attached bacteria in the open NW Mediterranean Sea (0-1000 m). *Biogeosciences,* 5 (6), 1573-1586.

Middelboe, M., Borch, N.H., Kirchman, D.L. (1995) Bacterial utilization of dissolved free amino acids, dissolved combined amino acids and ammonium in the Delaware Bay estuary: effects of carbon and nitrogen limitation. *Marine Ecology Progress Series*, 128:109–120.

Mironov, D., Terzhevik, A., Kirillin, G., Jonas, T., Malm, J., & Farmer, D. (2002). Radiatively driven convection in ice-covered lakes: Observations, scaling, and a mixed layer model. *Journal of Geophysical Research: Oceans*, 107 (C4), 7-1. doi:10.1029/2001JC000892

Mitamura, O., Kawashima, M., Maeda, H. (2000) Urea degradation by picophytoplankton in the euphotic zone of Lake Biwa. *Limnology*, 1:19–26.

Modi, N., Ganguly, S., Bárcena-Uribarri, I., Benz, R., van den Berg, B., & Kleinekathöfer, U. (2015). Structure, Dynamics, and Substrate Specificity of the OprO Porin from Pseudomonas aeruginosa. *Biophysical journal*, 109 (7), 1429-38.

Moore, M.V., Hampton, S.E., Izmest'eva, L.R., Silow, E.A., Peshkova, E.V., Pavlov, B.K. (2009). Climate change and the world's "sacred sea"—Lake Baikal, Siberia. *Bioscience*, 59 (5), 405-417

Morris, R.M., Longnecker, K., & Giovannoni, S.J. (2006). Pirellula and OM43 are among the dominant lineages identified in an Oregon coast diatom bloom. *Environmental Microbiology*, 8 (8), 1361-70.

Neef, A., Amann, R., Schlesner, H., Schleifer, K. (1998). Monitoring a widespread bacterial group: in situ detection of planctomycetes with 16S rRNA-targeted probes. *Microbiology*. 144 (12): 3257-3266. doi:10.1099/00221287-144-12-3257

Newton, R., Jones, S., Eiler, A., McMahon, K., & Bertilsson, S. (2011). A guide to the natural history of freshwater lake bacteria. *Microbiology and Molecular Biology Reviews*: Mmbr, 75 (1), 14-49. doi:10.1128/MMBR.00028-10

Nikrad, M., Cottrell, M., & Kirchman, D. (2012). Abundance and single-cell activity of heterotrophic bacterial groups in the western arctic ocean in summer and winter. *Applied and Environmental Microbiology*, 78 (7), 2402-2409. doi:10.1128/AEM.07130-11

Nikrad, M., Cottrell, M., & Kirchman, D. (2014). Uptake of dissolved organic carbon by gammaproteobacterial subgroups in coastal waters of the west Antarctic Peninsula. *Applied and Environmental Microbiology*, 80 (11), 3362-3368. doi:10.1128/AEM.00121-14

O'Brien, W.J., Hershey, A.E., Hobbie, J.E., Hullar, M.A., Kipphut, G.W., Miller, M.C. *et al.* (1992). Control mechanisms of arctic lake ecosystems: A limnocorral experiment. *Hydrobiologia: The International Journal of Aquatic Sciences*, 240 (1-3), 143-188. doi:10.1007/BF00013459

O'Reilly, C.M., Sharma, S., Gray, D.K., Hampton, S.E., Read, J.S., Rowley, R.J. *et al*. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters*, 42 (24), 773-10. doi:10.1002/2015GL066235

Okazaki, Y., Fujinaga, S., Tanaka, A., Kohzu, A., Oyagi, H., Nakano, S. (2017). Ubiquity and quantitative significance of bacterioplankton lineages inhabiting the oxygenated hypolimnion of deep freshwater lakes. *The ISME Journal*. doi:10.1038/ismej.2017.89

Orsi, W., Smith, J., Liu, S., Liu, Z., Sakamoto, C., Wilken, S., Santoro, A. (2016). Diverse, uncultivated bacteria and archaea underlying the cycling of dissolved protein in the ocean. *The ISME Journal*, 10 (9), 2158-2173. doi:10.1038/ismej.2016.20

Özkundakci, D., Gsell, A.S., Hintze, T., Täuscher, H., Adrian, R. (2016). Winter severity determines functional trait composition of phytoplankton in seasonally ice-covered lakes. *Global Change Biology*, 22 (1), 284-298. doi:10.1111/gcb.13085

Pacheco. R., Moel. M., Segrè. D. (2019) Costless metabolic secretions as drivers of interspecies interactions in microbial ecosystems. *Nature Communication.* Vol.10, 103. https://doi.org/10.1038/s41467-018-07946-9.

Parks, D., Imelfort, M., Skennerton, C., Hugenholtz, P., & Tyson, G. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25 (7), 1043-1055. doi:10.1101/gr.186072.114

Parveen, B., Mary, I., Vellet, A., Ravet, V., and Debroas, D. (2013) Temporal dynamics and phylogenetic diversity of free-living and particle-associated Verrucomicrobia communities in relation to environmental variables in a mesotrophic lake. *FEMS Microbiolgy Ecology*, 83: 189–201.

Paver, S.F. and Kent, A.D. (2010) Temporal Patterns in Glycolate-Utilizing Bacterial Community Composition Correlate with Phytoplankton Population Dynamics in Humic Lakes. *Microbial Ecology*, 60: 406–418.

Pizzetti, I., Fuchs, B., Gerdts, G., Wichels, A., Wiltshire, K., Amann, R. (2011). Temporal variability of coastal Planctomycetes clades at Kabeltonne station, North Sea. *Applied and Environmental Microbiology*, 77 (14), 5009-5017. doi:10.1128/AEM.02931-10

Price, M. N., Dehal, P. S., Arkin, A. P. (2010). FastTree 2-- Approximately maximum-likelihood trees for large alignments. *PloS one*, 5 (3), e9490. doi:10.1371/journal.pone.0009490

Rabus, R., Gade, D., Helbig, R., Bauer, M., Glöckner, F., *et al.* (2002). Analysis of N-acetylglucosamine metabolism in the marine bacterium Pirellula sp. strain 1 by a proteomic approach. *Proteomics*, 2 (6), 649-655. doi:10.1002/1615-9861(200206)2:63.0.CO;2-R

Requena, J. (2012). Stress Response in Microbiology. *Caister Academic Press.* Norfolk, UK. ISBN: 978-1-908230-04-1

Rinta-Kanto, J., Sun, S., Sharma, S., Kiene, R., Moran, M. (2012). Bacterial community transcription patterns during a marine phytoplankton bloom. *Environmental Microbiology*, 14 (1), 228-239. doi:10.1111/j.1462-2920.2011.02602.x

Ron, Eliora Z. (2012). "Bacterial Stress Response". In Rosenberg, Eugene; DeLong, Edward F.; Lory, Stephen; Stackebrandt, Erko; Thomson, Fabiano. *Prokaryotes : a handbook on the biology of bacteria* (4th ed.). Berlin: Springer. pp. 589–603.

Rousk, J., Bengtson, P. (2014). Microbial regulation of global biogeochemical cycles. *Frontiers in Microbiology,* 5: 103. doi:10.3389/fmicb.2014.00103

Salonen, K., Leppäranta, M., Viljanen, M., Gulati, R. (2009). Perspectives in winter limnology: Closing the annual cycle of freezing lakes. *Aquatic Ecology*, 43 (3), 609-616.

Salcher, M., Posch, T., Pernthaler, J. (2013). *In situ* substrate preferences of abundant bacterioplankton populations in a prealpine freshwater lake. *The ISME Journal,* vol.7, 896–907.

