Microbial Ecology of Phototrophs
in Iron-rich Boreal Shield Lakes

by

Jackson Makoto Tsuji

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Biology

Waterloo, Ontario, Canada, 2020

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner   Dr. Clara Chan
                                           Associate Professor, University of Delaware

Supervisor(s)   Dr. Josh D. Neufeld
                                           Professor, University of Waterloo

Internal Members   Dr. Laura A. Hug
                                           Assistant Professor, University of Waterloo

                                           Dr. Andrew C. Doxey
                                           Associate Professor, University of Waterloo

Internal-external Member   Dr. Sherry Schiff
                                           Professor, University of Waterloo

Author's Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Chapter 2

Under my supervision as undergraduate researchers, Norman Tran and Rachel Beaver assisted with enrichment culturing of *Chlorobia* members. Katja Engel extracted genomic DNA. A version of this chapter has been published as:

> Tsuji JM, Tran N, Schiff SL, Venkiteswaran JJ, Molot L, Tank M, Hanada S, Neufeld JD. Anoxygenic photosynthesis and iron–sulfur metabolic potential of *Chlorobia* populations from seasonally anoxic Boreal Shield lakes. ISME J 2020:1-16. doi:10.1038/s41396-020-0725-0

Chapter 3

Collection of lake samples was done as part of a larger sampling team including Rachel Henderson, Jennifer Mead, Jared Wolfe, Kate Thompson, Emily Barber, Eric McQuay, Kai Liu, Jason Venkiteswaran, Sherry Schiff, and Richard Elgood. Lake dissolved oxygen concentrations for 2016 were provided by the IISD-ELA, and lake surface areas were also provided by the IISD-ELA from the analyses of Ken Beaty, Sandra Rolph, and Kathryn Funk performed in 1998 and 2002. Under my supervision as a co-op student, Nicolette Shaw performed several of the DNA and RNA extractions presented in this work with assistance from Xinda Lu. Several DNA samples were also extracted by Emilie Spasov. Geochemical and stable isotopic measurements were performed by Kai Liu and Richard Elgood. Following collaborative experimental design and sample collection, Kate Thompson performed phototrophy-related rate measurements.

Chapter 4

Marcus Tank assisted with enrichment cultivation. Under my supervision as a co-op student, Nicolette Shaw also assisted with enrichment cultivation, extracted high molecular weight DNA, assisted with electron microscopy, and assisted with construction of the pigment infrared fluorescence detection device. A version of this chapter is available as a pre-print on the bioRxiv server as:

> Tsuji JM, Shaw NA, Nagashima S, Venkiteswaran JJ, Schiff SL, Hanada S, Tank M, Neufeld JD. Anoxygenic phototrophic *Chloroflexota* member uses a Type I reaction center. *bioRxiv* 2020. doi:10.1101/2020.07.07.190934

Abstract

The early Earth oceans of the Archaean Eon (approx. 3.8-2.5 billion years ago) are thought to have been predominantly anoxic with ferruginous waters containing high levels of dissolved ferrous iron and low levels of sulfate. How primary production operated in these ferruginous waters, where early life would have evolved, remains poorly understood. Based on the global occurrence of massive banded iron formations that can be dated to as early as the Mesoarchaean Era, many speculate that biological processes helped to drive the oxidation of iron in the Earth's oceans, which today are sulfate-rich and iron-poor. Along with oxygenic photosynthesis, where produced molecular oxygen can oxidize iron abiotically, alternative microbial processes such as photoferrotrophy, where ferrous iron serves as the direct electron donor for phototrophy, have been implicated in the deposition of banded iron formations. Other forms of anoxygenic phototrophy, such as phototrophic sulfide oxidation, could have also served as a means of primary production in early ferruginous waters. However, the diversity, ecology, and relative importance of anoxygenic forms of phototrophy in Archaean oceans remain poorly characterized given the current global prevalence of molecular oxygen.

Boreal Shield lakes number in the millions across northern regions. Although most Boreal Shield lakes mix and re-oxygenate seasonally, an estimated 15% of these lakes develop anoxic bottom waters over the summer or winter months that gain high levels of dissolved iron while having low levels of sulfate. As such, these Boreal Shield lakes can potentially complement permanently anoxic and ferruginous lakes that have been used to gain insights about the properties of ferruginous waters but are rare globally. The research presented in this thesis aimed to characterize the diversity and ecology of phototroph communities in the ferruginous water columns of seasonally anoxic Boreal Shield lakes. By performing a multi-year lake survey at the International Institute for Sustainable Development Experimental Lakes Area (IISD-ELA) combining high-throughput DNA/RNA sequencing, enrichment cultivation, physicochemical characterization, and rate measurements, the goal of this research was to clarify the metabolic diversity of phototrophy, physicochemical controls on phototrophs, and the relative importance of different modes of phototrophy in Boreal Shield lake anoxic zones with implications for modern ecology and for study of early Earth biogeochemistry.

Metagenomes were sequenced from preliminary DNA samples collected from Lakes 227 and 442 at the IISD-ELA to analyze the functional gene content of the lake microbial communities

(Chapter 2). Metagenomes were assembled and contigs binned to recover high-completeness and low-contamination draft genome bins. The genetic potential for sulfide oxidation, indicated by the *dsrA* gene, was found in three of the four recovered and curated genome bins classified to the *Chlorobia* class, but the fourth genome bin was found to contain the *cyc2* candidate marker gene for iron oxidation and was unique to Lake 227. Two new species of *Chlorobia* were enriched in sulfide-containing medium from Lakes 227 and 304. Recovered draft genome bins of both novel *Chlorobia* members contained both the *dsrA* and *cyc2* genes, indicative of the potential for both ferrous iron and sulfide oxidation. The Lake 304 strain, provisionally named, "*Candidatus* Chlorobium canadense", was subsequently incubated in ferrous iron-containing medium yet did not show signs of photoferrotrophic activity. Nevertheless, robust methods for the detection of *Chlorobia*-associated *cyc2* homologs were developed, and analysis of reference genomes indicated that the presence of the *cyc2* gene is associated with photoferrotrophic potential. Genomic potential for iron- and sulfur-cycling processes were also detected among several other recovered genome bins outside the *Chlorobia*. Overall, evidence for sulfide-oxidizing phototrophy was found in the lakes despite low sulfate levels, along with genomic potential for photoferrotrophy, and "*Ca.* Chl. canadense" was enriched to allow for future study of the functional role of *cyc2* among the *Chlorobia*.

Expanding on initial results from cultivation and metagenome analyses, nine Boreal Shield lakes at the IISD-ELA with diverse physicochemical characteristics were surveyed in the summer and/or fall from 2016-2018 to understand the broader distribution of phototrophs among ferruginous lake systems (Chapter 3). Lake anoxic zones samples generally grouped into two categories based on microbial community composition as inferred from 16S rRNA gene and metagenome data. One of those categories was associated with high relative abundances and diversities of anoxygenic phototrophs within the *Chlorobia* and *Chloroflexota* phyla and included three lakes. Lake categories based on microbial community composition generally matched categories determined via hierarchical clustering based on lake physicochemical parameters. Compared to the other surveyed lakes, the three detected "phototroph-rich" lakes tended to have low surface area:depth ratios with high dissolved organic carbon content, high dissolved iron content, and low levels of sulfate in their anoxic waters. Anoxic water column metagenomes from the three lakes contained high relative abundances of *Chlorobia* members along with *Chlorobia*-associated *cyc2* and *dsrA* genes. Metatranscriptome sequencing revealed that *Chlorobia* members comprised nearly 50% of total gene expression in the upper anoxic zones of these three lakes based on the *rpoB* single-copy taxonomic marker gene. Among *Chlorobia* members, the *dsrA* gene for phototrophic sulfide oxidation was highly expressed,

whereas the *cyc2* gene was expressed at lower levels for two lakes (Lakes 221 and 304) and minimally expressed for the third (Lake 227). Rate incubation data from Lake 227 suggested that oxygenic phototrophy was the dominant contributor to iron oxidation even 1 m below the measured oxic/anoxic zone boundary. Measured iron oxidation rates exceeded measured rates of iron reduction at all measured regions of the upper anoxic water column. Sulfide-oxidizing phototrophy was also detected and was tightly coupled to sulfate reduction. Photoferrotrophy was not detected but may have occurred at rates less than the observed iron reduction rate. Metatranscriptome data also revealed high expression of the particulate methane monooxygenase (*pmoA*) gene deep in the anoxic zone of Lake 227, implying that novel forms of methane oxidation occur in this lake. Altogether, the data demonstrate that Boreal Shield lakes commonly contain high relative abundance and diverse phototroph communities in their seasonal anoxic zones. In addition, the lake anoxic zones appear to host active and cryptic sulfur cycles, along with the potential for iron oxidation/reduction and methane cycling.

Enrichment cultivation in a ferrous iron-containing medium from Lake 227 allowed for the recovery and characterization of a novel anoxygenic phototrophic *Chloroflexota* member, provisionally named "*Candidatus* Chlorohelix allophototropha" (Chapter 4). This organism performs phototrophy using a distinct fourth clade of Type I photosynthetic reaction center (RCI) protein, despite placing sister taxonomically to Type II reaction center (RCII)-utilizing *Chloroflexota* members. "*Ca.* Chx. allophototropha" contains chlorosomes, uses bacteriochlorophyll *c*, and encodes the FMO protein like other RCI-utilizing phototrophs in the *Chlorobiales* and *Chloracidobacterales* orders. "*Ca.* Chx. allophototropha" also encodes the potential for carbon fixation using the Calvin-Benson-Bassham (CBB) cycle, unlike all known RCI-utilizing phototrophs. The discovery of "*Ca.* Chx. allophototropha", as the first representative of a novel *Chloroflexota* order (i.e., "*Ca.* Choroheliales"), sheds light on longstanding questions about the evolution of photosynthesis, including the origin of chlorosomes among RCII-utilizing *Chloroflexota* members. The *Chloroflexota* is now the only phylum outside the *Cyanobacteria* containing genomic potential for both major classes of photosynthetic reaction center and can thus serve as an additional system for exploring fundamental questions about the evolution of photosynthesis.

Acknowledgements

It's hard to know where to start in thanking those who have provided help and support for this thesis. This thesis represents the culmination of years of interdisciplinary scientific work but also flows from the collective input of many people into my life. I'll try my best to do justice to both sides.

On the scientific side, I have been astounded by how much support and training I have received to further my project and develop myself as a researcher. Back in 2015, I remember having a conversation with Dr. Josh Neufeld, then my undergraduate thesis supervisor, and being told about a new collaborative research venture with several of his colleagues in earth science to study iron-rich lakes in northern Canada. He told me that the project would push boundaries and challenge me but that he thought I was up to the task of taking it on in his lab as a grad student. I'm so glad I accepted his offer! Despite being the sole microbiology student on a team of limnologists and stable isotope geochemists and having just come out of undergrad, I was immediately welcomed onto the team and my input was valued. Thanks to Dr. Sherry Schiff, Dr. Jason Venkiteswaran, and Richard Elgood for "showing me the ropes" of lake sampling and guiding me on the major research questions in a field that was very new to me. You could have easily kept me as a lab microbiologist who did molecular work and grew bacteria, but you brought me out to the field at the IISD-ELA and asked me how to design several aspects of our study, which was both terrifying and exciting.

As our project to understand phototrophy in lake anoxic zones developed, we realized that additional training and resources would be needed to fulfill our goals. I am abundantly thankful to my supervisor Josh for finding some incredible opportunities for me in that regard. By connecting with Prof. Mike Jetten at Radboud University, I was welcomed as the first international participant in the Soehngen Institute of Anaerobic Microbiology (SIAM) "Talent Grant" program to come to The Netherlands and learn anaerobic microbial cultivation skills. Dr. Cornelia Welte at Radboud welcomed me as her trainee and even arranged opportunities to travel to Germany to learn from her former PhD supervisor who grew anoxygenic phototrophs. I am thankful to Mike, Cornelia, and the whole team at Radboud and SIAM for their generosity in what proved to be a very formative trip for my research. Josh connected me with the Joint Genome Institute (JGI) to meet some our growing sequencing needs, and together he and I wrote a grant application for metagenome sequencing of Boreal Shield lake water column samples. That application was funded, and when visiting the JGI facility in their annual user meeting and explaining our lake survey, they offered to more than double

On the personal side, thank you to my family for believing in me, encouraging me, and always being there for me. You helped me develop from a nervous young boy who stuttered to a PhD candidate who gives talks at international conferences. Thanks for building into me both skills and character and for modelling for me what it looks like to care for others. And thanks for listening to long explanations of my research and truly taking the time to learn about what I do! Thanks to Josh, my supervisor, for fostering such a helpful learning environment in the lab. Instead of stepping in and fixing problems when things looked too challenging for me, you provided the space and graciousness for me to work out solutions and grow as a person. I don't know how much you recognize that your approach to let me "spread my wings and fly" made for one of the best educational experiences of my entire life. For Sherry, thanks for welcoming me into your lab not just as a collaborator but as one of your own students. I hope to show the same generosity you've shown me to others. In my broader sphere, thanks to Aaron and Joanna T., Josh C., Alex L., Janice A., Bill and Carri T., Ruth S., Jeannie W., Shuping Z., Grace and Darren B., and Janet Y. for providing such support for me over the long years we've met and prayed together and for helping to shape for me a backbone of how to live not just as a researcher but as a member of my community. Lastly, I'm thankful to M. Berry who taught me how to think and write about modern literature (your input truly shaped how this thesis was written) and to J. Jamieson who sparked in me a deep interest about biology. To these and so many others, I express my sincere gratitude.

In my early years as a grad student, I thought my thesis work might be some kind of "magnum opus" for my life in science, but now I wonder if it is just a next step on a much longer journey of discovery and learning. Nevertheless, it's been a great ride. I present this work as something I hope will be valuable to the scientific community and as a product of the input I've received from many others.

Dedication

To David, Reta Ann, and Lillian – may you grow in wonder of the amazing world we live in.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANI	Average nucleotide identity
BChl *c*	Bacteriochlorophyll *c*
BLAST	Basic local alignment search tool
BLASTN	BLAST-nucleotide
BLASTP	BLAST-protein
Ca.	*Candidatus*
CBB cycle	Calvin-Benson-Bassham cycle
cbbL	Gene encoding the large subunit of RuBisCO
CCD	Charge-coupled device
CH_4	Methane
Chl.	*Chlorobium*
Chx.	*Chlorohelix*
CO_2	Carbon dioxide gas
csmA	Gene encoding bacteriochlorophyll *c* binding protein
DCMU	Diuron or 3-(3,4-dichlorophenyl)-1,1-dimethylurea
DIC	Dissolved inorganic carbon
DNA	Deoxyribonucleic acid
DO	Dissolved oxygen
DOC	Dissolved organic carbon
EDTA	Ethylenediaminetetraacetic acid
EET	Extracellular electron transfer
FMO protein	Fenna-Matthews-Olson complex protein
fmoA	Gene encoding the FMO protein
GTDB	Genome Taxonomy Database
GTDB-Tk	GTDB-Toolkit
HMM	Profile hidden Markov model

IISD-ELA	International Institute for Sustainable Development Experimental Lakes Area
IR	Infrared
LED	Light-emitting diode
MAG	Metagenome-assembled genome
mcrA	Gene encoding the alpha subunit of methyl coenzyme M reductase
N_2	Dinitrogen gas
NCBI	National Center for Biotechnology Information
NIR	Near infrared
PAR	Photosynthetically active radiation
PBS	Phosphate-buffered saline
PCC	Porin-cytochrome *c*
PCoA	Principal coordinate analysis
PCR	Polymerase chain reaction
pmoA	Gene encoding the alpha subunit of particulate methane monooxygenase
pscA	Gene encoding the core subunit of RCI
PSI	Photosystem I
PSII	Photosystem II
RCI	Type I reaction center
RCII	Type II reaction center
RNA	Ribonucleic acid
rpoB	Gene encoding the beta subunit of prokaryotic RNA polymerase
rRNA	Ribosomal RNA
SDS	Sodium dodecyl sulfate
SEM	Scanning electron microscopy
sp.	species
SSU	Small subunit
TEM	Transmission electron microscopy

Chapter 1

Introduction to phototrophy and the iron cycle

1.1 Literature review

1.1.1 The microbial iron cycle in aquatic systems: past and present perspectives

Iron is among the most abundant elements in the Earth's crust and essential for living organisms, yet iron biogeochemical cycling remains poorly understood (Boyd and Ellwood 2010). Key to the iron cycle is that two major redox states of iron exist in nature, ferrous iron (Fe^{2+}) and ferric iron (Fe^{3+}), and these tend to be highly soluble and poorly soluble, respectively, in waters of circum-neutral pH (Weber, Achenbach and Coates 2006). Both forms are highly redox active. Given that the surface of the modern Earth is predominantly oxidized due to the global prevalence of molecular oxygen, most modern freshwater and marine environments contain scarce amounts of iron in their upper waters. Only in the lower water column, if waters become hypoxic or anoxic, does dissolved ferrous iron begin to accumulate, for example due to release from sediments. Other ions are also released into aquatic systems upon hypoxia or anoxia that can influence the iron geochemistry of the system, including sulfide (S^{2-}), a corrosive product of sulfate-reducing bacteria, which reacts with ferrous iron when at sufficient concentration to produce insoluble iron sulfide precipitates (Amenabar *et al.* 2017). Due to these combined properties, iron tends to be a limiting nutrient in aquatic surface environments on the modern Earth, with organisms using aggressive strategies such as siderophores for iron accumulation (Boiteau *et al.* 2016). In waters having low dissolved oxygen levels, dissolved ferrous iron can potentially accumulate with peak concentration being influenced by other water column factors.

Recent advances in iron geomicrobiology have revealed a diverse array of iron-linked microbial metabolisms that occur in nature, and such findings have reshaped classical interpretations of the iron cycle (Melton *et al.* 2014; Bryce *et al.* 2018). Ferric iron can serve as an electron acceptor coupled to microbial oxidation of organic compounds or molecular hydrogen (Lovley 1991), performed by iron-reducing bacteria, or can serve as an electron acceptor coupled to ammonia oxidation in a recently discovered process called "Feammox" (Clément *et al.* 2005; Huang and Jaffé 2015, 2018). Conversely, ferrous iron can serve as an electron donor for microaerophilic iron-oxidizing bacteria, which rely on trace levels of oxygen as their electron acceptor to oxidize iron

1

aerobically (Emerson and Moyer 1997), or by acidophilic and aerobic iron-oxidizing bacteria (Temple and Colmer 1951). Ferrous iron can also be used as an electron donor for anaerobic metabolism by nitrate-reducing iron-oxidizing bacteria, which use nitrate as an electron acceptor (Straub *et al.* 1996). In a unique metabolic process, photoferrotrophic bacteria also oxidize ferrous iron anaerobically but do so in a normally non-thermodynamically favourable reaction, coupling iron oxidation to the reduction of dissolved inorganic carbon (e.g., bicarbonate or carbon dioxide) using light energy (Ehrenreich and Widdel 1994). Given these diverse possible microbial iron redox processes, which complement a suite of abiotic reactions reviewed elsewhere (Melton *et al.* 2014), the iron cycle, especially in redox transition zones, can potentially be highly dynamic and closely linked to other biogeochemical cycles. In this new and exciting period of iron cycling research, with access to new tools and an expanded understanding of iron biogeochemical pathways (Melton *et al.* 2014), the impetus for researchers exploring iron biogeochemistry is to foster highly interdisciplinary and integrative collaborative networks to reflect the newfound complexity of the field (Raiswell and Canfield 2012).

The role of diverse iron-cycling bacteria in nutrient cycling is increasingly apparent among global environments. Following their recent discovery in acidic wetland soils, signatures of Feammox bacteria have also been found in anammox bioreactors (Li *et al.* 2018), with the suggestion that the process could be more metabolically diverse in terms of ammonia oxidation products than previously recognized. Methane oxidation has recently been found to be coupled to ferric iron reduction by "*Ca.* Methanoperedens ferrireducens", with implications that such activity might be widespread globally (Cai *et al.* 2018). Iron-reducing bacteria in freshwater systems have been identified as having the potential for mercury methylation, meaning that these bacteria could contribute to key aspects of water quality management (Kerin *et al.* 2006; Bravo *et al.* 2018). In addition, ferric iron microfossils deposited by microaerophilic iron-oxidizing bacteria, usable as tracers of past and present activity of these microorganisms, have been found in lake and marine sediments and possibly fossils, hinting at the ecological importance of these bacteria through time (Chan *et al.* 2011). Nitrate-reducing and iron-oxidizing bacteria have been found in lake sediments and have been enriched from various aquatic systems and contaminated sites (Melton, Schmidt and Kappler 2012; Bryce *et al.* 2018). Photoferrotrophs have similarly been detected in sediments (Melton, Schmidt and Kappler 2012) and have also been found in the ferruginous water columns of several permanently anoxic lakes (reviewed below). Moreover, emerging evidence for widespread cryptic iron cycling (Berg *et al.* 2016; Kappler

2

and Bryce 2017) implies that rapid iron metabolism may occur even in environments with low measured iron concentrations.

Despite an expanded known diversity of iron-cycling microorganisms, exploration of the modern ecology of these microorganisms has been limited by the poorly understood genetics of microbial iron metabolism. Limited availability of diagnostic genetic methods hinders the rapid evaluation of iron-cycling metabolic potential in natural environments via nucleic acid sequencing. In one view, genetic knowledge of iron metabolism will remain largely incomplete because many iron cycling processes operate fundamentally differently at the gene level compared to more classically studied nutrient cycling processes such as those involved in nitrogen or sulfur cycling (Melton *et al.* 2014). Because iron metabolism is mediated by diverse and seemingly unrelated heme-binding cytochromes, and some iron metabolic processes are performed electrochemically (e.g., using nanowires) or entirely unconnected to the cell via the export of intermediary compounds, the likelihood of finding a well-defined set of marker genes may be low (Melton *et al.* 2014).

Nevertheless, recent work has begun to group genes involved in iron metabolism into broad classes. Three-gene clusters encoding porin cytochrome *c* complexes – structures that span the cell membrane to perform EET – have been implicated in iron metabolism, along with single genes encoding fused cytochrome-porin proteins that similarly facilitate EET (He *et al.* 2017). Iron-reducing bacteria often encode high numbers of heme- or multi heme-binding cytochromes, which could potentially serve as a diagnostic genomic marker (Methé *et al.* 2003; Cai *et al.* 2018). In addition, the recently published tool FeGenie attempts to summarize current knowledge about diverse iron cycling gene markers and provides an automated means for their detection in genomes (Garber *et al.* 2020). Such recent work offers promise that at least some diagnostic markers for iron metabolism exist at the genetic level. Future methods of iron metabolism detection, such as the use of machine learning techniques to detect whole-genome signatures of iron metabolism, might also have promise compared to traditional marker gene-based approaches. At the same time, the metabolic flexibility of many iron-cycling organisms (e.g., those that perform EET) means that even the presence of specific genes cannot necessarily be used to confirm iron metabolic activity. Many of the candidate gene systems that have been identified also still require verification at the biochemical level to understand their function and properties.

The diversity and ecology of iron-cycling bacteria is also of great interest for early Earth research. The Earth's oceans of the Archaean Eon (approx. 3.8-2.5 billion years ago) are thought to

have been predominantly anoxic and ferruginous, containing high levels of dissolved ferrous iron and low levels of sulfate (Poulton and Canfield 2011). Recent research suggests that ferruginous ocean conditions were even prevalent well into the Proterozoic Eon (approx. 2.5-0.54 billion years ago; Fakhraee *et al.* 2019). Such conditions differ markedly from the mostly oxygenated, sulfate-rich, and iron-poor conditions of modern oceans and suggests that the iron cycle was key to early microbial life. Geochemical and stable isotopic evidence from modern banded iron formations, many of which are dated to the Archaean, suggests that both iron oxidation and iron reduction processes were highly active in the late Archaean (Johnson *et al.* 2008; Konhauser *et al.* 2017). To explain these iron cycling processes, recent work on Archaean ocean biogeochemistry has increasingly incorporated knowledge of modern biological iron cycling rather than relied on abiotic factors alone. Along with oxygenic photosynthesis helping to drive the Great Oxidation Event at the end of the Archaean, iron-reducing bacteria are thought to have played an important role in mineral transformations associated with marine sediments (Johnson *et al.* 2008). In addition, the roles of alternative forms of microbial iron oxidation have been explored, including microaerophilic iron oxidation and photoferrotrophy (Posth, Konhauser and Kappler 2013). Microaerophilic iron-oxidizing bacteria could have accelerated iron oxidation after early oxygenic phototrophs began to generate molecular oxygen (Chan, Emerson and Luther 2016). Photoferrotrophy could have operated prior to the advent of oxygenic photosynthesis and provided a means of both iron oxidation and autotrophic carbon fixation for surrounding microbial life, including methanogens, with implications for global climate including modulation of the global methane cycle (Kappler *et al.* 2005; Thompson *et al.* 2019).

Stable iron isotopic evidence from analysis of Mesoarchaean banded iron formations (deposited approx. 3.8 billion years ago), combined with modelling, suggests that ferrous iron oxidation occurred in the Mesoarchaean water column at rates best explained by photoferrotrophy (Czaja *et al.* 2013). In addition, one model considering the combined activity of anoxygenic and oxygenic phototrophs in the water column suggests that, following the advent of oxygenic photosynthesis, anoxygenic photosynthesis processes such as photoferrotrophy could have delayed the oxidation of Earth to timelines better matching the rock record, compared to if oxygenic photosynthesis had operated alone (Ozaki *et al.* 2019). Any insights about early Earth microbial communities are highly applicable to astrobiology, given the possibility that other planets may possess geochemistry similar to that of the early Earth (Ozaki *et al.* 2018). Overall, early Earth research benefits from increased understanding of the modern microbial iron cycle and the understanding of the microbial ecology of iron-rich environments.

1.1.2 The diversity and evolution of chlorophototrophy

Phototrophy, the biological process of light energy conversion, is globally influential today and is thought to have had substantial impacts on the biogeochemistry of the planet over geologic time. Phototrophy is performed among living organisms in two major functional ways: rhodopsin-based and chlorophyll-based phototrophy (Thiel, Tank and Bryant 2018). Although rhodopsin-based phototrophy is widespread among both bacteria and archaea, due to its simplicity as a way to supplement cells with energy, chlorophyll-based phototrophy, or chlorophototrophy, has the potential for more complex uses of light energy including for electron transfer and fully autotrophic growth (Blankenship 2014). The diversity and evolution of chlorophototrophy is an area of much research interest and lies at the intersection of diverse fields including microbiology, biochemistry, genetics, biophysics, and geology.

Oxygenic chlorophototrophy is the most prolific and active form of phototrophy on a global scale and contributes an estimated 50 Pg carbon yr^{-1} to net primary production in aquatic systems (Raven 2009). The process is the only form of chlorophototrophy found outside of bacteria, being mediated by chloroplasts in eukaryotes in addition to bacterial members of the *Cyanobacteria* phylum. Oxygenic chlorophototrophy operates using two tandem photosynthetic reaction centers, PSII and PSI, which together convert light energy to a biologically usable form and successively excite electrons obtained from oxidizing water (Fischer, Hemp and Johnson 2016). Those electrons are used to generate reducing equivalents for the cell that can subsequently be used to fix carbon using the CBB cycle (Tang, Tang and Blankenship 2011).

Although less well characterized than oxygenic chlorophototrophy, various forms of anoxygenic chlorophototrophy exist that rely on alternative electron donors to water. Anoxygenic chlorophototrophs have been found to use sulfide, thiosulfate, ferrous iron, nitrite, arsenic(III), manganese(II), or molecular hydrogen as photosynthetic electron donors, and some are also capable of cyclic electron flow (Colbeau and Vignais 1981; Ehrenreich and Widdel 1994; Griffin, Schott and Schink 2007; Kulp *et al.* 2008; Johnson *et al.* 2013; Overmann and Garcia-Pichel 2013). Anoxygenic chlorophototrophs can be either aerobic (Koblížek 2015) or anaerobic with respect to their oxygen metabolism. They fix carbon using several different possible pathways, and some are reported to be photoheterotrophic (Tang, Tang and Blankenship 2011).

Unlike oxygenic chlorophototrophy, which aside from eukaryotes only occurs among *Cyanobacteria* members, anoxygenic chlorophototrophy is found among members of at least seven

bacterial phyla having unique photosynthetic structures and properties (Thiel, Tank and Bryant 2018; Ward, Cardona and Holland-Moritz 2019). All known anoxygenic chlorophototrophs rely on a single type of photosynthetic reaction center for light energy conversion, either a Type I reaction center (RCI), which is structurally and functionally analogous to PSI, or a Type II reaction center (RCII) that is structurally and functionally analogous to PSII (Fischer, Hemp and Johnson 2016). Known PSI and RCI primary sequences, as well as PSII and RCII primary sequences, have low but discernable homology, implying that the proteins have shared evolutionary origins (Sadekar, Raymond and Blankenship 2006). Remote structural homology between PSI/RCI and PSII/RCII imply that these general classes of photosynthetic reaction centers also together have a distant but common ancestor (Sadekar, Raymond and Blankenship 2006). Among anoxygenic chlorophototrophs, prior to my thesis research[1], members of the *Bacteroidota*, *Acidobacteriota*, and *Firmicutes* phyla were known to use RCI for phototrophy, and members of the *Chloroflexota*, *Proteobacteria*, *Gemmatimonadota*, and *Eremiobacterota* phyla were thought to use solely RCII (Thiel, Tank and Bryant 2018; Ward, Cardona and Holland-Moritz 2019). How photosynthetic reaction centers evolved and became distributed among extant microbial life remains a major research question, with many stating that lateral gene transfer played a substantial role in the spread of phototrophy-related genes to these diverse microorganisms (Martin, Bryant and Beatty 2018).

The timing of the evolutionary origin of various forms of phototrophy has been substantially discussed and debated (Olson and Blankenship 2005). It has been assumed previously that anoxygenic photosynthesis arose prior to oxygenic photosynthesis, but recent geochemical and biochemical data have challenged this view (Cardona 2019). Mounting evidence from the rock record suggests that low levels of molecular oxygen were present in selected areas of the planet well before the Great Oxidation Event approx. 2.5 billion years ago (Anbar *et al.* 2007); some place the earliest traces of atmospheric molecular oxygen at approx. 3 billion years ago (Crowe *et al.* 2013). Biochemical and structural analyses of photosynthetic reaction centers also indicate that oxygenic versions of reaction centers diverged from their anoxygenic counterparts early in the evolution of life and well before the Great Oxidation Event (Cardona 2018; Cardona *et al.* 2018; Cardona and Rutherford 2019). The development of anoxygenic chlorophototrophic metabolisms is also debated. Some argue that the earliest forms of anoxygenic chlorophototrophy likely oxidized sulfide and only

---

[1] Based on the work of Chapter 4 of this thesis, some *Chloroflexota* members are now also known to use RCI for anoxygenic chlorophototrophy.

later developed the capability to use ferrous iron as an electron donor (Martin, Bryant and Beatty 2018). Also, the importance of manganese(II) as a photosynthetic electron donor through geologic time has been subject to much discussion (Johnson *et al.* 2013; Jones and Crowe 2013; Daye *et al.* 2019). Nevertheless, it is likely that, at least by the late Archaean, multiple chlorophototrophic modes had evolved with the potential to operate contemporaneously.

1.1.3 Archaean ocean analogues: phototrophy in ferruginous water columns

To better understand the types of biogeochemical processes possible in Earth's early oceans, modern aquatic environments with geochemistry analogous to Archaean oceans have been identified and explored for their geochemical and microbial properties. As part of this effort, permanently anoxic (meromictic) and ferruginous lakes have been identified and studied as Archaean ocean analogues (Koeksoy *et al.* 2016). Such lakes are generally thought to be rare globally (Walter *et al.* 2014), and to date, seven such meromictic and ferruginous lakes have been studied in an Archaean ocean context (Figure 1.1). Study of these Archaean ocean analogues has allowed for a first glimpse into the nature of phototrophy and primary production in ferruginous environments (Llirós *et al.* 2015; Lambrecht 2019).

Anoxygenic phototrophic microbial communities have been detected in each of the studied meromictic and ferruginous lakes. However, the functional role of those phototroph communities has been less clear. In Lake Matano (Indonesia), anoxygenic phototrophs belonging to the *Chlorobia* class and related to the photoferrotroph *Chl. ferrooxidans* were detected in the illuminated chemocline (Crowe *et al.* 2008). However, rates of photoferrotrophy could not be measured in subsequent incubations. Instead, only sulfide-oxidizing phototrophy was measurable despite low sulfide levels in the water column (Crowe *et al.* 2014a). In Lake Pavin (France), although evidence for partial iron oxidation was detected in the water column based on stable iron isotope data, abiotic processes could potentially explain the observed iron isotope values, such as incomplete iron oxidation in the lake oxycline and precipitation of iron phosphates (Busigny *et al.* 2014). Anoxygenic phototrophs were detected in this lake but not assigned a definitive function (Berg *et al.* 2016).

Rate incubations in Kabuno Bay of Lake Kivu (Democratic Republic of the Congo) revealed that photoferrotrophy occurred in the upper anoxic zone of this lake at rates of up to 100 μM Fe^{2+} d^{-1}, and the novel photoferrotrophic species *Chl. phaeoferrooxidans*, a member of the *Chlorobia* class, was isolated from the water column (Llirós *et al.* 2015). A second novel species of photoferrotroph,
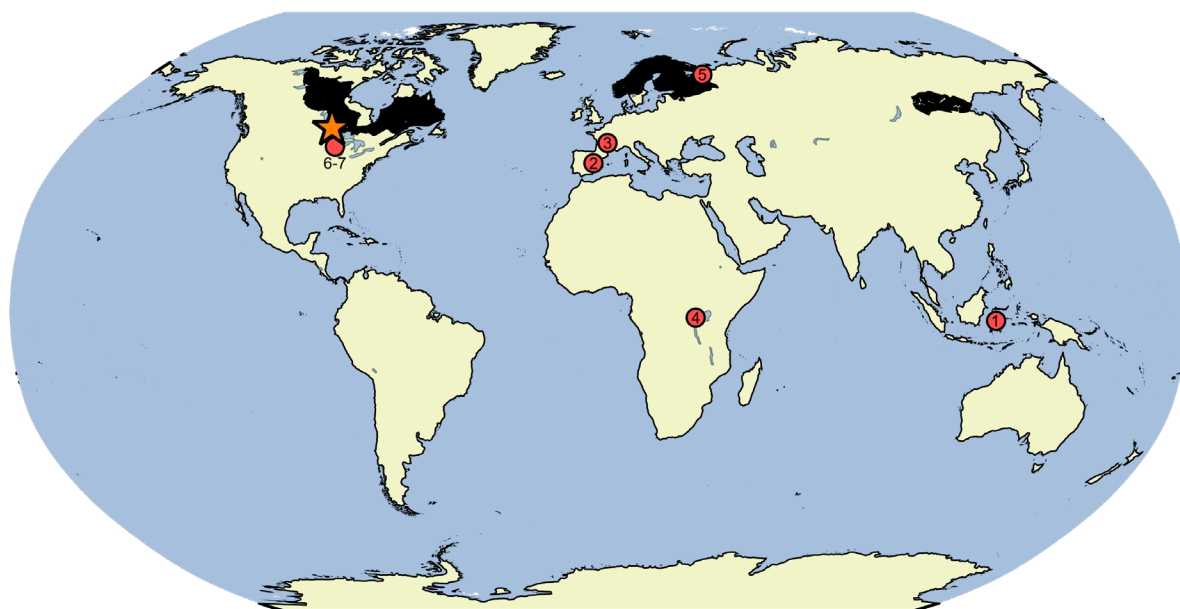
Figure 1.1 Study site and the global context of ferruginous lakes. Meromictic and ferruginous lakes that have been studied to date as analogues of Archaean ocean environments are marked by red circles. The Boreal Shield (black) likely contains additional meromictic and ferruginous lakes (e.g., 5) along with millions of seasonally anoxic ferruginous lakes as surveyed in this work. The IISD-ELA is marked with an orange star. 1, Lake Matano (Indonesia); 2, Lake La Cruz (Spain); 3, Lake Pavin (France); 4, Kabuno Bay of Lake Kivu (Democratic Republic of the Congo); 5, Lake Svetloe (Russia); 6-7, Brownie and Canyon Lakes (U.S.A.). Spatial data courtesy of ESRI and DMTI Spatial.

Chl. sp. strain BLA1, was also recently isolated from the water column of Brownie Lake (U.S.A.; Lambrecht 2019). Recent study of Lake Svetloe (Russia) also revealed populations of *Chlorobia* in the lake water column based on DNA sequencing, although their mode of phototrophy was unclear (Savvichev *et al.* 2017). Lastly, in Lake La Cruz (Spain), photoferrotrophic iron oxidation was measured, but rates were low – estimated at 0.2-1.4 μM Fe^{2+} d^{-1} – and microscopy and rate-based data indicated that other phototrophic processes, including phototrophic sulfide oxidation, likely dominated phototrophy in the water column (Walter *et al.* 2014).

Beyond phototrophy, the surrounding non-phototrophic microbial consortia in the ferruginous water column of some of the above lakes have also been examined. In Kabuno Bay of Lake Kivu, potential iron- and sulfate-reducing bacteria were found in the water column, along with potential methanogens and methanotrophs (Llirós *et al.* 2015). The ferruginous water column of Lake Pavin was found to have cryptic sulfur cycling activity, mediated by sulfate-reducing bacteria, as well as

8

possible cryptic iron cycling (Berg *et al.* 2019). Several lakes were found to have active methane cycles in their anoxic zones. In Lake Svetloe, DNA sequencing suggested that methanogens were present in the anoxic zone, and yet evidence for methane oxidation was also detected based on rate measurements, mediated by an unknown microbial consortium (Kadnikov *et al.* 2019). Somewhat similarly, Lake Matano's anoxic zone was found to have abundant methane with the potential for slow rates of methane oxidation, potentially linked to the reduction of manganese(IV) or ferric iron (Crowe *et al.* 2011).

Overall, observations from meromictic and ferruginous lakes indicate the presence of photoferrotrophs in several of these Archaean ocean analogues, with potential cryptic elemental cycles fuelling other processes in the water column, such as sulfide-oxidizing phototrophy. Phototrophs can exist in the context of a larger consortium of iron-, sulfur-, and methane-cycling microorganisms. Oxygenic phototrophs in the above waters could also impact phototroph communities. The data from these lakes also suggest substantial variability in phototrophic microbial communities among Archaean ocean analogue systems and points to the need for better diagnostic tests at the molecular level for iron metabolic potential. Although meromictic and ferruginous lakes are globally scarce, their study has provided key initial insights into the microbial community structure and function of ferruginous environments, relevant to study of the early Earth, while also raising several questions for future work.

1.2 Boreal Shield lakes as Archaean ocean analogues

Boreal Shield lakes occur in northern region globally on Precambrian Shield terrain (Figure 1.1), and an estimated 15% of such lakes develop anoxic bottom waters seasonally in the summer and/or winter months (Keller 2007; Schiff *et al.* 2017). Although the water columns of most Boreal Shield lakes mix and reoxygenate in the spring and/or fall, due to their underlying geology, many Boreal Shield lakes develop ferruginous conditions in their bottom waters once anoxic (Cook 1984). My thesis research began with the novel hypothesis that seasonally anoxic Boreal Shield lakes can serve as Archaean ocean analogues and supplement existing research at meromictic and ferruginous lakes. In contrast to meromictic and ferruginous lakes, which are rare, Boreal Shield lakes are estimated to number in the millions globally (Keller 2007; Teodoru *et al.* 2009), opening substantial new opportunities for research on ferruginous water columns. The abundance of Boreal Shield lakes could also mean that, in addition to relevance for Archaean ocean biogeochemistry and astrobiology,

understanding the ecology of ferruginous water columns could prove relevant for modern climate research in an era of global change (Anas, Scott and Wissel 2015).

In initial work preceding this thesis, the water columns of two Boreal Shield lakes at the IISD-ELA, Lakes 227 and 442, were sampled over three years (2011, 2013, and 2014), and samples were collected for geochemical and stable isotopic measurements, as well as for DNA sequencing (see full report in Appendix A). Both lakes were found to develop high levels of dissolved ferrous iron and low levels of sulfate and sulfide in their bottom waters over the summer months following the onset of anoxia (Figure A1). Differences in stable carbon isotopic values ($\delta^{13}C$) between dissolved inorganic carbon, particulate organic carbon, and methane indicated the possibility for carbon fixation below the oxic-anoxic zone boundaries of the lakes as well as the possibility of methane incorporation into particulate organic carbon in the anoxic zone (Figure A2). Dissolved and particulate iron isotope values indicated the possibility for partial iron oxidation in the lake water column (Figure A2), and low phosphate levels in the lakes ruled out the possibility for phosphate-driven ferrous iron precipitation as suggested previously in Lake Pavin. Microbial community analysis indicated that, below the oxic-anoxic zone boundary where low but measurable light penetration occurred (Supplementary Figure A1), high relative abundance populations of *Chlorobium* occurred, along with populations of potential iron- and sulfate-reducing bacteria, methanotrophs, and other sulfur-cycling microorganisms (Figure A3). Moreover, because both lakes were found to reoxygenate regularly at depth (Figure A4; Supplementary Figure A7), the observed anaerobic microbial communities must have recolonized the anoxic water column each year following seasonal mixing. These exciting data showed numerous parallels with existing research at meromictic and ferruginous lakes and suggested that Boreal Shield lake water columns were usable as Archaean ocean analogues.

1.3 Thesis outline

Based on these initial data, the goal of my thesis research was to characterize the diversity and ecology of phototrophic microbial communities in the anoxic zones of iron-rich Boreal Shield lakes. I combined metagenomics, enrichment culturing, biogeographical approaches, and activity measures such as metatranscriptomics and *ex situ* incubations, to understand the functional properties of anoxic zone phototrophs, controls on their distribution among Boreal Shield lakes, and the potential roles of phototrophs in mediating biogeochemical transformations in the lake anoxic water columns. I sought to distinguish between iron- and sulfur-oxidizing forms of phototrophy in particular and to test for the

presence of photoferrotrophy in the anoxic waters of the lakes. Through performing these first-ever characterizations of the interplay of anoxic zone phototrophs with boreal lake iron cycles, this study sought to expand sparse knowledge of the iron cycle in these globally abundant lake systems and to explore the potential applications of Boreal Shield lakes for understanding the microbial ecology of the Earth's early oceans.

This thesis is divided into three main research chapters, nested between this introduction and a concluding section. In each of the three chapters, a different approach was used to understand the novel iron geomicrobiology of Boreal Shield lakes at the IISD-ELA, a long-term ecological reference site in northwestern Ontario, Canada (Figure 1.1). In Chapter 2, metagenomics and enrichment culturing were combined to explore the potential for iron and sulfur cycling among *Chlorobia* members (i.e., "green sulfur bacteria"), which represented the highest relative abundance phototrophs detected in the lake anoxic zones. Two reference lakes, Lakes 227 and 442, were the main focus of that work, along with Lake 304, which was sampled for enrichment culturing. The usage of the *cyc2* candidate gene marker for photoferrotrophy was explored among members of the *Chlorobia* class, and an enriched population was tested for photoferrotrophic activity. The work of Chapter 3 was a direct continuation of concepts from Chapter 2. Here, results are reported from an expanded Boreal Shield lake survey encompassing nine lakes, with a gradient of physicochemical conditions, over three years. Biogeographical approaches were used to determine key physicochemical drivers of microbial community composition and their impacts on phototrophic consortia in the lakes. Activity measures were also used to probe the interactions of phototrophs with the iron and sulfur cycle in the anoxic zones of selected lakes. In Chapter 4, I report the enrichment of "*Ca.* Chx. allophototropha", a phototrophic member of the *Chloroflexota* phylum, enriched from Lake 227, that uses a novel class of Type I photosynthetic reaction center. This strain highlights the importance of examining poorly studied ferruginous environments and their resident phototrophs. The physiological and genomic properties of the strain are examined, along with their implications for the diversity and evolution of photosynthesis. Through characterizing the phototrophs in the ferruginous water columns of Boreal Shield lakes, this research provides first insights into the iron/sulfur geomicrobiology of globally abundant and seasonally ferruginous water bodies, with prehistoric and modern ecological relevance.