Sanschagrin, S., Yergeau, E. (2014). Next-generation sequencing of 16S ribosomal RNA gene amplicons. *Journal of visualized experiments: JoVE*, (90), 51709. doi:10.3791/51709

Schauer, K., Rodionov, D.A., de Reuse, H. (2008). New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? *Trends in Biochemical Science*. 33: 330–338.

Schleifer K.H., Schmid, M., Twachtmann, U., Klein, M., Wagner, M. *et al.* (2000). Molecular evidence for genus level diversity of bacteria capable of catalyzing anaerobic ammonium oxidation. *Systematic and Applied Microbiology*, 23 (1), 93-106. doi:10.1016/S0723-2020(00)80050-8

Schlesner, H., Stackebrandt, E. (1986). Assignment of the genera planctomyces and pirella to a new family planctomycetaceae fam. nov. and description of the order planctomycetales ord. nov. Systematic and Applied Microbiology, 8 (3), 174-176. doi:10.1016/S0723-2020(86)80072-8

Schlesner, H., Rensmann, C., Tindall, B., Gade, D., Rabus, R., Pfeiffer, S., Hirsch, P. (2004). Taxonomic heterogeneity within the Planctomycetales as derived by DNA-DNA hybridization, description of Rhodopirellula baltica gen. nov., sp. nov., transfer of Pirellula marina to the genus Blastopirellula gen. nov. as Blastopirellula marina comb. nov. and emended description of the genus Pirellula. *International Journal of Systematic and Evolutionary Microbiology*, 54, 1567-1580.

Schloss, P., Westcott, S., Ryabin, T., Hall, J., Hartmann, M., Hollister, E., *et al.* (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, 75 (23), 7537-7541. doi:10.1128/AEM.01541-09

Segata, N., Börnigen, D., Morgan, X., Huttenhower, C. (2013). PhyloPhlAn is a new method for improved phylogenetic and taxonomic placement of microbes. *Nature Communications*, 4 (1). doi:10.1038/ncomms3304

Seitzinger, S. P., Harrison, J. A., Dumont, E., Beusen, A., Bouwman, A.F. (2005). Sources and

delivery of carbon, nitrogen, and phosphorus to the coastal zone: An overview of Global Nutrient Export from Watersheds (NEWS) models and their application. *Global Biogeochemical Cycles,* 19: GB4S01. doi:10.1029/2005GB002606

Sharma, A., Becker, J., Ottesen, E., Bryant, J., Duhamel, S., Karl, D., *et al.* (2014). Distinct dissolved organic matter sources induce rapid transcriptional responses in coexisting populations of Prochlorococcus, Pelagibacter and the OM60 clade. *Environmental Microbiology*, 16 (9), 2815-2830. doi:10.1111/1462-2920.12254

Shaw, P.J., Chapron, C., Purdie, D.A., Rees, A.P. (1998) Impacts of phytoplankton activity on dissolved nitrogen fluxes in the tidal reaches and estuary of the Tweed, UK. *Marine Pollution Bulletin*, 37:280–294.

Shi, Y., Tyson, G., & DeLong, E. (2009). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature*, 459 (7244), 266-269. doi:10.1038/nature08055

Simon, M., Rosenstock, B. (1992) Carbon and nitrogen sources of planktonic bacteria in Lake Constance studied by the composition and isotope dilution of intracellular amino acids. *Association for the Sciences of Limnology and Oceanography,* 37:1496–1511.

Solomon, S., (Atmospheric chemist). (2007). Climate change 2007: The physical science basis: Contribution of working group I to the fourth assessment report of the intergovernmental panel on climate change. *New York: Cambridge University Press.*

Solomon, C., Collier, J., Berg, G., & Glibert, P. (2010). Role of urea in microbial metabolism in aquatic systems: A biochemical and molecular review. *Aquatic Microbial Ecology*, 59 (1), 67-88. doi:10.3354/ame01390

Sommer, U., Adrian, R., De Senerpont Domis, L., Elser, J., Gaedke, U., Ibelings, B., *et al.* (2012). Beyond the plankton ecology group (PEG) model: Mechanisms driving plankton succession. *Annual Review of Ecology, Evolution, and Systematics*, 43, 429-448. https://doi.org/10.1146/annurev-ecolsys-110411-160251

Stewart, F., Ottesen, E., & DeLong, E. (2010). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *The ISME Journal*, 4 (7), 896-907. doi:10.1038/ismej.2010.18

Stocker, R., Seymour, J. (2012). Ecology and physics of bacterial chemotaxis in the ocean. *Microbiology and Molecular Biology Reviews*, 76 (4), 792-812. doi:10.1128/MMBR.00029-12

Stottlemyer, R., Toczydlowski, D. (1999). Seasonal change in precipitation, snowpack, snowmelt, soil water and streamwater chemistry, Northern Michigan. *Hydrological Processes*, 13 (14-15), 2215-2231. doi:10.1002/(SICI)1099-1085(199910)13:14/153.0.CO;2-V

Straile, D. (2002). North Atlantic Oscillation synchronizes food-web interactions in central European lakes. Proceedings. *Biological Sciences - The Royal Society*. 269. 391-395. doi:10.1098/rspb.2001.1907.

Strous, M., Fuerst, J., Kramer, E., Logemann, S., Muyzer, G., Van, D., *et al.* (1999). Missing lithotroph identified as new Planctomycete. *Nature*, 400 (6743), 446-449.

Strous, M., Kraft, B., Bisdorf, R., & Tegetmeyer, H. E. (2012). The binning of metagenomic contigs for microbial physiology of mixed cultures. *Frontiers in microbiology*, 3, 410. doi:10.3389/fmicb.2012.00410

Sukharev, S., Betanzos, M., Chiang, C., Guy, R. (2011). The gating mechanism of the large mechanosensitive channel MscL. *Nature,* vol.409, 720–724.

Sutcliffe, I.C. (2010) A phylum level perspective on bacterial cell envelope architecture. *Trends in Microbiology*. 18, 464–470.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution*, 30 (12), 2725-2729.

Tang, K., Jiao, N., Liu, K., Zhang, Y., Li, S., Kirchman, D. (2012). Distribution and functions of TonB-dependent transporters in marine bacteria and environments: Implications for dissolved organic matter utilization. *PLoS One*, 7 (7), 41204. doi:10.1371/journal.pone.0041204

Tanoue, E., Nishiyama, S., Kamo, M., & Tsugita, A. (1995). Bacterial membranes: Possible source of a major dissolved protein in seawater. *Geochimica Et Cosmochimica Acta*, 59 (12), 2643-2648. doi:10.1016/0016-7037(95)00134-4

Teeling, H., Fuchs, M., Becher, D., Klockow, C., Bennke, C., *et al.* (2012). Substrate-Controlled Succession of Marine Bacterioplankton Populations Induced by a Phytoplankton Bloom. *Science* (New York, N.Y.). 336. 608-11. 10.1126/science.1218344.

Teeseling, M., Mesman, R., Kuru, E., Espaillat, A., Cava, F., Brun, Y., *et al.* (2015). Anammox Planctomycetes have a peptidoglycan cell wall. *Nature Communications*, 6, 1-6. doi:10.1038/ncomms7878

Tekniepe, B., Schmidt, J., & Starr, P. (1981). Life cycle of a budding and appendaged bacterium belonging to morphotype IV of the Blastocaulis-Planctomyces group. *Current Microbiology: An International Journal*, 5 (1), 1-6. doi:10.1007/BF01566588

Thomson, D. (2009). Climate change: Shifts in season. *Nature*, 457 (7228), 391-2. doi:10.1038/457391a

Tran P., Ramachandran A., Khawasik O., Beisner B.B., Rautio M., Huot Y., Walsh D.A. (2018). Microbial life under ice: Diversity and in situ activity of Verrucomicrobia in seasonally ice-covered lakes. *Environmental Microbiology*, 20 (7). doi: 10.1111/1462-2920.14283

Tsementzi, D., Poretsky, R., Rodriguez-R, L.M., Luo, C., Konstantinidis, K.T. (2014). Evaluation of metatranscriptomic protocols and application to the study of freshwater microbial communities. *Environmental Microbiology Reports*, 6 (6), 640-655.