Chapter 2[2]

Anoxygenic photosynthesis and iron-sulfur metabolic potential of *Chlorobia* populations from seasonally anoxic Boreal Shield lakes

2.1 Introduction

Precambrian Shield regions in northern areas globally contain millions of small humic lakes (Keller 2007; Teodoru *et al.* 2009; Anas, Scott and Wissel 2015). Many of these Boreal Shield lakes develop a seasonally anoxic water column that becomes ferruginous, dominated by dissolved ferrous iron compared to sulfate and sulfide (Poulton and Canfield 2011; Schiff *et al.* 2017). However, little is known about the iron and sulfur microbiology of seasonally anoxic and ferruginous Boreal Shield lake systems. Several studies have examined the role of phototrophic members of the class *Chlorobia* (formerly phylum *Chlorobi*; also called "green sulfur bacteria") in boreal lake waters (Karhunen *et al.* 2013), because these organisms are commonly found at high relative abundances in the illuminated and anoxic water columns of boreal lakes. *Chlorobia* members are typically assumed to use a reduced sulfur compound, such as sulfide or thiosulfate, as their photosynthetic electron donor (Imhoff 2014a, 2014b). However, recent cultivation-based discoveries have expanded the known metabolic diversity of anoxygenic photosynthesis to include the photosynthetic oxidation of compounds such as ferrous iron, arsenate, and nitrite, and these alternative metabolic modes of photosynthesis challenge classical interpretations of how anoxic ecosystems function (Ehrenreich and Widdel 1994; Griffin, Schott and Schink 2007; Kulp *et al.* 2008; Thiel, Tank and Bryant 2018).

Among the *Chlorobia* class, diverse ferrous iron-oxidizing phototrophs, also called photoferrotrophs, have been cultivated recently. Previously, the only known photoferrotroph in this group was *Chl. ferrooxidans* strain KoFox, which was enriched from freshwater lake sediments (Heising *et al.* 1999). Recently, *Chl. phaeoferrooxidans* strain KB01, a bacteriochlorophyll *e*-containing member of the class *Chlorobia*, was isolated from the anoxic water column of the meromictic and ferruginous Kabuno Bay (Llirós *et al.* 2015). In addition, *Chl.* sp. strain N1 was enriched from marine sediments (Laufer *et al.* 2017). Both *Chl. ferrooxidans* and *Chl.*

phaeoferrooxidans oxidize ferrous iron as their sole photosynthetic electron donor, assimilating sulfur as sulfate (Heising *et al.* 1999; Llirós *et al.* 2015). *Chl.* sp. N1 is capable of oxidizing either ferrous iron or sulfide as the photosynthetic electron donor and can also utilize organic compounds (Laufer *et al.* 2017). A fourth recently isolated photoferrotroph, *Chl.* sp. BLA1, was shown to also be capable of oxidizing sulfide, although genome data was not yet available for this strain (Lambrecht 2019). Phylogenetically, *Chl. ferrooxidans* and *Chl. phaeoferrooxidans* are sister groups to one another, whereas *Chl.* sp. N1 is closely related to *Chl. luteolum*, a sulfide-oxidizing species of *Chlorobia* hypothesized to be able to switch to photoferrotrophic growth (Frigaard and Bryant 2008; Llirós *et al.* 2015; Laufer *et al.* 2017). These cultivation-based discoveries have highlighted the potential metabolic diversity and flexibility of *Chlorobia* members. Discovery of additional photoferrotrophs has also stimulated discussion around the potential involvement of photoferrotrophs in early Earth microbial communities, such as the ferruginous water columns of the Earth's oceans in the Archaean Eon (approx. 3.8-2.5 billion years ago; Llirós *et al.* 2015; Koeksoy *et al.* 2016; Camacho *et al.* 2017).

Despite growing evidence for the metabolic diversity of anoxygenic photosynthesis, the ecology of photoferrotrophy and other alternative photosynthetic modes remains poorly understood. One of the key challenges in assessing the modern importance of photoferrotrophy has been a limited understanding of the biochemistry and genetics of iron oxidation. Recent work has begun to unravel the molecular basis for iron oxidation in photoferrotrophs and other neutrophilic iron-oxidizing bacteria (Melton *et al.* 2014; He *et al.* 2017). Genes encoding for PCC protein complexes, such as *mtrABC*, *pioAB*, and *mtoAB*, have been identified as potentially useful markers for iron oxidation in some species (Melton *et al.* 2014; Kato *et al.* 2015; He *et al.* 2017; Gupta *et al.* 2019). Additionally, genes encoding for outer-membrane monoheme c-type cytochromes, such as *cyc2*, have received recent interest for their role in EET, with proteomic and RNA-seq studies indicating a role for such genes in microaerophilic iron oxidation (Castelle *et al.* 2008; Barco *et al.* 2015; McAllister *et al.* 2020). Among the *Chlorobia*, genome sequences available for *Chl. ferrooxidans*, *Chl. phaeoferrooxidans*, *Chl.* sp. N1, and *Chl. luteolum* show that their genomes encode *cyc2* homologues, unlike all other sulfide-oxidizing *Chlorobia* members (Frigaard and Bryant 2008; Crowe *et al.* 2017; Thompson *et al.* 2017; Bryce *et al.* 2019), providing preliminary evidence that *cyc2* might serve as a genomic marker for photoferrotrophy within the class *Chlorobia*. Confirming the validity of candidate iron oxidation gene markers such as *cyc2* for photoferrotrophic metabolism would greatly expand the potential to survey for photoferrotrophy and other forms of EET in nature.

Here, genome-resolved metagenomics and enrichment cultivation are combined to probe the metabolic diversity of anoxygenic phototrophs in seasonally anoxic and ferruginous Boreal Shield lakes, particularly examining the potential for photoferrotrophy. Following up initial 16S rRNA gene sequencing and iron isotope data that led to the hypothesis that photoferrotrophy is an important photosynthetic process in the water columns of two Boreal Shield lakes (Appendix A), draft genome bins of key populations of *Chlorobia*, the dominant anoxygenic phototrophs, are recovered from the same two lakes. In addition, two novel species of *Chlorobia* are enriched from the lakes and from a third Boreal Shield lake, allowing for assessment of the genetic and functional potential of these species for phototrophic sulfide and iron oxidation. Lastly, the usage of *cyc2* as a functional gene marker for photoferrotrophy is explored. Through this first usage of genome-resolved metagenomics to explore the metabolic diversity of *Chlorobia* in a ferruginous water column, this research aims to extend sparse knowledge of iron-sulfur microbiology in ferruginous lake systems, with implications for understanding both Archaean microbial communities and modern ferruginous environments.

2.2 Materials and methods

2.2.1 Lake sampling, metagenome sequencing, assembly, and binning

An initial round of lake sampling was performed previously from 2011-2014 at the International Institute for Sustainable Development Experimental Lakes Area (IISD-ELA; near Kenora, Canada), which is located at 49°40′N, 93°45′W (Appendix A). Lake 227, an experimentally eutrophied lake, was studied, along with Lake 442, a nearby and pristine reference lake (Schindler *et al.* 1971; Campbell 1994). The IISD-ELA sampling site, lake geochemistry, and sample collection methods were described in detail previously (Appendix A). Based on preliminary 16S rRNA gene data, six water column genomic DNA samples collected from Lakes 227 and 442 were selected for re-sequencing via shotgun metagenomics. All six samples were collected at or beneath the oxic-anoxic zone boundary, at depths where low light penetration occurs, and where high relative abundances of anoxygenic phototrophs, dominated by populations of *Chlorobia* had been detected previously. Samples from Lake 227 were selected at 6 m and 8 m in 2013 and 2014, and samples from Lake 442 were selected at 16.5 m in 2011 and 13 m in 2014. The sequencing library was prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs; Ipswich, Massachusetts, U.S.A.) and was sequenced on a single lane of a HiSeq 2500 (Illumina; San Diego, California, U.S.A.) in Rapid Run Mode with 2x200 base paired-end reads, generating 26.3 to

39.5 million reads per sample. Library preparation and sequencing was performed by the McMaster Genome Facility (Hamilton, Ontario, Canada).

Raw metagenome reads were quality-trimmed, assembled, binned, and annotated using the ATLAS pipeline, version 1.0.22 (White III *et al.* 2017; Kieser *et al.* 2020). The configuration file with all settings for the pipeline is available in the code repository associated with this work. To improve genome bin completeness and reduce bin redundancy, QC processed reads from related samples (i.e., L227 2013 6 m and 8 m, L227 2014 6 m and 8 m, and L442 samples) were co-assembled and re-binned using differential abundance information via a simple wrapper around ATLAS, *co-assembly.sh*, which is available in the atlas-extensions GitHub repository at https://github.com/jmtsuji/atlas-extensions. Briefly, the wrapper combines QC processed reads from the original ATLAS run for samples of interest, re-runs ATLAS on the combined reads, maps QC processed reads from the original samples onto the co-assembly, and then uses the read mapping information to guide genome binning. Version 1.0.22-coassembly-r3 of *co-assembly.sh* was used, relying on identical settings to the original ATLAS run, except that MEGAHIT was used for sequence assembly in place of metaSPAdes (Li *et al.* 2015; Nurk *et al.* 2017), MetaBAT2 version 2.12.1 was used for genome binning (Kang *et al.* 2019), and, for the L227 coassembly, a contig length threshold of 2200 was used.

2.2.2 Identification of *cyc2* genes and *Chlorobia* genome bins

Assembled contigs were assessed for the presence of *cyc2* gene homologues by building a custom HMM. The predicted primary sequences of *cyc2* were recovered from the four available genomes of *Chlorobia* known to possess the gene, *Chl. ferrooxidans*, *Chl. phaeoferrooxidans*, *Chl.* sp. N1, and *Chl. luteolum*, as well as from the genomes of reference microaerophilic iron-oxidizing bacteria as described elsewhere (He *et al.* 2017). A cytochrome 572 gene from *Leptospirillum* sp. (EDZ39515.1) was omitted from the reference dataset, due to its high divergence from other sequences, to build a more robust alignment for the *cyc2* clade relevant to the *Chlorobia* (Jeans *et al.* 2008). Collected sequences were aligned using Clustal Omega, version 1.2.3 (Sievers *et al.* 2011), and the alignment was used to build a profile HMM using the *hmmbuild* function of HMMER3, version 3.2.1 (Eddy 2011). Recovered *cyc2* genes over the course of the study were added to the alignment and HMM. The final HMM is provided with sequence alignments in the code repository associated with this work.

Recovered *cyc2* genes were compared to *cyc2* reference sequences by building a maximum-likelihood phylogeny. Due to poor sequence homology across much of the C-terminal end of the *cyc2* gene, phylogenetically uninformative residues in the alignment were masked using Gblocks, version 0.91b, with the flags "-t=p -b3=40 -b4=4 -b5=h", reducing the alignment size from 609 to 223 residues (Talavera and Castresana 2007). The maximum likelihood phylogeny was then prepared from the masked sequence alignment via IQ-TREE, version 1.6.10 (Nguyen *et al.* 2015), using the LG+F+I+G4 sequence evolution model as determined by the ModelFinder module of IQ-TREE (Kalyaanamoorthy *et al.* 2017). To build the consensus tree, 1000 bootstrap replicates were performed, each requiring ~100-110 tree search iterations for phylogeny optimization.

Genome bins were dereplicated using dRep version 2.0.5 (Olm *et al.* 2017) with default settings. All genome bins from individual assemblies and co-assemblies were pooled for dereplication to help ensure that the highest quality bins, whether from individual assemblies or co-assemblies, would be retained for downstream analyses. Genome bins of *Chlorobia* with > 90% completeness and < 10% contamination based on CheckM statistics (Parks *et al.* 2015) were selected for further study. These genome bins were imported into Anvi'o version 4 (Eren *et al.* 2015) and were manually examined for contigs improperly binned based on read mapping, tetranucleotide frequencies, and contig taxonomic classification. Import of ATLAS data into Anvi'o was performed using the *atlas-to-anvi.sh* script, version 1.0.22-coassembly-r4, available in the *atlas-extensions* GitHub repository at https://github.com/jmtsuji/atlas-extensions. Contigs in curated genome bins of *Chlorobia* were then ordered via the Mauve Contig Mover development snapshot 2015-02-13 for Linux (Rissman *et al.* 2009) using the *Chl. luteolum* genome to guide ordering. To assess bin quality, tRNA genes were predicted using Prokka v1.13.3 (Seemann 2014).

2.2.3 Enrichment cultivation, sequencing, and assembly

Lake 227 and Lake 442 were sampled again in June 2016 and July 2017, along with the nearby Lake 304 in July 2017. Basic sampling information for each lake is summarized in Table 2.1. Lake 304 has been described previously and also has seasonally anoxic bottom waters (Armstrong and Schindler 1971; Schindler 1974). The lake was experimentally fertilized with phosphate in 1971-72 and 1975-76, with ammonium in 1971-74, and with nitrate in 1973-74, but the lake has not been manipulated since then and rapidly returned to its non-eutrophied state once additions ceased, having a water residence time of ~2.7 years (Schindler 1985, 2012; Curtis and Schindler 1997). Lake 304 was added to sampling efforts for enrichment culturing to explore the broader distribution of

Table 2.1. Summary of basic physical, sampling, and historic parameters for Lakes 227, 442, and 304.

Lake	Maximum depth (m)	Mixing status	Experimentally eutrophied	Approx. oxic-anoxic zone boundary depth during sampling (m)	Depths sampled (m)	Key physicochemistry references
227	10	Mono to dimictic	1969-present	5	6, 8	(Schindler *et al.* 1971; Schiff *et al.* 2017)
442	17	Dimictic	Never	12	13, 15, 16.5	(Campbell 1994; Schiff *et al.* 2017)
304	6	Dimictic	1971-1976	5	6	(Armstrong and Schindler 1971; Schindler 1974, 1985)

photoferrotrophy among IISD-ELA lakes, because preliminary 16S rRNA gene sequencing data (not shown) indicated high relative abundances of *Chlorobia* populations in this lake. Anoxic water was collected from Lakes 227 and 304 at a depth of 6 m and Lake 442 at 15 m, where trace levels of light are generally detectable (i.e., in the range of 0.01-1 µmol photons m^{-2} s^{-1} between 400-700 nm wavelengths, as commonly measured when field sampling). Lake water was pumped to the surface using a gear pump and directly injected into N_2-filled 160 mL glass serum bottles that were sealed with blue butyl rubber stoppers (Bellco Glass; Vineland, New Jersey, U.S.A.), using a secondary needle to vent N_2 gas until the bottles were full. Water was kept cold (~4°C) and dark after collection and during shipping.

Enrichment cultures were grown using sulfide-containing Pfennig's medium, prepared as described by Imhoff (Imhoff 2014a, 2014b), or ferrous iron-containing freshwater medium, prepared as described by Hegler and colleagues (Hegler *et al.* 2008). The ferrous iron-containing medium contained 8 mM ferrous chloride, without filtration of precipitates, and used trace element solution SLA (Imhoff 2014a). Initial enrichments contained 10-20% lake water at a total volume of 50 mL and were inoculated into 120-160 mL glass serum bottles sealed with blue butyl rubber stoppers (Bellco Glass) or black rubber stoppers (Geo-Microbial Technology Company; Ochelata, Oklahoma, U.S.A.). The remaining headspace was flushed with a 90:10 N_2:CO_2 gas mix and left at a pressure of 150 kPa. Several bottles additionally had anoxic DCMU (Sigma-Aldrich; St. Louis, Missouri, U.S.A.) added to a concentration of 50 µM to block activity of oxygenic phototrophs (Clavier and Boucher 1992; Milucka *et al.* 2015). Bottles were incubated at 22°C and ~50 cm away from far red PARSource PowerPAR LED Bulbs (LED Grow Lights Depot; Portland, Oregon, U.S.A.) as the light source.

Cultures were maintained and purified in a variety of ways. Following initial enrichment, cultures were transferred with 1-10% inoculum into fresh media 2-4 times to continue to promote growth of the target phototroph. Dilutions to extinction were then performed in liquid culture with dilution factors ranging from 10^{-2} to 10^{-7} to eliminate contaminating organisms. Later, deep agar dilution series was performed on the same cultures to further enrich the target organisms (see below). Cultures were then transferred back to growth in liquid to yield higher cell biomass, with 5-10% inoculum typically being used in transfers between liquid cultures.

Two enrichment cultures survived repeated subculture and contained green sulfur bacteria based on pigment and marker gene sequence analyses. Biomass from these cultures was collected by centrifugation, and genomic DNA was extracted using the DNeasy UltraClean Microbial Kit (Qiagen; Venlo, The Netherlands). Genomic DNA was prepared into metagenome sequencing libraries using

the Nextera DNA Flex Library Prep Kit (Illumina), and the library was sequenced on a fraction of a HiSeq 2500 (Illumina) lane in High Output Run Mode with 2x125 base paired end reads. Library preparation and sequencing was performed by The Centre for Applied Genomics (TCAG; The Hospital for Sick Children, Toronto, Canada), generating ~6 million total reads per sample. Metagenomes were assembled using ATLAS version 1.0.22 without co-assembly. Genome bins of *Chlorobia* were manually refined as described above, using Anvi'o version 5 (Eren *et al.* 2015) via *atlas-to-anvi.sh* commit 99b85ac; contigs of curated bins were subsequently ordered, and tRNA genes counted, as described above.

2.2.4 Comparative genomics of *Chlorobia* genomes

Refined genome bins of *Chlorobia*, which belonged to the *Chlorobium* genus, were compared to genomes of reference strains from the *Chlorobiaceae* family. Genomes of all available type strains (as of 2017) and photoferrotrophs from the family were downloaded from the NCBI (for details, see the code repository associated with this work). Average nucleotide identity between genomes was calculated using FastANI version 1.1 (Jain *et al.* 2018). The phylogenetic relationship between genomes was determined by constructing a concatenated ribosomal protein alignment based on the rp1 set of 16 ribosomal protein genes using GToTree version 1.1.10 (Hug *et al.* 2016; Lee 2019). IQ-TREE version 1.6.9 (Nguyen *et al.* 2015) was used to construct the maximum likelihood phylogeny from the alignment. The LG+F+R4 model of sequence evolution was identified as optimal by the IQ-TREE ModelFinder module [33], and phylogeny construction used 1000 bootstrap replicates, each requiring ~100-110 tree search iterations for optimization.

The presence or absence of genes implicated in iron or sulfur cycling metabolism in the collection of *Chlorobiaceae* genomes was assessed using bidirectional BLASTP (Altschul *et al.* 1990). Genes were selected from the genomes of *Chl. ferrooxidans* and *Chl. clathratiforme* based on the genome comparison of Frigaard and Bryant (Frigaard and Bryant 2008). Predicted primary sequences of these genes were used to identify putative homologues across other genomes of *Chlorobia* using the BackBLAST pipeline, version 2.0.0-alpha2 (doi:10.5281/zenodo.3465955) (Bergstrand *et al.* 2016). The e-value cutoff for BLASTP hits was set at 10^{-40}, based on empirical testing of gene clusters known to be present or absent in reference genomes of *Chlorobia*, and the identity cutoff was set to 20%. Selected genes (e.g., *qmoA*, *dsrJ*) were omitted due to poor homology among reference genomes. Genes associated with photosynthesis and carbon fixation were similarly assessed using BackBLAST, except that most reference genes were selected from the genome of *Chl.*

tepidum according to Bryant and colleagues (Bryant *et al.* 2012) and Tourova and colleagues (Tourova *et al.* 2013).

To compare the ribosomal protein phylogeny and *cyc2* phylogeny for *Chlorobia* genomes containing *cyc2*, the amino acid sequence alignments used to construct the full phylogenies were subsetted to six relevant taxa and re-aligned. Maximum likelihood sequence phylogenies were constructed using IQ-TREE as described above for the full phylogenies. The tanglegram plot was visualized using Dendroscope version 3.5.10 (Huson and Scornavacca 2012).

2.2.5 Metagenome taxonomic and functional profiling

Relative abundances of genome bins within metagenome data were estimated by read mapping. The QC processed metagenome reads from each sample were iteratively mapped to all dereplicated genome bins (> 75% completeness, < 25% contamination) using bbmap version 38.22 (BBMap – Bushnell B. – sourceforge.net/projects/bbmap/) to determine the proportion of read recruitment to each genome bin. To minimize read mapping from closely related strains, bbmap was run with the "*perfectmode*" flag so that only identical reads would map; all other settings were identical to those in the ATLAS config file. Read recruitment was expressed in terms of the number of mapped metagenome reads to a genome bin divided by the total number of metagenome reads that mapped to assembled contigs. Overall, bin relative abundances are likely underestimated based on bbmap settings, which prevent SNPs from being detected, but are likely overestimated based on the calculation of read recruitment to assembled reads rather than total reads.

As a cross-comparison to the above genome bin-based method, pre-assembled metagenome reads were directly assessed for gene relative abundances. Open reading frames were predicted for QC processed metagenome reads using FragGeneScanPlusPlus commit 299cc18 (Singh *et al.* 2018). Predicted open reading frames from the pre-assembled reads were then used with MetAnnotate development release version 0.9.2 (Petrenko *et al.* 2015) to scan for genes of interest using the HMM queries mentioned above and to classify the hits taxonomically. MetAnnotate performed taxonomic classification using the USEARCH method against the RefSeq database (release 80; March 2017). The default e-value cutoff of 10^{-3} was used for assigning taxonomy to hits. Relative abundances of phylotypes were calculated and visualized based on MetAnnotate results using *metannotate_barplots.R* version 1.1.0, available at https://github.com/jmtsuji/metannotate-analysis, with a HMM e-value cutoff of 10^{-10} to accommodate the shorter lengths of HMM hits.

Taxonomy and functional gene information were then overlaid onto relative abundance data. Genome bin taxonomy was determined using the GTDB-Tk, version 0.2.2, relying on GTDB release 86, version 3 (Parks *et al.* 2018; Chaumeil *et al.* 2020). As such, GTDB taxonomy names are used throughout this manuscript. Genomes were tested for the presence of the *cyc2* and *dsrA* functional genes using the HMM developed in this study and an HMM available on FunGene (Fish *et al.* 2013), respectively. Gene amino acid translations were predicted for all genome bins using prodigal version 2.6.3 (Hyatt *et al.* 2010), via the GTDB-Tk, and amino acid translations were searched via HMMs using the MetAnnotate pipeline, version 0.9.2 (Petrenko *et al.* 2015). An e-value cutoff of 10^{-40} was used to filter low-quality gene hits.

Genome bins were also probed for their broader iron- and sulfur-cycling genetic potential. To assess the potential for iron cycling, FeGenie (Garber *et al.* 2020) commit 30174bb was used to scan the genome bins using default settings. Results were filtered to only those in the "iron oxidation" or "iron reduction" categories. To assess for sulfur-cycling genetic potential, amino acid predictions of the genome bins, generated by the GTDB-Tk, were scanned for the *aprAB*, *dsrAB*, and *soxB* genes using HMMs downloaded from FunGene (Fish *et al.* 2013) using hmmsearch version 3.1b2 (Eddy 2011). For known gene clusters (e.g., *aprA* and *aprB*), hits from the genome bins were verified to be directly adjacent to one another along the genome as expected.

2.2.6 Assessment of ferrous iron oxidation potential of *Chlorobia* enrichments

After initial sequencing, cultures containing *Chlorobia* spp. continued to be purified in the laboratory. A single culture, which was enriched on sulfide-containing medium, survived continued cultivation and was provisionally named "*Candidatus* Chlorobium canadense strain L304-6D" (ca.na.den'se N.L. neut. adj. *canadense* from or belonging to Canada). "*Ca*. Chl. canadense" was purified through multiple rounds of incubation in deep agar dilution series and picking of isolated green colonies (Pfennig 1978). For deep agar shakes, Pfennig's medium or modified *Chloracidobacterium thermophilum* Midnight medium (Tank and Bryant 2015) containing 1-3 mM buffered sulfide feeding solution were used (Imhoff 2014a). Eventually, in preparation for growth in ferrous iron-containing medium, cultures of "*Ca*. Chl. canadense" were transferred to growth in liquid freshwater medium as described by Hegler and colleagues (Hegler *et al.* 2008) (here termed, "Hegler freshwater medium") but containing 0.5-1 mM buffered sulfide feeding solution, 10-20 µM ferrous chloride, and 0.5 mg/L resazurin to monitor the medium's redox status.

To test their ability to oxidize ferrous iron phototrophically, "*Ca.* Chl. canadense" cells were inoculated into Hegler freshwater medium containing low levels of ferrous iron and lacking sulfide. The base Hegler freshwater medium (before addition of ferrous iron or sulfide) was aliquoted into 120 mL glass serum bottles under flow of sterile 90:10 N_2:CO_2 gas, with 50 mL of medium per bottle. Bottles were sealed with black rubber stoppers (Geo-Microbial Technology Company) and were bubbled for at least nine minutes with additional 90:10 N_2:CO_2 gas to purge trace oxygen. The headspace of each bottle was left at 150 kPa pressure. A sterile solution of ferrous chloride was added to each bottle to reach a ferrous iron content of 100 μM, and to some bottles, a sterile and anoxic solution of the chelator EDTA was added to a concentration of 120 μM as done elsewhere (Peng *et al.* 2019) to explore whether this chelator could enhance photoferrotrophic iron oxidation. Bottles were incubated at room temperature in the dark overnight to allow for complexation of ferrous iron and EDTA. All media was confirmed to have a pH of 6.5-7.

"*Ca.* Chl. canadense" was grown in 100 mL Hegler freshwater medium using a total of approx. 2 mM sulfide, fed incrementally in 0.4 mM doses, until sulfide was completely oxidized and the resazurin in the bottle went slightly pink. At the same time, a culture of *Chl. ferrooxidans* KoFox was grown in the same freshwater medium containing approx. 8 mM ferrous chloride until ferrous iron was completely oxidized. The entire 100 mL "*Ca.* Chl. canadense" culture was pelleted by centrifugation at 7000 x g for 13 minutes, and the pellet was washed twice with unamended Hegler freshwater medium. The 200 mg of wet biomass recovered was suspended in 1 mL of unamended freshwater medium, and 0.1 mL was inoculated into each relevant incubation bottle. Similarly, 100 mL of the *Chl. ferrooxidans* reference culture was pelleted by centrifugation at 7000 x g for 5 minutes, washed twice, and suspended in 1 mL of unamended freshwater media, and 0.1 mL was added to relevant bottles. Only the surfaces of *Chl. ferrooxidans* cell pellets, which contained mostly green cells and not brown ferric iron precipitate, were suspended with each wash, allowing for 3 mg of nearly iron-free wet biomass to be recovered. Once cultures were inoculated, they were transferred to a 22°C incubator without shaking where they received 30 μmol photons m^{-2} s^{-1} white light from a mix of fluorescent (F48T12/CW/VHO; Osram Sylvania; Wilmington, Massachusetts, U.S.A.) and incandescent (60W) bulbs.

Cultures were sampled regularly over a 21-day incubation period to monitor iron concentrations and iron oxidation states. At each sampling time point, an aliquot of culture was removed from each bottle using a sterile 90:10 N_2:CO_2-flushed syringe and 25 G needle, and 330 μL

of this culture aliquot was acidified immediately with 30 μL of 6 N hydrochloric acid. Acidified samples were stored for no more than two days at 4°C before being assessed for iron species concentrations via the ferrozine assay (Stookey 1970), with 5% ascorbic acid being used as the reducing agent to determine total iron species (Verschoor and Molot 2013). As one exception, the acidified samples collected on day 14 were stored at 4°C for eight days before being assayed, yet no clear abnormalities of iron concentrations were observed for these samples compared to other samples in the time series.

To assess the purity of the culture, a positive control of "*Ca.* Chl. canadense" grown in a sulfide-containing medium, using the same inoculum as for the photoferrotrophy test, was examined using confocal laser scanning microscopy. A 5 mL aliquot of culture was pelleted, resuspended in 50 μL of supernatant, and visualized as a wet mount on a Zeiss LSM 700 confocal laser scanning microscope. To detect the autofluorescence of bacteriochlorophyll *c*/*d*/*e*-containing cells, the sample was excited using a 488 nm laser, and light emissions from 600-800 nm were measured. Transmitted light was also measured to provide information about the non-fluorescent portions of the slide.

In addition, for a previous subculture, genomic DNA was extracted from a pellet of cell biomass using the DNeasy UltraClean Microbial Kit, and 16S rRNA gene amplicon sequencing was performed, as described by Kennedy and colleagues (Kennedy *et al.* 2014), to identify the key contaminants in the culture. As an exception from the protocol by Kennedy and colleagues, the primers 515f and 926r (targeting the V4-V5 hypervariable region) were used for amplification (Parada, Needham and Fuhrman 2016), and PCR was performed in singlicate. During PCR, samples were incubated in the thermocycler at 95°C for 10 minutes, then incubated for 35 cycles of 95°C for 30 seconds, 50°C for 30 seconds, and 68°C for 1 minute, and finally incubated at 68°C for 7 minutes before being held at 12°C. Sequencing data was processed using QIIME2, version 2019.10.0 (Bolyen *et al.* 2019), within the AXIOME3 pipeline (https://github.com/neufeld/AXIOME3), development commit e35959d. Specifically, DADA2 was used to trim primer regions from raw sequencing data, merge paired end (2x250 base) reads, perform sequence denoising, and generate an ASV table (Callahan *et al.* 2016). Taxonomic classification of ASVs was performed using QIIME2's scikit learn-based taxonomy classifier against SILVA release 132, trimmed to the V4-V5 region, as a reference database (Quast *et al.* 2013), and this information was overlaid on the ASV table. The SILVA classifier was trained using QIIME2, version 2019.7.0, which relies on the same version of scikit-learn as 2019.10.0.

2.3 Results

2.3.1 Recovery of *Chlorobium* genome bins

Binning of assembled contigs from the lake and enrichment culture metagenomes, followed by dereplication of highly similar bins and manual curation, allowed for the recovery of six highly complete, low-contamination genome bins that classified within the genus *Chlorobium* (Table 2.2). Three of the genome bins recovered from lake metagenomes had best representatives selected by dRep from the Lake 227 (2013 sample) co-assembly, and one had its best representative from the Lake 442 (2011/2014 sample) co-assembly. In addition, a single *Chlorobium* genome bin was recovered from each enrichment culture. Recovered bins had at least 90.1% completeness and a maximum of 2.8% contamination based on analysis by CheckM (Parks et al. 2015). Recovered genome bins were between 1.9 to 2.6 Mb in size, within the approximate length range of reference genomes associated with the family *Chlorobiaceae* (2.0-3.3 Mb; average 2.7 Mb), and were represented by 13 to 208 contigs. Other lower quality genome bins classified as belonging to the *Chlorobia* were also recovered (see the code repository associated with this work), but these were not considered for gene pathway analysis in order to minimize the risk of including false positives. In particular, L442 Bin 74 was excluded from the set of curated bins due to its high contig count (282) and elevated predicted contamination (4.1%) and strain heterogeneity (25%).

The constructed HMM for the *cyc2* gene enabled the recovery of five potential *cyc2* homologues associated with *Chlorobia* from assembled contigs (Figure 2.1). Three of these *cyc2* homologues were found in the set of six manually curated genome bins described above, and another homologue was detected in one of the lower quality genome bins of *Chlorobia* (L227 2014 Bin 92). The final homologue was detected on a short contig in the same lower quality genome bin and was removed from subsequent analyses due to its lack of gene neighbours. Detected *cyc2* homologues were full-length genes, and their predicted primary sequences were detected in the HMM search with a maximum e-value of ~10-140. All recovered *cyc2* homologues contained the heme-binding CXXCH motif near the N-terminus of their predicted primary sequence, and all homologues were predicted by I-TASSER (Yang et al. 2015) to contain a porin-like beta barrel structure for membrane binding (Figure 2.1A; one example shown; He et al. 2017). Aside from the closely related genomes of *Chl*. *ferrooxidans* and *Chl*. *phaeoferrooxidans* (Figure 2.2), all other genomes showed substantial genome re-arrangement around the *cyc2* homologue (Figure 2.1B). However, detected *cyc2* homologues were always adjacent to a *c5* family cytochrome (Figure 2.1B). These *c5* family

Table 2.2 Quality statistics for genome bins of *Chlorobia* recovered in this study. Completeness and contamination were calculated using CheckM (see Methods). Note that no ribosomal RNA genes were recovered for the bins.

Bin ID	Contigs	Length (Mb)	N50 (kb)	L50	GC content (%)	Genes	tRNA genes	Completeness (%)	Contamination (%)
L227 2013 Bin 55	208	2.38	16.7	42	48.5	2397	39	97.4	0.1
L227 2013 Bin 56	140	2.09	27.6	21	43.9	2009	38	96.7	0.6
L227 2013 Bin 22	57	2.55	79.5	12	45.9	2471	37	98.5	2.8
L442 Bin 64	128	1.9	28.6	22	48.3	1871	35	90.1	0.0
L227 enr. S-6D Bin 1	13	2.5	306.7	3	46.8	2385	44	99.5	1.0
Ca. Chl. canadense L304-6D	15	2.61	271.5	3	48.7	2504	45	99.4	0.0

cytochromes had low sequence identity to one another (i.e., as low as 30%) yet also contained a conserved CXXCH motif (Figure 2.3). All recovered *cyc2* homologues grouped monophyletically with *cyc2* predicted primary sequences belonging to reference genomes of *Chlorobia* compared to the *cyc2* of other reference strains (Figure 2.1C). Overall, the combination of HMM search specificity, genomic context, predicted gene motifs, and phylogenetic placement is strong evidence that the identified cytochrome genes are *cyc2* homologues.

2.3.2 Enrichment cultivation of Chlorobia

Enrichment cultivation was attempted from anoxic lake water using both sulfide- and ferrous iron-containing anoxic media. For Lake 442, one enrichment in a ferrous iron-containing medium grew with evidence of light-driven ferrous iron oxidation, but 16S rRNA gene sequencing showed that the culture was dominated by a member of the genus *Rhodopseudomonas*, which did not

Figure 2.1 Recovered *cyc2* homologues of *Chlorobia* from metagenomes. **A**, Homology model of the Cyc2 protein from *Chl. phaeoferrooxidans* KB01, generated by I-TASSER based on sequence homology (with leading 18 signal peptides trimmed according to SignalP; Petersen *et al.* 2011) and visualized using UCSF Chimera (Pettersen *et al.* 2004). The porin-like beta barrel structure is highlighted in dark blue, and the location of the single N-terminal heme-binding motif on the outer membrane portion of the structure is indicated by a black arrow. **B**, Genomic context of the recovered *cyc2* homologues. A 10 kb region is shown of assembled contigs surrounding the homologues. The *cyc2* and adjacent *c5* family cytochrome genes are highlighted in green and pink, respectively. Other shared genes between contigs are highlighted in yellow. Remaining genes are coloured grey if they lacked a functional annotation (e.g., no significant BLASTP hit, or hit to a domain of unknown function) or are coloured purple if their closest BLASTP hit corresponded to a gene associated with the *Chlorobiaceae* family. One gene in the 10 kb region had its closest BLASTP hit to a gene associated with the *Comamonadaceae* family (*Betaproteobacteria*; ~51% amino acid identity); this gene is highlighted in blue. **C**, Sequence comparison of recovered *cyc2* homologues to *cyc2* genes of known iron-oxidizing microorganisms. The phylogenetic tree was built from a 223 residue masked Cyc2 amino acid sequence alignment (see Methods). Bootstrap values over 50/100 are shown, and the scale bar represents the proportion of residue changes along the alignment. Due to the uncertain evolutionary history of *cyc2*, the phylogeny is midpoint rooted. A green box highlights the monophyletic *Chlorobia* clade. Adjacent to the phylogeny, a subset of the unmasked Cyc2 sequence alignment is shown (positions 40-80 of 609; N-terminus end). Alignment positions having 100% sequence identity in the predicted heme-binding site (CXXCH motif) are marked with red asterisks. Other positions with 100% sequence identity are marked with black asterisks. Note that L227 2014 Bin 92 (name marked with an asterisk) was included in this figure for comparison of its *cyc2* gene despite the bin quality being lower than the main *Chlorobia* genome bin set.

represent a detectable genus in environmental sequencing data. Given the known metabolic versatility of *Rhodopseudomonas* spp. and the negligible contribution of members of this genus to the studied lake microbial communities, this culture was not pursued further. No members of the *Chlorobia* were ever detected in any enrichment culture from Lake 442.

For Lakes 227 and 304, growth was observed for both media tested. Enrichments grown in sulfide-containing medium developed green colour after ~12-15 weeks of incubation and contained populations of *Chlorobia* based on Sanger sequencing of both the V3-V4 region of the 16S rRNA
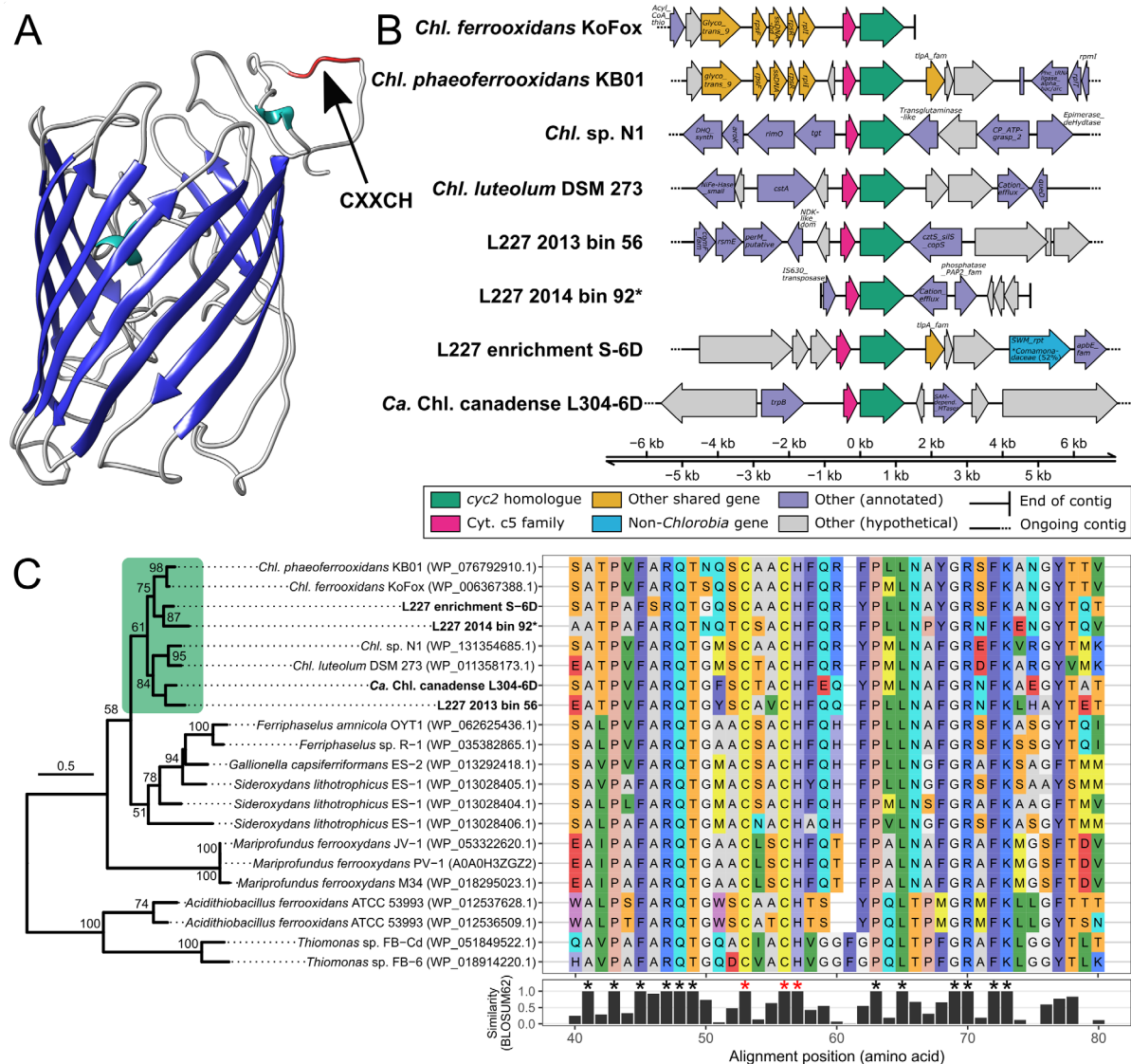
gene and the partial *dsrAB* gene. Gene regions were amplified using *Chlorobia*-targeted PCR primers 341f/GSB822r (Overmann, Coolen and Tuschak 1999) and PGdsrAF/PGdsrAR (Mori et al. 2010), respectively. However, the appearances of cultures grown in ferrous iron-containing medium for these lakes differed from those of reference photoferrotrophic *Chlorobia* members. Instead of forming a reddish-brown ferric iron precipitate, the enrichments blackened, potentially indicative of sulfate reduction and metal sulfide formation. After blackening, enrichments developed a green colour (Figure 2.4). Suspecting that sulfate reduction was occurring in the ferrous iron enrichment bottles to support sulfide-oxidizing phototrophy, one Lake 227 enrichment was transferred into sulfide-containing media and still developed green pigmentation, albeit slowly (Figure 2.4). Amplification and Sanger sequencing of the partial *dsrAB* gene cluster from this enrichment showed that the gene sequence only differed by one ambiguous base from, and thus was essentially identical to, a corresponding Lake 227 sulfide enrichment. Ferrous iron enrichments eventually stopped being followed due to long growth times, low biomass, and instability of the cultures. Enrichments on sulfide-containing media were continued instead.

Metagenome sequencing of the two successful enrichments of *Chlorobia* grown on sulfide (L227 enrichment S-6D and L304 enrichment S-6D) showed that genomes bins of the *Chlorobia* from both enrichments encoded *cyc2* homologues. Although L227 enrichment S-6D ceased to grow in laboratory culture during the early enrichment process, the L304 enrichment S-6D continued to be purified and was named "*Candidatus* Chlorobium canadense strain L304-6D". At the time of running the ferrous iron oxidation test (below), the "*Ca*. Chl. canadense" culture consisted of >90% cells of *Chlorobia* based on epifluorescence microscopy (Figure 2.5). Other contaminating cells potentially represented chemoorganoheterotrophs and sulfate reducing bacteria based on the taxonomic classifications of ASVs from previous 16S rRNA gene amplicon sequencing data (see the code repository associated with this work). In total, for "*Ca*. Chl. canadense", 13 subcultures were performed from the initial lake water inoculum until the iron oxidation experiment was performed.
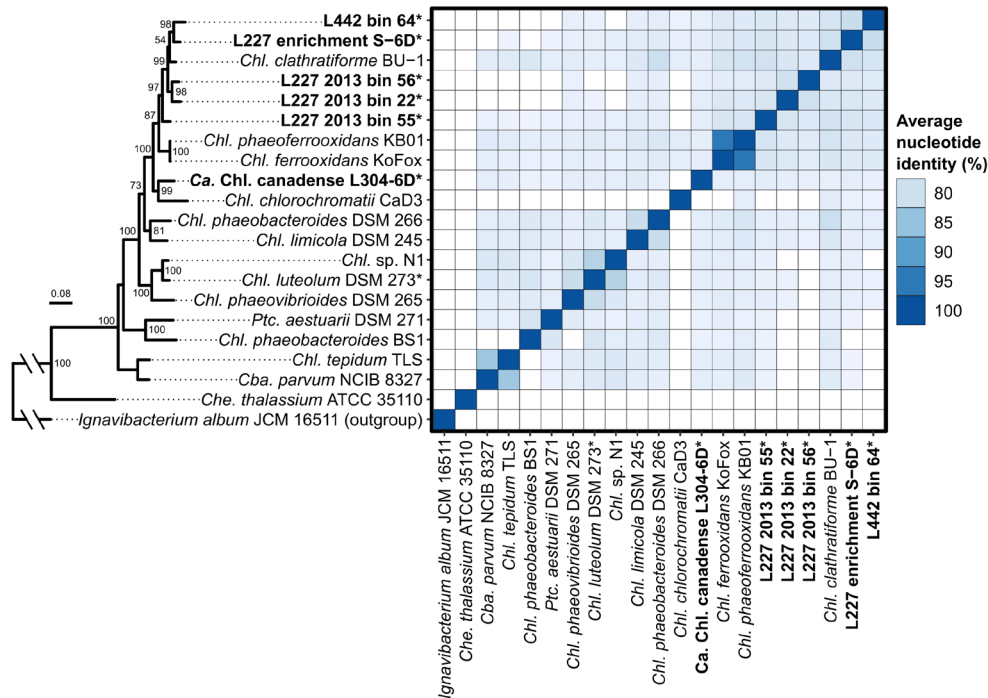
Figure 2.2 Average nucleotide identity between recovered genome bins and reference genomes of *Chlorobia*. The same ribosomal protein tree as in Figure 2.6 is shown to the left of the heatmap.
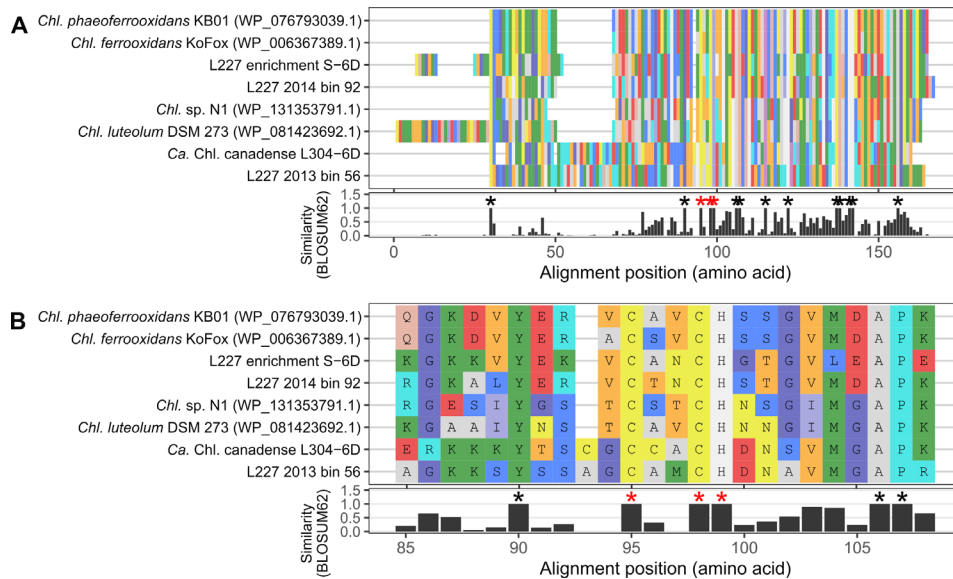
Figure 2.3 Multiple sequence alignments of *c5* family cytochromes adjacent to *cyc2* genes in the genomes of *Chlorobia* explored in this study. A, Graphical representation of the full *c5* family cytochrome sequence alignment. **B**, Expanded view of the same alignment in the region surrounding the heme-binding CXXCH motif. Black asterisks indicate regions in the alignment with 100% conserved residues, and red asterisks indicate the conserved amino acids in the CXXCH motif.

Media controls:
Left: Pfennig's medium (S2-)

Right: Freshwater medium from Hegler and colleagues, 2008 (Fe^{2+})

Reference culture of *Chl. ferrooxidans* KoFox grown on Hegler medium

L304 enrichment S-6D
Inoculated Aug. 8th, 2017

Subcultured Nov. 23rd, 2017 (~3.5 months later)

Typical growth time of ~3 weeks

Cells ~0.8-1 x 1.2-1.5 µm in size

Named "*Ca.* Chl. canadense L304-6D"

| Aug. 24th, 2017 | Nov. 23rd, 2017 | Healthy subculture | Simple stain (dry mount); bright field microscopy |

L227 enrichment S-6D
Dates of initial inoculation and subculture same as above

Typical growth time of ~4-8 weeks (slower growing)

Cells ~0.7-0.9 x 1-1.5 µm in size

Strain was lost in May 2018
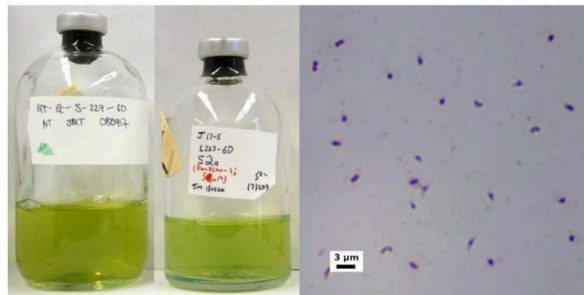
| Nov. 23rd, 2017 | Healthy subculture | Simple stain (dry mount); bright field microscopy |

L227 enrichment Fe-6E
Inoculated into Hegler medium (Fe^{2+}) on Aug. 8th, 2017

Developed greenish-black colouration

Eventually transitioned to sulfide-containing medium (Pfennig's), although growth was poor; *dsrA* sequence confirmed identical to L227-S-6D above
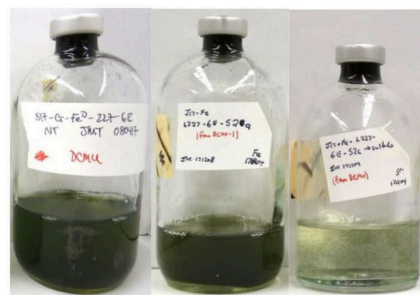
| Oct. 19th, 2017 | Subculture, Hegler | Subculture, Pfennig |

Figure 2.4 Photographs of enrichment cultures of *Chlorobia* in ferrous iron- and sulfide-containing media.
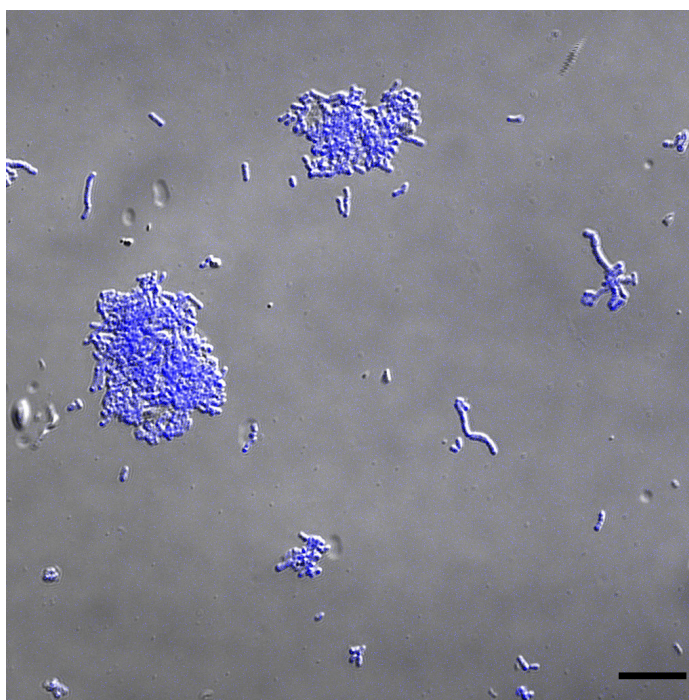
Figure 2.5 Epifluorescence microscopy image of the "*Ca.* Chl. canadense" positive control culture used in the ferrous iron oxidation test. Blue colour in the image represents autofluorescence at 600-800 nm wavelengths (presumably from bacteriochlorophyll *c*/*d*/*e*-containing cells of *Chlorobia*) when excited using a 488 nm laser, and grey colour represents transmitted light from the 488 nm laser. The scale bar represents 10 μm.

2.3.3 Metabolic diversity and phylogeny of *Chlorobia* genome bins

Recovered genome bins of *Chlorobia* were genetically diverse with respect to the surveyed photosynthesis-associated genes (Figure 2.6). One of the lake-recovered bins (L227 2013 Bin 56) lacked all assessed genes involved in sulfide, elemental sulfur, and thiosulfate oxidation but contained the *cyc2* gene homologue. Conversely, two other lake-recovered bins (L227 2013 Bin 55 and L442 Bin 64) contained all assessed genes in the sulfide oxidation pathway but lacked the *cyc2* gene. Although the final lake-recovered bin (L227 2013 Bin 22) lacked both *cyc2* and genes for sulfur oxidation, this lack of key metabolism genes might be due to incomplete genome binning. Lastly, the two enrichment culture bins encoded both *cyc2* and the sulfide oxidation gene pathway, similarly to *Chl.* sp. N1 and *Chl. luteolum*. Like some cultured photoferrotrophs, L227 2013 Bin 22 and the "*Ca.* Chl. canadense" genome also encoded putative homologues to a bacterioferritin hypothesized by Frigaard and Bryant (Frigaard and Bryant 2008) to play a role in photoferrotrophy. Genes for

31
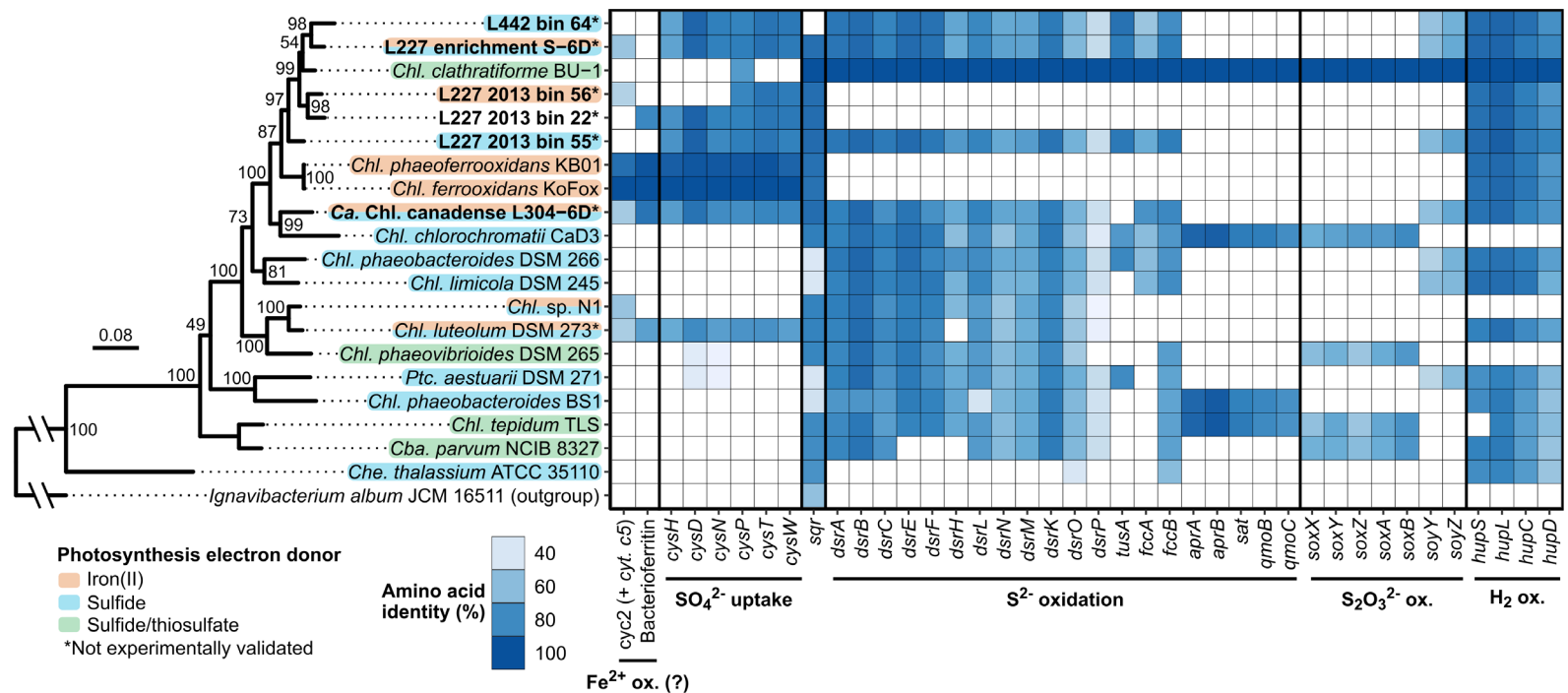
Figure 2.6 Iron- and sulfur-oxidizing genetic potential in recovered genome bins of *Chlorobia* compared to reference strains. The left side of the figure shows a maximum-likelihood phylogenetic tree of the *Chlorobia* based on concatenated ribosomal protein amino acid sequences (see methods). Bootstrap values over 50/100 are shown, and the scale bar represents the percentage of residue changes across the 2243 amino acid alignment. Species of *Chlorobia* are shaded based on their known or hypothesized metabolic potential. Genome bins of *Chlorobia* recovered from this study are bolded. On the right side, a heatmap is shown displaying the presence/absence of genes associated with iron and sulfur metabolism among *Chlorobia* based on bidirectional BLASTP. Heatmap tiles are shaded based on the percent amino acid identity of an identified gene compared to the reference sequence (*Chl. ferrooxidans* for putative iron metabolism-associated genes, and *Chl. phaeoclathratiforme* for sulfur metabolism-associated genes). Although the cytochrome *c5* family gene upstream of *cyc2* was not searched for directly due to poor sequence homology, the gene was verified manually to be adjacent to all hits of *cyc2*.

32
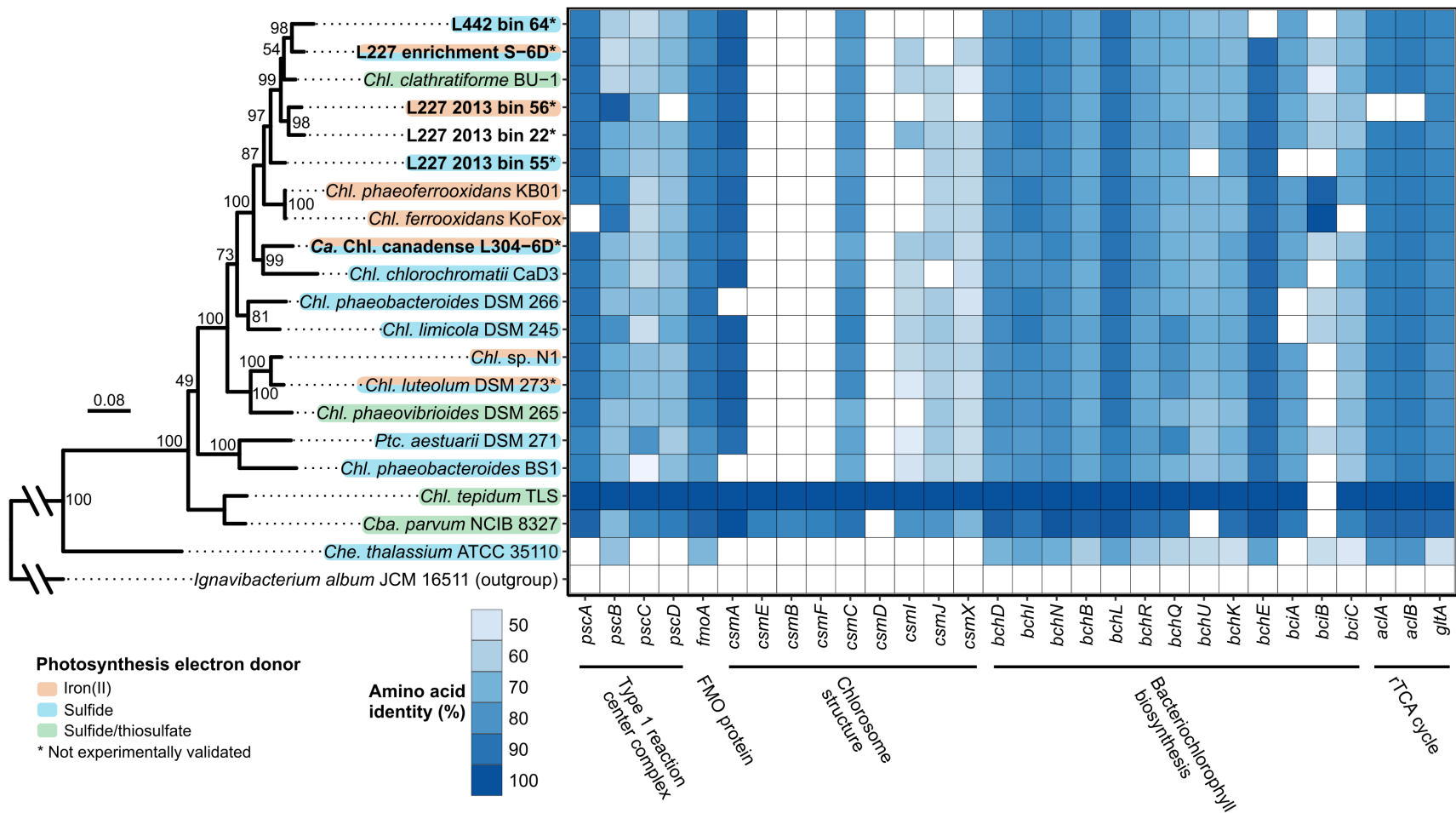
Figure 2.7 Genomic potential for photosynthesis and carbon fixation in recovered genome bins of *Chlorobia* compared to reference strains. The figure layout and phylogenetic tree are identical to Figure 2.6. All query sequences for bidirectional BLASTP were derived from the genome of *Chl. tepidum* TLS except for BciB, which was derived from the genome of *Chl. ferrooxidans* KoFox.

hydrogen oxidation were detected in all of the genome bins (Figure 2.6), along with marker genes for photosynthesis and bacteriochlorophyll biosynthesis (Figure 2.7). Marker genes for carbon fixation (via the reverse TCA cycle) were detected in all genome bins, with the exception of L227 2013 Bin 56, which was missing one of the two assessed gene clusters for the process (*aclAB*; Figure 2.7). Although L227 2013 Bin 56 was also missing half of the *cys* gene pathway for sulfate uptake (Figure 2.6), the *cys* gene cluster occurred at the very end of a contig for this genome bin, implying that the bin likely ought to contain the complete *cys* pathway but that an assembly break occurred.

Phylogenetic analysis of recovered *Chlorobium* genome bins based on concatenated ribosomal protein sequences showed that all four lake-recovered bins and the bin from the Lake 227 enrichment formed a monophyletic subgroup within the *Chlorobiaceae*, including *Chl. phaeoclathratiforme* BU-1 (Figure 2.6). Directly basal to this subgroup was a branch containing *Chl. ferrooxidans* and *Chl. phaeoferrooxidans*. By comparison, the bin of "*Ca.* Chl. canadense" grouped sister to *Chl. chlorochromatii*, separately from the above group and from the group containing *Chl. luteolum* and *Chl.* sp. strain N1. Comparison of the concatenated ribosomal protein phylogeny to the Cyc2 phylogeny revealed some congruency between the two phylogenies for potentially photoferrotrophic members of the *Chlorobia*, with the exception of L227-S-6D and "*Ca*. Chl. canadense" (Figure 2.8). Overall, all six recovered *Chlorobium* genome bins appear to represent novel species, having an ANI of < 79.6% compared to the genomes of available type strains (Figure 2.2).

2.3.4 Functional and taxonomic profiling

Read mapping to genome bins showed that several distinct populations of *Chlorobia* were relevant to the lake environment (Figure 2.9). Bins containing *cyc2* (no *dsrA*) and bins containing *dsrA* (no *cyc2*) were both found at similar relative abundances in Lake 227 samples in 2013 and 2014 (Figure 2.9A). Generally, genome bins of *Chlorobia* containing *cyc2* accounted for 0.6-2.8% of the Lake 227 microbial community data in the assessed samples, whereas bins of *Chlorobia* containing *dsrA* accounted for 1.7-4.1% of the data. Analyses of unassembled metagenome read data indicated similar relative abundances (Figure 2.10). In contrast, only genome bins encoding sulfide oxidation were detected in Lake 442 metagenome assemblies. Even when unassembled metagenome data were scanned, only a single HMM hit from the Lake 442 metagenomes matched *cyc2* classified as belonging to a member of *Chlorobia*, compared to 27 000-37 000 total *rpoB* hits for these same metagenomes, showing that the lack of *cyc2* genes detected in genome bins of *Chlorobia* from L442 was not a result of incomplete assembly or genome binning. One genome bin of *Chlorobia* overlapped between L227 and L442 (L442 2011 16.5m Bin 8); the rest were distinct between the two
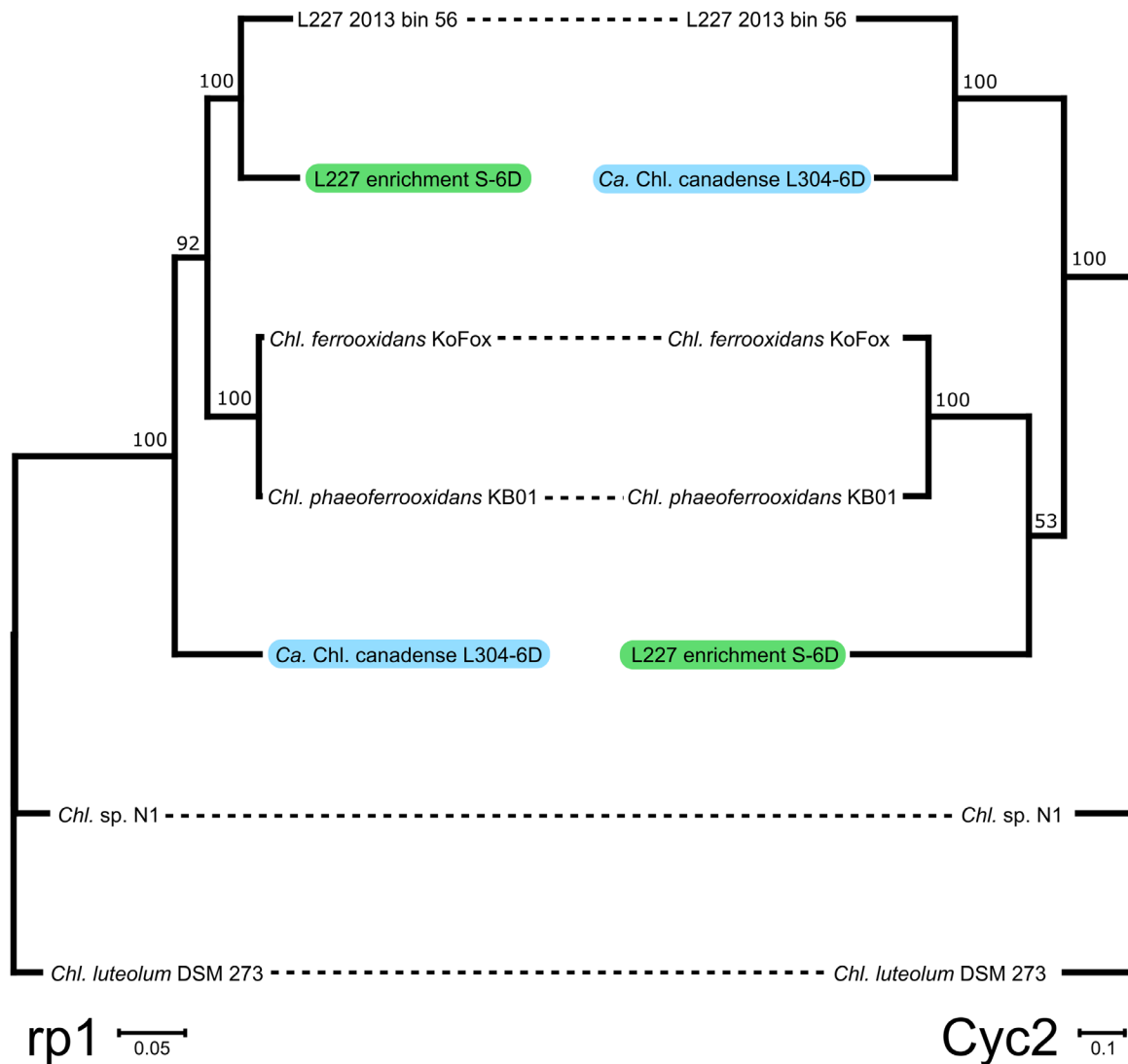
Figure 2.8 Tanglegram comparing the concatenated ribosomal protein phylogeny (Figure 2.6) and Cyc2 phylogeny (Figure 2.1C) among *Chlorobia*. The two tips whose placements differ between the two phylogenies are highlighted in blue or green.

lakes. In addition, the bins of *Chlorobia* recovered from the enrichment culture metagenomes were not detected at high relative abundances in any of the six lake metagenomes.

Several non-*Chlorobia* genome bins were also found to contain the *cyc2* or *dsrA* gene through sequence searches (Figure 2.9B). Genome bins containing potential *cyc2* homologues grouped into the *Rhodoferax* or *Gallionella* genera known to include iron-cycling bacteria (Finneran, Johnsen and Lovley 2003; Kato *et al.* 2015) or into the *Steroidobacteraceae* or *Methylomonadaceae* families. The genome classified to the *Gallionella* genus, L442 Bin 70, also encoded a *mtoAB* gene cluster having 75.5% and 48.5% identity at the amino acid level to the *mtoA*/*mtoB* genes of the microaerophilic iron-oxidizer *Gallionella capsiferriformans* ES-2 (Emerson *et al*. 2013; not shown). Given that both genomes associated with the *Steroidobacteraceae* (L442 Bin 75 and L442 Bin 84) also encoded the PCC-type *mtrABC* operon implicated in iron reduction (Figure 2.11; He *et al.* 2017), it is possible that these genomes represent novel, neutrophilic iron-reducing bacterial populations related to the acidophile *Acidibacter ferrireducens* (Falagán and Johnson 2014). The genome bins classified to the *Methylomonadaceae* occurred at high relative abundances in the lake metagenomes, having a cumulative relative abundance of up to ~17% in Lake 227 (Figure 2.9C), and encoded *pmoA* genes,

Figure 2.9 Bubble plot showing predicted relative abundances of recovered genome bins in the lake and enrichment culture environments. The size of each bubble corresponds to the relative abundance of a genome bin or taxon within metagenome data based on mapping of assembled reads. Each bubble is also labelled with its percent relative abundance for clarity. A, Bins of *Chlorobia* coloured by their metabolic potential based on functional gene markers described in this study. Bolded names correspond to the higher-quality bins described in Table 2.2 and Figure 2.6. In parentheses beside each name are bin quality statistics reported by CheckM – the predicted % completion, % contamination, and % strain heterogeneity, respectively. The displayed bins all classify to the *Chlorobium* genus based on GTDB taxonomy. B, Non-*Chlorobia* bins (≥ 0.01% relative abundance) found to have the same functional gene markers (*cyc2* or *dsrA*) in their genome sequences. Quality statistics are reported as in panel A. On the right side of the panel, the GTDB family and genus classifications of each bin are shown. C, Family-level taxonomic composition of the metagenomes based on GTDB classifications of genome bins with ≥ 1% relative abundance. The phylum of each family is displayed on the right side of the panel. For unresolved or non-Latin family names, the order (and class, if needed) is displayed in parentheses beside the family name for clarity. D, Assembly statistics for the metagenome data. The total number of quality-trimmed metagenome reads is represented by the total height of each bar. Reads that mapped to the filtered sequence assembly (i.e., excluding short contigs, see Methods) are highlighted in blue and are considered "assembled" reads. Assembled read totals were used to determine relative abundances of genome bins in panels A-C. Figure 2.10 shows relative abundances based on predictions from unassembled metagenome reads for comparison.
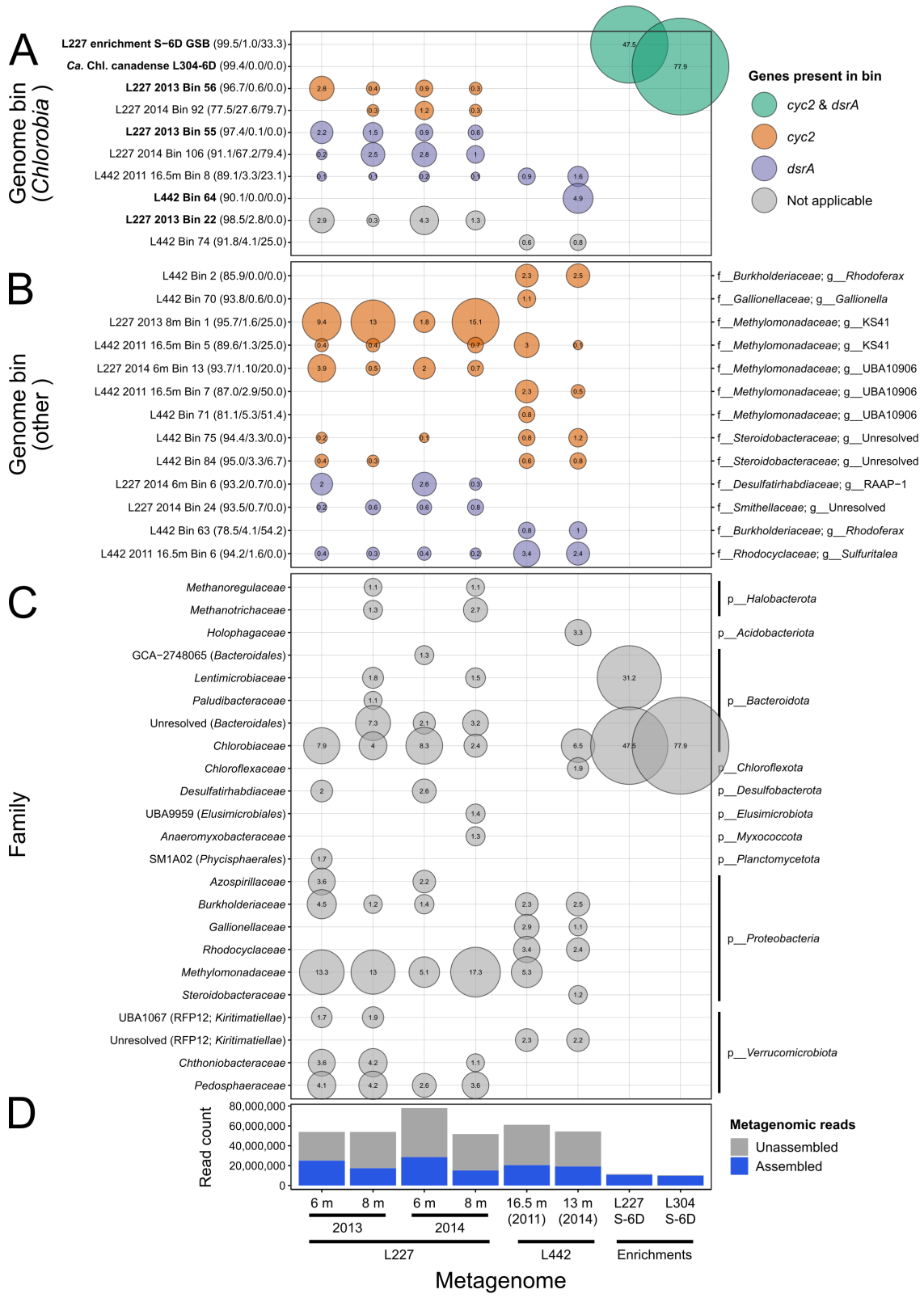
with the exception of L442 2011 16.5m Bin 5. These genome bins encoded a "repCluster2" *cyc2* variant based on analysis by FeGenie, unlike the "repCluster1" variants detected in genomes of *Chlorobia* and of *Gallionella*, but they encoded no detectable *cyc1* gene adjacent to the *cyc2* sequences as seen in the "repCluster2" *cyc2* sequences of *Rhodoferax* (Figure 2.11). Three other genome bins belonging to the phyla *Myxococcota* and *Verrucomicrobiota* encoded PCC-type gene clusters (Figure 2.11). Several genomes, including all detected genomes of *Chlorobium*, were found based on FeGenie to encode a "repCluster3" *cyc2* variant (Figure 2.11). This distant homologue of the photoferrotrophy-associated "repCluster1" *cyc2* variant (having ~20% amino acid identity to "repCluster1" sequences) could potentially play a role in EET but has not been associated with photoferrotrophy in reference cultures and was not detected by the HMM developed in this study.

Genome bins encoding *dsrA*, which is involved in either sulfide oxidation or sulfate reduction, included a bin related to known chemolithotrophic sulfide-oxidizing bacteria of the genus *Sulfuritalea* and a bin related to heterotrophic sulfate-reducing bacteria of the family *Desulfatirhabdiaceae* (Kuever 2014; Watanabe, Kojima and Fukui 2014). A bin classified to the family *Smithellaceae* of the order *Syntrophales*, along with a bin classified to the genus *Rhodoferax*, also encoded *dsrA*. Three of these four genomes bins were also found to encode the *aprAB* operon involved in sulfite metabolism (Figure 2.11). The two bins related to the *Sulfuritalea* and *Rhodoferax* genera were found to encode the *soxB* gene involved in thiosulfate oxidation (Figure 2.11), which suggests the potential for sulfur compound oxidation (Frigaard and Bryant 2008). No other genome bins in the dataset were found to encode these key sulfur-cycling gene markers except for two genome bins that could not be confidently classified as sulfur cyclers due to limited gene content (Figure 2.11).

Chlorobia members appeared to represent the dominant phototrophs in the lake anoxic zone samples. In total, genome bins classified to the *Chlorobia* represented as high as 8.3% of lake microbial communities based on read mapping (Figure 2.9C). Predicted relative abundances of populations of *Chlorobia* were 1.2-1.8 times higher when unassembled reads were assessed directly (Figure 2.10), with a predicted maximum relative abundance of 12.3%, showing that high relative abundance is not a result of assembly bias (Figure 2.9D). The only other potential chlorophototroph detected in the dataset at >1 % relative abundance was a single genome bin classified to the family *Chloroflexaceae* (L442 Bin 82; 92.9/1.2% completeness/contamination), which made up 1.9% of the L442 microbial community in 2014 at 15 m depth (Figure 2.9C) and contained the *bchL* and *pufM* photosynthesis genes (data not shown). This genome bin was undetected in all other samples.
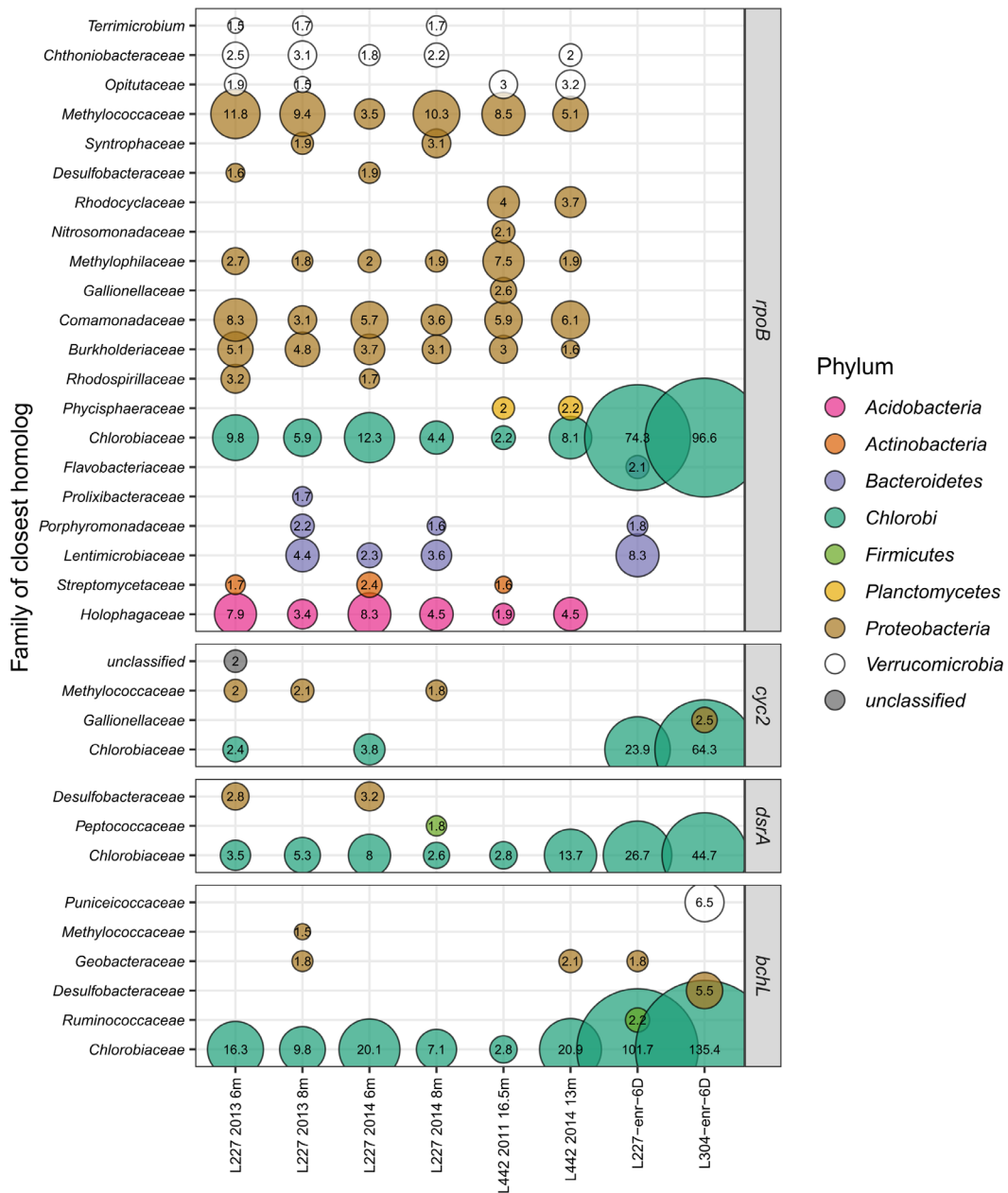
Figure 2.10 Bubble plot showing taxonomic and functional profiling of unassembled metagenome data, in the same general layout as Figure 2.9. HMM hit counts were normalized to the total hits of *rpoB*, a single-copy taxonomic marker gene, within each sample, such that HMM hits are expressed as a percentage of *rpoB*. All families (NCBI taxonomy) with > 1.5% of normalized hits to the HMMs are shown. Although not shown, normalized *Chlorobiaceae*-associated *cyc2* hits for Lake 227 at 8 m were 0.9% for both 2013 and 2014.