Twiss, M., McKay, R., Bourbonniere, R., Bullerjahn, G., Carrick, H., Smith, R., Wilhelm, S. (2012). Diatoms abound in ice-covered Lake Erie: An investigation of offshore winter limnology in Lake Erie over the period 2007 to 2010. *Journal of Great Lakes Research*, 38 (1), 18-30. doi:10.1016/j.jglr.2011.12.008

Urbach, E., Vergin, K., Young, L., Morse, A., Larson, G., Giovannoni, S. (2001). Unusual bacterioplankton community structure in ultra-oligotrophic Crater Lake. *Limnology and Oceanography*, 46 (3), 557-572. doi:10.4319/lo.2001.46.3.0557

van Niftrik, L., Devos, D. P. (2017). Editorial: Planctomycetes-Verrucomicrobia-Chlamydiae Bacterial Superphylum: New Model Organisms for Evolutionary Cell Biology. *Frontiers in microbiology*, 8, 1458. doi:10.3389/fmicb.2017.01458

Verpoorter, C., Kutser, T., Seekell, D.A. & Tranvik, L.J. (2014). A global inventory of lakes based on high-resolution satellite imagery. *Geophysical Research Letters*, 41 (18), 6396-6402.

Wang, J., Jenkins, C., Webb, R., & Fuerst, J. (2002). Isolation of Gemmata-like and Isosphaera-like Planctomycete bacteria from soil and freshwater. *Applied and Environmental Microbiology*, 68(1), 417-22.

Wang, Y., Qian, P.Y. (2013) Conserved Regions in 16S Ribosome RNA Sequences and Primer Design for Studies of Environmental Microbes. *Encyclopedia of Metagenomics.* Springer, New York, NY

Wang, X., Sharp, C., Jones, G., Grasby, S., Brady, A., & Dunfield, P. (2015). Stable-isotope

probing identifies uncultured planctomycetes as primary degraders of a complex heteropolysaccharide in soil. *Applied and Environmental Microbiology*, 81 (14), 4607-4615. doi:10.1128/AEM.00055-15

Ward, N., Rainey, F., Hedlund, B., Staley, J., Ludwig, W., Stackebrandt, E. (2000). Comparative phylogenetic analyses of members of the order Planctomycetales and the division Verrucomicrobia: 23S rRNA gene sequence analysis supports the 16S rRNA gene sequence-derived phylogeny. *International Journal of Systematic and Evolutionary Microbiology*, 50, 1965-1972.

Wegner, C., Richter-Heitmann, T., Klindworth, A., Klockow, C., Richter, M., Achstetter T., Glöckner F.O., Harder, J. (2013). Expression of sulfatases in Rhodopirellula baltica and the diversity of sulfatases in the genus Rhodopirellula. *Marine Genomics*, 9, 51-61. doi:10.1016/j.margen.2012.12.001

Weiner, R., Taylor, L., Henrissat, B., Hauser, L., Land, M., *et al.* (2008). Complete genome sequence of the complex carbohydrate-degrading marine bacterium, Saccharophagus degradans strain 2-40t. *PLoS Genetics*, 4 (5). doi:10.1371/journal.pgen.1000087

Weyhenmeyer, G.A., Livingstone, D.M., Meili, M., Jensen, O., Benson, B. & Magnuson, J.J. (2011). Large geographical differences in the sensitivity of ice-covered lakes and rivers in the northern hemisphere to temperature changes. *Global Change Biology*, 17 (1), 268-275. doi:10.1111/j.1365-2486.2010.02249.

Wheeler, P.A., Kirchman, D.L. (1986) Utilization of inorganic and organic nitrogen by bacteria in marine systems. *Association for the Sciences of Limnology and Oceanography*, 31:998–1009.

White, D. (2007). The physiology and biochemistry of prokaryotes. *New York: Oxford University Press.*

Williams, P.J., Berman, T., Holm-Hansen, O. (1976) Amino acid uptake and respiration by marine heterotrophs. *Marine Biology*, 35:41–47.

Woebken, D., Teeling, H., Wecker, P., Dumitriu, A., Kostadinov, I., Delong, E., Glöckner, F. (2007). Fosmids of novel marine Planctomycetes from the Namibian and Oregon coast upwelling systems and their cross-comparison with Planctomycete genomes. *The ISME Journal*, 1 (5), 419-35.

Worden, A., Follows, M., Giovannoni, S., Wilken, S., Zimmerman, A., Keeling, P.J. (2015).

Rethinking the marine carbon cycle: Factoring in the multifarious lifestyles of microbes. *Science*, 347 (6223).

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., & Xu, Y. (2012). DbCAN: A web resource for automated carbohydrate-active enzyme annotation. *Nucleic Acids Research*, 40 (W1), 451. doi:10.1093/nar/gks479

Yoon, J., Jang, J.H., Kasai, H. (2013). Algisphaera agarilytica gen. nov., sp nov., a novel representative of the class Phycisphaerae within the phylum Planctomycetes isolated from a marine alga. *Antonie van Leeuwenhoek*. 105. 317-324. doi:10.1007/s10482-013-0076-1.

Zehr, J.P., Axler, R.P., Goldman, C.R. (1985) Heterotrophic mineralization of amino acid nitrogen in subalpine Castle Lake, Calf. *Marine Chemistry*, 16:343–350.

Zeigler, A., Allen, E., Badger, J., McCrow, J., Paulsen, I., Elbourne, L., *et al.* (2012). Influence of nutrients and currents on the genomic composition of microbes across an upwelling mosaic. *The ISME Journal*, 6 (7), 1403-1414. doi:10.1038/ismej.2011.201

Zhou, J., Bruns, M., Tiedje, J.M. (1996). DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology*, 62 (2), 316-32

Supplementary Figures

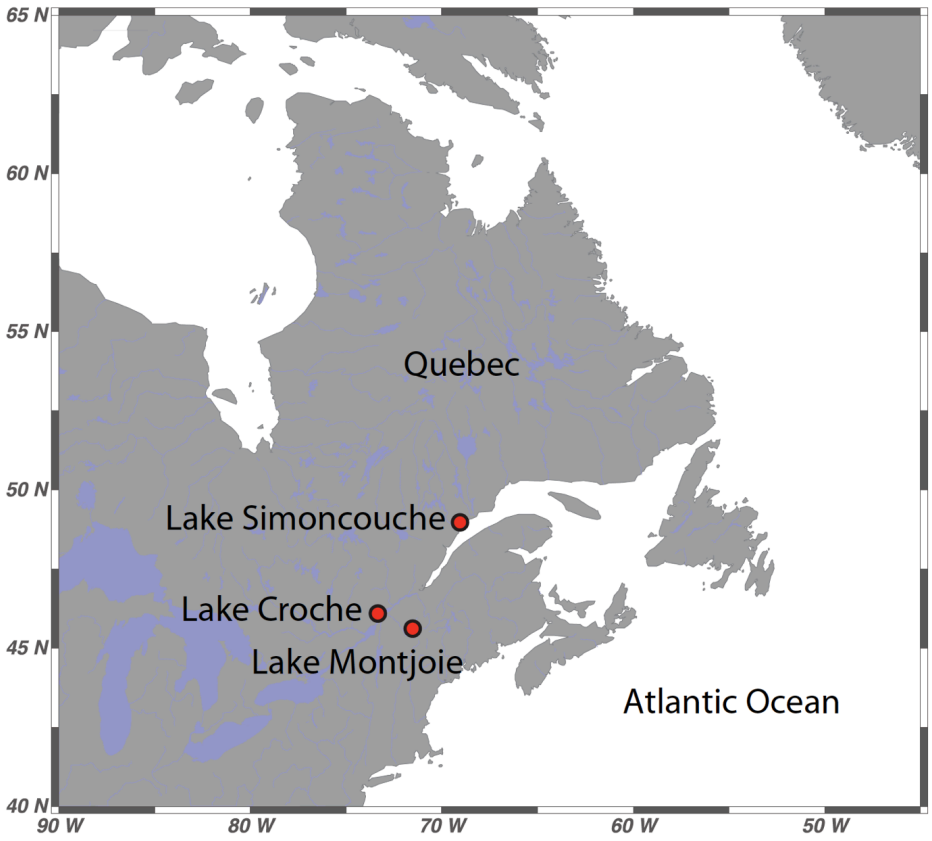

Figure S1.
The locations of Lakes Croche, Montjoie and Simoncouche in temperate and boreal regions of Quebec, Canada.