39
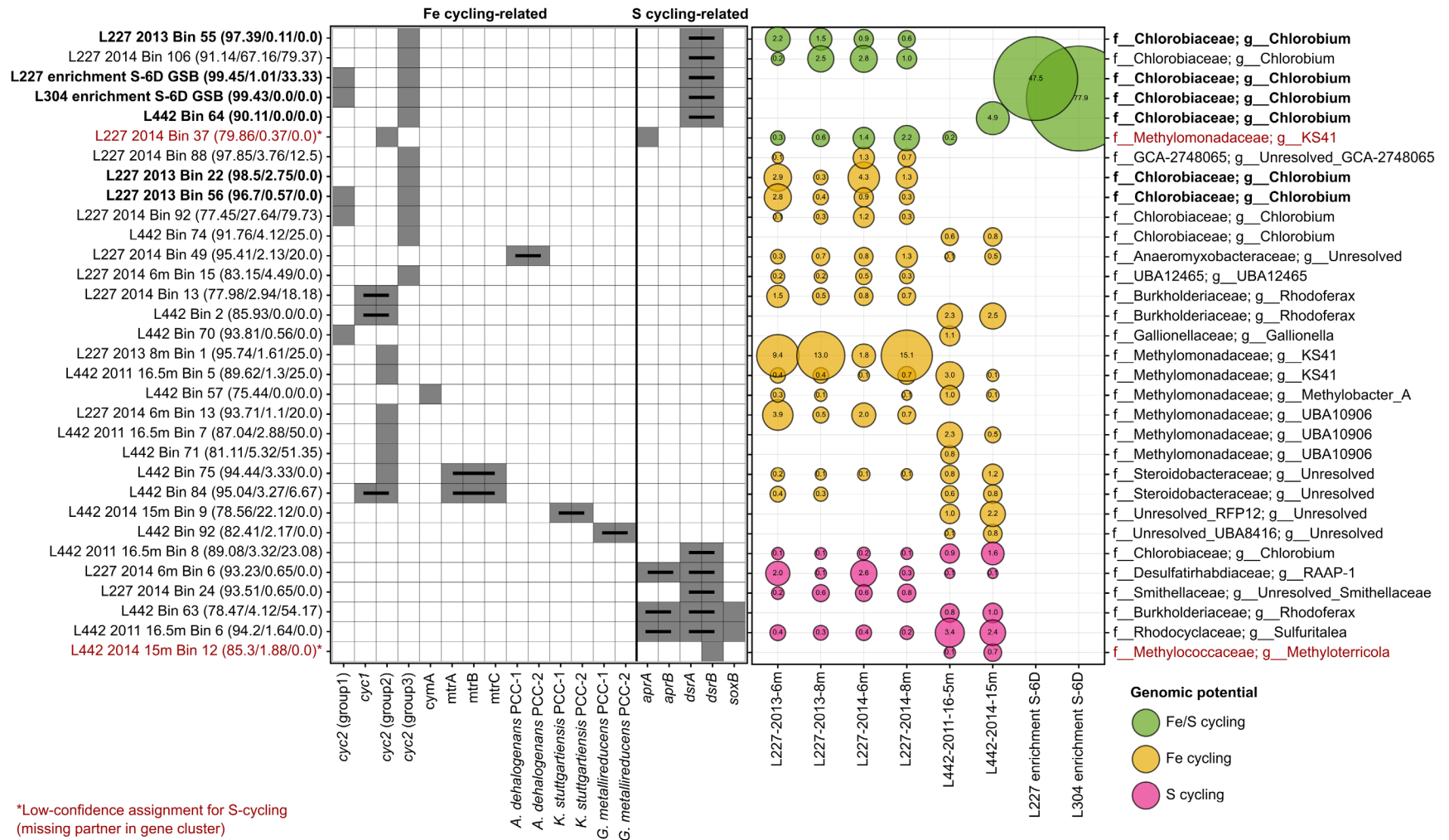
Figure 2.11 Genome bins recovered in this study with iron/sulfur-cycling potential. The left heatmap shows the presence/absence of genes in each genome bin, with thick lines connecting gene clusters. Relative abundances (≥0.05%) of genomes bins in metagenomes (displayed as in Figure 2.9A-B) are shown in a bubble plot on the right.
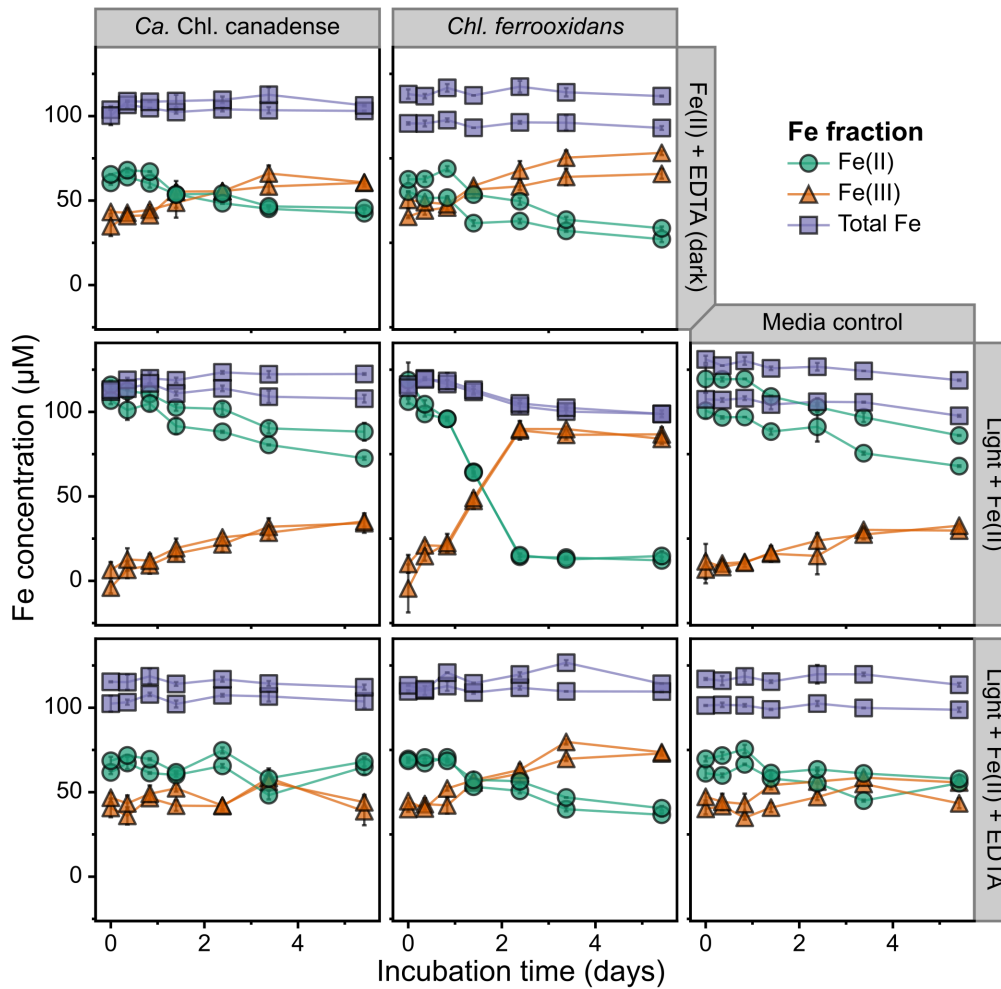
Figure 2.12 Iron oxidation activity of two *Chlorobia* cultures under different growth conditions over time. The concentrations of ferrous iron, total iron, and ferric iron (determined by the subtraction of ferrous iron from total iron) are shown over a 10-day incubation period for "*Ca.* Chl. canadense" (enriched in this study) and *Chl. ferrooxidans*. The activity of the cultures in iron-amended media is shown when incubated in the dark with EDTA, in the light, and in the light with EDTA. Each panel depicts the results of biological duplicates, with error bars representing the standard deviation of technical duplicates in the ferrozine assay. The same results are shown for the full 21-day incubation period in Figure 2.13.
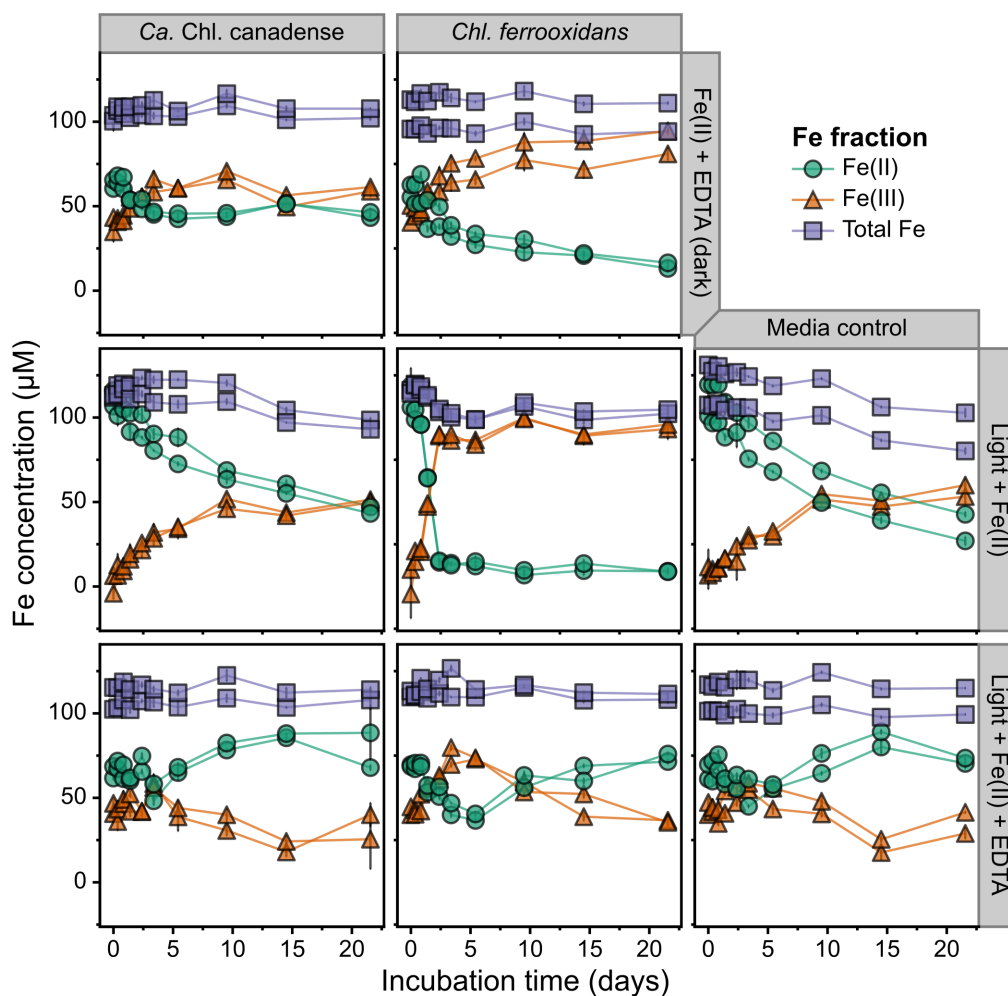
Figure 2.13 Iron oxidation activity of cultures of *Chlorobia* over an extended incubation period of 21 days. The figure layout is identical to Figure 2.12.

2.3.5 Assessment of ferrous iron oxidation potential of "*Ca.* Chl. canadense"

Cultures of *Chl. ferrooxidans* showed the expected behaviour when incubated in ferrous iron containing medium. After a short lag phase, both *Chl. ferrooxidans* cultures containing 100 μM ferrous chloride (no EDTA) and exposed to light began oxidizing ferrous iron, and the cultures had nearly completely oxidized all ferrous iron within two days of the initial setup (Figure 2.12). Similar bottles incubated in the dark showed consistent iron oxidation rates of ~5 μM d^{-1} that corresponded to rates observed in uninoculated media controls; this iron oxidation was inferred to be abiotic, for example due to small amounts of oxygen contamination during sampling. No ferrous iron oxidation outside of this baseline effect was observed in the *Chl. ferrooxidans* cultures supplemented with 120

µM EDTA, in contrast to the stimulatory effect of EDTA reported for *Rhodopseudomonas palustris* TIE-1 (Peng *et al.* 2019), which uses a different gene system (*pioAB*) for EET (Gupta *et al.* 2019). The EDTA-amended medium consistently appeared to have ~50% ferrous iron oxidation at the beginning of the experiment, based on the ferrozine assay, but this apparent oxidation has been observed previously with EDTA and could potentially be due to ineffective binding of ferrozine to EDTA-chelated iron (Verschoor and Molot 2013).

In contrast to *Chl. ferrooxidans*, "*Ca.* Chl. canadense" showed no photoferrotrophic iron oxidation activity over the course of the experiment (Figure 2.12; Figure 2.13), and the iron oxidation profiles of the cultures closely matched those of the uninoculated media control bottles. A reference bottle of Hegler freshwater medium prepared in the same batch as those in the main experiment but amended with 600 µM buffered sulfide feeding solution, in place of ferrous iron, received the same inoculum of "*Ca.* Chl. canadense". Complete oxidation of sulfide in the bottle was observed within five days of inoculation and incubation under light (not shown), indicating that the media and inoculum were not the cause of the lack of observed photoferrotrophic activity.

2.4 Discussion

In this study, the diversity of anoxygenic photosynthesis was explored in the ferruginous water columns of Boreal Shield lakes and the potential for photoferrotrophy was evaluated using a combination of genome-resolved metagenomics and enrichment cultivation. Although photoferrotrophy could not be induced in laboratory cultures, robust homologs of the *cyc2* gene, associated with photoferrotrophy in the genus *Chlorobium*, were detected, along with evidence for broader iron cycling and cryptic sulfur cycling in the lakes. Overall, this research provides genomic context to previous 16S rRNA gene data supporting the existence of *Chlorobia*-driven photoferrotrophy in Boreal Shield lakes (Schiff *et al.* 2017), adding nuance to the interpretation of these data while expanding sparse knowledge of the diversity of *cyc2* within the *Chlorobia*. This research reinforces the need for additional genetic work to understand the function and regulation of *cyc2* in photoferrotrophy.

Several novel variants of *cyc2* were detected in the genome bins of *Chlorobia* recovered in this study from lake and enrichment culture metagenomes (Figure 2.1). The *cyc2* gene, which is putatively implicated in ferrous iron oxidation among microaerophilic iron-oxidizing bacteria (McAllister *et al.* 2020), has been hypothesized to play a role in photoferrotrophy based on its presence in the genomes

of cultured photoferrotrophs within the *Chlorobia* and its absence in the genomes of strict sulfide-oxidizing *Chlorobia* members (Frigaard and Bryant 2008; Crowe *et al.* 2017). The detection of four additional high-confidence homologues of *cyc2* in this study doubles the known diversity of the *cyc2* gene among members of the *Chlorobia* compared to cultured representatives. Furthermore, the robust detection method developed in this work combining sequence composition, genomic context, and phylogenetic placement allows for the identification of additional *cyc2* homologs, compared to related cytochromes, in future research. The detection of diverse *Chlorobia*-associated *cyc2* genes in this research suggests that *cyc2*-encoding *Chlorobia* members, although only cultivated in the laboratory recently, play an important role in seasonally anoxic Boreal Shield lake microbial communities based on their high relative abundances in lake metagenomes (Figure 2.9A).

Homologs of *cyc2* also appear to be associated with other bacteria potentially involved in iron cycling in Lakes 227 and 442. One genome bin classified to the genus *Gallionella*, L442 Bin 70, encoded *cyc2* and a *mtoAB* homolog and thus could represent a microaerophilic iron-oxidizing bacterium. In addition, genomes such as L442 Bin 2 classified to the genus *Rhodoferax* (related to *Rhodoferax ferrireducens*) and L442 Bins 75 and 84 belonging to the *Steroidobacteraceae* (related to *Acidibacter ferrireducens*) encoded an alternative *cyc1/cyc2* gene cluster and could correspond to known iron-reducing bacteria. Other bacterial genomes encoding PCC-type gene systems involved in EET were also detected, hinting at a broader iron-based biogeochemical cycle in the lakes. At the same time, genome bins classified to the *Methylomonadaceae* that are unrelated to known iron-cycling bacteria also encoded an alternative *cyc2* variant. It is possible that the *cyc2* variant might be involved in a form of EET for these potential methanotrophs, which are found in the anoxic zones of Lakes 227 and 442 despite the oxygen requirement for particulate methane monooxygenase, as recently observed for other lakes (van Grinsven *et al.* 2019, 2020). However, EET by bacterial methanotrophs has not been reported in the literature previously, leaving the role of these *cyc2* homologs unclear. Altogether, the broader detection of genes involved in iron cycling in lake metagenome data suggests that iron metabolism could potentially play an important biogeochemical role in the seasonally anoxic and ferruginous water columns of these lakes.

Along with *cyc2*, the genomic potential for sulfide oxidation was detected among genome bins of *Chlorobia* recovered in this study. A higher number of genome bins of *Chlorobia* encoded the genetic potential for sulfide oxidation (based on the marker gene *dsrA*) than encoded *cyc2*, and in addition, *cyc2* associated with *Chlorobia* members was entirely undetectable in Lake 442

metagenome data. Combined with observations of blackening of ferrous iron-containing medium during initial enrichment culturing before phototrophic growth (Figure 2.4), these genomic data suggest that sulfide is an important electron donor for anoxygenic phototrophs in ferruginous Boreal Shield lake waters. At least one genome bin was detected at high relative abundance in Lake 227 metagenomes that likely represented a sulfate-reducing bacterium (Figure 2.9C; Figure 2.11) and might thus provide the required sulfide to sustain phototrophic sulfide oxidation, even if sulfide is at low levels overall in the lakes. Such "cryptic sulfur cycling", where sulfur turnover is rapid at low measured concentrations of sulfur redox species, has been observed previously in other sulfide-limited systems (Kappler and Bryce 2017), including the permanently anoxic and ferruginous Lake La Cruz (Walter *et al.* 2014), and might have also played an important role in Archaean oceans once sufficient organic carbon was available. In addition, because all detected genomes of *Chlorobia* encoded the *hupLS* genes (Figure 2.6), it is possible that oxidation of molecular hydrogen, as observed in laboratory cultures of *Chlorobia* (Heising *et al.* 1999), could also play a role in the metabolism of the populations of *Chlorobia* in the lakes.

Photoferrotrophy could not be induced in "*Ca.* Chl. canadense" (Figure 2.12), which was enriched in a sulfide-containing medium, despite the fact that this organism encoded a *cyc2* homolog and a bacterioferritin in its genome (Figure 2.6). The lack of observed photoferrotrophic growth in "*Ca.* Chl. canadense" could be due to several reasons. Firstly, it is possible that, although *cyc2* is associated with reference photoferrotrophic *Chlorobia* members, the *cyc2* gene plays a peripheral role in the metabolism of these organisms and is not directly involved in photoferrotrophy. Secondly, *cyc2* could represent a single step in a larger iron oxidation gene pathway that is missing in the "*Ca.* Chl. canadense" genome. For example, the *c5* family cytochrome detected directly upstream of all *cyc2* homologs identified among genomes of *Chlorobia* (Figure 2.1B) could potentially be involved in iron electron transfer. Unlike the primary sequences of all other *c5* family cytochromes adjacent to *cyc2* in reference genomes or environmental genome bins, the *c5* family cytochrome of "*Ca.* Chl. canadense" contains an additional cysteine residue in the heme-binding motif in its primary sequence (i.e., CCXCH rather than the typical CXXCH; Figure 2.3) that could have deleterious consequences for iron oxidation (Allen *et al.* 2009).

As a third possibility, *cyc2* could be associated with alternative cellular roles, including electron exchange with humic substances as hypothesized elsewhere (He *et al.* 2019) or even direct electron uptake from solid-phase conductive substances, as has been observed in the photoferrotroph

Rps. palustris TIE-1 (Guzman *et al.* 2019). However, direct links between Cyc2 and oxidation of alternate substrates to dissolved ferrous iron have not been demonstrated experimentally. Lastly, the *cyc2* gene product in "*Ca.* Chl. canadense" might allow for iron oxidation but expression might not be favourable under the tested laboratory conditions. This option would align with preliminary reports that the *cyc2*-encoding but sulfide-oxidizing *Chl. luteolum* is capable of changing to photoferrotrophic growth (Kate Thompson and Sean Crowe, personal communication), despite previous difficulties inducing this behaviour (Laufer *et al.* 2017). The "*Ca.* Chl. canadense" culture can be used to probe the regulation and function of *cyc2* within the class *Chlorobia* in future work, especially given its atypical cytochrome *c5* sequence and given that all other *cyc2*-encoding and cultured *Chlorobia* members to date have photoferrotrophic activity. However, this present study is unable to draw conclusive links between *cyc2* and photoferrotrophy.

Despite being physically nearby (~13.5 km apart) and both having ferruginous bottom waters, Lakes 227 and 442 differed substantially in the microbial community composition of their anoxic water columns based on metagenome data. Most genome bins examined in this study (Figure 2.9A-B) were specific to one of the two lakes based on read mapping counts. Such single-lake specificity is lost when summarizing data at higher taxonomic levels (e.g., Figure 2.9C) and was not observed for the same DNA samples when analyzed previously using 16S rRNA gene amplicon sequencing and clustering of sequences at a 97% identity threshold (Appendix A). Population-level analyses showed that *cyc2* was not encoded by Lake 442 genome bins of *Chlorobia* despite being encoded by genome bins of *Chlorobia* in Lake 227 metagenome data. It is possible that physico-chemical differences between the lakes, such as ferrous iron and sulfate concentration, hydrogen and bicarbonate concentrations (observed to impact photoferrotrophy rates in *Rps. palustris* TIE-1; Croal *et al.* 2009), lake bathymetry, anoxia timing, pH, light quality, or dissolved organic carbon levels, could drive this and other microbial community differences between the two lakes. Understanding key factors governing the favourability of *cyc2* among Boreal Shield lakes could prove valuable for understanding the ecology and function of *cyc2*.

The strong iron isotope fractionation observed between the oxic and anoxic zone boundaries of Lakes 227 and 442 previously (Appendix A) appears to occur regardless of whether *Chlorobia*-affiliated *cyc2* genes are detectable. It is possible that microaerophilic iron oxidation could contribute to the observed iron isotope fractionation in Lake 442 given the detection of a *cyc2*-encoding genome bin within the *Gallionella* in Lake 442 metagenomic data (Figure 2.9B). Partial iron reduction by

iron-reducing bacteria might also contribute to the observed fractionation. However, other yet-unknown iron cycling processes might also be at work. Sampling for iron isotopes at higher spatial and temporal resolution, combined with knowledge of additional iron-cycling processes in the lakes, could allow for the sources of the observed isotopic fractionation in both lakes to be determined, with implications for reconstructing ancient biogeochemical cycles from the rock record.

The research presented in this study provides insight into the metabolic diversity and ecology of anoxygenic photoautotrophs in seasonally anoxic Boreal Shield lakes. High relative abundance genome bins of *Chlorobia* in lake metagenomes can encode the *cyc2* gene associated with iron oxidation, but phototrophic sulfide oxidation appears to be more widespread environmentally and can be detected via enrichment culturing. Although photoferrotrophic growth could not be induced, the novel *cyc2*-encoding species "*Ca.* Chl. canadense" was enriched, and this species could be used in future research exploring the function and regulation of *cyc2* as genetic studies continue to progress in linking the *cyc2* gene product to its cellular role (Chan *et al.* 2018). Probing the metabolic diversity of anoxygenic phototrophs in Boreal Shield lakes could lead to novel inferences about photosynthesis in early Earth oceans, including the role of cryptic sulfur cycling compared to photoferrotrophic activity. Novel phototrophic metabolisms, such as recently reported phototrophic oxidation of manganese(II) (Daye *et al.* 2019), could also be explored with implications for both modern and prehistoric biogeochemical cycling. Altogether, the findings presented in this study serve as an important basis for future work probing the biogeography and activity of anoxygenic phototrophs in natural waters. Understanding the metabolic diversity of phototrophy in Boreal Shield lakes will enhance comprehension of modern boreal ecosystems with additional implications for the evolution of early life on Earth.

2.5 Data and code availability

Raw metagenome reads for the freshwater lake and enrichment culture metagenomes are available in the NCBI sequence read archive under BioProjects PRJNA518727 and PRJNA534305, respectively. Six curated genome bins of *Chlorobia* are available under the same BioProjects, along with all assembled contigs from lake metagenomes. The complete set of uncurated genome bins were uploaded to a Zenodo repository at doi:10.5281/zenodo.2720705. Code for downloading these data and performing the analyses presented in this paper, along with intermediate data files, are available at https://github.com/jmtsuji/Chlorobia-cyc2-genomics (doi:10.5281/zenodo.3228523).

Chapter 3

Biogeography and activity of chlorophototrophs in the ferruginous water columns of Boreal Shield lakes

3.1 Introduction

Ferruginous conditions have likely been widespread among aquatic systems throughout much of Earth's history (Poulton and Canfield 2011). Such conditions, characterized by anoxic waters with high levels of dissolved ferrous iron and comparatively low levels of sulfate and sulfide, were likely dominant in the Earth's oceans throughout the Archaean Eon, ~3.8-2.5 billion years ago (Crowe *et al.* 2014b), and well into the Proterozoic Eon, ~2.5-0.54 billion years ago (Fakhraee *et al.* 2019). The modern oceans are nearly fully oxygenated and sulfate-rich, and most modern and anoxic freshwater systems that have been studied develop euxinic (sulfide-dominated) rather than ferruginous waters (Diao *et al.* 2017; Posth *et al.* 2017; Wang and Benoit 2017; Savvichev *et al.* 2018). A small number of modern and permanently anoxic (i.e., meromictic) ferruginous lakes have been studied in detail, and these lakes have been found to host unique biogeochemical cycles potentially relevant to study of the early Earth or life on other planets (Koeksoy *et al.* 2016). Photoferrotrophy, the light-driven and direct microbial oxidation of iron (Camacho *et al.* 2017), has been confirmed in the anoxic waters of at least three such lakes (see below), along with active methane cycling (Crowe *et al.* 2011; Lambrecht *et al.* 2019) and iron isotopic signatures relevant to understanding the deposition of banded iron formations (Busigny *et al.* 2014). However, meromictic and ferruginous lakes are thought to be rare globally (Walter *et al.* 2014), severely limiting the ability to understand the microbial diversity and biogeochemical properties associated with ferruginous waters.

Photoautotrophy is a key process for energy flow through ecosystems. Aquatic chlorophototrophs are estimated to be responsible for around half of modern global primary production (Field *et al.* 1998), and aquatic photosynthesis is thought to have directly or indirectly fueled the majority of the oxidation of the oceans in Earth's early history (Lyons, Fike and Zerkle 2015). The dominant modern form of chlorophototrophy is oxygenic phototrophy, where water is used as the photosynthetic electron donor and is oxidized to molecular oxygen, and this process is performed by members of the *Cyanobacteria* phylum and by chloroplast-containing eukaryotes (Fischer, Hemp and Johnson 2016). However, anoxygenic phototrophy, which relies on alternative

photosynthetic electron donors including sulfide, ferrous iron, manganese(II), arsenic(III), and nitrite, is also widespread in aquatic ecosystems and is associated with diverse microbial taxa having unique photophysiologies (Ehrenreich and Widdel 1994; Griffin, Schott and Schink 2007; Kulp *et al.* 2008; Overmann and Garcia-Pichel 2013; Thiel, Tank and Bryant 2018; Daye *et al.* 2019). Anoxygenic phototrophs could form important biogeochemical links between the carbon cycle and cycling of other elements. For example, purple sulfur bacteria have been found to block the upwelling of phosphorus to inhibit surface algal bloom formation (Wang and Benoit 2017), and activity of anoxygenic phototrophs could also facilitate cryptic elemental cycles where nutrients are rapidly recycled to fuel growth of microbial consortia (Berg *et al.* 2016). Nevertheless, the role of diverse anoxygenic phototrophs in nutrient cycling in aquatic systems, as well as their interaction with oxygenic phototrophs, remains poorly explored.

Study of meromictic and ferruginous lakes already suggests that anoxygenic photosynthesis is complex in its ecology in ferruginous waters. In Lake Matano (Indonesia), photoferrotrophy was predicted to occur in the deep and ferruginous water column, but metagenome data showed the potential for phototrophic sulfide oxidation, and rate incubations were unable to demonstrate photoferrotrophic activity at quantifiable levels (Crowe *et al.* 2008, 2014a). Combined with microscopy data, rate measurements taken from water column samples of Lake La Cruz (Spain) revealed that multiple phototrophic modes, including photoferrotrophy, likely occurred simultaneously in the water column (Walter *et al.* 2014). Novel species of photoferrotrophs were cultivated from Kabuno Bay of Lake Kivu (Democratic Republic of the Congo) and Brownie Lake (U.S.A.), and rate incubation experiments performed in Kabuno Bay indicated that most phototrophy in the anoxic zone was performed by photoferrotrophs (Llirós *et al.* 2015; Lambrecht 2019). Anoxygenic photosynthesis was later estimated using rate incubations and stable isotopic measures to supply up to 89% of fixed carbon in Kabuno Bay's anoxic waters (Morana *et al.* 2016). Variable results from each of these field sites suggests that additional work is needed to understand the structure and function of phototrophic microbial communities in iron-rich aquatic systems.

Beyond intensely studied yet rare meromictic and ferruginous lakes, millions of Boreal Shield lakes exist in northern regions globally that likely develop ferruginous waters during seasonal anoxia (Appendix A). Such lakes offer the opportunity to study the microbial ecology and biogeochemistry of ferruginous waters in an unprecedented manner, with each lake being able to serve as a semi-independent ecosystem replicate across the range of physicochemical conditions possible for lakes in

the Boreal Shield. Although the microbial ecology of boreal lakes has been examined previously (Peura *et al.* 2018; Tran *et al.* 2018; Crevecoeur *et al.* 2019), the iron/sulfur biogeochemistry of Boreal Shield lakes, with implications for the ecology of ferruginous waters, remains mostly unexplored (Appendix A). Here, nine Boreal Shield lakes at the IISD-ELA are surveyed to probe the biogeography and activity of phototrophs in ferruginous water columns. By combining geochemical and stable isotopic measures, 16S rRNA gene amplicon sequencing, metagenomics and metatranscriptomics, and rate measurements, this study represents the first comprehensive and multi-lake survey of microbial communities in ferruginous waters. This study aims to clarify the diversity of phototrophic microbial communities in Boreal Shield lakes, physicochemical drivers of phototrophy in ferruginous waters, and the impact of phototrophs on the iron and sulfur biogeochemistry of the lakes.

3.2 Materials and methods

3.2.1 Study site, sample collection, and physicochemical measurements

The IISD-ELA sampling site has been described in detail previously (Armstrong and Schindler 1971; Brunskill and Schindler 1971; Schindler 1971; Schindler and Fee 1974; see also Appendix A). Eight lakes within the IISD-ELA that were known or suspected to develop seasonal anoxia based on previous field data were selected for the survey, along with a permanently oxic reference lake. These lakes were sampled in the summer or fall from 2016 until 2018 across four main sampling events. Sampling in June 2016, September 2016, and September 2017 involved full water column profiles of the previously studied reference Lakes 227 and 442, along with top/bottom sampling of the epilimnion and hypolimnion of the other selected lakes. A full water column profile of Lake 227 including samples for RNA extraction was taken in September 2017. In July 2018, specific lakes of interest based on 2016-2017 data (namely, Lakes 227, 221, and 304) were surveyed again with full water column profiles and additional tests of phototroph activity. A single depth in the mid anoxic zones of Lakes 221 and 304 was sampled for RNA, and samples from Lake 227 were taken for rate measurements (as described below).

Profiles of temperature, dissolved oxygen, and light penetration were collected when sampling. Temperature and dissolved oxygen were generally measured using an EXO multi-parameter sonde (Xylem; Rye Brook, New York, U.S.A.) equipped with temperature and oxygen probes. For selected water column profiles, a HQ40D Portable Multi Meter for Water (Hach; Loveland, Colorado, U.S.A.)

was used instead. Boreal Shield lakes at the IISD-ELA are known to be poorly buffered with circum-neutral water column pH, and so pH was not measured. Light profiles were measured using a LI-192 underwater quantum sensor attached to a LI-250 light meter or LI-1400 data logger (LI-COR Biosciences; Lincoln, New England, U.S.A). Light attenuation coefficients were then calculated from light profile data as described previously (Schanz, 1985), and coefficient values were averaged across all measured depths in the water column for each profile.

Lake water column samples were collected using a closed system gear pump and line. The sampling line was pre-flushed with water at the desired depth prior to sampling to allow sufficient flow time for water to be fully replaced twice through the line. Samples for measuring concentrations of sulfate, ammonium, total dissolved iron, and dissolved organic carbon were collected and analyzed as described previously, along with samples for determining concentrations and carbon stable isotopic values of dissolved inorganic carbon and methane, and other samples for assessing carbon/nitrogen composition and stable isotopic values of particulate organic matter (Appendix A). In addition, samples were collected for analyzing concentrations of dissolved ferrous and ferric iron. Water was pumped directly through a pre-flushed 0.45 µm inline filter and injected into 10-20 mL acid washed glass vials pre-flushed with dinitrogen gas and sealed with blue butyl rubber stoppers (Bellco Glass; Vineland, New Jersey, U.S.A.). Vials were pre-filled with 6 N trace metal grade hydrochloric acid to immediately acidify the samples to a final concentration of 0.5 N acid. Samples were kept chilled (approx. 4°C) until measurement of iron species concentrations using the ferrozine assay (Stookey 1970; Viollier *et al.* 2000).

Samples for water column DNA were collected by pumping water through sterile 0.22 µm Sterivex polyvinyl fluoride filters (Merck Millipore; Burlington, Massachusetts, U.S.A.). When collecting water column RNA, the same filters were used as for DNA, but filters were immediately filled with 1.8 mL of DNA/RNA Shield (Zymo Research; Irvine, California, U.S.A.) once packed and purged of residual water. Filters were collected and subsequently extracted and analyzed in triplicate for RNA samples. Sterivex filters were kept chilled after collection until being frozen (at -20°C) upon return to the sampling camp later that day. Filters were then shipped chilled to the University of Waterloo and were kept frozen (at -20°C) until processing.

3.2.2 Nucleic acid extraction

For each Sterivex filter used for DNA collection, the Sterivex filter case was opened, and the filter membrane was carefully removed using a scalpel blade that had been flame-sterilized to destroy contaminating DNA. Each filter membrane was cut in half lengthwise along the filter; one half was stored frozen to provide a backup sample, and DNA was extracted from the other half using the DNeasy PowerSoil or DNeasy PowerSoil HTP 96 Kit (Qiagen; Venlo, The Netherlands). Extractions were performed according to the kit protocol, and the optional 10 min incubation at 70°C after adding Solution C1 was performed to enhance cell lysis. Mechanical lysis was performed for samples in PowerBead Tubes using a FastPrep-24 instrument (MP Biomedicals; Santa Ana, California, U.S.A.) set at 4.5 m/s for 45 s, and mechanical lysis was performed for samples in PowerBead DNA Plates using a mixer mill MM 400 (Retsch; Haan, Germany) set at 30 Hz for 10 min. Resulting DNA concentrations were then quantified using a Nanodrop spectrophotometer (Thermo Fisher Scientific; Waltham, Massachusetts, U.S.A.) or using the Qubit dsDNA HS Assay Kit with Qubit 2.0 fluorometer (Thermo Fisher Scientific).

RNA extraction was performed using the ZymoBIOMICS DNA/RNA Miniprep Kit (Zymo Research) with initial steps of the protocol modified slightly to accommodate the volume of DNA/RNA Shield associated with each filter. Sterivex filters filled with DNA/RNA Shield were thawed, and the DNA/RNA Shield was pumped out of the filter cases and saved for downstream use. Filter cases were then opened and filters excised as described above. Each filter half was cut into small pieces. One half was stored in a clean 2 mL microfuge tube along with half of the collected DNA/RNA Shield solution and frozen as a backup sample. The other filter half, along with the remaining DNA/RNA Shield, was added into a dry ZR BashingBead Lysis Tube for extraction. Mechanical lysis was performed using a FastPrep-24 instrument (MP Biomedicals). Lysis tubes were shaken at 6.5 m/s for 60 s twice, and tubes were allowed to cool on ice for at least 60 s between mechanical lysis rounds. After centrifugation, the entire supernatant volume was transferred to a new microfuge tube, and one volume of DNA/RNA Lysis Buffer was added to the tube and mixed. RNA extraction was then performed according to the DNA & RNA Parallel Purification protocol in the kit manual using in-column DNase I treatment. Only RNA (and not DNA) extracts were saved because corresponding Sterivex filters for DNA were collected and processed for the same lake depths using the protocol described above. The resulting RNA extracts were quantified using a Nanodrop spectrophotometer (Thermo Fisher Scientific) and Qubit RNA Assay Kit with Qubit 2.0 fluorometer

(Thermo Fisher Scientific). Extracts were also run on a 1% agarose gel stained with GelRed (Biotium; Fremont, California, U.S.A.) to confirm that rRNA of the expected size lengths was visible.

3.2.3 16S rRNA gene amplicon sequencing data processing

All extracted DNA samples were amplified using the universal prokaryotic 16S rRNA gene primers 515F-Y and 926R (Quince *et al.* 2011; Parada, Needham and Fuhrman 2016). Samples collected in 2016 were amplified, prepared into sequencing libraries, and sequenced on a MiSeq (Illumina; San Diego, California, U.S.A.) by the Joint Genome Institute (Lawrence Berkeley National Laboratory, Berkeley, U.S.A.) in a sequencing run described previously (Spasov *et al.* 2020). The other batch of samples (from 2017 and 2018 field sampling) was processed at the University of Waterloo. Gene amplification using primers with attached index sequences and sequencing adaptors was performed as described previously (Bartram *et al.* 2011; Kennedy *et al.* 2014). The thermocycler profile had an initial denaturation at 95°C for 3 min, followed by 35 cycles of 95°C for 30 s, 50°C for 30s, and 68C for 1 min, followed by a final extension of 68°C for 7 min. In addition, the PCR Master Mix contained 1x ThermoPol Reaction Buffer (New England Biolabs; Ipswich, Massachusetts, U.S.A.), 0.6 mg/mL bovine serum albumin, 200 µM dNTPs, 0.2 µM forward and reverse primer, and 0.625 U Hot Start *Taq* DNA Polymerase (New England Biolabs). A total of 24 µL Master Mix, plus 1 µL of sample, was used for each amplification reaction.

Triplicate amplifications performed for each sample were pooled, and samples were then run on a 1% agarose gel stained with GelRed (Biotium) for relative quantification. All samples were then pooled at normalized concentrations based on quantification results, and non-template controls and samples without amplification were included in pooling. The pooled library was run in a 1% agarose gel stained with ethidium bromide and the amplicon band was cut from the gel and purified using the Wizard SV Gel and PCR Clean-Up System (Promega; Madison, Wisconsin, U.S.A.). Based on quantification using the Qubit dsDNA HS Assay Kit with a Qubit 2.0 fluorometer (Thermo Fisher Scientific), the final sequencing library was prepared at 5.8 pM concentration, including 15% of spiked PhiX Control (Illumina). The library was sequenced on a MiSeq (Illumina) at the University of Waterloo using a 2x250 cycle MiSeq Reagent Kit v2 (Illumina) and demultiplexed using the MiSeq Reporter software.

All 16S rRNA gene sequencing data were processed together via the AXIOME pipeline, commit 1ec1ea6 (https://github.com/neufeld/axiome3; Lynch *et al.* 2013), which relied on QIIME2,

version 2019.10 (Bolyen *et al.* 2019). Default pipeline settings were used, except that the "cutadapt trim-paired" module of QIIME2 had to be run after data import, using forward and reverse primer sequences as parameters along with the "--p-discard-untrimmed" flag, to allow for the different sequencing run batches to be trimmed to consistent amplicon regions. Because of this, trim options for the "dada2 denoise-paired" module were set to zero. Briefly, within AXIOME, paired-end reads were merged and denoised using DADA2 (Callahan *et al.* 2016) and the resulting ASVs were classified using the QIIME2 naïve Bayes classifier (Bokulich *et al.* 2018), which had been trained against the SILVA SSU database, release 138 (Quast *et al.* 2013; Glöckner *et al.* 2017). The resulting ASV table was rarefied to 12,000 sequences where rarefication was required for downstream statistical analyses.

3.2.4 Metagenome and metatranscriptome sequencing and data processing

Metagenomes were sequenced in two separate batches. Samples of DNA from the June and September 2016 sampling events were prepared into sequencing libraries and sequenced by the Joint Genome Institute (Lawrence Berkeley National Laboratory). Only selected samples representing the hypolimnia of lakes were sequenced. Metagenome library preparation was performed using the Nextera XT DNA Library Preparation Kit including library amplification steps. Resulting libraries were sequenced on a HiSeq 2500 (Illumina) with 2x150 base paired end reads, generating 88.2 to 273.5 million reads per sample.

Selected DNA samples from 2017 and 2018 field sampling were prepared for metagenome sequencing and sequenced by the McMaster Genome Facility (McMaster University; Hamilton, Ontario, Canada). Sample DNA was sheared to the desired insert size using a Covaris ultrasonicator (Covaris; Woburn, Massachusetts, U.S.A.), and metagenome sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) including library amplification steps. The resulting library was sequenced on a single lane of a HiSeq 2500 (Illumina) with 2x250 base paired end reads, generating 41.2 to 128.2 million reads per sample.

All metagenome data was processed using the ATLAS pipeline, version 2.1.4 (Kieser *et al.* 2020). Default settings were used for the run except that the minimum percent identity threshold for read mapping (via "contig_min_id") was set to 99% and only MaxBin2 and MetaBAT2 were used as binning algorithms (Wu, Simmons and Singer 2016; Kang *et al.* 2019). To enhance genome binning quality, the six lake metagenomes from Chapter 2 were also included, along with a single

metagenome from the aphotic zone of the nearby and meromictic Lake 111 (available in the supplementary data associated with this work), which was also sampled in July 2018. The Lake 111 metagenome had limited environmental metadata and so was not analyzed further in the context of this work. The entire ATLAS pipeline, including quality control on raw reads, metagenome assembly of individual samples, metagenome binning, dereplication of bins from all samples and bin analysis, and gene clustering and annotation, was run end-to-end. For the genome binning step, all metagenome samples from the same lake were summarized in the same "BinGroup", allowing for read mapping information between each sample from the same lake to be used as differential abundance information to guide genome binning. After running ATLAS, the resulting dereplicated genome bins were taxonomically classified using the GTDB-Tk, version 0.3.2 (Chaumeil *et al.* 2020), which relied on the GTDB, release 89 (Parks *et al.* 2018). All genome bins had a minimum percent completeness of 50% and maximum percent contamination of 10% based on CheckM (Parks *et al.* 2015), according to default ATLAS settings.

In addition, selected replicate RNA samples from 2017 and 2018 field sampling were sequenced. These samples corresponded to a subset of the metagenomes sequenced from 2017 and 2018. Metatranscriptome library preparation and sequencing was performed by the McMaster Genome Facility (McMaster University). The Ribo-Zero rRNA Removal Kit (Bacteria; Illumina) was used to deplete rRNA genes in the samples prior to library preparation. Library preparation was then performed using the NEBNext Ultra II RNA Library Prep Kit for Illumina (New England Biolabs). Consistent RNA amounts for all samples, following rRNA depletion, were used as input and without directional RNA selection. The resulting library was sequenced on a single lane of a HiSeq 2500 (Illumina) in Rapid Run Mode with 2x250 bp paired end reads; this was the same sequencing run as for the 2017-2018 metagenomes described above. Sequencing generated 11.1 to 40.6 million reads per replicate. Metatranscriptome data were then processed using the ATLAS pipeline, version 2.2.0 (Kieser *et al.* 2020), up to the end of the "qc" module.

3.2.5 Unassembled metagenome and metatranscriptome read analyses

Quality controlled and unassembled metagenome and metatranscriptome reads were searched for taxonomic and functional genes of interest. Short peptide sequences were predicted from nucleotide read data using FragGeneScanPlusPlus commit 299cc18 (Singh *et al.* 2018). These peptide sequences were then queried with HMMs of interest using MetAnnotate version 0.92 (Petrenko *et al.* 2015). The HMMs used included the HMM for the *cyc2* candidate gene marker for iron oxidation

developed previously (Chapter 2) and HMMs downloaded from the FunGene database (Fish *et al.* 2013) for the *rpoB* single-copy taxonomic marker gene, *dsrA* gene implicated in sulfide metabolism, and *pmoA* and *mcrA* genes implicated in methane cycling. In addition, HMMs were downloaded from the Pfam database (Finn *et al.* 2016) for Type I (PF00223) and Type II (PF00124) photosynthetic reaction center genes, the *fmoA* gene involved in bacteriochlorophyll *a* binding (PF02327), and the nickel-dependent hydrogenase large subunit gene *hupL* (PF00374). A HMM was also downloaded from the TIGRFAM database (Haft *et al.* 2013) for the (bacterio)chlorophyll synthesis gene *bchL* (TIGR01281). MetAnnotate was run with default settings, and the RefSeq database (O'Leary *et al.* 2016), release 80 (March 2017), was used for taxonomic classification of hits. Raw search results and taxonomic classifications were summarized and statistically processed using the metannoviz library, version 1.1.0 (https://github.com/metannotate/metannoviz; doi:10.5281/zenodo.3963673). An e-value cutoff for short read search hits of 10^{-10} was applied, and read counts were normalized by HMM length before being expressed as a proportion of length-normalized *rpoB* hits.

3.2.6 Genome bin analyses

Functional gene content of recovered genome bins associated with the *Chlorobia* class was analyzed using the BackBLAST pipeline (Bergstrand *et al.* 2016), version 2.0.0-alpha4 (https://github.com/leeBergstrand/BackBLAST_Reciprocal_BLAST; doi:10.5281/zenodo.3465954), using the same queries and reference genome collection as described previously (Chapter 2), except that the genome of *Chl.* sp. strain BLA1 was added to the reference set (Lambrecht 2019). In addition, the entire genome bin collection was searched for genes associated with Type II or Type I photosynthetic reaction centers using hmmsearch, version 3.1b2 (Eddy 2011), with the same HMMs described above for unassembled read-based analyses.

3.2.7 Multivariate statistics and microbial diversity measures

Lakes were grouped by geographic and physicochemical parameters using hierarchical clustering. Physicochemical values from the lakes were averaged (e.g., in the case of light attenuation constants or isotopic differences between hypolimnion/epilimnion $\delta^{13}C$ values), or maximum measured values were used (e.g., in the case of total dissolved iron or sulfate), to produce a single value per each parameter per lake. Parameters were then normalized to a scale of -1 to 1 using the MinMaxScaler function in scikit-learn version 0.23.1 (Pedregosa *et al.* 2011). Normalization relied on simple multiplicative and translation-based transformations. A hierarchical clustering dendrogram

was then produced using seaborn version 0.10.1 (doi:10.5281/zenodo.592845) with default settings, except that the Ward linkage method was used.