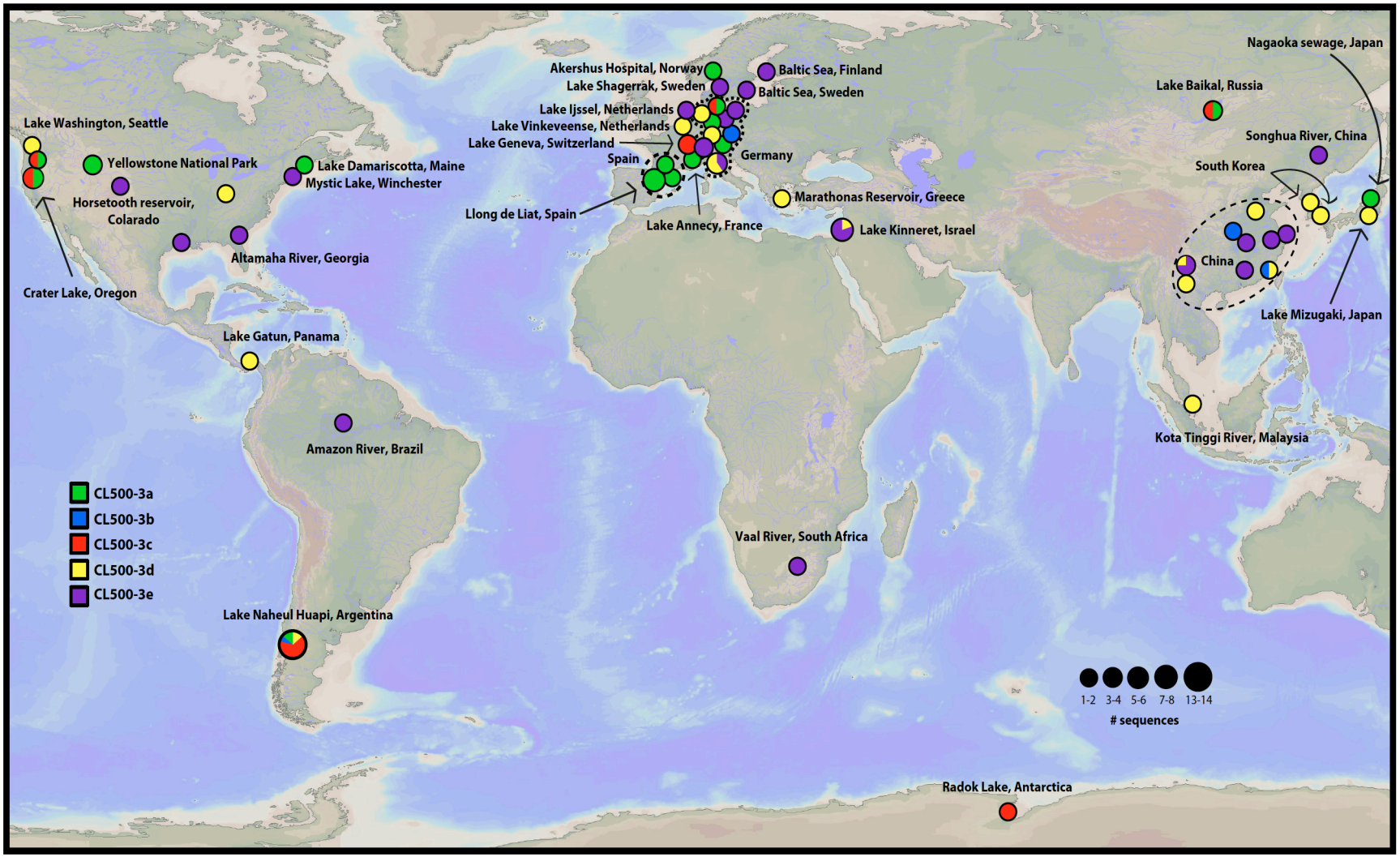

Figure S2.
Global distribution of the uncultivated CL500-3 clade of Phycisphaerae, based on publicly available 16S rRNA gene data.

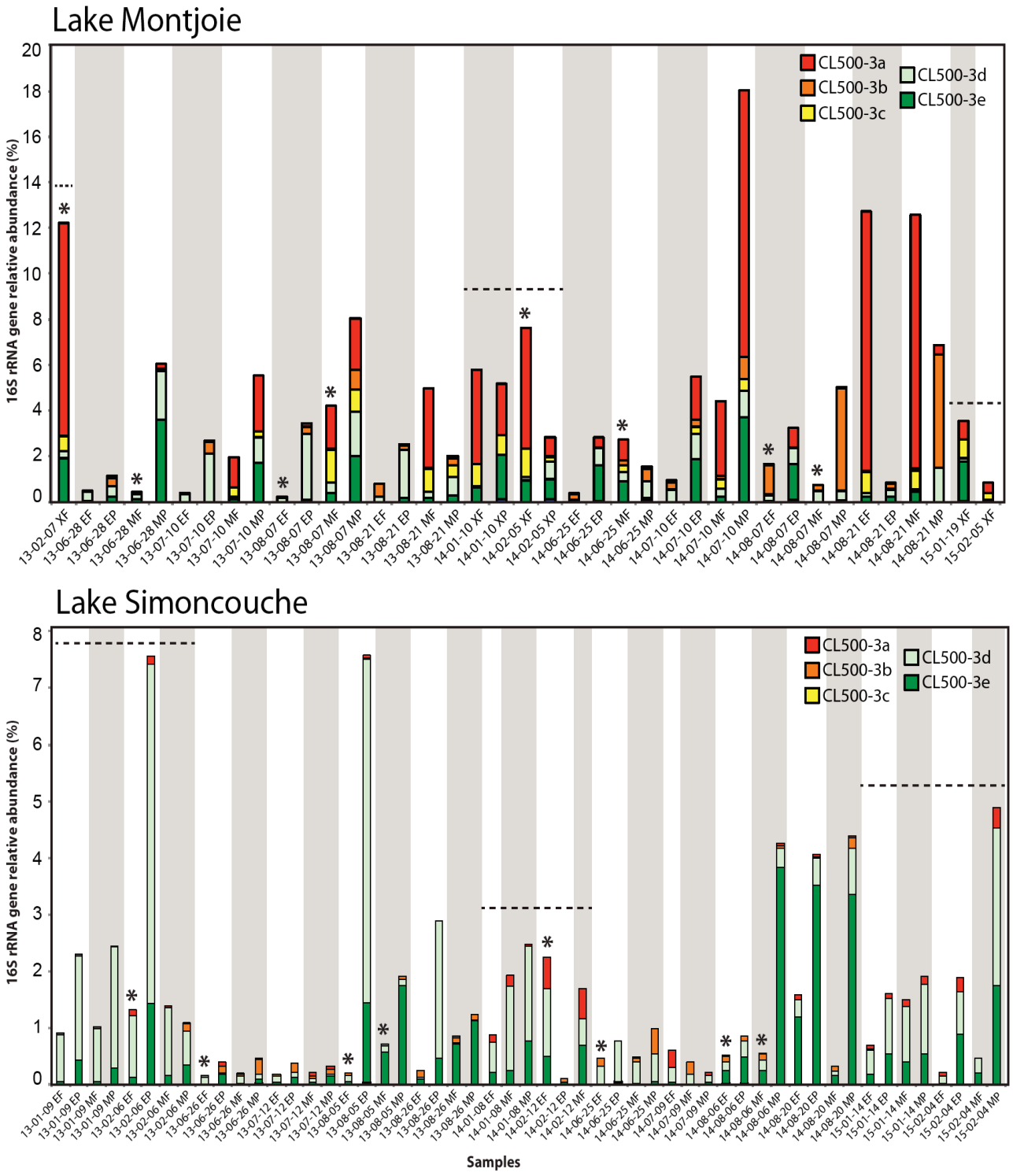

Figure S3.
Stacked bar plots demonstrating the percent relative abundance of the CL500-3 subclades, represented by the 16S rRNA amplicon data, for each sample taken through January 2013-February 2015. Top plot shows the environmental 16S rRNA data for Lake Montjoie, while bottom plot is for Lake Simoncouche. Grey and white panels depict free-living (F) and particle-associated (P) fractions taken on the same sampling date. Winter samples are depicted by a dashed line. Metagenomic data is available for the samples depicted by a star. Sampling date are in YY/MM/DD. Epilimnion, metalimnion and mixed samples are depicted by E, M and X, respectively.

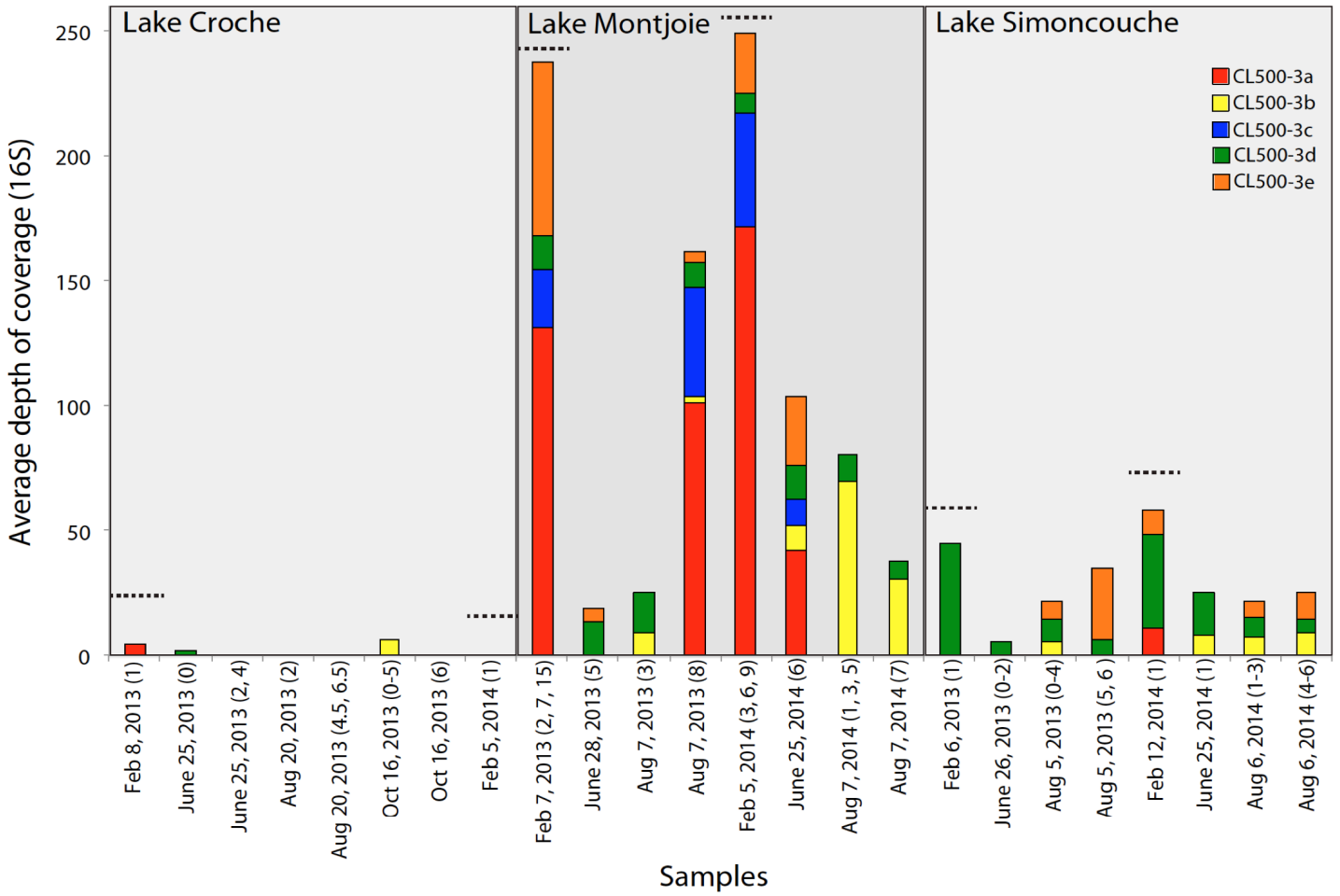

Figure S4.
Stacked bar plot demonstrating the relative abundance of the CL500-3 subclades, represented by the average depth of coverage of the 16S rRNA metagenomic sequences. Lake Croche, Lake Montjoie and Lake Simoncouche are represented by the three panels in the figure. Winter samples are depicted by a dashed line. Numbers in brackets after each sampling date represent the depth (or multiple depths) where the samples were taken. When looking at samples taken on the same date, the epilimnion samples are depicted by the shallow depths, while metalimnion samples are deeper.

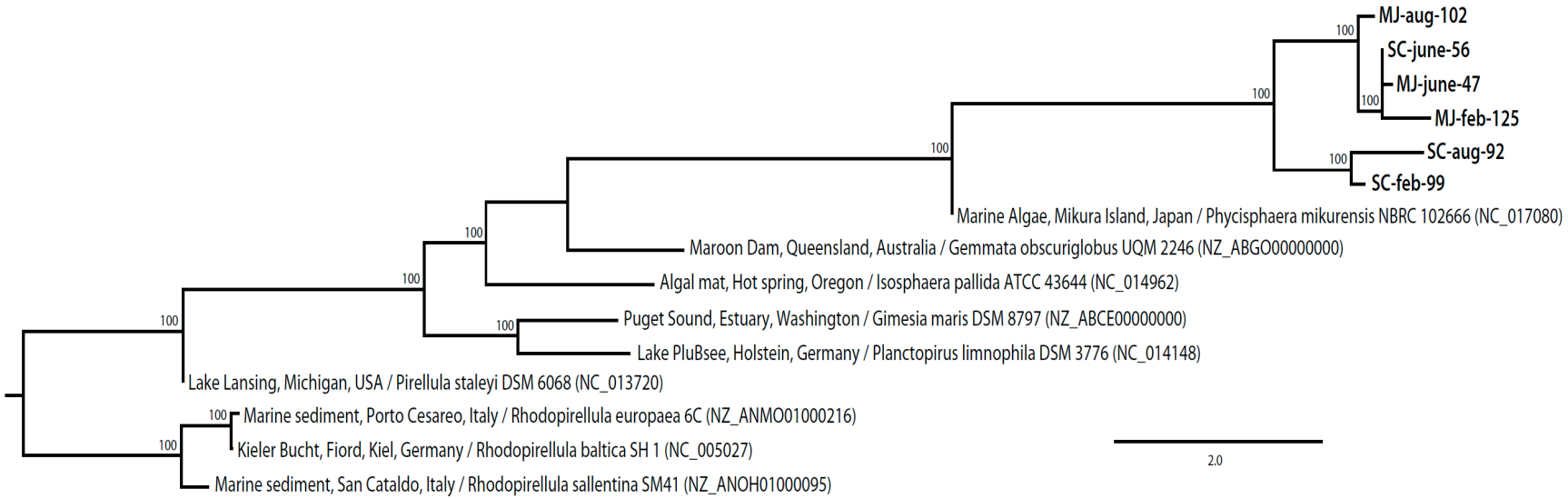

Figure S5.
Phylophlan protein phylogeny, based on hundreds of conserved proteins, demonstrating the phylogenetic relationship between the six CL500-3 MAGs and 9 Planctomycetes reference genomes.