Principal coordinate analyses were performed using scikit-bio version 0.5.6 (The scikit-bio development team – http://scikit-bio.org). For ASV data, the rarefied ASV table was used as input to calculate a dissimilarity matrix based on Bray Curtis distances, and the dissimilarity matrix was then used to produce the PCoA. Numeric metadata were then normalized from 0 to 1 using the MinMaxScaler function in scikit-learn version 0.23.1 (Pedregosa *et al.* 2011) and metadata was used to calculate the features for the PCoA biplot. For metagenome data, relative abundances of MAGs were calculated by mapping reads from each metagenome to the full set of MAGs obtained from bin dereplication. Relative abundances were expressed for each metagenome/MAG pair as the number of unassembled reads from the metagenome that mapped to the MAG divided by the total number of unassembled reads from the metagenome. A 0.01% minimum abundance threshold was then applied for all MAG-based relative abundances to eliminate false counts due to incorrect read mapping. Before calculation of the dissimilarity matrix, relative abundances of MAGs within each metagenome were then normalized to sum to 100% to avoid bias in the ordination. The dissimilarity matrix, PCoA, and biplot features were generated as described above for ASV data. Distance matrices for both ASV and MAG-based data were analyzed for statistically significant sample groupings using the PERMANOVA test (Anderson 2001) as implemented in scikit-bio, with default settings used except that 10 000 test permutations were run.

The presence/absence of MAGs in metagenome samples was calculated using a minimum abundance threshold of 0.3%. Relative abundances of MAGs were calculated by dividing the number of unassembled reads from a given metagenome that mapped to the MAG by the number of unassembled reads that mapped to all assembled contigs from that metagenome dataset.

3.2.8 Iron/sulfur cycling rate measurements

Water was collected from the upper and illuminated anoxic zone of Lake 227 in July 2018 for rate measurements. Three depths were sampled: 3.88 m, which was just beneath the oxic/anoxic zone boundary as detected by sonde, along with 4.25 m and 5.00 m. Water was pumped and overflowed into 200 mL glass serum bottles (four bottles per depth), and bottles were immediately sealed with blue butyl rubber stoppers (Bellco Glass) and kept chilled. Upon return to the sampling camp, 10 mL of headspace from each bottle was replaced with dinitrogen gas, and water was amended to a final

concentration of ~20 μM ferrous iron using ferrous chloride. Bottles were then shipped, chilled, to The University of British Columbia. Water was either left unamended or was amended with a final concentration of 50 μM DCMU (Sigma-Aldrich; St. Louis, Missouri, U.S.A.) to block oxygenic phototroph activity (Clavier and Boucher 1992). Water was then anoxically distributed into 50 or 25 mL sealable glass syringes that were pre-fitted with contactless Trace Range Oxygen Sensor spots (TROXSP5 – PyroScience; Aachen, Germany) and syringes were incubated at 10°C under 60 W fluorescent lights with ~20 µmol m^{-2} s^{-1} intensity in the 400-700 nm wavelength range. Samples were taken from syringes at regular time intervals to measure iron species concentrations using the ferrozine assay (Stookey 1970; Viollier *et al.* 2000) and sulfide concentrations using a spectrophotometric method (Cline 1969). Dissolved oxygen levels were also measured using a FireStingO2 (FSO2-4 – PyroScience; Aachen, Germany) oxygen meter. Once ferrous iron was fully oxidized or after the mid-point of the experiment, syringes were transferred to the dark but continued to be sampled regularly until the end of the experiment.

3.3 Results and discussion

The surveyed lakes had diverse physicochemical properties despite their close geographic proximity (Figure 3.1). Due to shallow surficial till depth and topography, lakes in the Boreal Shield can be small in area and densely clustered (Figure 3.1A), and yet nearly adjacent lakes can have substantial physicochemical differences. Two surveyed lakes, Lakes 221 and 224, were separated by less than 1 km but had nearly the maximum and minimum measured upper-limit concentrations of dissolved organic carbon within the survey, respectively (Figure 3.1B). Additionally, Lakes 221 and 222, which were adjacent and connected, had the maximum and minimum upper-limit concentrations of sulfate, respectively, that were measured when surveyed. This heterogeneity of lake properties resulted in samples spanning a gradient of parameters, such as the timing of seasonal anoxia over the course of the field season (Figure 3.1C), and allowed exploration of a wider microbial diversity of Boreal Shield lake systems along with drivers of lake microbial community composition (Figure 3.1D).

Although the properties of each surveyed lake were unique, lakes could generally be categorized as shallow (<6 m max. depth) with high dissolved organic carbon and thus high light attenuation coefficients, such as Lakes 221, 304, and 222, or as deeper lakes (>11 m max. depth) that had low to moderate dissolved organic carbon levels, such as Lakes 442, 373, 224, 626, and 239.
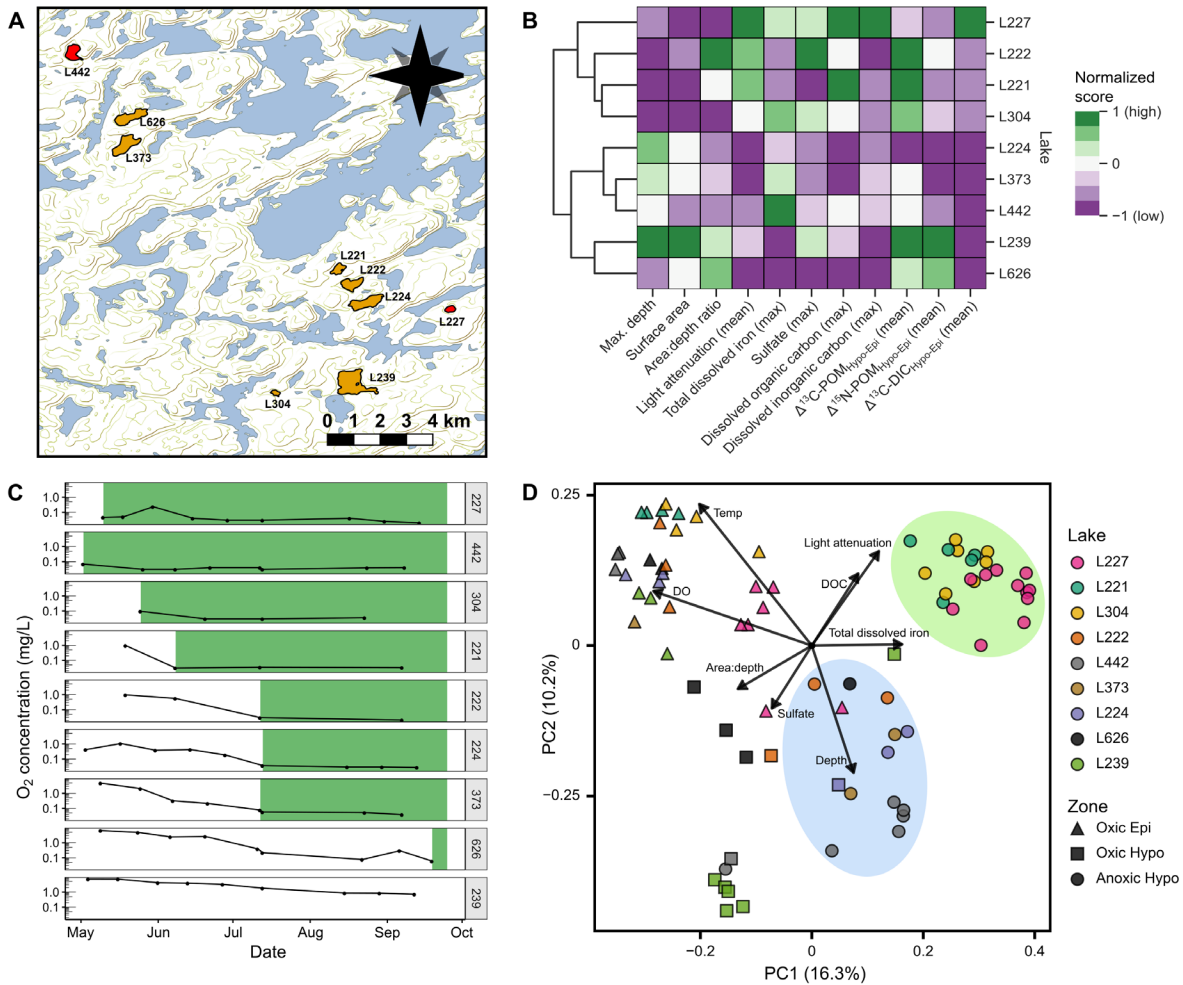
Figure 3.1 Geography and physicochemical properties of the surveyed lakes. A, Map of the nine surveyed lakes at the IISD-ELA. The two lakes surveyed via metagenomics previously are highlighted in red, and the other seven are highlighted in yellow. **B**, Heatmap summarizing key geographic and physicochemical parameters of the surveyed lakes. Values for each attribute, summarized in Table 3.1, have been normalized on a scale of -1 (lowest value; e.g., most shallow lake) to +1 (highest value; e.g., deepest lake). The results of hierarchical clustering of the lakes based on these normalized data are shown as a dendrogram. **C**, Timing of anoxia development in the nine surveyed lakes during the 2016 field season. For each time point, the lowest concentration of dissolved oxygen measured in a lake during water column profiling is shown. The panel for each lake is highlighted green following the first measurement of anoxic lake water. **D**, PCoA biplot summarizing the diversity of collected lake samples based on relative abundances of 16S rRNA gene ASVs. Arrows indicate the magnitude and direction of correlation of each geographic or physicochemical parameter (available in the code repository associated with this work) with ASV data. The green and blue ovals correspond to the observed groupings of anoxic zone microbial communities from small humic lakes and larger lakes, respectively, as described in the text. Spatial data are courtesy of ESRI and DMTI Spatial.

Table 3.1 Physico-chemical properties of the surveyed lakes. The parameters used to generate Figure 3.1B are shown, along with methane concentrations, for which no data were available for Lake 373, preventing use in dendrogram construction. For parameters where mean or maximum values are shown for each lake (see column labels), these summary statistics were calculated using measurements taken for the lake during the entire 2016-2018 sampling campaign. Isotopic differences (e.g., $\Delta^{13}C\text{-POM}_{Hypo\text{-}Epi}$) represent the difference in isotopic value (e.g., $\delta^{13}C\text{-POM}_{Hypo}$ - $\delta^{13}C\text{-POM}_{Epi}$) between the top (epilimnion) and bottom (hypolimnion) depths that were sampled and measured for a given lake at a given sampling time point. The averages of all measured isotopic differences for each lake are shown.

Lake	Max. depth (m)	Surface area ($m^2 \times 10^4$)	Area:depth ratio ($m \times 10^4$)	Light attenuation (mean; m^{-1})	Total dissolved iron (max; μM)	Sulfate (max; μM)	Methane (max; μM)
L227	10.0	5.0	5.0	1.74	222.9	23.3	914.6
L222	6.0	16.4	27.3	1.43	84.5	29.7	0.5
L221	5.5	9.0	16.4	1.35	73.3	14.2	103.0
L304	6.0	3.6	6.0	1.09	231.6	23.5	94.3
L224	25.0	25.9	10.4	0.34	102.5	17.8	0.1
L373	20.0	27.3	13.7	0.40	207.0	17.4	N/A
L442	17.0	16.0	9.4	0.65	310.4	19.8	135.3
L239	30.0	54.3	18.1	0.77	15.6	25.2	0.0
L626	11.0	25.9	23.5	0.47	19.6	15.1	0.3
Min.	5.5	3.6	5.0	0.34	15.6	14.2	0.0
Max.	30.0	54.3	27.3	1.74	310.4	29.7	914.6
Mean	14.5	20.4	14.4	0.92	140.8	20.7	156.0

Table 3.1, cont.

Lake	Dissolved organic carbon (max; μM C)	Dissolved inorganic carbon (max; μM C)	$\Delta^{13}C\text{-POM}_{Hypo\text{-}Epi}$ (mean; ‰)	$\Delta^{15}N\text{-POM}_{Hypo\text{-}Epi}$ (mean; ‰)	$\Delta^{13}C\text{-DIC}_{Hypo\text{-}Epi}$ (mean; ‰)
L227	1271	2232	-14.01	-5.5	6.76
L222	833	428	0.58	-1.9	-12.16
L221	1310	624	-2.29	-4.86	-12.08
L304	865	752	-2.94	-3.1	-9.68
L224	304	595	-23.03	-7.18	-13.85
L373	434	909	-9.72	-6.24	-14.52
L442	759	1044	-12.07	-5.34	-16.23
L239	686	514	-0.23	3.96	-15.02
L626	410	274	-8.72	2.33	-13.52
Min.	304	274	-23.03	-7.2	-16.23
Max.	1310	2232	0.58	4.0	6.76
Mean	764	819	-8.05	-3.1	-11.14
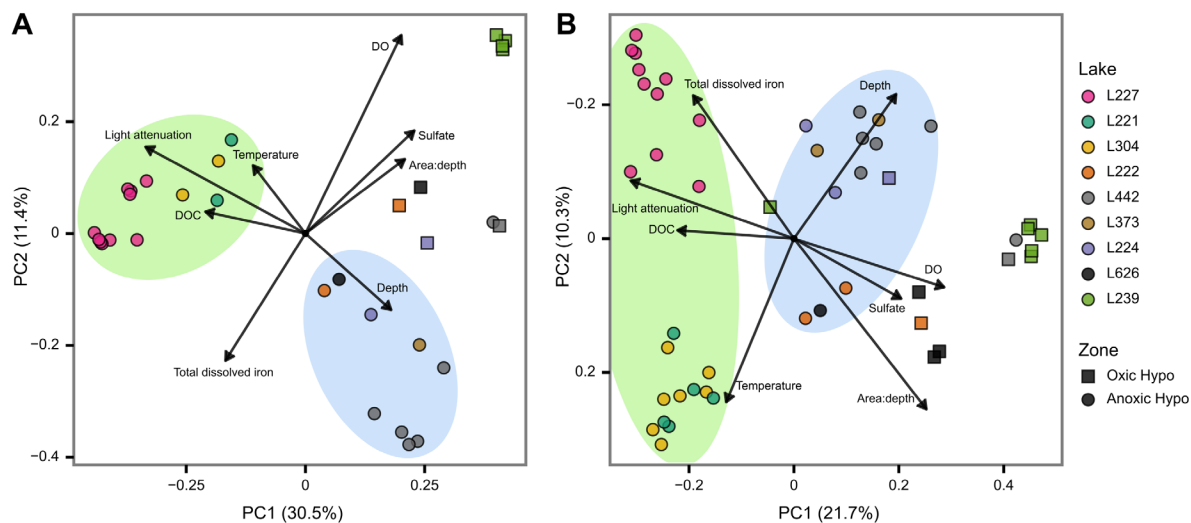
Figure 3.2 Principal coordinate analyses of samples collected from the hypolimnia of the surveyed lakes. A, Biplot based on relative abundance of MAGs. **B**, Biplot based on relative abundances of 16S rRNA gene ASVs showing all hypolimnion samples presented in Figure 3.1D. The formatting of both panels matches that of Figure 3.1D. Green and blue ovals correspond to the statistically significant groupings of anoxic zone microbial communities from small humic lakes and larger lakes, respectively, as described in the text.

Indeed, hierarchical clustering of normalized lake property data (Figure 3.1B) separated the lakes into these broad-level categories. Lake 227, which has been experimentally manipulated with the addition of phosphates and/or nitrogen compounds for over 40 years (Schindler 1974; Schindler *et al.* 2008), also grouped with Lakes 221, 304, and 222 yet clustered somewhat separately, potentially due to the lake's high levels of dissolved organic carbon, dissolved inorganic carbon, methane, and other nutrients. Additionally, Lakes 239 and 626 grouped somewhat separately from Lakes 442, 373, and 224. Lakes 239 and 626 may have clustered distinctly because of their high surface area:depth ratios or because both lakes had limited to no development of anoxic conditions during the survey period (Figure 3.1C).

Lake microbial community composition generally corresponded to observed physicochemical groupings of the lakes (Figure 3.1D). In a PCoA plot of lake samples based on relative abundances of ASVs, samples at a broad level appeared to group first by redox status (i.e., oxic versus anoxic), then by lake water column zone (e.g., epilimnion versus hypolimnion), and then by lake (e.g., Lake 442 versus Lake 373). Redox status is thought to be a key driver of microbial community composition in many environments globally (Ramírez-Flandes, González and Ulloa 2019), and the redox balance of a given water column sample is also generally related to its position in the water column. Interestingly,

anoxic hypolimnion samples appeared to separate into two groups: those belonging to Lakes 227, 221, and 304, and those belonging to Lakes 222, 442, 373, 224, and 626 (Figure 3.1D). This apparent clustering exactly matched highest level clustering predictions based on physicochemical data, with the exception of Lake 222, which had been grouped with Lakes 221 and 304 (Figure 3.1B). Ordinations of only hypolimnion samples generally supported the observed clustering (Figure 3.2). A PCoA plot based on the relative abundances of MAGs (Figure 3.2A) maintained well-separated L227/221/304 and L222/442/373/224/626 groups, which were found to be statistically significant groupings when tested along with a third "oxic hypolimnion" group (PERMANOVA, n=32, F=8.30, p<0.001). Although a PCoA plot based on hypolimnion samples within the ASV dataset indicated that Lake 227 samples grouped somewhat separately from Lake 221 and 304 samples (Figure 3.2B), such a grouping would match observations of lake clustering based on physicochemical data. The same sample groupings as tested for the MAG-based ordination were still found to be statistically significant groups for hypolimnion samples based on ASV data (PERMANOVA, n=48, F=7.23, p<0.001).

To further explore functional-level differences between microbial communities in the lake anoxic zones, phototrophic microbial community members and their genomic potential were profiled using metagenome data (Figure 3.3). Samples from Lakes 227 and 442 had been observed previously (Chapter 2) to contain high relative abundances of potentially phototrophic members of the *Chlorobia* class (also called "green sulfur bacteria"). Within the broader set of surveyed lakes, only Lakes 221 and 304, which had clustered with Lake 227, also had such high relative abundances of *Chlorobia* members (Figure 3.3A). *Chlorobia* members reached up to 33.5% and 25.7% relative abundance in Lake 304 based on ASV and MAG-based data, respectively, and up to 34.6% and 12.5% relative abundance in Lake 221 based on ASV and MAG-based data, respectively. (Data tables for ASV and MAG-based data are available in the data repository associated with this work.) Lakes 222 and 373 were also found to contain populations of *Chlorobia*, only at lower relative abundances of 1.0% and 4.9%, respectively, based on ASV data. In all of Lakes 227, 221, and 304, *Chlorobia*-associated *dsrA* genes, involved in sulfide oxidation, and *Chlorobia*-associated *cyc2* genes, potentially involved in photoferrotrophy or EET, were detectable. As observed previously (Chapter 2), even with additional years of sampling, no evidence was detected for *Chlorobia*-associated *cyc2* genes in Lake 442, which grouped into the second lake anoxic zone category.

Diverse anoxygenic and oxygenic phototrophs were also found in the studied lakes (Figure 3.3B). In Lakes 227, 221, and 304, 11 unique MAGs classified to the *Chlorobia* class were detected with genomic potential for phototrophy (based on encoding the *pscA* and/or *fmoA* genes). Interestingly, only four of these MAGs were shared between lakes when examined at the 0.3% relative abundance threshold; the other seven were unique as high relative abundance community members in each specific lake. Meanwhile, in Lakes 222 and 442, only three genomes bins of *Chlorobia* were detected with phototrophic potential, none of which were found at above 0.3% relative abundance in Lakes 227, 221, and 304. Although relative abundances of *Chlorobia* members in Lake 222 samples were too low for accurate quantification of *cyc2* based on unassembled read data, the detection of a *cyc2*-encoding genome bin of *Chlorobia* (*Chl.* bin ELA265) that was recovered from assembled contigs of Lake 222 metagenomes is evidence that *Chlorobia* members in Lake 222 may possess this cytochrome gene. Relative abundances of *Chlorobia* were low in Lake 373 for samples sent for metagenome sequencing (e.g., approx. 0.03% for Sept. 2016, 20 m depth), preventing evaluation of functional gene content. Five of the 14 total recovered MAGs of *Chlorobia* possessed the *cyc2* gene (Figure 3.4), and one of these (*Chl.* bin ELA225) encoded both the *cyc2* and the sulfide oxidation gene pathway, like what was observed previously for "*Ca.* Chl. canadense", which was enriched from the Lake 304 anoxic zone (Chapter 2). Four of the detected MAGs had at least 99.4% ANI to previously studied genome bins of *Chlorobium* from Lakes 227 and 442 (Chapter 2) and likely represented the same populations (Figure 3.5), although none of the recovered MAGs appeared to represent the same species as "*Ca.* Chl. canadense" (Figure 3.4). Several of the other genome bins of *Chlorobia* appeared to be fragmented based on having partial sulfate uptake or sulfide oxidation gene pathways (Figure 3.4). These genome bins are thus likely not reliable indicators of whether the populations they represent truly encode *dsrA* or *cyc2*, but their diversity and the detection of both *dsrA* and *cyc2* genes in the lake metagenomes implies that the potential functional diversity observed among populations of *Chlorobia* in Lake 227 previously (Chapter 2) also applies in Lakes 221, 304, and possibly 222.

The overall diversity of phototroph populations in Boreal Shield lakes has only been subject to limited study previously (Sinclair *et al.* 2017; Peura *et al.* 2018). Seven unique populations of potentially phototrophic members of the *Chloroflexia* class were detected across the surveyed lakes, all of which were found in Lakes 221, 304, 222, and 442. All but one of those populations (*Oscillochloris* bin ELA048) was unique to a single lake at the 0.3% relative abundance threshold. In particular, Lake 221 hosted diverse *Chloroflexia* members, including two canonical RCII-encoding

Oscillochloris members and two RCI-utilizing members of the highly novel and recently discovered "*Ca.* Chloroheliales" order. The latter two genome bins were similar to the genomes of the two enriched "*Ca.* Chloroheliales" members from Lake 227 (Chapter 4). Specifically, "*Ca*. Chloroheliales bin ELA729" had 99.4% ANI to "*Ca.* Chloroheliales bin L227-5C", and "*Ca*. Chlorohelix bin ELA319" had 87.45% ANI to "*Ca.* Chlorohelix allophototropha L227-S17" but had a 96% identical PscA-like predicted primary sequence. This finding represents the first reported detection of phototrophic "*Ca.* Chloroheliales" members in a natural environmental sample. In addition, at least eight genome bins of hypolimnion-dwelling *Cyanobacteria* members were recovered. Such genome bins were mostly detected in the second observed category of lake hypolimnia including Lakes 442, 373, 224, 626, and 239, and over half were found in multiple lakes. In Lake 227, the two detected members, classified to the *Pseudanabaenaceae* and *Nostocaceae* families, were unique to that lake and could have been enriched during artificial nutrient amendment. *Cyanobacteria* MAGs composed up to 2.7% of anoxic hypolimnion metagenome data from the lakes, excluding Lake 227, based on read mapping. All surveyed lakes also contained *Proteobacteria* members with potential for bacteriochlorophyll synthesis (Figure 3.3A). These members belonged mostly to the *Alpha-* and *Betaproteobacteria*, appeared most abundant in oxic hypolimnion samples (Figure 3.6), and likely represented aerobic anoxygenic phototrophic bacteria (Koblížek 2015; Kasalický *et al.* 2018) based on corresponding MAG classifications. A single MAG classified to the *Gemmatimonadota* phylum, detected in Lake 227 epilimnion metagenomes, also encoded RCII genes (i.e., *pufL* and *pufM*) with 91.6% and 82.7% identity at the predicted amino acid level, respectively, to the *pufL* and *pufM* genes of the aerobic anoxygenic phototroph *Gemmatimonas phototrophica* (Zeng *et al.* 2014). Altogether, these findings indicate that observed differences in overall microbial communities between lakes in the two hypolimnion categories also extend to differences in the composition and functional potential of phototroph communities.
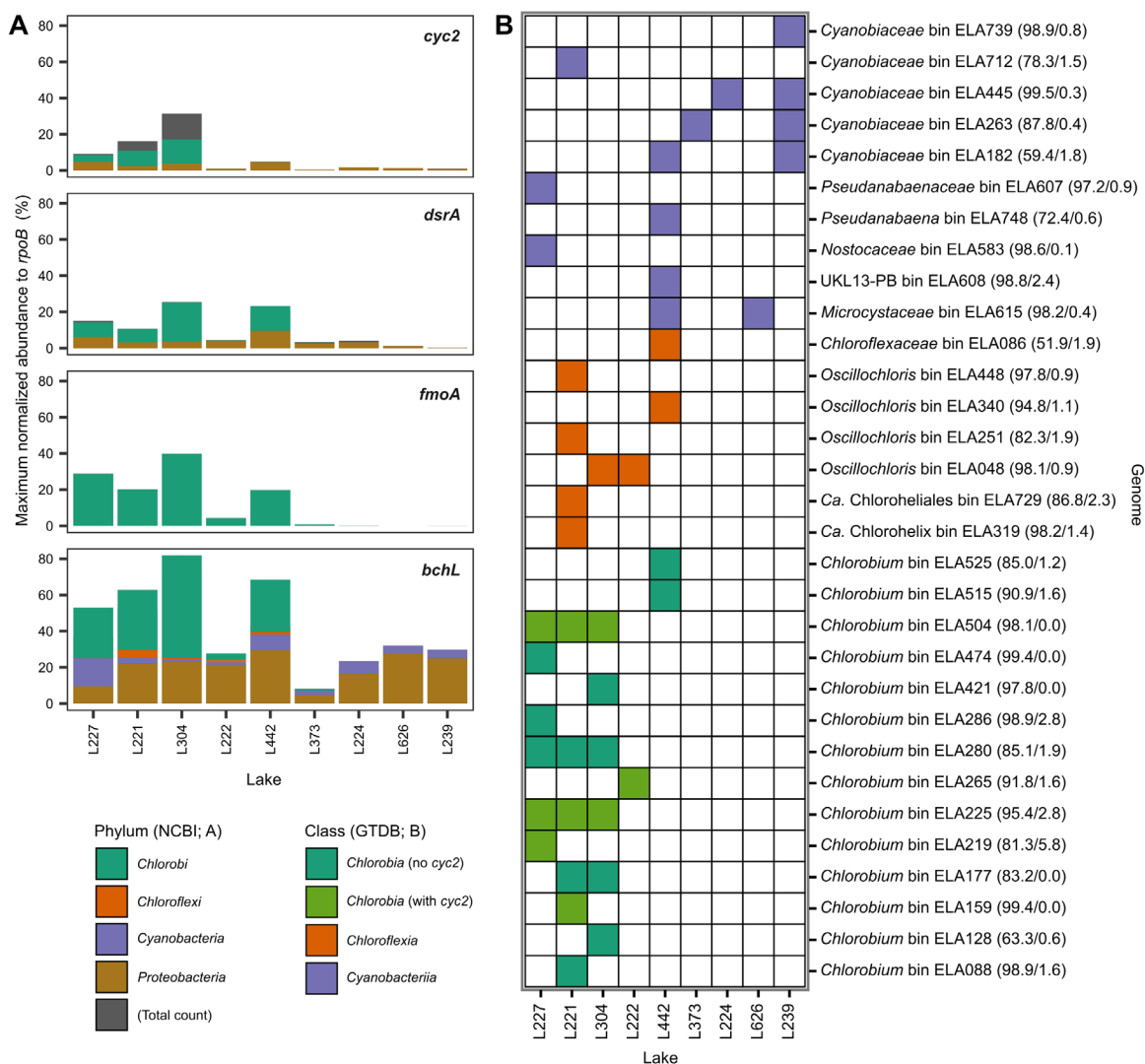
Figure 3.3 Phototrophic microbial consortia among the surveyed lakes. A, Bar plot showing the normalized abundances of functional genes involved in phototrophy within the nine surveyed lakes. The maximum normalized abundances of selected phyla across all hypolimnion metagenomes for a given lake are shown. **B**, Heatmap showing the presence/absence of phototroph MAGs in hypolimnion samples from each lake. A minimum relative abundance cutoff of 0.3% was used based on read mapping to metagenomes. Only genomes from the three classes listed were shown. The presence of the *cyc2* gene among MAGs of *Chlorobia* was determined based on Figure 3.4. To the right of each MAG name, the percent completeness and contamination of the MAG, as estimated by CheckM, are shown.
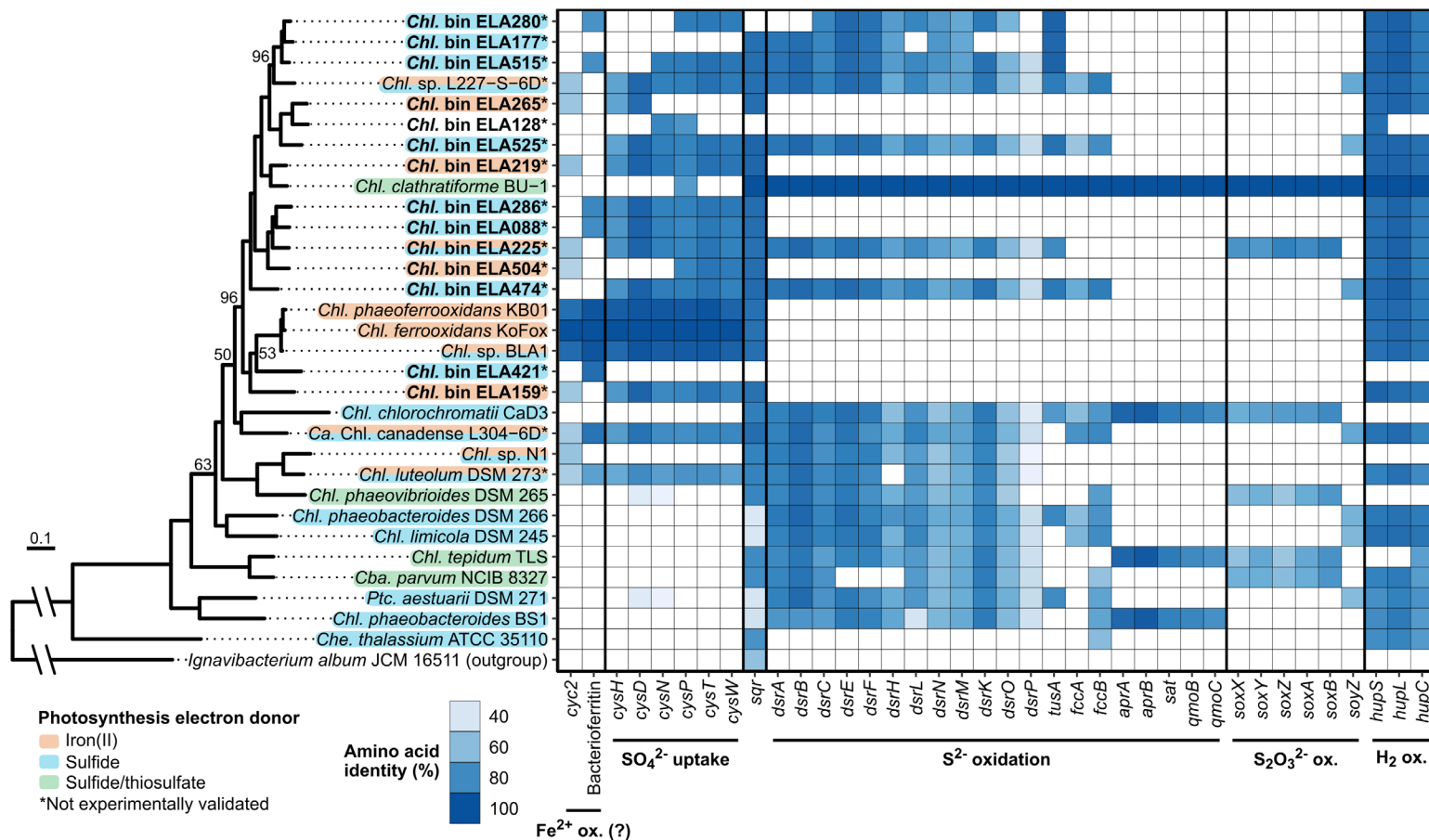
Figure 3.4 Genomic potential of recovered MAGs of *Chlorobia* for usage of metabolic electron donors. The maximum likelihood phylogenetic tree is based on a concatenated set of 74 core bacterial proteins, and the scale bar represents the percent amino acid change. Bootstrap values of 100/100 are omitted for clarity. The heatmap shows the percent amino acid identity of successful bidirectional BLASTP hits to the query primary sequence (either belonging to *Chl. ferrooxidans* or *Chl. clathratiforme*).

67
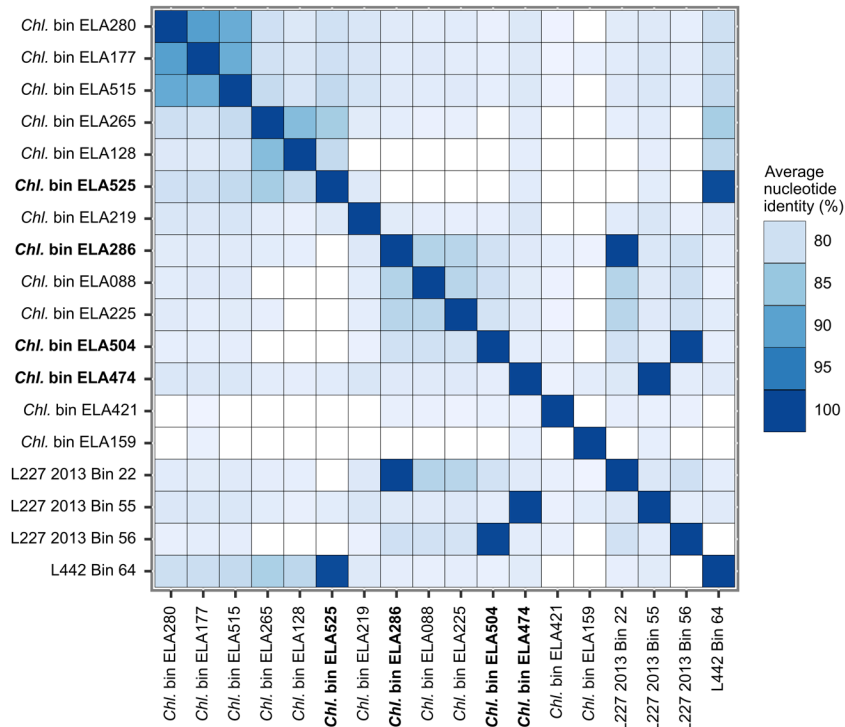
Figure 3.5 Average nucleotide identity of recovered MAGs of *Chlorobia*. The MAGs recovered in this study (i.e., the top 14 MAGs on the y-axis from *Chl.* bin ELA280 to *Chl.* bin ELA159) are compared to the environmental MAGs recovered in a previous metagenome study of Lakes 227 and 442 (i.e., the bottom four MAGs on the y-axis; see Chapter 2). Bolded MAGs from this study are a nearly perfect ANI match to the MAGs from the previous study (Chapter 2).

The observed separation between phototroph communities in the surveyed lake hypolimnia could be due to several reasons. The "phototroph-rich" hypolimnia of Lakes 227, 221, and 304 developed anoxia early in each field season (e.g., Figure 3.1C), and samples from these hypolimnia were distinguished from other lake hypolimnion samples most directly in PCoA biplots based on their surface area:depth ratios, dissolved organic carbon, light attenuation coefficients (Figure 3.1D). Lake topography and basin shape play key roles in anoxia development, given that steeply sloping lakes of sufficient depth, with protection from wind-driven mixing, will become thermal stratified and thus allow for depletion of dissolved oxygen in their bottom waters. In addition, high DOC lakes will readily absorb light in their surface waters, which further facilitates thermal stratification and anoxia development. For Lakes 221 and 304, the temperatures of the lake hypolimnia are higher than observed in other lake systems (i.e., ~8-10°C rather than ~4-5°C), likely due to the shallow depths of these lakes, which could also accelerate the growth of deep phototroph populations. In contrast, Lake 222 might not have developed high relative abundance populations of phototrophs due to its high

surface area:depth ratio, which enhances vertical water column mixing by wind and slows the development of anoxia. It is possible that, given sufficient time or artificial nutrient amendment, the hypolimnion of Lake 222 might develop similar phototrophic consortia as observed in other anoxygenic phototroph-rich lakes.

At the same time, the total dissolved iron and sulfate concentrations in the lakes could also play a role in distinguishing between these two hypolimnion types. Sulfate and total dissolved iron levels between lake samples were generally negatively correlated with one another based on biplot features (Figure 3.1, Figure 3.2), and the biplot features for these parameters were somewhat aligned with the separation of hypolimnion groups. In the context of early Earth history, the ratio of available sulfur to highly reactive iron (S:Fe$_{HR}$) is thought to be a key indicator of an aquatic environment's tendency to develop ferruginous compared to euxinic conditions (Poulton and Canfield 2011). Over time, most Boreal Shield lakes that develop anoxia will gain ferruginous bottom waters due to a high flux of ferrous iron from the sediments, given their underlying geology (Campbell and Torgersen 1980; Schiff *et al.* 2017). However, in the upper anoxic water column where rapid sulfate reduction might occur among Boreal Shield lakes (see below), it is unclear whether sufficient iron could be retained to support metabolisms such as photoferrotrophy compared to sulfide-oxidizing phototrophy. For at least one Lake 222 timepoint (Sept. 2017), sampling within the upper anoxic zone revealed a relatively high sulfate concentration of 29.7 µM compared to a total dissolved iron concentration of 16.3 µM, which could have allowed sulfur cycling processes to dominate iron cycling processes at that depth. Future work probing the fluxes of dissolved iron and sulfate in the upper lake anoxic zone, combined with study of the corresponding rates of microbial iron/sulfur cycling processes, could allow for the role of dissolved iron and sulfate in governing lake microbial community structure to be more carefully explored.

Seeking to better understand the function and activity of microbial communities in phototroph-rich anoxic zones, Lakes 227, 304, and 221 were revisited in September 2017 and July 2018 to characterize the potential microbial activity of these phototroph-rich lake systems (Figure 3.7). Metatranscriptome sequencing on RNA samples collected from the anoxic hypolimnia of all three lakes revealed that phototrophs in the water columns were highly active, accounting for 40-60% of all detected reads for the single-copy marker gene *rpoB* (Figure 3.7A). Expression of the *fmoA* gene involved in bacteriochlorophyll *a* binding among *Chlorobia* members was estimated at 5-10 times
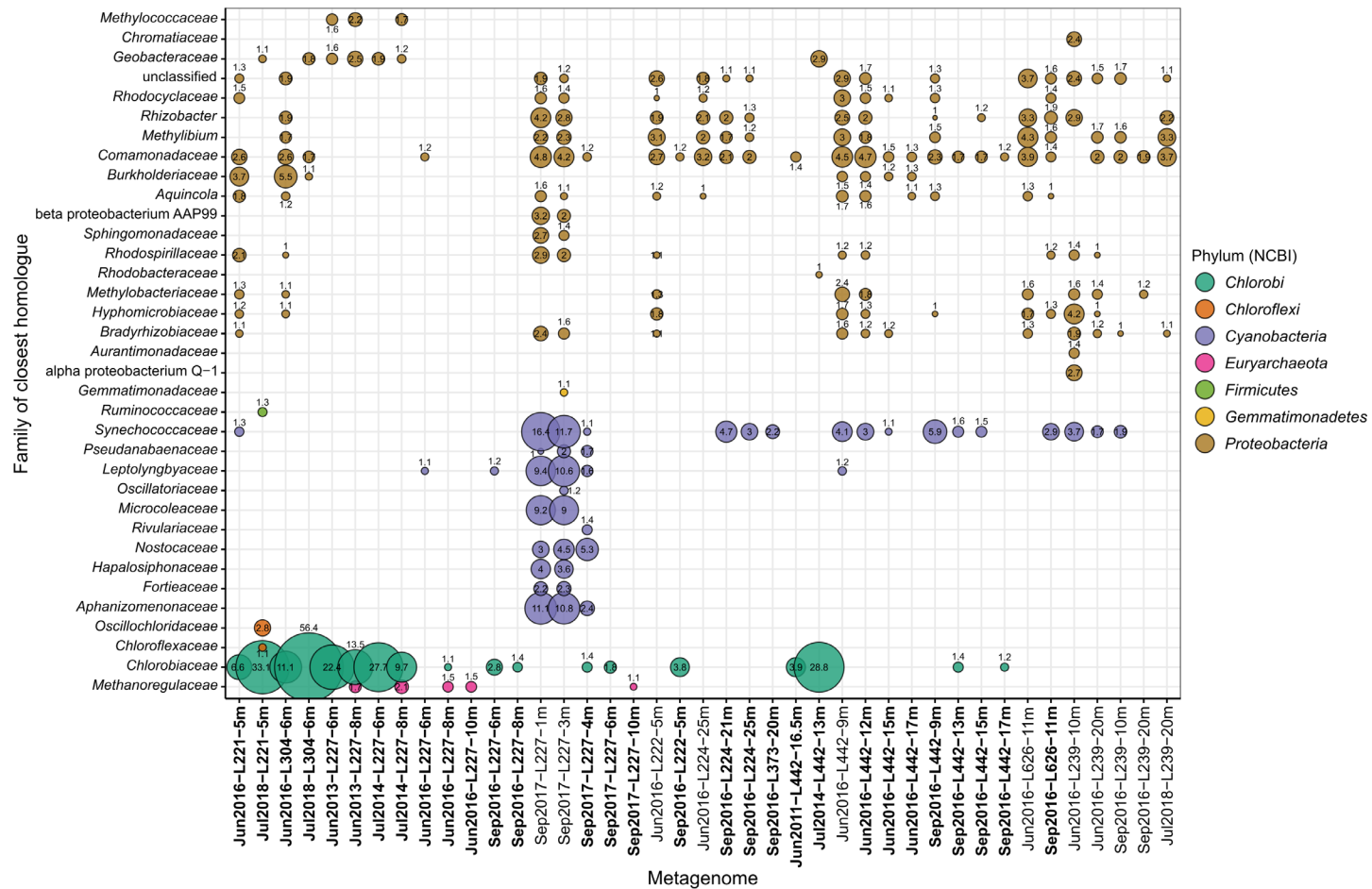
Figure 3.6 Functional potential for (bacterio)chlorophyll synthesis among lake metagenomes. The bubble plot shows the relative abundance, normalized to *rpoB*, of each microbial family (NCBI taxonomy) within lake metagenomes based on the abundance of the *bchL* gene in unassembled metagenome reads. All metagenomes in bold represent samples collected from lake anoxic zones.

70

higher than baseline community expression levels. Both the *cyc2* gene and *dsrA* genes were detected as expressed in all three lakes, with consistently higher levels of *dsrA* gene expression compared to *cyc2*. In Lake 227, where *Chlorobia* members appeared highly active, *cyc2* expression was at low levels nearly indistinguishable from false positive hits, whereas *dsrA* expression was the highest of any observed lake, implying that *cyc2* expression was not favorable at the time of sampling (September 2017) for Lake 227. Lakes 221 and 304 (July 2018) had higher *Chlorobia*-associated *cyc2* expression by comparison. *Chlorobia*-associated Ni-Fe hydrogenase gene expression was detected in all three lakes, along with some expression of *Chloroflexota*-associated hydrogenase genes in Lake 221. Lastly, expression of PSI genes by *Cyanobacteria* members were detectable in all three lakes, implying oxygenic phototroph activity even in the anoxic and ferruginous lake waters that were sampled. Lake 227 in particular had high *psaA/psaB* expression levels nearly five-fold that of baseline community expression levels.