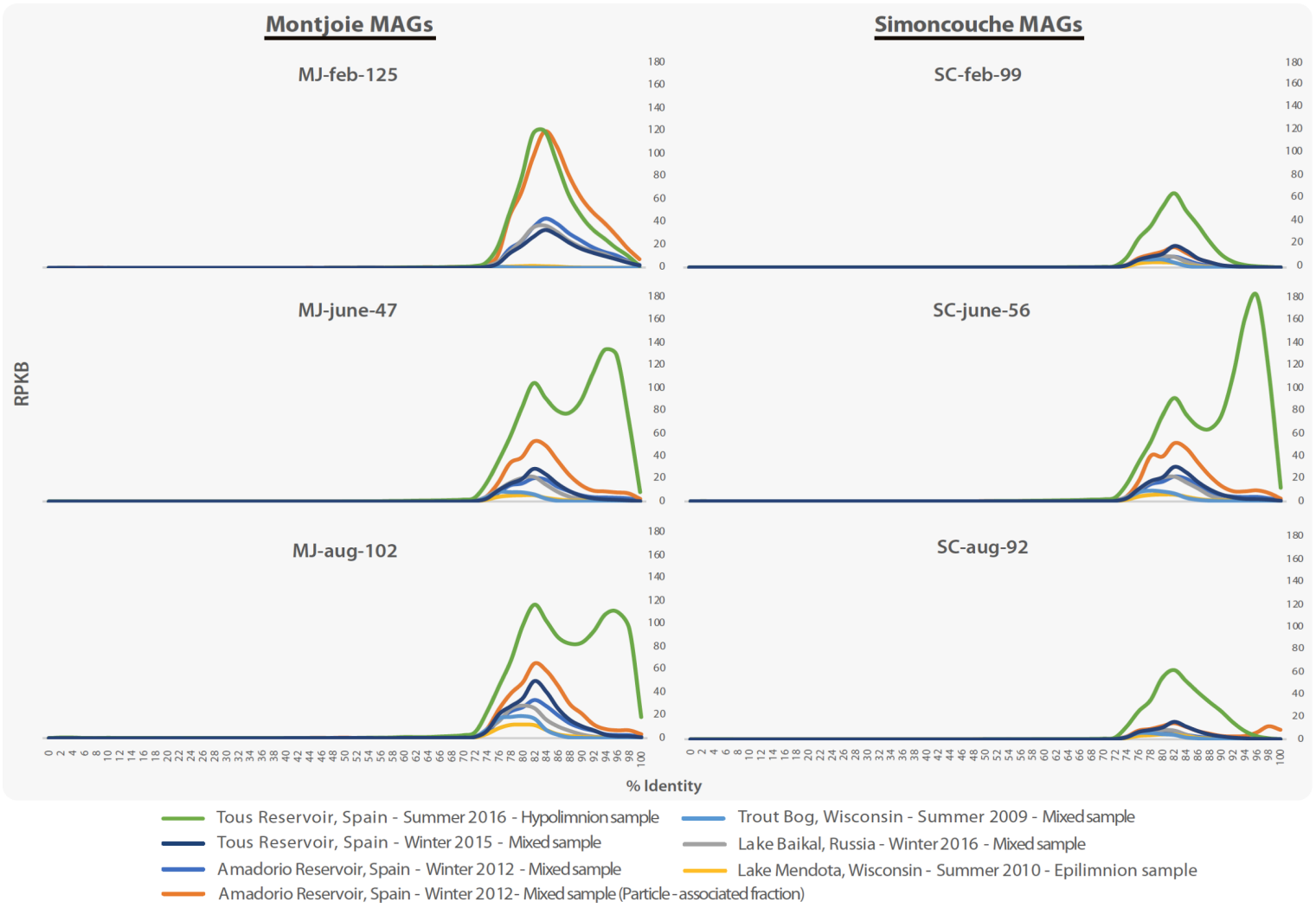

Figure S6.
Metagenomes originating from several freshwater environments, including two Spanish reservoirs (Tous and Amadorio), two Wisconsin Lakes (Trout and Mendota), as well as from under the ice in Lake Baikal, were recruited to the six CL500-3 MAGs. Each panel represents the reads recruited per kilobase of transcript per billion reads per sample (RPKB) at a given percent identity, for the six CL500-3 MAGs. X-axis represents the percent identity at which the raw metagenomic reads from other freshwater lakes are recruiting to the CL500-3 MAGs.

Supplementary Tables

Sample ID	Lake	Date (YYMMDD)	Strata	Size Fraction	Ice thickness (cm)	Depth (m)	16S rRNA data	Meta-genome data	Meta-transcriptome data
C_130111_EP	Croche	130111	E	P	58	0	x		
C_130111_MF	Croche	130111	M	F	58	N/A	x		
C_130111_MP	Croche	130111	M	P	58	N/A	x		
C_130208_EF	Croche	130208	E	F	60	0.5	x	x	x
C_130208_EP	Croche	130208	E	P	60	0.5	x		
C_130208_MF	Croche	130208	M	F	60	N/A	x		
C_130208_MP	Croche	130208	M	P	60	N/A	x		
C_130625_EF	Croche	130625	E	F	N/A	0	x	x	x
C_130625_EP	Croche	130625	E	P	N/A	0	x		
C_130625_MF	Croche	130625	M	F	N/A	2, 4	x	x	x
C_130625_MP	Croche	130625	M	P	N/A	2, 4	x		
C_130709_EF	Croche	130709	E	F	N/A	0.5	x		x
C_130709_EP	Croche	130709	E	P	N/A	0.5	x		
C_130709_MP	Croche	130709	M	P	N/A	3.5	x		
C_130709_MF	Croche	130709	M	F	N/A	3.5			x
C_130807_EF	Croche	130807	E	F	N/A	1.5	x		
C_130807_EP	Croche	130807	E	P	N/A	1.5	x		
C_130807_MF	Croche	130807	M	F	N/A	4, 6	x		
C_130807_MP	Croche	130807	M	P	N/A	4, 6	x		
C_130820_EF	Croche	130820	E	F	N/A	1.5	x	x	
C_130820_EP	Croche	130820	E	P	N/A	1.5	x		
C_130820_MF	Croche	130820	M	F	N/A	4.5, 6.5	x	x	
C_130820_MP	Croche	130820	M	P	N/A	4.5, 6.5	x		
C_131016_EF	Croche	131016	E	F	N/A	0 to 5		x	x
C_131016_MF	Croche	131016	M	F	N/A	6		x	
C_140106_EP	Croche	140106	E	P	60	0.5	x		
C_140106_MF	Croche	140106	M	F	60	4, 8	x		
C_140106_MP	Croche	140106	M	P	60	4, 8	x		
C_140205_EF	Croche	140205	E	F	54	0.5	x	x	
C_140205_EP	Croche	140205	E	P	54	0.5	x		
C_140205_MF	Croche	140205	M	F	54	4, 6	x		
C_140205_MP	Croche	140205	M	P	54	4, 6	x		
C_140625_EF	Croche	140625	E	F	N/A	1	x		
C_140625_EP	Croche	140625	E	P	N/A	1	x		
C_140625_MF	Croche	140625	M	F	N/A	3.75	x		

C_140625_MP	Croche	140625	M	P	N/A	3.75	x			
C_140709_EF	Croche	140709	E	F	N/A	1	x			
C_140709_EP	Croche	140709	E	P	N/A	1	x			
C_140709_MF	Croche	140709	M	F	N/A	4.25	x			
C_140709_MP	Croche	140709	M	P	N/A	4.25	x			
C_140806_EF	Croche	140806	E	F	N/A	1	x			
C_140806_EP	Croche	140806	E	P	N/A	1	x			
C_140806_MF	Croche	140806	M	F	N/A	4	x			
C_140806_MP	Croche	140806	M	P	N/A	4	x			
C_140820_EP	Croche	140820	E	P	N/A	1.5	x			
C_140820_MF	Croche	140820	M	F	N/A	4.75	x			
C_150114_EF	Croche	150114	E	F	35	0.5	x			
C_150114_EP	Croche	150114	E	P	35	0.5	x			
C_150114_MF	Croche	150114	M	F	35	4.5, 8	x			
C_150114_MP	Croche	150114	M	P	35	4.5, 8	x			
C_150204_EF	Croche	150204	E	F	43	0.5	x			
C_150204_EP	Croche	150204	E	P	43	0.5	x			
C_150204_MF	Croche	150204	M	F	43	4.5, 7.5	x			
C_150204_MP	Croche	150204	M	P	43	4.5, 7.5	x			
M_130207_XF	Montjoie	130207	E	F	30	N/A	x	x		x
M_130628_EF	Montjoie	130628	E	F	N/A	1.5	x			
M_130628_EP	Montjoie	130628	E	P	N/A	1.5	x			
M_130628_MF	Montjoie	130628	M	F	N/A	4.5	x	x		
M_130628_MP	Montjoie	130628	M	P	N/A	4.5	x			
M_130710_EF	Montjoie	130710	E	F	N/A	1, 3.5	x			
M_130710_EP	Montjoie	130710	E	P	N/A	1, 3.5	x			
M_130710_MF	Montjoie	130710	M	F	N/A	6	x			x
M_130710_MP	Montjoie	130710	M	P	N/A	6	x			
M_130807_EF	Montjoie	130807	E	F	N/A	3	x	x		
M_130807_EP	Montjoie	130807	E	P	N/A	3	x			
M_130807_MF	Montjoie	130807	M	F	N/A	8	x	x		
M_130807_MP	Montjoie	130807	M	P	N/A	8	x			
M_130821_EF	Montjoie	130821	E	F	N/A	2.5	x			x
M_130821_EP	Montjoie	130821	E	P	N/A	2.5	x			
M_130821_MF	Montjoie	130821	M	F	N/A	7.5	x			x
M_130821_MP	Montjoie	130821	M	P	N/A	7.5	x			
M_140110_XF	Montjoie	140110	X	F	41	3, 11	x			
M_140110_XP	Montjoie	140110	X	P	41	3, 11	x			
M_140205_XF	Montjoie	140205	X	F	54	N/A	x	x		x
M_140205_XP	Montjoie	140205	X	P	54	N/A	x			

M_140625_EF	Montjoie	140625	E	F	N/A	1, 3	x		
M_140625_EP	Montjoie	140625	E	P	N/A	1, 3	x		
M_140625_MF	Montjoie	140625	M	F	N/A	6	x	x	x
M_140625_MP	Montjoie	140625	M	P	N/A	6	x		
M_140710_EF	Montjoie	140710	E	F	N/A	N/A	x		
M_140710_EP	Montjoie	140710	E	P	N/A	N/A	x		
M_140710_MF	Montjoie	140710	M	F	N/A	N/A	x		
M_140710_MP	Montjoie	140710	M	P	N/A	N/A	x		
M_140807_EF	Montjoie	140807	E	F	N/A	1, 3, 5	x	x	x
M_140807_EP	Montjoie	140807	E	P	N/A	1, 3, 5	x		
M_140807_MF	Montjoie	140807	M	F	N/A	7	x	x	x
M_140807_MP	Montjoie	140807	M	P	N/A	7	x		
M_140821_EF	Montjoie	140821	E	F	N/A	2, 4, 6	x		
M_140821_EP	Montjoie	140821	E	P	N/A	2, 4, 6	x		
M_140821_MF	Montjoie	140821	M	F	N/A	9	x		
M_140821_MP	Montjoie	140821	M	P	N/A	9	x		
M_150119_XF	Montjoie	150119	X	F	40	4, 8, 12	x		
M_150205_XF	Montjoie	150205	X	F	N/A	N/A	x		
S_130109_EF	Simoncouche	130109	E	F	N/A	1	x		x
S_130109_EP	Simoncouche	130109	E	P	33	1	x		
S_130109_MF	Simoncouche	130109	M	F	33	N/A	x		
S_130109_MP	Simoncouche	130109	M	P	33	N/A	x		
S_130206_EF	Simoncouche	130206	E	F	57	1	x	x	x
S_130206_EP	Simoncouche	130206	E	P	57	1	x		
S_130206_MF	Simoncouche	130206	M	F	57	N/A	x		
S_130206_MP	Simoncouche	130206	M	P	57	N/A	x		
S_130626_EF	Simoncouche	130626	E	F	N/A	0 to 2	x	x	x
S_130626_EP	Simoncouche	130626	E	P	N/A	0 to 2	x		
S_130626_MF	Simoncouche	130626	M	F	N/A	3, 4 ,5, 6	x		
S_130626_MP	Simoncouche	130626	M	P	N/A	3, 4 ,5, 6	x		
S_130712_EF	Simoncouche	130712	E	F	N/A	0, 1, 2, 3	x		x
S_130712_EP	Simoncouche	130712	E	P	N/A	0, 1, 2, 3	x		
S_130712_MF	Simoncouche	130712	M	F	N/A	4, 5, 6	x		
S_130712_MP	Simoncouche	130712	M	P	N/A	4, 5, 6	x		
S_130805_EF	Simoncouche	130805	E	F	N/A	0,1,2,3,4	x	x	
S_130805_EP	Simoncouche	130805	E	P	N/A	0,1,2,3,4	x		
S_130805_MF	Simoncouche	130805	M	F	N/A	5, 6	x	x	
S_130805_MP	Simoncouche	130805	M	P	N/A	5, 6	x		
S_130826_EF	Simoncouche	130826	E	F	N/A	0,1,2,3	x		x
S_130826_EP	Simoncouche	130826	E	P	N/A	0,1,2,3	x		