Lake 227 was sampled again in July 2018 to measure rates of phototrophic iron and sulfur oxidation (Figure 3.7B). Consistent with metatranscriptome data, the sampled lake water appeared to have high activity of oxygenic phototrophs despite the lack of measurable dissolved oxygen content in the water during sampling. When unamended with the oxygenic phototrophy inhibitor DCMU, lake water at 4.25 m as well as at 3.88 m (Figure 3.8) had a rapid increase in dissolved oxygen levels, as well as rapid oxidation of ferrous iron, within the first 24 hours of incubation. When moved to the dark, in the 4.25 m incubation, ferric iron was reduced completely to ferrous iron, presumably due to the activity of iron-reducing bacteria that could tolerate brief exposure to dissolved oxygen, which reached a peak concentration of 4.4 µM in that incubation. By contrast, no geochemical changes were observed in any of the DCMU-amended treatments (Figure 3.8), which block expression of oxygenic but not anoxygenic phototrophs, during light exposure. In the dark, however, sulfide accumulation was observed for the incubation representing 5.00 m depth, implying that sulfate-reducing bacteria were active at this depth. Moreover, the dark-associated sulfide increase also implies that cryptic and phototrophic sulfide oxidation occurred in the light for this incubation, or else sulfide levels would have increased consistently throughout the experiment. These findings imply that sulfide-oxidizing phototrophy was occurring at 5.00 m depth even if sulfide was non-measurable. Unfortunately, because no substantial ferric iron was present at the start of any of the incubations, it is impossible to distinguish whether cryptic iron oxidation/reduction might have also been occurring.
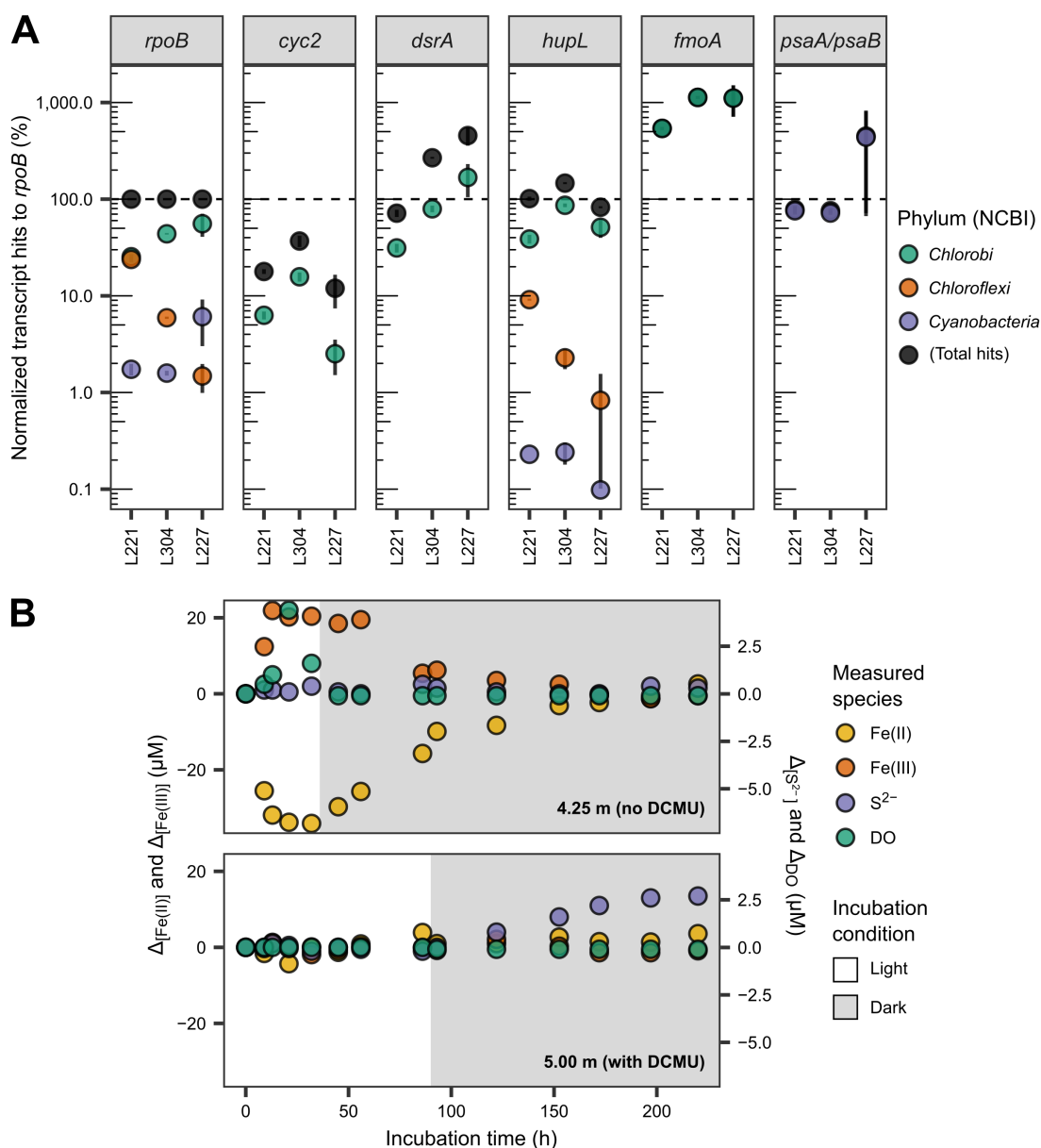
Figure 3.7 Phototroph gene expression and activity among Lakes 221, 304, and 227. A, RNA expression levels of phototrophy functional genes. Transcript hits for each gene and taxonomic group are normalized to total metatranscriptome *rpoB* hits. Error bars show the standard deviation of hits from triplicate metatranscriptomes. Lake 227 was sampled in September 2017, and Lakes 221 and 304 were sampled in July 2018. **B,** Geochemical changes in incubated anoxic water, sampled from Lake 227 in July 2018, over time. Samples were kept in the light until the grey zone of the figure, where they were moved to the dark. Concentrations are expressed as changes from initial levels. All initial concentrations were at assay detection limits except for ferrous iron, which ranged in initial concentration from 31.9 to 42.1 µM. Only two example profiles are shown here, one without and one with Diruon (DCMU) added. The full set of geochemical profiles are included in Figure 3.8.
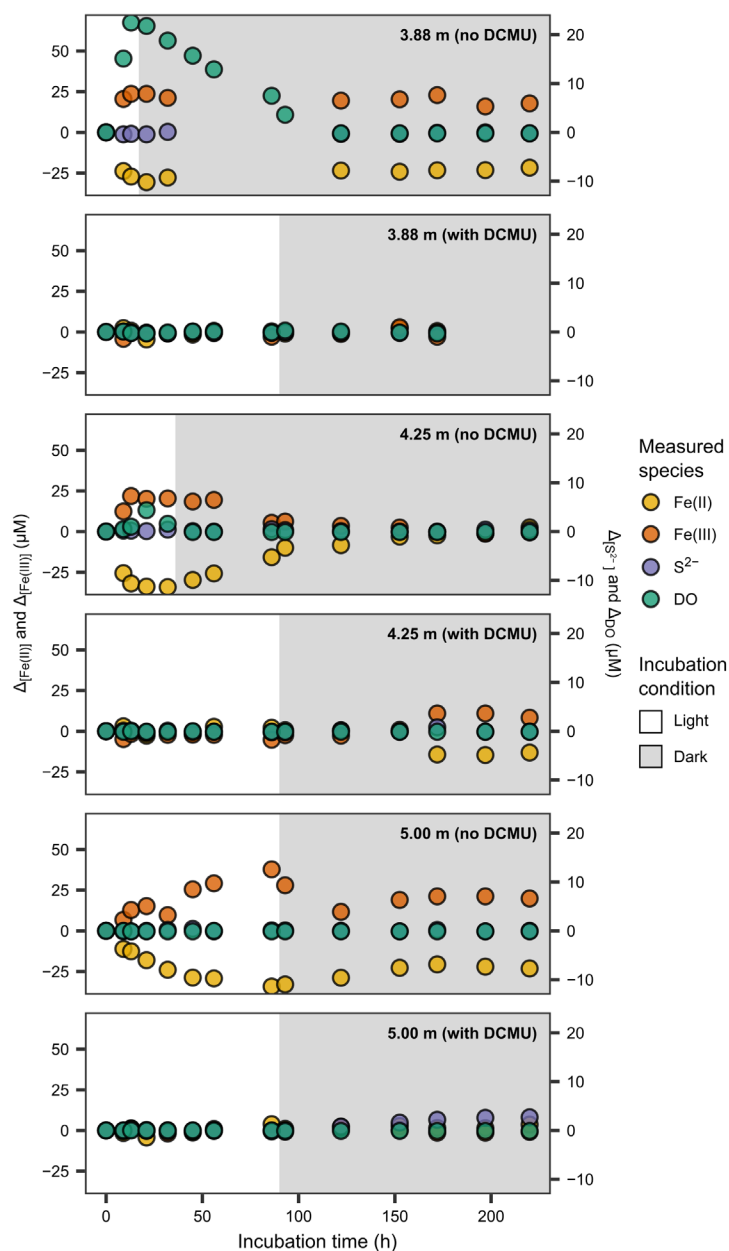
Figure 3.8 Geochemical changes for incubated Lake 227 water from all sampled depths. Each panel is laid out in the same manner as Figure 3.7B. Initial concentrations of ferrous iron in the incubations ranged from 30.1 to 42.1 µM, and other initial measured concentrations were at assay detection limits. For the 3.88 m and 4.25 m incubations spiked with Diuron, the dissolved oxygen sensors were not calibrated, and so the initial dissolved oxygen concentrations are unknown. However, the lack of ferrous iron oxidation in these incubations implies that dissolved oxygen was at trace levels or absent.

The detection of rapid ferrous iron oxidation by *Cyanobacteria* members in the Lake 227 anoxic zone has important implications for early Earth research. Various laboratory experiments simulating the period around the Great Oxidation Event have been performed to understand oxygenic phototroph activity in the context of ferruginous waters (Koeksoy *et al.* 2016). Microbial mats of iron-rich hot springs have also been examined (Planavsky *et al.* 2009). However, no previous study has used a natural pelagic system as a Great Oxidation Event analogue for growth and activity of oxygenic phototrophs. Incubation data indicate that rates of ferrous iron oxidation exceeded rates of ferric iron reduction, meaning that iron-rich floc material could be generated by the activity of *Cyanobacteria* members and precipitate to the bottom of the lake in a manner analogous to that hypothesized for the deposition of banded iron formations (Johnson *et al.* 2008). Decreasing activity of oxygenic phototrophs was also observed moving downward in the water column. For example, at 5.00 m depth, the non-DCMU amended incubation had slow ferrous iron oxidation and no detectable accumulation of dissolved oxygen, implying that oxygen production was being tightly coupled to iron oxidation (Figure 3.8). Such tight coupling could have been facilitated by co-occurring microaerophilic iron-oxidizing bacteria, as has been observed previously (Field *et al.* 2016), or, although less likely, could have been driven by a novel form of direct iron oxidation by oxygenic phototrophs. Such depth-dependent changes in phototrophy could be applied to a highly reducing Archaean ocean setting to better understand factors controlling phototrophy rates in ancient waters.

Moreover, the combined metatranscriptome and phototrophy rate data provide important insights into the phototrophic microbial consortia of Boreal Shield lakes. Although the functional role of *cyc2* among *Chlorobia* members is still not validated, the detected expression of *cyc2* in Lakes 221 and 304, as well as the detection of additional genome bins encoding *cyc2* yet lacking the typical sulfide oxidation gene pathway, leaves open the possibility that photoferrotrophy occurs in phototroph-rich anoxic hypolimnia. Nevertheless, the data presented here imply that phototrophic sulfide oxidation was the more favourable phototrophic process. Expression of *dsrA* was consistently higher than *cyc2* expression, and any potential photoferrotrophy in Lake 227 can be constrained as slower than the rate of iron reduction at the time of sampling. Also, genomic potential for phototrophic sulfide oxidation is more broadly distributed among Boreal Shield lakes, such as Lake 442. Such findings agree with observations from Lake La Cruz and Lake Matano, where photoferrotrophy, if present, did not appear to be the dominant process (Crowe *et al.* 2014a; Walter *et al.* 2014), and it is possible that photoferrotrophy is not a highly favourable mode of phototrophy in many Boreal Shield lakes. At the same time, the data presented here demonstrate the presence of a cryptic sulfur cycle in the upper anoxic waters of Lake 227, hinting that sulfur cycling is an important

contributor to biogeochemical cycling in Boreal Shield lakes and matching observations in Lake Pavin (Berg *et al.* 2019). The combined metagenome and activity data from this study also leave open the possibility that cryptic iron cycling is occurring in these systems.

Methane also represented an important factor for Earth's early development and is important in the context of modern global change (Crowe *et al.* 2011). Phototrophy is directly linked to the methane cycle given that organic carbon inputs from primary production fuel methanogenesis (Thompson *et al.* 2019), and others have observed aerobic methane oxidation directly coupled to oxygenic phototrophy (Milucka *et al.* 2015). Metatranscriptome data from across the water column of Lake 227 in September 2017 suggests that active methane cycling occurs in this lake, including potentially novel forms of methane oxidation (Figure 3.9). Aerobic methane oxidation activity is typically thought to peak at or just below the oxic-anoxic zone interface of lakes (Mayr *et al.* 2019; Thottathil, Reis and Prairie 2019), but substantial relative abundances of gammaproteobacterial methanotroph populations (see the data repository associated with this work) and high expression of the particulate methane monooxygenase gene *pmoA* (Figure 3.9) were detected throughout the Lake 227 anoxic water column. Hits of *pmoA* sequences in Lake 227 anoxic zone metatranscriptome data consistently had closest similarity to monooxygenases in the genomes of known gammaproteobacterial methanotrophs in the RefSeq database, making it unlikely that these hits were false positives. Even at 8 m and 10 m depth, 4-6 m below the oxic-anoxic zone interface and 1-3 m below the zone where light was measurable (at >0.01 µmol PAR photons m^{-2} s^{-1}), *pmoA* gene expression was close to two orders of magnitude higher than baseline community *rpoB* expression levels and also exceeded expression of the *mcrA* gene associated with methanogenesis or reverse methanogenesis (Welte 2018). Similarly to observed previously (Appendix A), the δ^{13}C-POM stable isotopic value became increasingly negative with depth (Figure 3.9), and when compared to the δ^{13}C-DIC stable isotopic value, exceeded the known fractionation factor (approx. 4-13‰) for phototrophic carbon fixation via the rTCA cycle used by *Chlorobia* members (Thomazo *et al.* 2009; Tang, Tang and Blankenship 2011). Thus, it is possible that methane was being incorporated into the particulate organic carbon fraction in the lake anoxic zone, for example as biomass from microbial metabolism. Others have recently reported observations of methane oxidation by "aerobic" methanotrophs in the anoxic zones of lakes (Oswald *et al.* 2016a; van Grinsven *et al.* 2019, 2020). It is possible that a yet-undiscovered mechanism for methane oxidation, for example via EET or supply of internal molecular oxygen, exists among these bacteria, which could shift understanding of global methane cycling models.
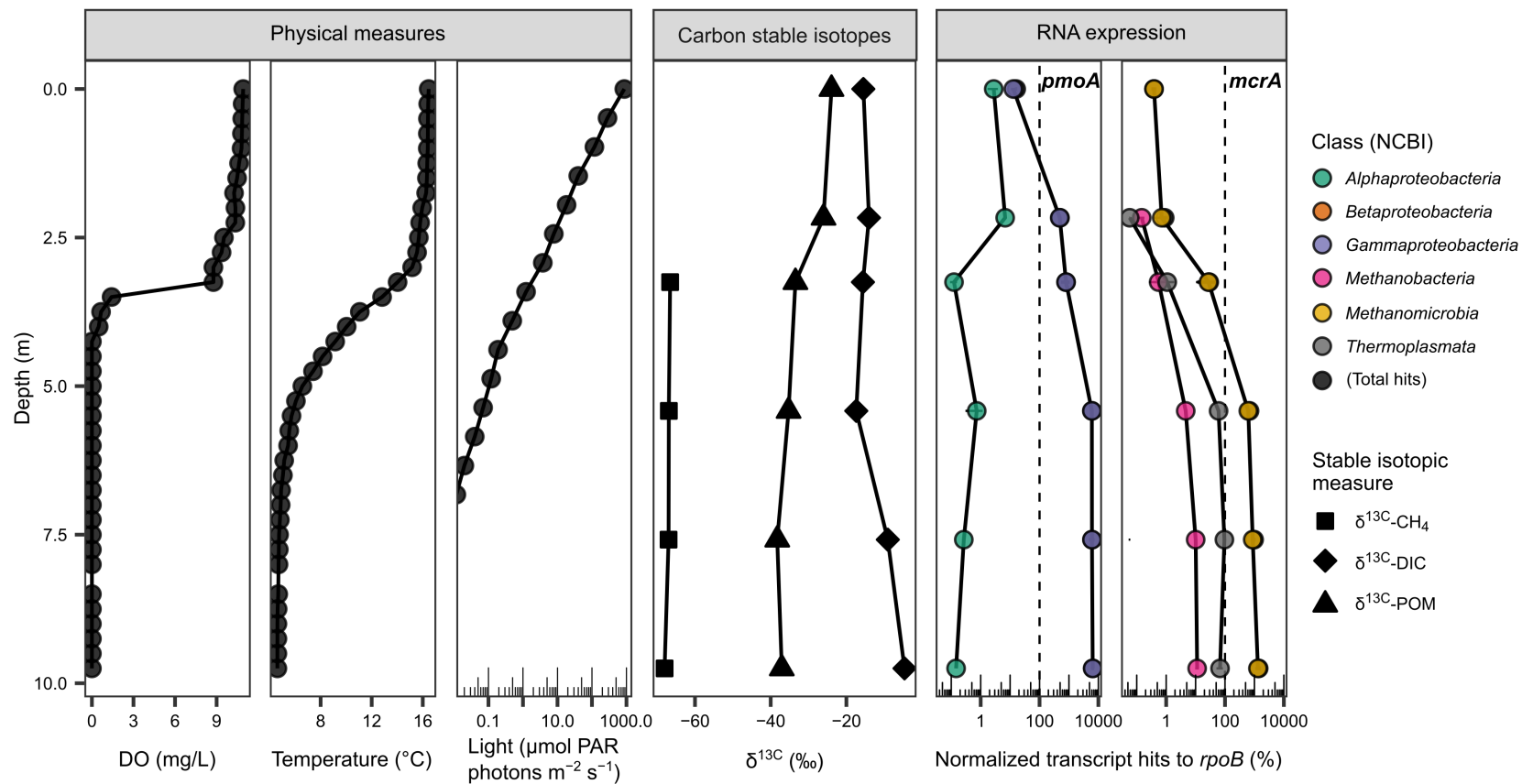
Figure 3.9 Methane cycling in Lake 227. Physical measures including dissolved oxygen, temperature, and light levels are shown across the Lake 227 water column in September 2017, along with carbon stable isotopes and RNA expression levels of key genes involved in methane cycling. Gene transcript hits are expressed as normalized relative abundances to total metatranscriptome *rpoB* expression. Error bars in RNA data show the standard deviation of hits from triplicate (or duplicate; 1 m and 8 m depth) metatranscriptomes.

3.4 Conclusion

This survey of nine Boreal Shield lakes at the IISD-ELA uncovered substantial diversity of anoxygenic and oxygenic phototroph populations in the ferruginous water columns of these systems. Selected Boreal Shield lake systems were shown to develop high relative abundance and diverse anoxic zone phototroph communities having members with potentially unique photophysiologies. Additionally, phototroph-rich lake anoxic zones were found to be distinguishable from the anoxic zones of other ferruginous lakes at the whole microbial community level. Basin shape and dissolved organic carbon levels might be key parameters governing the favourability of phototrophy in Boreal Shield lakes because these parameters influence lake anoxia timing and thus rates of sediment iron release. Levels of sulfate and dissolved iron in the upper anoxic water column might also play an important role in phototroph microbial community structure. Phototroph-rich lake systems were found to have unique properties, including the common occurrence of the *cyc2* candidate iron oxidation gene and the presence of novel RCI-encoding *Chloroflexota* members, as well as attributes potentially helpful for making inferences about early Earth ecosystems, such as cryptic sulfur cycling and the occurrence of rapid oxidation of ferrous iron, driven by oxygenic photosynthesis, in the lake anoxic zone. Future mining of survey data to explore methane cycling and the broader iron and sulfur cycles of the lakes, as well as to explore microbial novelty associated with underexplored ferruginous systems, will broaden sparse knowledge of iron biogeochemistry. The data presented in this study imply that millions of additional Boreal Shield lakes in northern regions globally contain such diverse and potentially novel phototrophic microbial communities.

3.5 Data and code availability

Metagenome data from 2016 are available in the NCBI Sequence Read Archive under accessions SRP142039-SRP142047, SRP142049, SRP142051-SRP142067, and SRP142069. Amplicon sequencing data from 2016 are available in the Joint Genome Institute Genome Portal under project IDs 1137806 and 1137809. Metagenome, metatranscriptome, and amplicon sequencing data from 2017-2018 are available under the NCBI BioProject accession PRJNA664486. Other raw data files supporting this work, including ASV tables, relative abundances and classifications of MAGs, and physico-chemical measurements, are available in a Zenodo repository at doi: 10.5281/zenodo.4038307.

Chapter 4[3]

Anoxygenic phototrophic *Chloroflexota* member uses a Type I reaction center

4.1 Introduction

The evolution of photosynthesis remains poorly understood (Olson and Blankenship 2005; Cardona 2019). Among chlorophyll-based phototrophic organisms, two distinct classes of photosynthetic reaction center, either quinone-type or Fe-S type, are used for light energy conversion by members of at least eight microbial phyla, yet the evolutionary history of these reaction centers is unclear (Sadekar, Raymond and Blankenship 2006; Hohmann-Marriott and Blankenship 2011; Thiel, Tank and Bryant 2018; Ward, Cardona and Holland-Moritz 2019). Members of a single microbial phylum, the *Cyanobacteria* (along with chloroplasts in eukaryotes), use both quinone and Fe-S type reaction centers in tandem as PSII and PSI, respectively, to generate molecular oxygen through the oxidation of water (Fischer, Hemp and Johnson 2016). Diverse anoxygenic phototrophs making up the remaining known phyla perform phototrophy using a single reaction center class, either RCII or RCI related to PSII or PSI, respectively (Thiel, Tank and Bryant 2018). Combined phylogenetic, biochemical, and physiological evidence suggests that many photosynthesis-associated genes, including reaction center genes, are ancient in origin and may have been transferred between distantly related microorganisms multiple times over evolutionary history (Bryant *et al.* 2012; Martin, Bryant and Beatty 2018). Such lateral gene transfer is presumed to be a key factor allowing for the existence of both oxygenic and anoxygenic photosynthesis (Soo, Hemp and Hugenholtz 2019). However, modern case studies of such distant lateral gene transfer events are limited, hindering understanding of the fundamental mechanisms allowing for chlorophototrophy to evolve (Zeng *et al.* 2014).

Among phototrophic microorganisms, anoxygenic phototrophs within the *Chloroflexota* phylum (formerly the *Chloroflexi* phylum(Hanada 2014; Parks *et al.* 2018); also called filamentous anoxygenic phototrophs or "green non-sulfur bacteria") have been particularly cryptic in their evolutionary origin. Although these bacteria use RCII for photosynthesis, several phototrophic *Chloroflexota* members also contain chlorosomes, a light-harvesting apparatus that is otherwise found

only in RCI-utilizing bacteria (Frigaard and Bryant 2004). How chlorosomes were adopted by both RCI-containing phototrophs and RCII-containing phototrophic *Chloroflexota* members remains highly speculative, with some suggesting interactions between RCI and RCII-utilizing bacteria in the evolutionary past (Hohmann-Marriott and Blankenship 2007; Bryant *et al.* 2012). However, concrete evidence for such interactions has never been demonstrated in nature. This research reports the cultivation of a highly novel phototroph within the *Chloroflexota* phylum that uses RCI and chlorosomes for phototrophy, provisionally classified as, "*Candidatus* Chlorohelix allophototropha strain L227-S17" (Chlo.ro.he'lix. Gr. adj. *chloros*, green; Gr. fem. n. *helix*, spiral; N.L. fem. n. *Chlorohelix*, a green spiral. al.lo.pho.to.tro'pha. Gr. adj. *allos*, other, different; Gr. n. *phos*, -*otos*, light; N.L. suff. -*tropha*, feeding; N.L. fem. adj. *allophototropha*, phototrophic in a different way). The physiological and genomic properties of the novel bacterium are described, as well as its implications for the evolution of photosynthesis.

4.2 Materials and methods

4.2.1 Enrichment cultivation of "*Ca.* Chx. allophototropha"

To culture novel anoxygenic phototrophs, Lake 227, a seasonally anoxic and ferruginous Boreal Shield lake at the IISD-ELA (near Kenora, Canada), was sampled. Lake 227 develops up to ~100 μM concentrations of dissolved ferrous iron in its anoxic water column (Appendix A), and anoxia is more pronounced than expected naturally due to the long-term experimental eutrophication of the lake (Schindler *et al.* 2008). The lake's physico-chemistry has been described in detail previously (Appendix A). Water was collected from the illuminated portion of the upper anoxic zone of Lake 227 in September 2017, at 3.88 and 5.00 m depth, and was transported anoxically and chilled in 120 mL glass serum bottles, sealed with black rubber stoppers (Geo-Microbial Technology Company; Ochelata, Oklahoma, U.S.A.), to the laboratory.

Water was supplemented with 2% v/v of a freshwater medium containing 8 mM ferrous chloride (Hegler *et al.* 2008) and was distributed anoxically into 120 mL glass serum bottles, also sealed with black rubber stoppers (Geo-Microbial Technology Company), that had a headspace of dinitrogen gas at 150 kPa final pressure. Bottles were spiked with a final concentration of 50 μM DCMU (Sigma-Aldrich; St. Louis, Missouri, U.S.A.) to block oxygenic phototrophic activity (Vandermeulen, Davis and Muscatine 1972) and to eliminate oxygenic phototrophs. Spiking was performed either at the start of the experiment or as needed based on observations of oxygenic

phototroph growth. Bottles were incubated at 10°C under white light (20-30 µmol photons m^{-2} s^{-1} at 400-700 nm wavelengths; blend of incandescent and fluorescent sources) or at 22°C under far red LED light (using a PARSource PowerPAR LED bulb; LED Grow Lights Depot; Portland, Oregon, U.S.A.). Cultures were monitored regularly for ferrous iron concentration using the ferrozine assay (Stookey 1970) and were amended with additional freshwater medium or ferrous chloride (e.g., in 0.1-1 mM increments) when ferrous iron levels dropped, presumably due to iron oxidation.

The "*Ca.* Chx. allophototropha" enrichment culture was gradually acclimatized to higher concentrations of freshwater medium through repeated feeding and subculture. Cultures were moved to room temperature (20-22°C) growth and could be grown under white light (fluorescent, halogen, or incandescent; same light intensity as above) or under far red LED light (using a PARSource PowerPAR LED bulb). A deep agar dilution series (Pfennig 1978) was used to eliminate contaminants from the culture. The same freshwater medium was used for this dilution series, amended with an acetate feeding solution (Imhoff 2014a). Ferrous chloride-containing agar plugs (10 mM concentration) were initially used at the bottom of tubes to form a ferrous iron concentration gradient to qualitatively determine best conditions for growth, with 0.125 g/L sodium thioglycolate being used as a reducing agent in the medium. Over multiple rounds of optimization, the growth medium was adjusted to contain 2 mM ferrous chloride (without sodium thioglycolate), 1.2 mM sodium acetate, and 50 mg/L resazurin, and to be a 1:1 dilution of the original freshwater medium, kept at a pH of ~7.5. In addition, cultures were generally incubated in soft agar (0.2-0.3%; w/v) to promote faster growth of "*Ca.* Chx. allophototropha" cells compared to growth in standard concentration agar (~0.8%; w/v) or in liquid medium (Gaisin *et al.* 2019).

Agar shake tubes were incubated under 740 nm LED lights for a period of time to eliminate purple phototrophic bacteria, because 740 nm light is not readily used for photosynthesis by these bacteria (Overmann and Garcia-Pichel 2013). Additionally, a higher medium pH of ~7.5-8.5 was used for a time to select against green phototrophic bacteria belonging to the *Chlorobia* class, because these bacteria are known to grow poorly in moderately basic conditions (Imhoff 2014b). Lastly, sodium molybdate was added to the medium to inhibit the activity of sulfate-reducing bacteria (Biswas *et al.* 2009).

Molybdate levels were optimized experimentally by testing against sulfate-reducing bacteria enriched from the "*Ca.* Chx. allophototropha" culture. Inoculum from the "*Ca.* Chx. allophototropha" culture was injected into Hungate tubes containing 10 mL of Sulfate Reducing Medium (M803;

HiMedia Laboratories; Mumbai, Maharashtra, India) and a 90:10 N_2:CO_2 headspace, and tubes were incubated for seven days at room temperature in the dark. The enrichment was then subcultured into additional Hungate tubes containing Sulfate Reducing Medium that were spiked with sodium molybdate. A molybdate concentration range of 0.2-200 mM was used, corresponding to a molybdate:sulfate ratio ranging from 1:100 to 10:1, given the 20 mM sulfate concentration in the medium. Tubes were incubated at room temperature in the dark for 14 days to monitor sulfate reducing bacterial activity, which was indicated by the development of a black iron-sulfide precipitate. Ultimately, a final concentration of 100 μM molybdate, in a 1:10 concentration ratio relative to sulfate, was used in "*Ca.* Chx. allophototropha" cultures.

4.2.2 Physiological characterization

Filaments of "*Ca.* Chx. allophototropha" cells were identified based on pigment autofluorescence using epifluorescence microscopy. An Eclipse 600 light microscope equipped for fluorescence microscopy (Nikon; Shinagawa, Tokyo, Japan) was used with a xenon light source. Autofluorescence in the wavelength range expected for bacteriochlorphyll *c* was detected using an excitation filter of 445(±45) nm, dichroic mirror of 520 nm, and emission filter of 715 (LP) nm (all from Semrock; Rochester, New York, U.S.A.), before light collection using an ORCA-Flash4.0 monochromatic camera (Hamamatsu Photonics; Hamamatsu City, Shizuoka, Japan). Examination of cells under epifluorescence microscopy enabled the growth of cells to be qualitatively assessed. Growth of "*Ca.* Chx. allophototropha" was also confirmed based on relative increases in cell pigmentation using a custom-built pigment fluorescence detection system (Figure 4.1) that was optimized for bacteriochlorophyll *c*. This system leveraged the known excitation and emission wavelengths of bacteriochlorophyll *c* in whole cells of 457-460 nm and 775 nm, respectively (Overmann and Garcia-Pichel 2013), to non-destructively monitor bulk pigment production in culture tubes. Excitation light was shone into an otherwise dark vessel containing the sample of interest, and a near-infrared camera pointed into an opening in the vessel was used to detect emitted light. A blue PARSource PowerPAR LED bulb (LED Grow Lights Depot; Portland, Oregon, U.S.A.), having an emission range of ~380-500 nm and peak emission wavelength of ~450 nm, was used as the light source for pigment excitation. For the near-infrared camera, a C615 HD portable 1080p webcam with autofocus (Logitech; Lausanne, Switzerland) was custom-modified to remove the infrared light filter in front of the CCD. A 770 nm near-infrared bandpass filter (Salvo Technologies; Seminole, Florida, U.S.A.), made of borosilicate glass and having a full-width at half maximum of 100 nm for input light
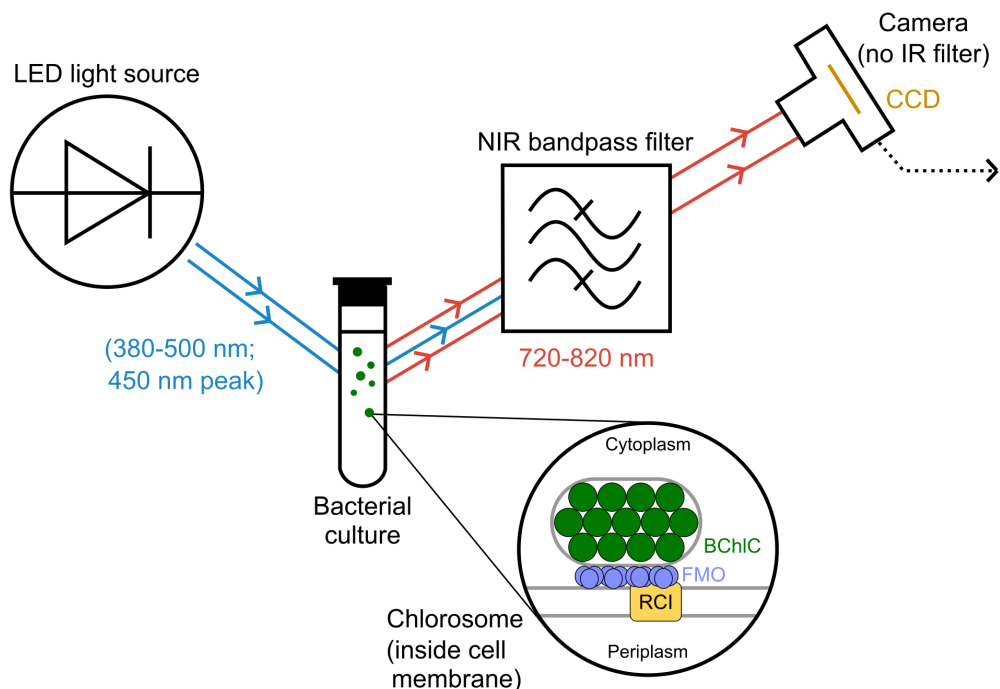
Figure 4.1 Schematic of a pigment fluorescence detection system optimized for bacteriochlorophyll *c*. Light is depicted as straight lines with arrows, and pigment-containing microbial biomass inside the tube is depicted as green dots. The fluorescently active chlorosome complex of an example Type I reaction center-containing cell is shown as an inset. An example photograph using the system is shown in Figure 4.3A.

wavelengths, was used in place of the infrared light filter and placed directly in front of the webcam to block non-target light wavelengths.

To perform TEM, cell biomass was picked from agar shake tubes, and residual agar surrounding cells was digested using agarase. One unit of β-agarase I (New England Biolabs; Ipswich, Massachusetts, U.S.A.) and 10 uL of 10x reaction buffer was added to 100 μL of cell suspension and incubated at 42°C for 1.5 h. Following cell pelleting and removal of supernatant, cells were then fixed for 2 h at 4°C in 4%/4% glutaraldehyde/paraformaldehyde (dissolved in 1x PBS) and stored at 4°C. Sample preparation and imaging was performed at the Molecular and Cellular Imaging Facility of the Advanced Analysis Center (University of Guelph, Ontario, Canada). Washed cell pellets were enrobed in 4% Noble agar (ThermoFisher Scientific; Waltham, Massachusetts, U.S.A.), and enrobed samples were cut into 1 mm cubes. The cubes were fixed in 1% (w/v) osmium tetroxide for 45 min and subjected to a 25-100% ethanol dehydration series. Samples were then subjected to stepwise infiltration using 25-100% LR White Resin (Ted Pella; Redding, California, U.S.A.),

82

transferred to gelatin capsules in 100% LR White Resin, and allowed to polymerize overnight at 60°C. An Ultracut UCT ultramicrotome (Leica Microsystems; Wetzlar, Germany) equipped with a diamond knife (DiATOME; Hatfield, Pennsylvania, U.S.A.) was used to cut 50 nm thin sections, which were then floated on 100-mesh copper grids and stained with 2% uranyl acetate and Reynold's lead citrate to enhance contrast. Sections were imaged under standard operating conditions using a Tecnai G2 F20 TEM (ThermoFisher Scientific; Waltham, Massachusetts, U.S.A.) that was running at 200 kV and was equipped with a Gatan 4k CCD camera (Gatan; Pleasanton, California, U.S.A.). For SEM, fixed cells were washed in 1x PBS and incubated in 1% (w/v) osmium tetroxide at room temperature for 30 min. Following incubation, cells were washed, deposited on an aluminum stub, dried, and then sputter coated with a gold:palladium mixture using a Desk V TSC Sample Preparation System (Denton Vacuum; Moorestown, New Jersey, U.S.A.). Prepared samples were imaged using a Quanta FEG 250 SEM (ThermoFisher Scientific) with a high-voltage setting of 10 kV and working distance of 9.9 mm.

Pigments were extracted from the "*Ca.* Chx. allophototropha" enrichment culture in acetone:methanol (7:2 v/v). Cells were picked from agar shake tubes and washed once using 10 mM Tris-hydrochloric acid (pH=8). The resulting pelleted cell material was suspended in 400 μL of a solution of 7:2 acetone:methanol and was mixed vigorously by vortex and pestle. Supernatant was evaporated until mostly dry and resuspended in 600 μL acetone. The absorbance of the extracted pigments was assessed between 350-1000 nm using a UV-1800 UV/Visible Scanning Spectrophotometer (Shimadzu; Kyoto, Japan) and a 1 mL quartz cuvette.

4.2.3 16S rRNA gene amplicon sequencing and analysis

The microbial community composition of early enrichment cultures was assessed using 16S rRNA gene amplicon sequencing with subsequent analysis of the obtained sequences. Genomic DNA was extracted from pelleted cell biomass using the DNeasy UltraClean Microbial Kit (Qiagen; Venlo, The Netherlands) according to the manufacturer protocol, with the addition of a 10 min treatment at 70°C after adding Solution SL for improved cell lysis. The V4-V5 hypervariable region of the 16S rRNA gene was then amplified from extracted DNA using the universal prokaryotic PCR primers 515F-Y (Parada, Needham and Fuhrman 2016) and 926R (Quince *et al.* 2011), using primers with attached index sequences and sequencing adaptors as described previously (Bartram *et al.* 2011; Kennedy *et al.* 2014; Cavaco *et al.* 2019). Amplification was performed in singlicate. Library pooling, cleanup, and sequencing on a MiSeq (Illumina; San Diego, California, U.S.A.) was

performed as described previously (Kennedy *et al.* 2014). Sequence data analysis was performed using QIIME2 (Bolyen *et al.* 2019) version 2019.10 via the AXIOME (Lynch *et al.* 2013) pipeline, commit 1ec1ea6 (https://github.com/neufeld/axiome3), with default analysis parameters. Briefly, paired-end reads were trimmed, merged, and denoised using DADA2 (Callahan *et al.* 2016) to generate an ASV table. Taxonomic classification of ASVs was performed using QIIME2's naïve Bayes classifier (Bokulich *et al.* 2018) trained against the SILVA SSU database (Quast *et al.* 2013; Glöckner *et al.* 2017), release 132. The classifier training file was prepared using QIIME2 version 2019.7.

4.2.4 Metagenome sequencing and isolation of high molecular weight genomic DNA

The functional gene content of early liquid enrichment cultures was assessed via shotgun metagenome sequencing. Genomic DNA was extracted as above, and library preparation and sequencing was performed at The Centre for Applied Genomics (TCAG; The Hospital for Sick Children, Toronto, Canada). The Nextera DNA Flex Library Prep Kit (Illumina) kit was used for metagenome library preparation, and libraries were sequenced using a fraction of a HiSeq 2500 (Illumina) lane with 2x125 base paired end reads to obtain 5.0-7.3 million total reads per sample.

To facilitate read cloud DNA sequencing (Chen *et al.* 2020), high molecular weight DNA was extracted from the "*Ca.* Chx. allophototropha" enrichment culture grown in agar shake tubes. High molecular weight genomic DNA was lysed and initially purified using a modified salting-out procedure (Manual CG000116, Rev A; 10x Genomics; Pleasanton, California, U.S.A.), and additional contaminants were then removed using a phenol:chloroform treatment modified from Zhou and colleagues (Zhou, Bruns and Tiedje 1996). Briefly, colonies were picked from agar, centrifuged at 1010 x g for 7 min, and washed once with PBS to remove excess agar. A lysis buffer consisting of 1.33 mL of 10 mM Tris-hydrochloric acid (pH=8), 400 mM sodium chloride, and 2 mM EDTA (pH=8) was added to the pelleted biomass. In addition, 0.53 mL of 10% SDS and 0.5 mL of 1 mg/mL Proteinase K (ThermoFisher Scientific; Waltham, Massachusetts, U.S.A.), 1% SDS, and 2 mM EDTA (pH=8) were added. Cells were then gently lysed for 18 hours at 37°C. Following lysis, 4.3 mL of 5 M sodium chloride was added, and the sample was centrifuged at 1010 x g for 5 min. The supernatant was transferred to 15 mL MaXtract High Density tubes (Qiagen; Venlo, The Netherlands), gently mixed with 1 volume Phenol:Chloroform:Isoamyl alcohol (25:24:1), and centrifuged at 1500 x g for 2 min. The resulting 11 mL of supernatant was transferred into 19.5 mL of ice-cold 100% ethanol, mixed by inversion, and aliquoted into 1.5 mL tubes for ethanol precipitation.

Tubes were centrifuged at 4°C and 6200 x g for 5 min to precipitate DNA. After removal of supernatant and drying of DNA pellets, the dry DNA pellets were serially resuspended into a single 30 μL aliquot of Tris-EDTA (10mM Tris-hydrochloric acid and 1 mM EDTA; pH=8) buffer.

Size selection was performed on the resulting DNA extract using pulsed-field gel electrophoresis and electroelution as described previously (Cheng *et al.* 2014). Briefly, electrophoresis was performed using a CHEF MAPPER Pulsed-Field Gel Electrophoresis system (Bio-Rad; Hercules, California, U.S.A.) run at 14°C, 5.5 V/cm, 1.0-6.0 s pulse, and 120° angle, for 16 hours. Electroelution of DNA from the excised gel fragment, targeting DNA strands longer than 25 kb, was performed at 120 V for 2 h in cellulose membrane dialysis tubing (12,000 MWCO with 33 mm average flat width; Sigma-Aldrich; St. Louis, Missouri, U.S.A.), with current reversed for 1 min before completion. Dialysis tubing was prepared in advance of electroelution by boiling in a 2% sodium bicarbonate and 1 mM EDTA (pH=8) solution for 10 min, then boiling in water for 10 min, and finally storing in 20% ethanol and 1 mM EDTA (pH=8) at 4°C until use. Electroeluted DNA solution was concentrated using Amicon Ultra-15 Centrifugal Filter Units (30 kDa MWCO; Millipore; Burlington, Massachusetts, U.S.A.), and DNA was washed via ethanol precipitation and resuspended in 20 μL of Tris-EDTA buffer.

Read cloud DNA sequencing library preparation was performed on the purified high molecular weight genomic DNA using the TELL-Seq WGS Library Prep Kit (Chen *et al.* 2020; Universal Sequencing Technology; Canton, Massachusetts, U.S.A.) following ultralow input protocol recommendations for small genomes. Library amplification was performed using 10 μL TELL-beads and 16 amplification cycles. The resulting library was sequenced using a MiSeq Reagent Kit v2 (300-cycle; Illumina) with 2x150 bp read length on a MiSeq (Illumina) to a depth of 19.6 million total reads.