S_130826_MF	Simoncouche	130826	M	F	N/A	4, 5, 6	x		x
S_130826_MP	Simoncouche	130826	M	P	N/A	4 ,5, 6	x		
S_140108_EF	Simoncouche	140108	E	F	30	0.5	x		x
S_140108_MF	Simoncouche	140108	M	F	30	3, 4, 5	x		
S_140108_MP	Simoncouche	140108	M	P	30	3, 4, 5	x		
S_140212_EF	Simoncouche	140212	E	F	45	1	x	x	x
S_140212_EP	Simoncouche	140212	E	P	45	1	x		
S_140212_MF	Simoncouche	140212	M	F	45	5	x		
S_140625_EF	Simoncouche	140625	E	F	N/A	1	x	x	x
S_140625_EP	Simoncouche	140625	E	P	N/A	1	x		
S_140625_MF	Simoncouche	140625	M	F	N/A	3, 4, 5	x		
S_140625_MP	Simoncouche	140625	M	P	N/A	3, 4, 5	x		
S_140709_EF	Simoncouche	140709	E	F	N/A	1, 2, 3	x		x
S_140709_MF	Simoncouche	140709	M	F	N/A	4, 5, 6	x		
S_140709_MP	Simoncouche	140709	M	P	N/A	4, 5, 6	x		
S_140806_EF	Simoncouche	140806	E	F	N/A	1, 2, 3	x	x	
S_140806_EP	Simoncouche	140806	E	P	N/A	1, 2, 3	x		
S_140806_MF	Simoncouche	140806	M	F	N/A	4, 5, 6	x	x	
S_140806_MP	Simoncouche	140806	M	P	N/A	4, 5, 6	x		
S_140820_EF	Simoncouche	140820	E	F	N/A	1,2,3,4	x		
S_140820_EP	Simoncouche	140820	E	P	N/A	1,2,3,4	x		
S_140820_MF	Simoncouche	140820	M	F	N/A	5.5	x		
S_140820_MP	Simoncouche	140820	M	P	N/A	5.5	x		
S_150114_EF	Simoncouche	150114	E	F	35	0.5	x		
S_150114_EP	Simoncouche	150114	E	P	35	0.5	x		
S_150114_MF	Simoncouche	150114	M	F	35	3, 4, 5	x		
S_150114_MP	Simoncouche	150114	M	P	35	3, 4, 5	x		
S_150204_EF	Simoncouche	150204	E	F	43	1	x		
S_150204_EP	Simoncouche	150204	E	P	43	1	x		
S_150204_MF	Simoncouche	150204	M	F	43	2,3,4,5,6	x		
S_150204_MP	Simoncouche	150204	M	P	43	2,3,4,5,6	x		
Total number of samples							143	24	24

Table S1.
Description of the 16s rRNA gene, metagenome and metatranscriptome datasets employed in this study. E = Epilimnion, M = Metalimnion, X = Mixed, F = Free-living, P = Particle-associated, N/A = Not available.

MAGs	MJ-aug-102	MJ-june-47	SC-june-56	MJ-feb-125	SC-aug-92	SC-feb-99
MJ-aug-102		89	92	75	73	71
MJ-june-47	88.7		66	75	70	70
SC-june-56	89.2	93.1		74	70	68
MJ-feb-125	73.4	75.4	74.9		69	66
SC-aug-92	56.5	55.0	53.7	53.7		78
SC-feb-99	58.4	55.8	54.3	54.5	73.7	

Table S2.
Average nucleotide identity (ANI) pairwise comparisons between the MAGs are shown on top, while the average amino acid identity (AAI) pairwise comparisons are on the bottom.

Freshwater Ecosystems	Samples	SRA Accession ID
Amadorio Reservoir, Spain - Winter 2012 - Mixed sample	Amadorio Resevoir - February 1, 2012 - 10m (0.1 micron)	SRR1173821
Amadorio Reservoir, Spain - Winter 2012 - Mixed sample (Particle-associated fraction)	Amadorio Resevoir - February 1, 2012 - 10m (5 micron)	SRR1173877
Lake Mendota, Wisconsin - Summer 2010 - Epilimnion sample	Lake Mendota - Epilimnion May 20, 2010 - July 6, 2010	SRR3960573, SRR3986823
Trout Bog, Wisconsin - Summer 2009 - Mixed sample	Trout Bog - June 3, 2009 - Epilimnion/Hypolimnion, Trout Bog - August 3, 2009 - Epilimnion/Hypolimnion	SRR4029376, SRR4029406, SRR4029375, SRR4029421
Tous Reservoir, Spain - Winter 2015 - Mixed sample	Tous Reservoir - Feburary 2015 - 12m, 25m	SRR4198666, SRR4198832
Tous Reservoir, Spain - Summer 2016 - Hypolimnion sample	Tous Reservoir - July 2015 - 13m	SRR5338504
Lake Baikal, Russia - Winter 2016 - Mixed sample	Lake Baikal - March 2016 - 5m, 20m	SRR5896114, SRR5896115

Table S3.
Description of metagenomes originating from several freshwater environments, including two Spanish reservoirs (Tous and Amadorio), two Wisconsin Lakes (Trout and Mendota), as well as from under the ice in Lake Baikal. Description includes sampling location and time data, along with an SRA accession number.

	KEGG metabolism	MJ-feb-125	SC-feb-99	MJ-june-47	SC-june-56	MJ-aug-102	SC-aug-92
Metabolism	Amino acid metabolism	55	98	83	67	31	79
	Biosynthesis of other secondary metabolites	0	4	2	1	3	8
	Carbohydrate metabolism	47	94	81	66	41	63
	Energy metabolism	27	34	29	13	9	20
	Enzyme families	16	17	23	14	15	23
	Glycan biosynthesis and metabolism	16	31	25	22	10	22
	Lipid metabolism	16	36	25	16	13	31
	Metabolism of cofactors and vitamins	38	43	57	35	18	37
	Metabolism of other amino acids	9	13	9	6	4	6
	Metabolism of terpenoids and polyketides	7	17	17	13	3	11
	Nucleotide metabolism	30	43	45	32	18	29
	Xenobiotics biodegradation and metabolism	0	5	1	1	0	1
	Unclassified	10	26	17	16	9	25
	Total	*271*	*461*	*414*	*302*	*174*	*355*
	% of KEGG in MAG	**52**	**51**	**53**	**50**	**46**	**48**
Genetic Information Processing	Folding, sorting and degradation	16	30	22	30	21	23
	Replication and repair	29	49	51	46	31	53
	Transcription	11	27	28	15	17	11
	Translation	79	127	89	74	64	87
	Unclassified	13	15	18	13	9	25
	Total	*148*	*248*	*208*	*178*	*142*	*199*
	% of KEGG in MAG	**29**	**27**	**27**	**30**	**38**	**27**
Environmental Information Processing	Membrane transport	60	107	87	76	37	79
	Signal transduction	12	21	20	8	2	25
	Signaling molecules and interaction	1	0	1	2	0	0
	Total	*73*	*128*	*108*	*86*	*39*	*104*
	% of KEGG in MAG	**14**	**14**	**14**	**14**	**10**	**14**
Cellular Processes	Cell motility	10	44	24	15	10	51
	Transport and catabolism	2	3	5	3	2	3
	Unclassified	15	27	22	18	8	22
	Total	*27*	*74*	*51*	*36*	*20*	*76*
	% of KEGG in MAG	**5**	**8**	**7**	**6**	**5**	**10**

Table S4.
Number of protein-encoding genes found within the CL500-3 MAGs assigned to specific KEGG metabolisms.

CAZymes	MJ-feb-125	SC-feb-99	MJ-june-47	SC-june-56	MJ-aug-102	SC-aug-92
MAG size (bp)	1233551	2473776	2002500	1593273	993108	2071391
Completeness (%)	45.19	73.24	69.81	54.55	40.65	45.76
Estimated genome size (bp)	2729699	3377630	2868500	2920757	2443070	4526641
Glycoside hydrolases (GH)	7	21	13	14	9	11
Estimated # of GH	15	29	19	26	22	24
Glycosyl transferases (GT)	11	16	20	20	12	27
Estimated # of GT	24	22	29	37	30	59
Polysaccharide lyases (PL)	2	0	2	2	2	2
Estimated # of PL	4	0	3	4	5	4
Carbohydrate esterases (CE)	12	19	15	7	4	16
Estimated # of CE	27	26	21	13	10	35
Carbohydrate binding modules (CBM)	5	10	14	12	4	5
Estimated # of CBM	11	14	20	22	10	11
Auxiliary Activities (AA)	0	2	0	1	1	3
Estimated # of AA	0	3	0	2	2	7
Dockerin	2	35	8	4	11	15
Estimated # of Dockerin	4	48	11	7	27	33

Table S5.
Actual and estimated numbers of CAZymes found in the CL500-3 MAGs.