4.2.5 Genome assembly and binning

Read cloud metagenome sequencing data for the "*Ca.* Chx. allophototropha" enrichment culture were demultiplexed and quality control was performed on index reads using the Tell-Read pipeline, version 0.9.7 (Universal Sequencing Technology), with default settings. Demultiplexed reads were then assembled using Tell-Link, version 1.0.0 (Universal Sequencing Technology), using global and local kmer lengths of 65 and 35, respectively. Metagenome data for this sample, together with metagenome data for an early enrichment of the same culture, were then partitioned into genome

bins and analyzed using the ATLAS pipeline, version 2.2.0 (Kieser *et al.* 2020). Unassembled read data demultiplexed via Picard version 2.21.6 (Picard Toolkit – Broad Institute – http://broadinstitute.github.io/picard/) were used as input into ATLAS (starting with the "qc" module). Following assembly of these data, the default scaffolds were replaced with scaffolds generated using Tell-Link, and the pipeline was then allowed to continue. The second set of unassembled short read data that was generated from DNA extracted from an early enrichment of the "*Ca.* Chx. allophototropha" culture was used to help guide genome binning of the Tell-Link scaffolds using MaxBin 2 version 2.2.4 (Wu, Simmons and Singer 2016) and MetaBAT version 2.12.1 (Kang *et al.* 2019).

The "*Ca.* Chx. allophototropha" genome bin was identified based on its taxonomic placement within the *Chloroflexota* phylum according to the GTDB-Tk version 0.3.3, which relied on GTDB release 89 (Parks *et al.* 2018; Chaumeil *et al.* 2020). This genome bin was manually curated to remove potentially incorrectly binned scaffolds. Genes predicted by Prodigal version 2.6.3 (Hyatt *et al.* 2010) were queried against the NCBI RefSeq protein database (O'Leary *et al.* 2016) using BLASTP (Altschul *et al.* 1990) (e-value cutoff = 10^{-10}), and the taxonomic lineages of gene hits was determined using the script *make-lineage-csv.py* (https://github.com/dib-lab/2018-ncbi-lineages, commit 63e8dc7). Any scaffolds with at least one hit to a subject sequence associated with the *Chloroflexota* phylum were kept, and the top five highest-scoring hits for each gene were considered. Remaining scaffolds were checked for their median coverage, length, and gene classifications, and genes were further assessed against the NCBI nr protein database via BLASTP. Short contigs containing no genes or genes consistently matching the same non-*Chloroflexota* phylum were discarded from the bin. Following curation, the genome bin was annotated using Prokka version 1.14.6 (Seemann 2014), and selected portions of the genome were checked for functional genes using the BlastKOALA webserver version 2.2 (Kanehisa, Sato and Morishima 2016).

Short read metagenome sequencing data generated from a second early enrichment culture was also analyzed. This second enrichment was lost early in the cultivation process but contained another member of the "*Ca.* Chloroheliales" (see results and discussion). The single enrichment culture metagenome was processed using ATLAS version 2.2.0 (Kieser *et al.* 2020) for read quality control, assembly and genome binning. A single genome bin was classified using the GTDB-Tk (Chaumeil *et al.* 2020) to the *Chloroflexota* phylum. To manually curate this genome bin, which was highly fragmented, all predicted protein sequences in the bin were taxonomically classified using the Kaiju

webserver (Menzel, Ng and Krogh 2016) with default settings. Any contigs having at least one hit classified within the *Chloroflexota* phylum were retained. Protein-coding sequences on other contigs were further screened using BLASTP (Altschul *et al.* 1990) against the NCBI RefSeq database (O'Leary *et al.* 2016) using an e-value cutoff of 10^{-10}. Contigs having median coverage values greater than or less than one standard deviation of the mean contig coverage value across the genome bin were assessed gene-by-gene for signs of mis-binning. Other contigs having a BLASTP hit with >80% identity to a RefSeq entry or containing a tRNA or rRNA gene (which was subsequently screened using BLASTN against the NCBI nr database, excluding uncultured subjects) were also examined. Following curation, the genome bin was annotated using Prokka (Seemann 2014) as above. All ATLAS settings for both runs are available in the code repository associated with this work.

To assess the community composition of the enrichment cultures, the relative abundances of each recovered genome bin from ATLAS were assessed by calculating the number of reads mapped to each bin divided by the total reads mapped to assembled scaffolds. To cross-validate these results, short protein fragments were predicted directly from unassembled read data using FragGeneScanPlusPlus (Singh *et al.* 2018) commit 9a203d8. These protein fragments were searched using a HMM for the single-copy housekeeping gene *rpoB*, and hits were assigned taxonomy using MetAnnotate (Petrenko *et al.* 2015) version 0.92 with default settings. The HMM for *rpoB* (2009 release) was downloaded from FunGene (Fish *et al.* 2013), and taxonomic classification was performed using the RefSeq database, release 80 (March 2017; O'Leary *et al.* 2016). Output of MetAnnotate was analyzed using the metannoviz R library, version 1.0.2 (https://github.com/metannotate/metannoviz; doi:10.5281/zenodo.3963673) with an e-value cutoff of 10^{-10}.

4.2.6 Identification of RCI-associated genes

To detect remote homologs of RCI-associated genes in the two "*Ca.* Chloroheliales"-associated genome bins, hmmsearch (Eddy 2011) version 3.1b2 was used to search predicted proteins from the bins using HMMs downloaded from Pfam (Finn *et al.* 2016). In particular, genes encoding the core Type I reaction center (*pscA*/*pshA*/*psaAB*; PF00223), a Type I reaction center-associated protein (*pscD*; PF10657), chlorosomes structural units (*csmAC*; PF02043, PF11098), and a chlorosome energy transfer protein (*fmoA*; PF02327), were queried. Detected homologs in the genome bins were assessed for their alignment to known RCI-associated genes encoded by phototrophic microorganisms belonging to other phyla. The tertiary structures of the detected homologs were also predicted using I-

TASSER (Yang *et al.* 2015), and phylogenetic placement was assessed (see below). Custom HMMs were also built for selected homologs. Primary sequences predicted for each gene of interest from the two genomes were aligned using Clustal Omega (Sievers *et al.* 2011) version 1.2.3, and HMMs were generated using hmmbuild (Eddy 2011) version 3.1b2. Custom HMMs and homology models generated by I-TASSER are available in the code repository associated with this work.

4.2.7 Assessment of genomic potential for photosynthesis within the *Chloroflexota* phylum

Representative genomes for the phylum *Chloroflexota* were downloaded from NCBI based on information in the GTDB (Parks *et al.* 2018), release 89. All genomes that represented a type species of *Chloroflexota* according to NCBI and the GTDB were downloaded, as well as genomes representing members of known phototrophic clades, namely the *Chloroflexaceae* family (Hanada 2014), "*Ca.* Thermofonsia" order (Ward *et al.* 2018), and "*Ca.* Roseiliniales" order (Tank *et al.* 2017). Selected genomes of known phototrophs that were deposited in other genome databases were also downloaded (see details in the code repository associated with this work). In addition, any genome bins listed in the GTDB that belonged to the "54-19" order (which represented the name of the "*Ca.* Chloroheliales" order in this database) were downloaded. Non-phototrophic lineages were subsequently pruned to one representative per genus, with the exception of non-phototrophic lineages closely related to phototrophic clades. In total, this left 58 genomes, including genomes of 28 known phototrophs. A concatenated core protein phylogeny was constructed for this genome collection using GToTree (Lee 2019) version 1.4.11. The 'Bacteria.hmm' collection of 74 single-copy marker genes was used, and a minimum of 30% of the marker genes were required in each genome for the genome to be kept in the final phylogeny. IQ-TREE (Nguyen *et al.* 2015) version 1.6.9 was used to build the phylogeny from the masked and concatenated multiple sequence alignment. ModelFinder (Kalyaanamoorthy *et al.* 2017) was used to determine the optimal evolutionary model for phylogeny building. Branch supports were calculated using 1000 ultrafast bootstraps (Hoang *et al.* 2018).

A collection of photosynthesis-associated genes, including genes for reaction centers, antenna proteins, chlorosome structure and attachment, bacteriochlorophyll synthesis, and carbon fixation, was selected based on the genome analyses of Tang and colleagues (Tang *et al.* 2011) and Bryant and colleagues (Bryant *et al.* 2012). Reference sequences for these genes were selected from genomes of well-studied representatives of the phylum, *Chloroflexus aurantiacus* (Pierson and Castenholz 1974), *Oscillochloris trichoides* (Keppen, Baulina and Kondratieva 1994), and *Roseiflexus castenholzii*

(Hanada *et al.* 2002). Bidirectional BLASTP (Altschul *et al.* 1990) was performed on these reference sequences against the entire *Chloroflexota* genome collection (above) to detect potential orthologs. The BackBLAST pipeline (Bergstrand *et al.* 2016), version 2.0.0-alpha3 (http://github.com/leeBergstrand/BackBLAST_Reciprocal_BLAST; doi:10.5281/zenodo.3697265), was used for bidirectional BLASTP, and cutoffs for the *e* value, percent identity, and query coverage of hits were empirically optimized to 10^{-3}, 20%, and 50%, respectively. Genes involved in iron oxidation or reduction were also searched using FeGenie (Garber *et al.* 2020), commit 30174bb, with default settings.

4.2.8 Phylogenetic assessment of photosynthesis-associated genes

Selected photosynthesis-associated genes were tested against genes from other phyla to examine the phylogenetic relationship of the novel "*Ca.* Chloroheliales" sequences to those of other phototrophic organisms. Genes for RCI/PSI (*pscA*/*pshA*/*psaAB*; Cardona 2018), bacteriochlorophyll *a* binding (*fmoA*; Olson 2005), chlorosome structure (*csmA*; Pedersen *et al.* 2010), (bacterio)chlorophyll synthesis (*bchIDH*/*chlIDH* and *bchLNB*/*chlLNB*/*bchXYZ*; Bryant *et al.* 2012), and RuBisCo (*cbbL*; Tabita *et al.* 2008), were examined. Genomes potentially containing these genes of interest were determined using a combination of automated detection via AnnoTree (Mendler *et al.* 2019) and descriptions in the literature. Annotree was also used to summarize the GTDB reference tree for downstream data visualization. Representative genomes from this initial genome set were selected from the GTDB, based on genome quality and taxonomic novelty, before being downloaded from NCBI. Potential orthologs of the genes of interest were identified in the downloaded genomes using bidirectional BLASTP, which was performed using the primary sequences of known reference genes as queries (see the code repository associated with this work) via BackBLAST (Bergstrand *et al.* 2016) version 2.0.0-alpha3. In addition, for some sequence sets (i.e., Type I reaction centers; Bch proteins; CbbL), additional reference sequences were manually added based on literature references (Tabita *et al.* 2008; Bryant *et al.* 2012; Cardona 2018). Sequence sets were then pruned to remove closely related sequences. Identified primary sequences were aligned using Clustal Omega (Sievers *et al.* 2011) version 1.2.3, and the sequence sets were further verified by manually inspecting the sequence alignments for unusual variants. Alignments were then masked from non-informative regions using Gblocks (Talavera and Castresana 2007) version 0.91b with relaxed settings (-t=p -b3=40 -b4=4 -b5=h) to preserve regions with remote homology (see results in Table 4.1). Maximum likelihood protein phylogenies were built from masked sequence alignments using IQ-TREE

Table 4.1 Properties of sequence alignments and phylogenetic trees presented in this study. The unmasked alignment length of the core protein set was not available due to the constraints of the software pipeline used to generate the final masked alignment.

Protein set	Alignment length (unmasked)	Alignment length (masked)	Evolutionary rate model used
Core proteins (*Chloroflexota* phylum)	N/A	11 613	LG+F+R6
Type I reaction center (PsaA/PsaB/PscA/PshA)	1292	548	LG+F+G4
FmoA	399	356	LG+G4
CbbL	644	412	LG+I+G4
CsmA	89	74	LG+F+G4
BchIDH/ChlIDH	3156	1702	LG+F+I+G4
BchLNB/ChlLNB/BchXYZ	2010	819	LG+F+I+G4

(Nguyen *et al.* 2015) version 1.6.9. Evolutionary rate model selection was performed using ModelFinder (Kalyaanamoorthy *et al.* 2017), and 1000 rapid bootstraps were used to calculate branch support values (Hoang *et al.* 2018). Rate models used are summarized in Table 4.1.

4.2.9 Searching for additional RCI and RCI-associated gene homologues

Preparation of the genome collection of *Chloroflexota* revealed that two existing genomes bins in the NCBI database placed taxonomically within the "*Ca.* Choroheliales" order. Unassembled read files for the environmental metagenomes associated with those bins were downloaded from the European Nucleotide Archive to identify any potentially unbinned but novel photosynthesis-associated genes. Short protein sequences were predicted directly from unassembled read data using FragGeneScanPlusPlus (Singh *et al.* 2018) commit 9a203d8. These short protein sequences were scanned using the two custom HMMs developed in this study (above) via hmmsearch (Eddy 2011) v3.1b2 using a relaxed e-value cutoff of 10^{-1}. The average coverage and read recruitment of each genome bin was also calculated by mapping the unassembled metagenomic reads onto the bins using bbmap.sh version 38.75 (BBMap – Bushnell B. – sourceforge.net/projects/bbmap/).

4.3 Results and discussion

Water was collected from the illuminated and anoxic water column of Lake 227, an iron-rich Boreal Shield lake, to enrich for resident bacterial phototrophs. After approximately two months of

incubation under light, enrichment cultures began showing ferrous iron oxidation activity, and 16S rRNA gene amplicon sequencing from some cultures indicated the presence of novel bacterial populations distantly related to known *Chloroflexota* members (see the code repository associated with this work). After several rounds of amendment with freshwater medium and ferrous chloride, the enrichments were subcultured in liquid medium in dilution series in an attempt to remove contaminating bacteria. However, filamentous phototrophs eventually ceased growth or were overtaken by purple phototrophic bacteria in these liquid culture bottles. Subsequent transfer of enrichment cultures to deep agar dilution series allowed for more vigorous growth of filamentous phototrophs. An early liquid enrichment culture, which was derived from water sampled at 3.88 m depth but had been stored at 4°C in the dark for five months rather than subjected to substantial liquid dilution series, allowed for best growth of filamentous phototrophs when used as deep agar dilution series inoculum. Further enrichment of this culture in deep agar dilution series eliminated a purple phototrophic bacterium related to the metabolically versatile *Rhodopseudomonas palustris* (Jiao *et al.* 2005) and a green phototrophic bacterium belonging to the *Chlorobium* genus. Given the subsequent detection of genes for RCI in the *Chloroflexota* member's genome, the bacterium was provisionally classified as, "*Candidatus* Chlorohelix allophototropha strain L227-S17". The specific steps performed during enrichment cultivation for the final culture of "*Ca.* Chx. allophototropha" are described in Appendix B.

Metagenome sequencing and assembly recovered the nearly closed genome of "*Ca.* Chx. allophototropha", which represented the only detectable phototroph and comprised 43.4% of raw metagenome sequences (Figure 4.2). The curated genome bin for "*Ca.* Chx. allophototropha" was 6.01 Mb in size, with an average GC content of 46.8%, and encoded a total of 5080 genes, including 49 tRNA genes and three complete and identical 16S ribosomal RNA (rRNA) and 23S rRNA genes. The genome was comprised of 36 scaffolds, with 98% of bases contained in three scaffolds greater than 50 kb in length. According to CheckM, the genome was 98.2% complete and 1.4% contaminated, indicating that the majority of the genome was captured during read cloud sequence assembly and binning. Highly complete genome bins were also recovered for the two major heterotroph populations remaining in the culture. These bins were classified within the *Geothrix* genus (which includes the iron-reducing bacterium *Geothrix fermentans*; Coates *et al.* 1999) and the *Lentimicrobiaceae* family and represented 38.3% and 16.6% of mapped assembled reads, respectively (Figure 4.2). Neither of these genome bins encoded genes for photosynthetic reaction centers. A homolog of the *omc* gene cluster (Leang, Coppi and Lovley 2003) implicated in iron reduction was
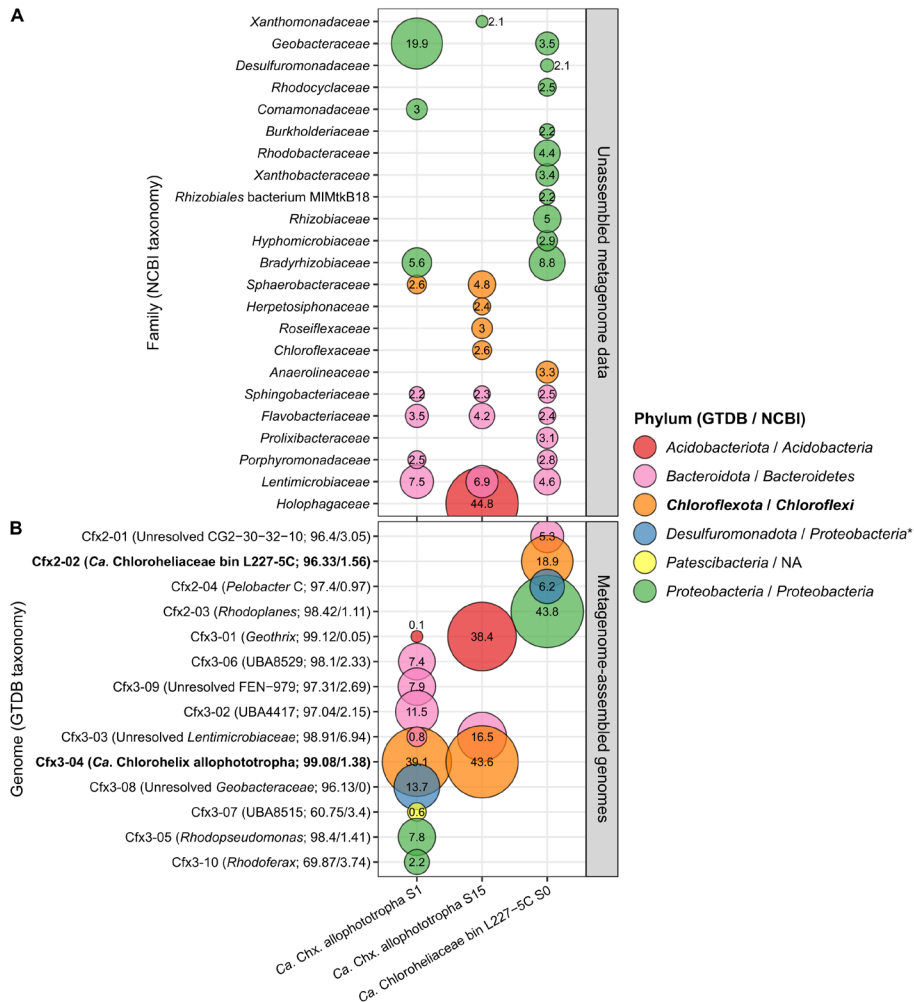
Figure 4.2 Composition of the "*Ca.* Choroheliales" enrichment cultures described in this study.
A, Bubble plot showing the relative abundances of taxa within "*Ca.* Chloroheliales" enrichment
culture metagenomes based on classification of partial RpoB fragments recovered from unassembled
reads. Bubbles are coloured based on phylum according to NCBI taxonomy, because reads were
classified against the RefSeq database. Microbial families with greater than 2% relative abundance
are shown. Two metagenomes were derived from subcultures #1 and 15 of the "*Ca.* Chx.
allophototropha' culture, and the third metagenome was derived from the initial enrichment of "*Ca.*
Chloroheliales bin L227-5C". **B**, Bubble plot showing the relative abundances of uncurated genome
bins recovered from the same metagenomes. Relative abundances are expressed as the percentage of
reads mapped to the genome bin compared to the total number of assembled reads. Bubbles are
coloured by phylum according to GTDB taxonomy and are shown for all entries greater than 0.05%
relative abundance. Listed beside each genome bin name are the GTDB genus that the bin was
classified to and the percent completeness and contamination of the bin according to CheckM.
Statistics shown for the "*Ca.* Chx. allophototropha" bin differ slightly compared to those mentioned
in the methods, which describes the statistics for the bin after manual bin curation.

detected in the genome bin classified to the *Geothrix* genus, although no close homologs to known genes involved in iron oxidation or reduction were detected in the "*Ca.* Chx. allophototropha" genome. Beyond these genome bins, the "*Ca.* Chx. allophototropha" culture was known to contain a low relative abundance sulfate-reducing bacterium that could be enriched in a sulfate-reducing bacterial medium (Methods).

Aside from "*Ca.* Chx. allophototropha", a second species of filamentous phototroph related to the novel strain was also enriched. A second enrichment culture was grown using water collected from 5.00 m depth from Lake 227, and this culture was found, based on 16S rRNA gene amplicon sequencing data, to contain a bacterium with 89.9% 16S rRNA gene sequence identity to the "*Ca.* Chx. allophototropha" ASV. This bacterium was provisionally classified as "*Ca.* Choroheliales sp. L227-5C". Subculture of the enrichment culture was attempted in deep agar dilution series, but the filamentous anoxygenic phototroph species was overtaken by purple phototrophic bacteria and failed to grow. Enrichment cultures were also set up for a neighbouring Boreal Shield lake, Lake 221, and the same two novel *Chloroflexota*-associated ASVs were detected in early enrichments based on 16S rRNA gene sequencing data. However, the organisms from this lake could not be successfully cultivated in the laboratory.

Despite not obtaining a long-term enrichment of "*Ca.* Choroheliales sp. L227-5C", genomic DNA was extracted from the initial enrichment culture of this bacterium to allow for metagenome sequencing and genome bin recovery. The recovered genome bin of "*Ca.* Choroheliales sp. L227-5C" was highly complete but fragmented. It had a completeness and contamination of 96.3% and 1.6%, respectively, as estimated by CheckM, but was composed of 546 contigs. The genome size was similar to that of "*Ca.* Chx. allophototropha", having a length of 6.31 Mb and an average GC content of 47.7 %. The genome encoded 4794 genes, including 79 tRNA genes and a full-length 16S and 23S rRNA gene, and represented 17.5% of the total mapped reads in the enrichment culture metagenome. Three non-*Chloroflexota* genome bins were also recovered from the metagenome, including a RCII-encoding phototroph, classified to the *Rhodoplanes* genus, that represented 43.8% of the mapped assembled reads (Figure 4.2) and likely corresponded to the purple phototrophic bacterium observed in the culture. Genomes bins classified to the *Pelobacter* genus and CG2-30-32-10 family (in the *Bacteroidales* order), that represented 6.2% and 5.3% of mapped assembled reads, respectively, were also recovered (Figure 4.2). Recovering this second genome related to "*Ca.* Chx. allophototropha" provided added robustness to subsequent genomic analyses of the novel strain.

Physiologically, "*Ca.* Chx. allophototropha" resembled other cultured phototrophs belonging to the *Chloroflexota* phylum (Figure 4.3). Cells grew well in soft agar (Figure 4.3A) and developed as long, spiralling filaments (Figure 4.3B), similar to those observed for *Chloronema giganteum* (Gich, Garcia-Gil and Overmann 2001), that were composed of cells with dimension of ~2-3 μm x 0.6 μm (Figure 4.3C-D). Although gliding motility of "*Ca.* Chx. allophototropha" cells was not observed, cells readily formed tight clumps (Figure 4.3D), like *Chloroflexus aggregans* (Hanada *et al.* 1995), that, when grown in liquid medium, would bind to glass surfaces or to media precipitates (not shown). In addition to the presence of light-harvesting chlorosomes (Figure 4.3E-F), "*Ca.* Chx. allophototropha" cells were pigmented, with the dominant pigment being bacteriochlorophyll *c* based on the 433 and 663 nm peaks (Figure 4.3G) observed in absorption spectrum data from acetone:methanol pigment extracts (Overmann and Garcia-Pichel 2013).

At the time when pigments were extracted, green phototrophic bacteria belonging to the *Chlorobia* class had not yet been fully eliminated from the "*Ca.* Chx. allophototropha" enrichment culture. To verify that the pigment absorption spectrum primarily reflected the pigments of "*Ca.* Chx. allophototropha" rather than pigments of these green phototrophic bacteria, leftover biomass picked from the same agar shake tube was examined using epifluorescence microscopy optimized for bacteriochlorophyll *c.* The autofluorescence of small, single cells (presumably *Chlorobia* members) was at least an order of magnitude lower than the autofluorescence of "*Ca.* Chx. allophototropha" cells. In addition, re-extraction of pigments from thin filaments of "*Ca.* Chx. allophototropha", which were carefully separated from other culture biomass under a dissecting microscope, showed the same peaks in the pigment absorption spectrum, although with higher background noise due to the lower cell concentration (Figure 4.4). Usage of bacteriochlorophyll *c* by "*Ca.* Chx. allophototropha" is also supported by epifluorescence microscopy data (Figure 4.3B-C) and imaging using a custom pigment fluorescence detection system (Figure 4.3A).

Despite physiological similarities to cultured *Chloroflexota* members, genome sequencing data revealed that "*Ca.* Chx. allophototropha" encoded a remote homolog of a *pscA*-like RCI gene (Figure 4.5) rather than the expected *pufL*/*pufM* RCII genes. A *pscA*-like gene was also detected in the genome bin of "*Ca.* Chloroheliales bin L227-5C". This detection of RCI genes in "*Ca.* Chx. allophototropha" makes the *Chloroflexota* phylum unique among all known phototroph-containing microbial phyla, which otherwise contain representatives that use strictly RCI or RCII for phototrophy (Figure 4.5A). Based on a maximum-likelihood amino acid sequence phylogeny

94
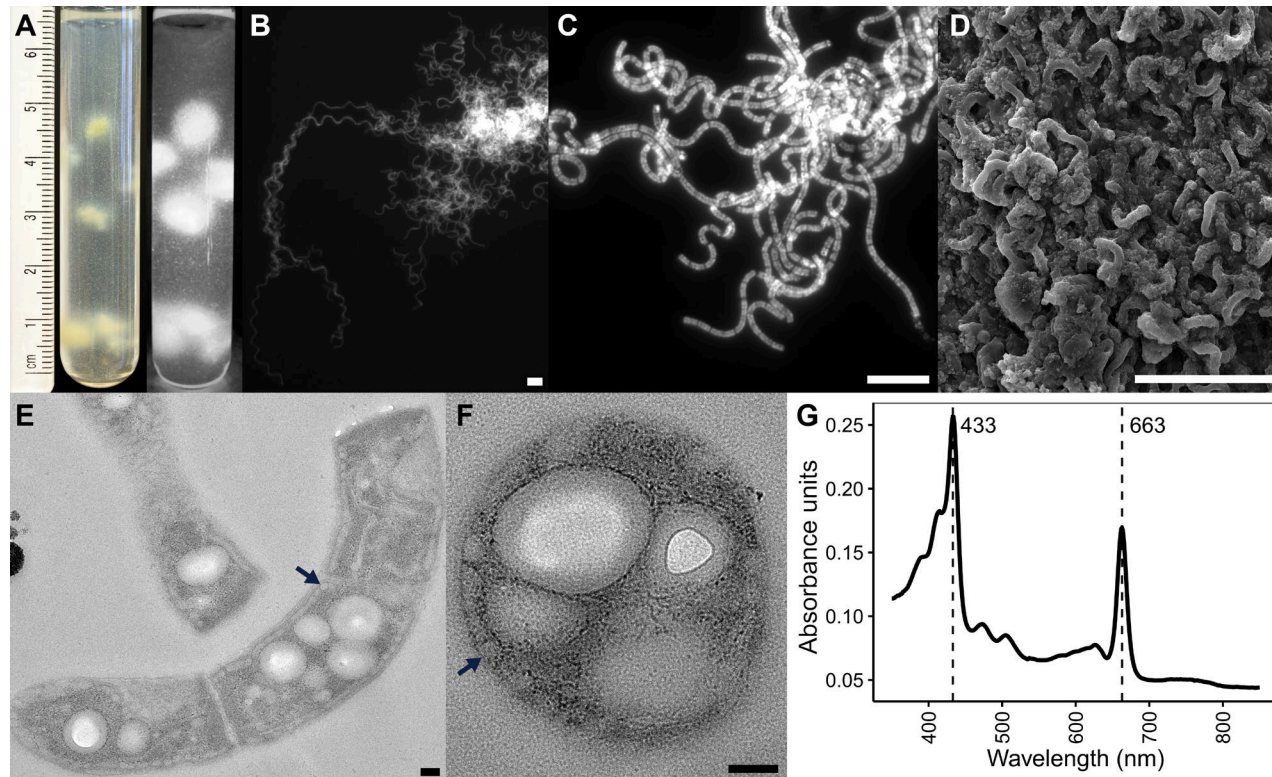
Figure 4.3 Physiology of the "*Ca.* Chx. allophototropha" culture. A, Example growth of the "*Ca.* Chx. allophototropha" enrichment in soft agar. The left panel is a photograph of a test tube with a ruler shown for scale, and the right panel shows the same tube when viewed under a pigment fluorescence detection system optimized for bacteriochlorophyll *c* (Figure 4.1) **B-C**, Epifluorescence microscopy images of "*Ca.* Chx. allophototropha" cells. Cell autofluorescence is shown, with light excitation and detection wavelengths optimized for bacteriochlorophyll *c*. **D,** Scanning electron microscopy image of an aggregate of "*Ca.* Chx. allophototropha" cells. **E-F**, Transmission electron microscopy images showing longitudinal and cross sections, respectively, of "*Ca.* Chx. allophototropha" cells. Example chlorosomes are marked with arrows. **G,** Absorption spectrum (350-850 nm) of an acetone:methanol extract of pigments from the "*Ca.* Chx. allophototropha" culture. Major absorption peaks are marked by dashed lines. Scale bars on the top row of images (B-D) represent 10 μm, and scale bars on the bottom row of images (E-F) represent 0.1 μm.
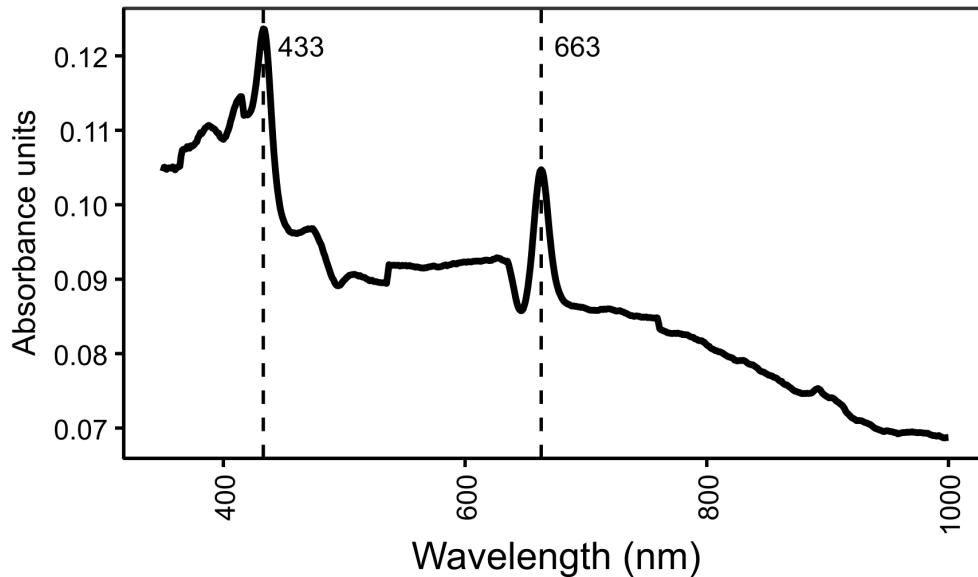
Figure 4.4 Absorption spectrum of pigments extracted from carefully separated individual filaments of "*Ca.* Chx. allophototropha" enrichment culture. The absorption spectrum of acetone:methanol pigment extracts from whole aggregates of mixed cells from the culture (i.e., containing '*Ca.* Chx. allophototropha' cells along with cells of heterotrophs and green phototrophic bacteria that had not been fully eliminated from the culture) is shown in Figure 4.3.

(Figure 4.5B; Figure 4.6), the PscA-like predicted protein of "*Ca.* Chx. allophototropha" represents a clearly distinct fourth class of RCI protein that places between clades for PscA-like proteins of *Chloracidobacterales* members and PshA proteins of *Heliobacterales* members. Homology modelling indicated that the "*Ca.* Chx. allophototropha" PscA-like protein contains the expected six N-terminal transmembrane helices for coordinating antenna pigments and five C-terminal transmembrane helices involved in core electron transfer (Gisriel *et al.* 2017; Figure 4.5C). In both recovered "*Ca.* Chloroheliales" genome bins, the novel *pscA*-like gene occurred on long scaffolds (3.29 Mb and 0.77 Mb for "*Ca.* Chx. allophototropha" and "*Ca.* Chloroheliaceae bin L227-5C", respectively) and was flanked by other genes that had high-ranking BLASTP hits to members of the *Chloroflexota* phylum (see the code repository associated with this work). The "*Ca.* Chx. allophototropha" genome did not contain the *pscBCD* photosystem accessory genes found among *Chlorobia* members (Bryant *et al.* 2012), nor a cyanobacterial *psaB*-like paralog, implying that "*Ca.* Chx. allophototropha" uses a homodimeric RCI complex (Fischer, Hemp and Johnson 2016) like other known anoxygenic phototrophs.
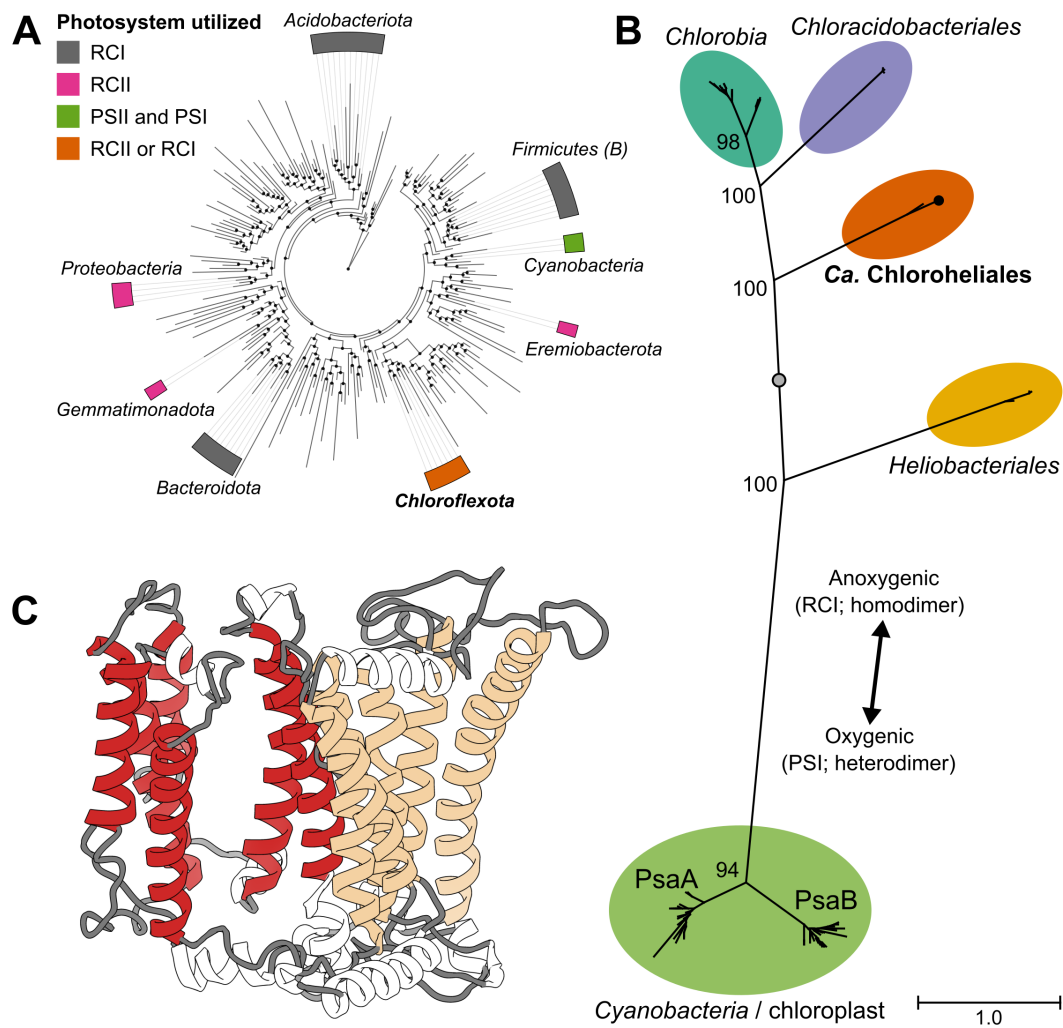
Figure 4.5 Functional novelty of the "*Ca.* Chlorohelix allophototropha" photosystem. A, Overview of the eight bacterial phyla containing known chlorophototrophs. A class-level summary of the GTDB bacterial reference tree (release 89) is shown, with chlorophototrophic phyla highlighted based on the photosynthetic reaction center utilized. **B**, Maximum likelihood phylogeny of Fe-S type reaction center primary sequences. Ultrafast bootstrap values for major branch points are shown out of 100 based on 1000 replicates, and the scale bar shows the expected proportion of amino acid change across the 548-residue masked sequence alignment (Table 4.1). Chlorophototrophic lineages are summarized by order name. The placement of the PscA-like sequence of "*Ca.* Chx. allophototropha" is indicated by a black dot. All clades above they grey dot utilize chlorosomes. **C**, Predicted tertiary structure of the novel "*Ca.* Chlorohelix allophototropha" PscA-like primary sequence based on homology modelling. The six N-terminal and five C-terminal transmembrane helices expected for RCI are coloured in red and tan, respectively.
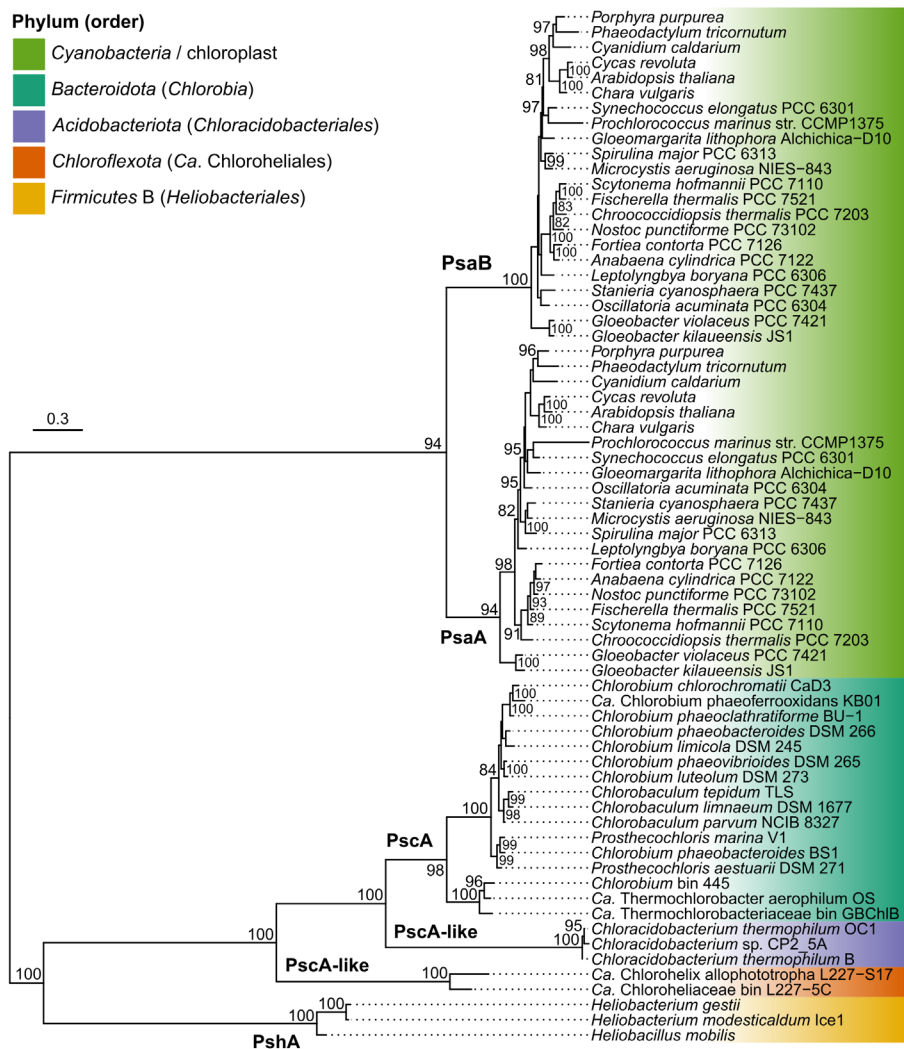
Figure 4.6 Maximum likelihood phylogeny of oxygenic and anoxygenic Type I reaction center predicted protein sequences. A simplified depiction of the same phylogeny is shown in Figure 4.5B. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 548-residue masked sequence alignment (Table 4.1).

Figure 4.7 Genomic potential for phototrophy among members of the *Chloroflexota* phylum. A maximum likelihood phylogeny of key members of the phylum is shown on the left based on a set of 74 concatenated core bacterial proteins. Ultrafast bootstrap values are shown out of 100 based on 1000 replicates, and the scale bar shows the expected proportion of amino acid change across the 11 613-residue masked sequence alignment (Table 4.1). The "*Ca.* Chloroheliales" clade is highlighted in orange. On the right, a heatmap shows the presence/absence of genes involved in photosynthesis or related processes based on bidirectional BLASTP. The intensity of the fill colour of each tile corresponds to the percent identity of the BLASTP hit. Query sequences were derived from the genomes of "*Ca.* Chlorohelix allophototropha", *Chloroflexus aurantiacus*, *Roseiflexus castenholzii*, or *Oscillochloris trichoides*, and the query organism used for each gene is marked by bolded outlines on the heatmap. GenBank accessions for sequences associated with the novel "*Ca.* Chloroheliales" members enriched in this work are summarized in Table 4.2, and raw results of bidirectional BLASTP are summarized in the code repository associated with this work. Orders containing potential phototrophs are labelled on the right of the heatmap. All genomes in the phylogeny encoded at least 50% of the core 74 protein set except for *Chloroflexi* RRmetagenome_bin16 and "*Ca.* Thermofonsia bin CP2-42A", which encoded 25 and 34 of the proteins, respectively.

99

Table 4.2. Genes potentially involved in phototrophy or carbon/nitrogen fixation among genomes of 'Ca. Chloroheliales' members recovered in this study. GenBank accession numbers for the predicted primary sequence of each gene are shown. Results correspond to those shown in Figure 4.7 except that homologs associated with the incomplete 3-hydroxypropionate bicycle are omitted for clarity. Raw results of bidirectional BLASTP are available in the code repository associated with this work.

Protein set	Gene	'Ca. Chx allophototropha'	'Ca. Chloroheliaceae bin L227-5C'
Type I reaction center-associated	fmoA	NWJ48165.1	NWJ94304.1
	pscA	NWJ45109.1	NWJ94303.1
Chlorosome structure and assembly	csmA	NWJ45202.1	NWJ94716.1
	csmM	NWJ45728.1	N/A
	csmY	NWJ47824.1	NWJ94536.1
(Bacterio)chlorophyll synthesis	bchI	NWJ45782.1	NWJ94300.1
	bchD	NWJ47387.1	NWJ94298.1
	bchH	NWJ47295.1	NWJ97069.1
	bchM	NWJ47294.1	NWJ95706.1
	bchJ	NWJ45810.1	NWJ97609.1
	bchE	NWJ45809.1	NWJ97608.1
	acsF	NWJ48920.1	N/A
	bchN	NWJ46338.1	NWJ94750.1
	bchB	NWJ46337.1	NWJ94751.1
	bchL	NWJ46336.1	NWJ94752.1
	bchF	NWJ46647.1	NWJ97746.1
	bchX	NWJ47811.1	NWJ94280.1
	bchY	NWJ47812.1	NWJ94279.1
	bchZ	NWJ47813.1	NWJ94278.1
	bchC	NWJ44534.1	NWJ98808.1
	bchG	NWJ46528.1	NWJ94994.1
	bchP	NWJ45618.1	NWJ97704.1
	bchU	NWJ44614.1	NWJ96331.1
	bchK	NWJ44593.1	NWJ96281.1
CBB (Calvin) cycle	cbbP	NWJ44408.1	NWJ96365.1
	cbbL	NWJ44404.1	NWJ96371.1
Nitrogen fixation	nifH	NWJ48757.1	NWJ96804.1
	nifD	NWJ48753.1	NWJ96800.1

Based on the GTDB classification system (Parks *et al.* 2018; Chaumeil *et al.* 2020) used throughout this work, "*Ca.* Chx. allophototropha" represents the first cultivated member of a novel order within the *Chloroflexota* phylum (Figure 4.7), which is provisionally named the "*Ca.* Choroheliales". This order placed immediately basal to the *Chloroflexales* order (Figure 4.7), which contains the canonical RCII-utilizing phototroph clade in the phylum that includes the *Chloroflexaceae*, *Roseiflexaceae*, and *Oscillochloridaceae* families (Hanada 2014). The *Chloroflexales* order also contains a basal non-phototrophic family, the *Herpetosiphonaceae*, which was formerly placed in its own order prior to reassignment in the GTDB (Parks *et al.* 2018). Searching the GTDB revealed two other genome bins that placed within the "*Ca.* Choroheliales" order (previously classified as the uncultured "54-19" order) that were assembled and binned in previous studies. One genome bin, "*Chloroflexi* bin 54-19", was recovered from an ammonium sulfate bioreactor metagenome (Kantor *et al.* 2015), and the other bin, '*Chloroflexi* RR_metagenome_bin16' (Ji *et al.* 2017), was recovered from a polar surface soil metagenome (Robinson Ridge, Antarctica). Neither genome bin contained detectable genes for chlorophototrophy (Figure 4.7). Additionally, "*Ca.* Chx. allophototropha"-like *pscA* or *fmoA* genes could not be detected in source metagenome data even when relaxed search thresholds (e.g., an e-value cutoff of 10^{-1}) were used. These data leave open the possibility that not all members of the "*Ca.* Choroheliales" are phototrophic. Nevertheless, the close phylogenetic placement of RCI-utilizing "*Ca.* Choroheliales" members and RCII-utilizing *Chloroflexales* members implies the potential for past genetic interaction between these phototroph groups.

The genome of "*Ca.* Chx. allophototropha" encoded numerous phototrophy-associated genes with remote homology to phototrophy genes of known organisms (Figure 4.7). A highly novel homolog of *fmoA*, encoding the FMO protein involved in energy transfer from chlorosomes to RCI, was detected, having only 21-26% identity to known FmoA sequences at the amino acid level yet having no other closely related homologs in the NCBI RefSeq database (Olson 2005; Figure 4.8). This finding supports that "*Ca.* Chx. allophototropha" relies on RCI and chlorosomes for phototrophic energy conversion and makes the *Chloroflexota* the third known phylum to potentially use the FMO protein in phototrophy (Bryant *et al.* 2007). Using both bidirectional BLASTP and HMM-based searches, a potential homolog of the key chlorosome structural gene *csmA* was also detected (Frigaard and Bryant 2004; Bryant *et al.* 2012) with ~33% identity at the predicted amino acid level to the CsmA primary sequence of *Chloroflexus aurantiacus* J-10-fl (Figure 4.7). This
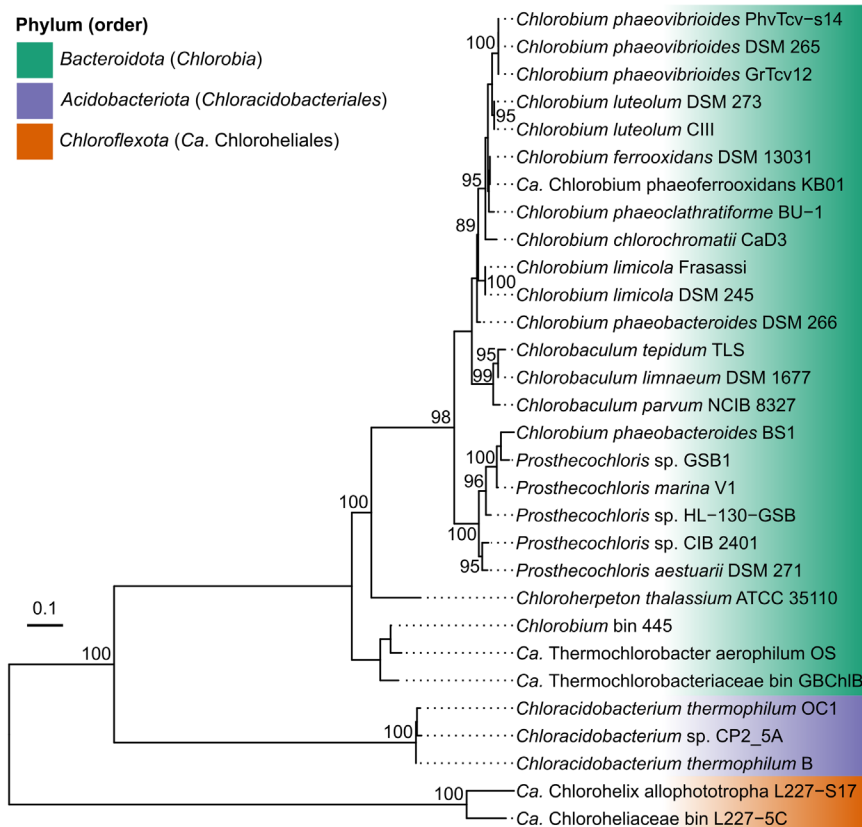
Figure 4.8 Maximum likelihood phylogeny of FMO protein (FmoA) sequences. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 356-residue masked sequence alignment (Table 4.1).

Figure 4.9 Structural properties of the predicted chlorosome protein CsmA in "*Ca.* Chx. allophototropha". **A,** Multiple sequence alignment of the CsmA primary sequence including representatives from known chlorosome-containing phyla. The entire CsmA primary sequence is shown, and a colour bar on the left of the alignment indicates the phylum and order associated with each sequence. A corresponding maximum likelihood phylogeny is shown in Figure 4.13. The BLOSUM62 sequence similarity score is shown underneath the alignment. Residues that are conserved across all sequences are marked with a black asterisk, and the H25 residue predicted to be involved in bacteriochlorophyll *a* binding is marked with a red asterisk. **B**, Homology model showing the possible tertiary structure of the "*Ca.* Chx. allophototropha" CsmA protein as predicted by I-TASSER. The tryptophan residue corresponding to the typical H25 residue is indicated by an arrow. **C**, Tertiary structure of the CsmA protein from *Chlorobium tepidum* (PDB accession 2K37); the H25 site is marked with an arrow.

103
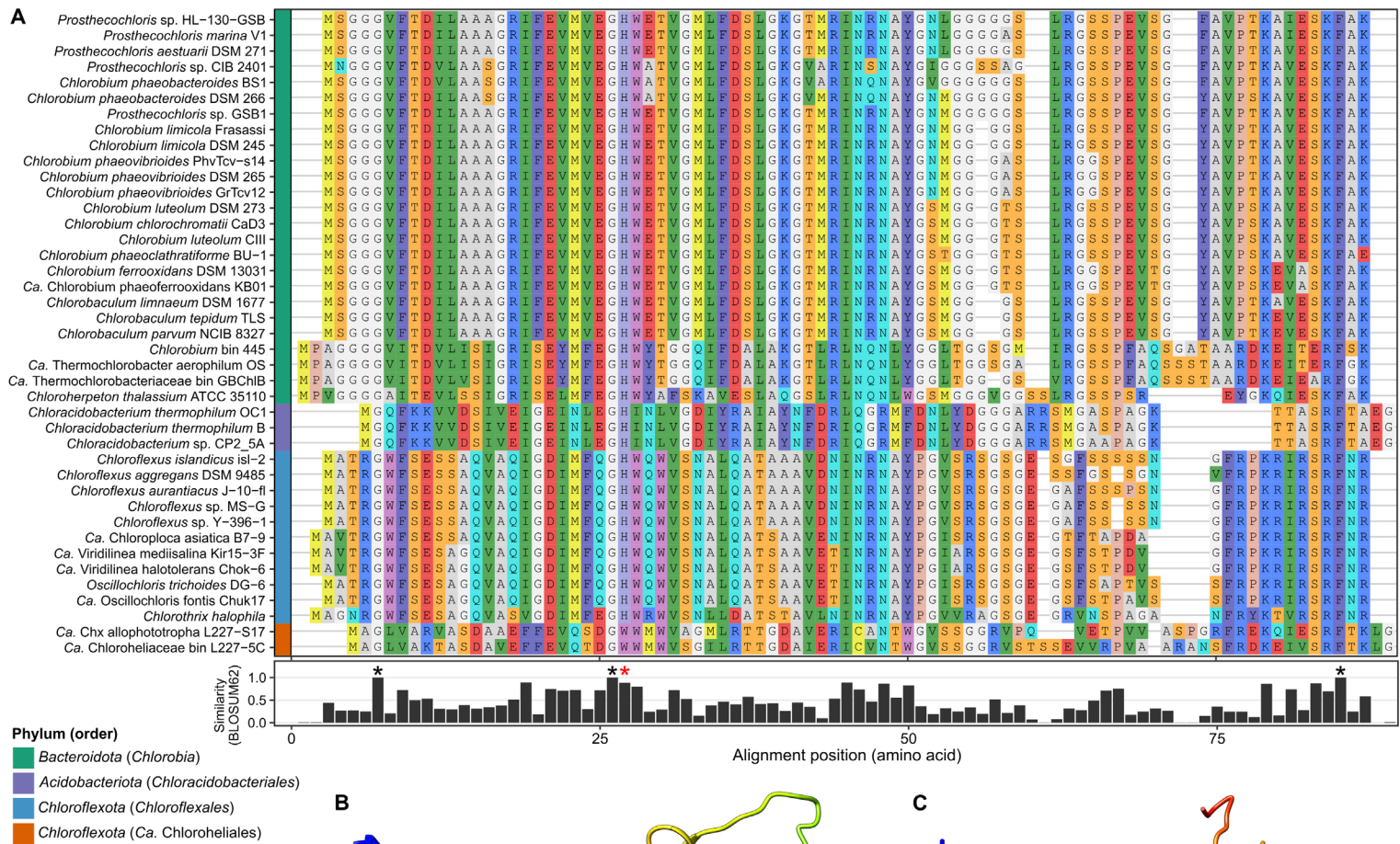
potential homolog had a similar predicted tertiary structure to the CsmA protein of *Chlorobium tepidum* (Pedersen *et al.* 2008) based on homology modelling and shared several highly conserved residues with other known CsmA sequences, although the potential homolog lacked the position-25 histidine hypothesized to be required for bacteriochlorophyll *a* binding (Pedersen *et al.* 2010; Figure 4.9). No other predicted open reading frames in the genome had close homology to known CsmA sequences, implying that this sequence represents a highly novel CsmA variant. Potential homologs of *csmM* and *csmY* (Bryant *et al.* 2012) were also detected in the genome (Figure 4.7; Table 4.2).

Similarly to some RCII-utilizing *Chloroflexota* members, the "*Ca.* Chx. allophototropha" genome encoded genes for the CBB cycle (Figure 4.7), including a deep-branching Class IC/ID "red type" *cbbL* gene (Tourova *et al.* 2006) representing the large subunit of RuBisCO (Figure 4.10). Carbon fixation via the CBB cycle has never before been reported for an RCI-utilizing phototroph, and this finding further highlights the novelty of "*Ca.* Chx. allophototropha" compared to known phototrophs. The novel genome did not encode the potential for carbon fixation via the 3-hydroxypropionate bicycle, which is thought to represent a more recent evolutionary innovation within the *Chloroflexales* order (Shih, Ward and Fischer 2017). The "*Ca.* Chx. allophototropha" genome also encoded the biosynthesis pathway for bacteriochlorophylls *a* and *c* (Bryant *et al.* 2012) and had the genomic potential for nitrogen fixation based on the presence of a *nifHI₁I₂BDK* gene cluster (Raymond *et al.* 2004; Figure 4.7). The genome bin of the second "*Ca.* Choroheliales" member enriched in this study (Figure 4.7; bin L227-5C) encoded a similar set of phototrophy-related gene homologs as "*Ca.* Chx. allophototropha", which further supports the robustness of the detection of these novel phototrophy-related genes.

Previously, the genetic content of RCII-utilizing phototrophic *Chloroflexota* members has been enigmatic because several *Chloroflexota*-associated phototrophy genes appear to be most closely related to genes of RCI-utilizing phototrophs. The results presented in this study suggest that "*Ca.* Chx. allophototropha" represents a missing transition form in the evolution of phototrophy among the *Chloroflexota* phylum that resolves this enigma by providing a direct link between RCI- and RCII-utilizing phototrophic taxa. Along with encoding chlorosomes, RCII-utilizing *Chloroflexota* members have been observed to encode bacteriochlorophyll synthesis genes most closely related to those of RCI-utilizing *Chlorobia* members (Bryant *et al.* 2012). According to maximum likelihood phylogenies of the (bacterio)chlorophyll synthesis proteins BchIDH/ChlIDH (Figure 4.11) and
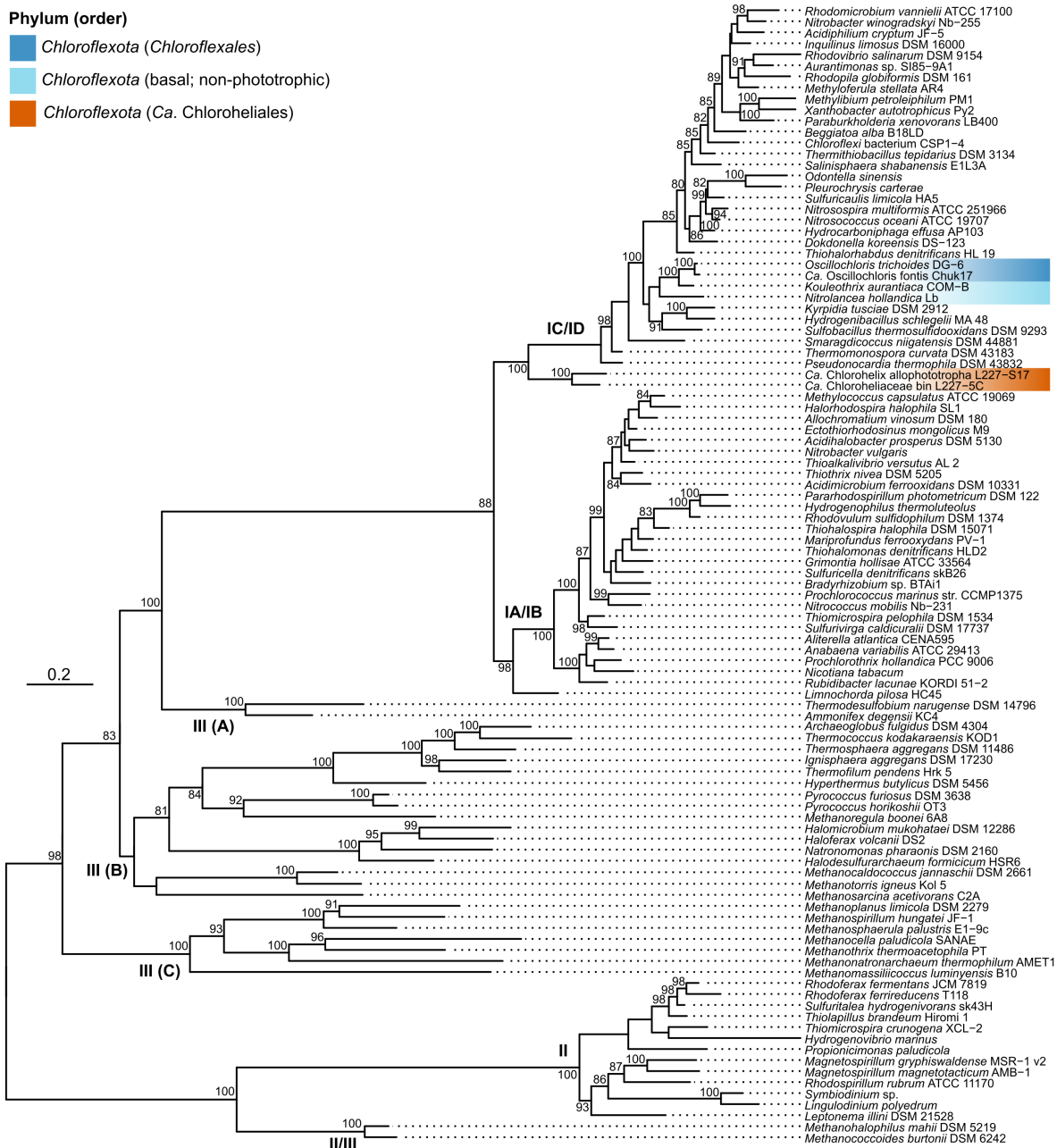
Figure 4.10 Maximum likelihood phylogeny of RuBisCO large subunit (CbbL) predicted protein sequences. Group I to III CbbL sequences are shown, and group IV sequences, which do not form proteins involved in carbon fixation, are omitted for conciseness. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 412-residue masked sequence alignment (Table 4.1).

Figure 4.11 Maximum likelihood phylogeny of BchIDH/ChlIDH protein sequences. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 1702-residue masked sequence alignment (Table 4.1). No sequences associated with the *Eremiobacterota* phylum are shown, because the complete BchIDH gene set could not be detected in any of the genome bins available from that phylum using bidirectional BLASTP.

BchLNB/ChlLNB/BchXYZ (Figure 4.12), "*Ca.* Chloroheliales" sequences place immediately adjacent to this sister grouping of RCII-utilizing *Choroflexota* and RCI-utilizing *Chlorobia* members. A phylogeny of the chlorosome structural protein CsmA also indicates closer evolutionary relatedness between "*Ca.* Chloroheliales" and RCII-utilizing *Chloroflexota* members compared to other RCI-utilizing anoxygenic phototrophs (Figure 4.13). The phylogenetic relatedness of RCI- and RCII-utilizing *Chloroflexota* members suggests a shared genetic history of phototrophy between these groups despite using different core reaction center components. Thus, it is possible that an ancestor of "*Ca.* Chx. allophototropha" received genes for RCI, chlorosomes, and bacteriochlorophyll synthesis, along with possible genes for carbon fixation, via lateral gene transfer and then provided the photosynthesis accessory genes needed by RCII-containing *Chloroflexales* members via either direct vertical inheritance or a subsequent lateral gene transfer event.

The cultivation of "*Ca.* Chx. allophototropha" also opens substantial possibilities for understanding the evolution of oxygenic photosynthesis. The *Chloroflexota* now represents the only phylum outside of the *Cyanobacteria* where both quinone and Fe-S type photosynthetic reaction centers can be found (Figure 4.5A), although in different organisms. Genes for RCII and RCI might be readily incorporated into a single member of the *Chloroflexota* phylum, for example given the apparent lateral transfer of RCII genes to basal *Chloroflexota* clades (Figure 4.7) described previously (Ward *et al.* 2018). One model for the origin of oxygenic photosynthesis suggests that a single organism received quinone and Fe-S type reaction centers via lateral gene transfer and then developed the capacity to use these two reaction center classes in tandem as PSII and PSI (Hohmann-Marriott and Blankenship 2011; Soo, Hemp and Hugenholtz 2019). The physiological effect of receiving both reaction center classes might now be testable in a laboratory setting. Ancestral state reconstructions could also yield insights into the genomic properties of the common ancestor of RCI-utilizing "*Ca.* Chloroheliales" and RCII-utilizing *Chloroflexales* members and whether this common ancestor was itself phototrophic. Generally, discovering "*Ca.* Chx. allophototropha" demonstrates the possibility for massive photosynthesis gene movements over evolutionary time, which is required by many evolutionary models but has been sparsely demonstrated in nature (Zeng *et al.* 2014; Martin, Bryant and Beatty 2018).

With the cultivation of "*Ca.* Chx. allophototropha", the *Chloroflexota* now represents the first known phylum of phototrophs having members that use either RCI or RCII for phototrophic energy conversion. "*Ca.* Chx. allophototropha" is unique among known phototrophs, because it encodes a distinct fourth class of RCI protein, forms the third known phototroph clade to utilize the FMO protein for energy transfer, and represents the only known RCI-utilizing phototroph to encode the CBB cycle for carbon fixation. Given that "*Ca.* Chx. allophototropha" could not be grown in freshwater medium amended with 0.5-3 mM sulfide in place of ferrous iron and that slow ferrous iron oxidation was observed in liquid cultures of "*Ca.* Chx. allophototropha" (not shown), this organism might also represent the first known photoferrotroph within the *Chloroflexota* phylum. Future research examining the biochemical properties of the reaction center and chlorosomes of "*Ca.* Chx. allophototropha", the iron metabolic potential of this strain, and the diversity and biogeography of "*Ca.* Chloroheliales" members, will provide fundamental insights into how photosynthesis operates and evolves. Cultivation of "*Ca.* Chx. allophototropha" allows the *Chloroflexota* to serve as a model system for exploring photosystem gene movements over evolutionary time.

4.4 Data and code availability

The three enrichment culture metagenomes, along with recovered genome bins, are available under the National Center for Biotechnology Information (NCBI) BioProject accession PRJNA640240. Amplicon sequencing data are available under the same NCBI BioProject accession. Custom scripts and additional raw data files used to analyze the metagenome and genome data are available at https://github.com/jmtsuji/Ca-Chlorohelix-allophototropha-RCI (doi:10.5281/zenodo.3932366).

Figure 4.12 Maximum likelihood phylogeny of predicted protein sequences of the paralogs BchLNB/ChlLNB and BchXYZ. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 819-residue masked sequence alignment (Table 4.1). One sequence among the *Chloroflexales* cluster for BchXYZ, associated with "*Ca.* Thermofonsia bin JP3-7", actually belongs to a genome bin that places in the basal clades of the *Chloroflexota* phylum (see Figure 4.7) and is thus marked with an asterisk.

Phylum (order)

- *Proteobacteria* (various)
- *Gemmatimonadota* (*Gemmatimonadales*)
- *Acidobacteriota* (*Chloracidobacteriales*)
- *Cyanobacteria* / chloroplast
- *Bacteroidota* (*Chlorobia*)
- *Chloroflexota* (*Chloroflexales*)*
- *Chloroflexota* (*Ca.* Chloroheliales)
- *Firmicutes* B (*Heliobacteriales*)
- *Eremiobacterota* (UBP12)

**BchLNB/
ChlLBC**

BchXYZ

0.2

Thiocapsa sp. KS1
Allochromatium vinosum DSM 180
Halorhodospira halophila SL1
Pseudohaliea rubra DSM 19751
Congregibacter litoralis KT71
Prochlorococcus marinus str. CCMP1375
Cyanobium gracile PCC 6307
Gemmatimonas phototrophica AP64
Thiorhodospira sibirica ATCC 700588
Methyloversatilis universalis FAM5
Rhodovibrio salinarum DSM 9154
Aquidulcibacter paucihalophilus TH1−2
Oceanibaculum indicum P24
Stappia stellulata DSM 5886
Rhodomicrobium vannielii ATCC 17100
Thalassobaculum litoreum DSM 18839
Rhodoferax fermentans JCM 7819
Rubritepida flocculans DSM 14296
Mongoliimonas terrestris MIMtkB18
Nevskia ramosa DSM 11499
Rhodospirillum rubrum ATCC 11170
Blastochloris viridis ATCC 19567
Afifella marina DSM 2698
Fulvimarina pelagi HTCC2506
Roseobacter litoralis Och 149
Erythrobacter longus DSM 6997
Rhodoblastus acidophilus DSM 137
Chloracidobacterium thermophilum OC1
Chloracidobacterium thermophilum B
Chloracidobacterium sp. CP2_5A
Nostoc punctiforme PCC 73102
Fortiea contorta PCC 7126
Anabaena cylindrica PCC 7122
Fischerella thermalis PCC 7521
Scytonema hofmannii PCC 7110
Chroococcidiopsis thermalis PCC 7203
Stanieria cyanosphaera PCC 7437
Microcystis aeruginosa NIES−843
Spirulina major PCC 6313
Cyanobacterium stanieri PCC 7202
Oscillatoria acuminata PCC 6304
Neosynechococcus sphagnicola sy1
Synechococcus elongatus PCC 6301
Dactylococcopsis salina PCC 8305
Prochlorothrix hollandica PCC 9006
Oscillatoria nigro−viridis PCC 7112
Gloeomargarita lithophora Alchichica−D10
Gloeobacter violaceus PCC 7421
Gloeobacter kilaueensis JS1
Chlorobium phaeobacteroides DSM 266
Chlorobium limicola DSM 245
Chlorobium phaeoclathratiforme BU−1
Ca. Chlorobium phaeoferrooxidans KB01
Chlorobium phaeovibrioides DSM 265
Chlorobium luteolum DSM 273
Chlorobium chlorochromatii CaD3
Chlorobaculum tepidum TLS
Chlorobaculum parvum NCIB 8327
Chlorobaculum limnaeum DSM 1677
Prosthecochloris sp. CIB 2401
Prosthecochloris aestuarii DSM 271
Prosthecochloris marina V1
Chlorobium phaeobacteroides BS1
Prosthecochloris sp. GSB1
Chloroherpeton thalassium ATCC 35110
Ca. Thermochlorobacter aerophilum OS
Chloroflexus sp. Y−396−1
Chloroflexus sp. MS−G
Chloroflexus aurantiacus J−10−fl
Chloroflexus aggregans DSM 9485
Chloroflexus islandicus isl−2
Ca. Viridilinea mediisalina Kir15−3F
Ca. Viridilinea halotolerans Chok−6
Ca. Chloroploca asiatica B7−9
Chloranaerofilum corporosum YNP−MS−B−OTU−15
Oscillochloris trichoides DG−6
Ca. Oscillochloris fontis Chuk17
Roseiflexus sp. RS−1
Roseiflexus castenholzii DSM 13941
Chlorothrix halophila
Ca. Chlorohelix allophototropha L227−S17
Ca. Chloroheliaceae bin L227−5C
Heliobacterium gestii
Heliobacterium modesticaldum Ice1
Heliobacillus mobilis
Ca. Eremiobacterota bin bog_1520
Ca. Eremiobacterota bin bog_1501
Ca. Eremiobacterota bin bog_1527
Ca. Eremiobacterota bin bog_1519
Chlorobium phaeoclathratiforme BU−1
Ca. Chlorobium phaeoferrooxidans KB01
Chlorobium chlorochromatii CaD3
Chlorobium phaeobacteroides DSM 266
Chlorobium limicola DSM 245
Chlorobium phaeovibrioides DSM 265
Chlorobium luteolum DSM 273
Prosthecochloris marina V1
Chlorobium phaeobacteroides BS1
Prosthecochloris sp. GSB1
Prosthecochloris sp. CIB 2401
Prosthecochloris aestuarii DSM 271
Chlorobaculum tepidum TLS
Chlorobaculum parvum NCIB 8327
Chlorobaculum limnaeum DSM 1677
Ca. Thermochlorobacteriaceae bin GBChlB
Ca. Thermochlorobacter aerophilum OS
Chloroflexus sp. Y−396−1
Chloroflexus sp. MS−G
Chloroflexus aurantiacus J−10−fl
Chloroflexus islandicus isl−2
Chloroflexus aggregans DSM 9485
Ca. Viridilinea mediisalina Kir15−3F
Ca. Viridilinea halotolerans Chok−6
Ca. Chloroploca asiatica B7−9
Oscillochloris trichoides DG−6
Ca. Oscillochloris fontis Chuk17
Chlorothrix halophila
Chloroflexi bin J036
Ca. Thermofonsia bin JP3−7 *
Roseiflexus sp. RS−1
Roseiflexus castenholzii DSM 13941
Ca. Chlorohelix allophototropha L227−S17
Ca. Chloroheliaceae bin L227−5C
Heliobacterium gestii
Heliobacterium modesticaldum Ice1
Heliobacillus mobilis
Chloracidobacterium thermophilum OC1
Chloracidobacterium thermophilum B
Chloracidobacterium sp. CP2_5A
Ca. Eremiobacterota bin bog_1492
Rhodospirillum rubrum ATCC 11170
Pararhodospirillum photometricum DSM 122
Rhodomicrobium vannielii ATCC 17100
Erythrobacter longus DSM 6997
Aquidulcibacter paucihalophilus TH1−2
Thiorhodospira sibirica ATCC 700588
Afifella marina DSM 2698
Rhodoblastus acidophilus DSM 137
Halorhodospira halophila SL1
Pseudohaliea rubra DSM 19751
Congregibacter litoralis KT71
Thiocapsa sp. KS1
Stappia stellulata DSM 5886
Oceanibaculum indicum P24
Blastochloris viridis ATCC 19567
Nevskia ramosa DSM 11499
Methyloversatilis universalis FAM5
Rhodoferax fermentans JCM 7819
Gemmatimonas phototrophica AP64
Roseobacter litoralis Och 149
Fulvimarina pelagi HTCC2506
Mongoliimonas terrestris MIMtkB18
Thalassobaculum litoreum DSM 18839
Rubritepida flocculans DSM 14296
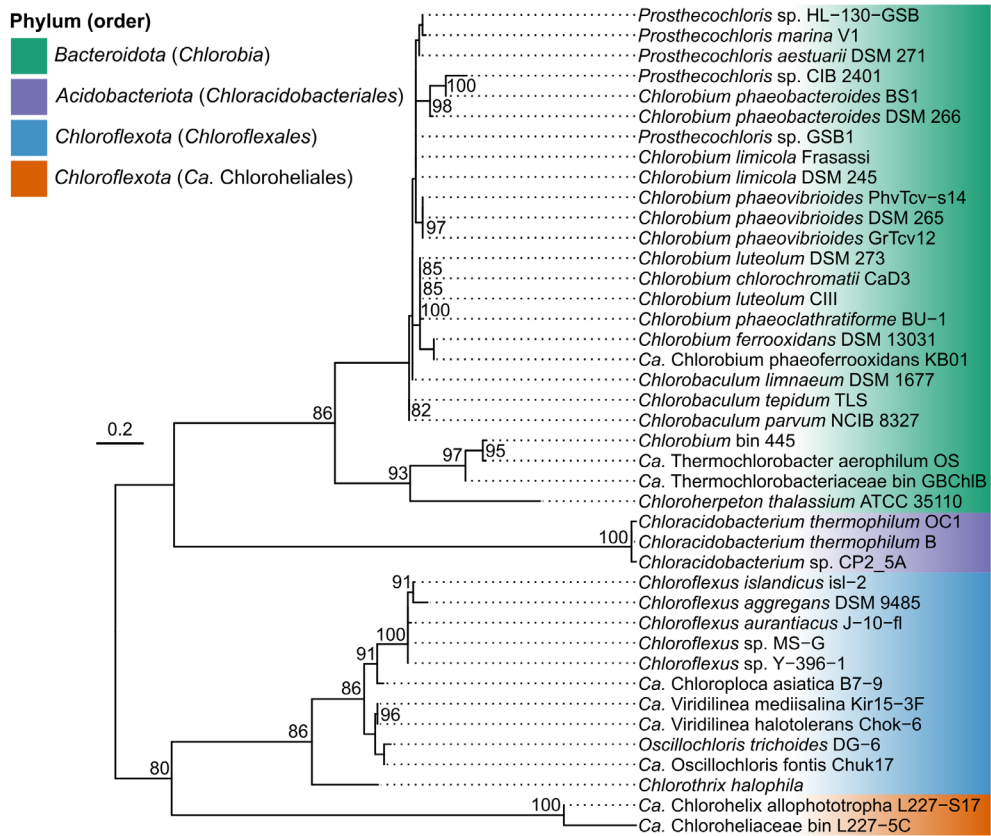Rhodovibrio salinarum DSM 9154

109

Figure 4.13 Maximum likelihood phylogeny of CsmA predicted protein sequences. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 74-residue masked sequence alignment (Table 4.1).

Chapter 5
Conclusion and outlook

5.1 Summary

Although oxygenic photosynthesis in aquatic environments receives substantial attention due to its important role in the global carbon and nitrogen cycles (Field *et al.* 1998), anoxygenic forms of photosynthesis are often overlooked. Anoxygenic phototrophs are typically assumed to be photoheterotrophic or to oxidize sulfide. Alternative modes of phototrophy, such as photoferrotrophy, are of great interest in the context of early Earth research and astrobiology (Posth, Konhauser and Kappler 2013), but the gap between these fields and modern aquatic ecology makes inferences about both the modern and ancient roles of alternative modes of phototrophy challenging to assess (Camacho *et al.* 2017). In my thesis work, I expand upon sparse knowledge of microbial biogeochemical cycling in ferruginous lakes, with geochemistry analogous to that of early Earth oceans (Koeksoy *et al.* 2016), by exploring the phototrophic microbial communities of globally abundant Boreal Shield lakes. Many of these lakes develop seasonal anoxia and ferruginous bottom waters and could complement research on meromictic and ferruginous lakes, which are rare, given that Boreal Shield lakes number in the tens of millions globally (Figure 1.1). My research shows the potential for metabolically diverse phototrophs in Boreal Shield lake water columns, along with active and cryptic sulfur cycling and the potential for oxygenic photosynthesis even in the lake anoxic zone. Potentially metabolically novel phototrophs found in multiple Boreal Shield lakes appear to thrive in lakes that readily develop anoxia and have a low surface area:depth ratio, high levels of iron, and high levels of dissolved organic carbon. Moreover, my research points to substantial taxonomic diversity of anoxygenic phototrophs in Boreal Shield lakes, including highly novel phototrophic *Chloroflexota* members that revise understanding of how photosynthesis evolved and functions.

5.1.1 Genetic potential for iron and sulfur cycling in Boreal Shield lake anoxic zones

The research presented in this thesis highlights the potential for population-level metabolic differences between *Chlorobia* members in Boreal Shield lake water columns. In contrast to the few and recently cultivated photoferrotrophic members of the *Chlorobia* class, four novel *Chlorobia*-associated *cyc2* variants were detected within metagenome data from Lakes 227, 442, and 304, which doubled the known diversity of this photoferrotrophy-associated cytochrome gene (Figure 2.1). Based

on comparative genomics, *cyc2* gene homologs appear to be unique to cultured *Chlorobia* members with the potential for iron oxidation, implying a functional link to photoferrotrophy (Figure 2.6). However, based on enriching a *cyc2*-encoding *Chlorobia* member from Lake 304, named "*Ca.* Chlorobium canadense*", and testing the iron oxidation potential of this strain, this work provided caution in assigning the functional role of *cyc2* to photoferrotrophy based on comparative genomic evidence alone. Unlike all other cultured *Chlorobia* members to date that encoded *cyc2*, "*Ca.* Chl. canadense" could not be induced to oxidize ferrous iron in the laboratory and instead oxidized sulfide (Figure 2.12). Several possible reasons exist for why photoferrotrophy was unable to be induced in "*Ca.* Chl. canadense", including that the *cyc2* gene product plays an alternative cellular role to photoferrotrophy or that tight metabolic controls regulate the expression of *cyc2* in these bacteria. "*Ca.* Chl. canadense" will prove valuable for determining the function and regulation of *cyc2* and its potential links to photoferrotrophy in laboratory culture.

The genomic potential for sulfide oxidation was detected among populations of *Chlorobia* in both Lakes 227 and 442 (Figure 2.9), and observations during enrichment cultivation pointed toward high activity of sulfate-reducing bacteria despite low sulfate levels in the lakes. Genomic potential for iron and sulfur cycling was also detected among several other populations in both lakes that included potential microaerophilic iron-oxidizing bacteria, sulfate reducing bacteria, and chemotrophic sulfur-oxidizing bacteria, highlighting the potential importance of iron and sulfur cycling in Boreal Shield lake systems (Figure 2.11). In addition, robust detection methods for *cyc2* were developed, incorporating a custom-built HMM, genome context information, and phylogenetic placement (Figure 2.1). These methods can be applied to additional genomes or metagenomes, expanding the sparse toolkit currently available for detection of genes potentially involved in iron cycling. Overall, this work played a foundational role for exploration of phototrophs and their surrounding microbial consortia in ferruginous water columns.

5.1.2 Diversity and drivers of phototrophy in Boreal Shield lakes

By surveying nine Boreal Shield lakes at the IISD-ELA, the research included in this thesis provided critical information on the broader distribution of phototrophs in Boreal Shield lake anoxic zones. Rather than being an anomaly detected in only a few Boreal Shield lakes with history of artificial nutrient addition, diverse and high relative abundance phototrophs were detected in a third of the studied lakes, including Lake 221, which has never been manipulated experimentally (Figure 3.3). Geographic and physicochemical data indicated that the surveyed lakes could be separated into two

broad categories – lakes that readily developed anoxia with a low surface area:depth ratio, high DOC, and high levels of dissolved iron, or clearer lakes with a higher surface area:depth ratio having variable iron levels (Figure 3.1B) and observed anoxia timings (Figure 3.1C). Microbial community data generally supported these lake groupings (Figure 3.1D) and lakes in the first group tended to be phototroph-rich, containing high relative abundances of anoxygenic phototrophs that had high observed diversity. It might be possible, in light of these findings, to broadly predict which Boreal Shield lakes are conducive to substantial phototroph growth based on basic knowledge of the lake's properties, such as its depth, surface area, and colour. All of the detected phototroph-rich lakes in the high DOC group also contained *Chlorobia*-associated *cyc2* genes (Figure 3.3), unlike what was observed for Lake 442 previously.

Based on probing the activity of phototrophs in the three detected "phototroph-rich" lakes, the research presented in this thesis indicated that phototrophs represented a substantial proportion (nearly half) of all gene expression in the upper anoxic zones of the lakes and that the *cyc2* gene among *Chlorobia* members was actively expressed in at least two of the lakes (Figure 3.7A). Confirming previous observations, rate-based evidence for sulfate reduction was found in the upper water column of Lake 227 that corresponded with sulfide oxidation activity of *Chlorobia* members and high expression of the *dsrA* marker gene for sulfide oxidation (Figure 3.7A-B). Such findings, combined with high expression of *Chlorobia*-associated *dsrA* genes in Lakes 304 and 221, point toward cryptic sulfur cycling playing a key role in shaping the phototroph microbial community structure of ferruginous Boreal Shield lake systems. The observations align with findings from other meromictic and ferruginous lake systems where photoferrotrophy was slow or undetectable compared to other forms of phototrophy (Crowe *et al.* 2014a; Walter *et al.* 2014) or where cryptic sulfur cycling was found (Berg *et al.* 2019). Phototrophic sulfide oxidation appeared to be the dominant anoxygenic phototrophic process among *Chlorobia* populations compared to photoferrotrophy at the times when lakes were sampled, although the possibility of photoferrotrophy coupled to cryptic iron cycling could not be excluded. Rate measurements and RNA sequences suggested that *Cyanobacteria* members were active in the lake anoxic zones and, at least in the case of Lake 227, were the primary drivers of phototrophic iron oxidation in the water column to a depth of at least 5 m (i.e., >1 m beneath the measurable oxic-anoxic zone boundary when sampling; Figure 3.7B). Phototrophic iron oxidation facilitated by *Cyanobacteria* members exceeded rates of iron reduction in incubations from the upper anoxic waters of Lake 227, and thus the Lake 227 water column might serve as a helpful analogue for

exploring microbial metabolisms and iron biogeochemistry in the Earth's oceans leading up to the Great Oxidation Event.

Lake 227 and other Boreal Shield lakes likely have an active and potentially highly novel methane cycle in their anoxic zones. Along with expression of the *mcrA* gene associated with methanogenesis or reverse methanogenesis by archaeal populations, expression of the particulate methane monooxygenase (*pmoA*) gene, associated with gammaproteobacterial methanotrophs thought to be strict aerobes, was detected in the deep, dark, and anoxic water column of Lake 227 (Figure 3.9). Beyond Lake 227, high relative abundances of related methanotrophs were detected in the deep anoxic zones of other surveyed lakes based on metagenome data (Figure 2.9). These data align with preliminary reports of methane oxidation by proteobacterial methanotrophs in lake anoxic zones or sediments, apparently not linked to oxygenic photosynthetic activity (Martinez-Cruz *et al.* 2017; van Grinsven *et al.* 2019, 2020), and could suggest highly novel aspects of the metabolism of these well-studied methanotrophic organisms that are yet to be uncovered.

5.1.3 Expanding understanding of the diversity and evolution of photosynthesis through the cultivation of "*Ca.* Chx. allophototropha"

As an unexpected finding from efforts to enrich iron-oxidizing phototrophs from Boreal Shield lakes, cultivation of "*Ca.* Chx. allophototropha" revealed the first known *Chloroflexota* member to use a Type I reaction center for photosynthesis and the first known case of RCI and RCII being used by different phototrophs within the same phylum. Although this bacterium appeared physiologically similar to known mesophilic and phototrophic *Chloroflexota* members (Figure 4.3), it encoded a distinct fourth clade of RCI protein most closely related to the RCI of *Chloracidobacterium* or *Heliobacterium* members (Figure 4.5). "*Ca.* Chx. allophototropha" was highly unique compared to known phototrophs, given that it represents the first known RCI-utilizing anoxygenic phototroph to encode the pathway for carbon fixation via the CBB cycle and represents the third known clade of phototroph to potentially use the FMO protein involved in bacteriochlorophyll *a* binding (Figure 4.7). The relatively close phylogenetic placement of "*Ca.* Chx. allophototropha" to the RCII-utilizing *Chloroflexaceae* family implied the potential for historic genetic interaction between these organisms. Moreover, consistent sister phylogenetic placement of photosynthesis-associated genes of "*Ca.* Chx. allophototropha" to those of *Chloroflexaceae* members suggests that lateral transfer of phototrophy genes, or descent from a common phototroph ancestor, occurred between these phototrophic groups in the evolutionary past. Thus, the data presented in my thesis could explain how RCII-utilizing

Chloroflexaceae members obtained chlorosomes that are typically only associated with RCI-utilizing phototrophs: they may have received such genes via an ancestor of "*Ca.* Chx. allophototropha" in the first-known genetic transfer event between RCI and RCII-utilizing organisms. In this way, "*Ca.* Chx. allophototropha" revises understanding of the diversity and evolution of photosynthesis and allows the *Chloroflexota* phylum to serve as a model system for exploring RCI/RCII interactions relevant to understanding the origin of oxygenic photosynthesis.

5.2 Outlook

The findings presented in this thesis open the possibilities for several novel avenues of research related to iron geomicrobiology and microbial ecology. In contrast to traditional perspectives that the hypolimnia of boreal lakes are mostly inactive sources of reduced compounds, my thesis research provides strong evidence that diverse microbial redox cycling, including cryptic sulfur turnover, occurs in the anoxic water columns of Boreal Shield lakes. Moreover, the fact that functional differences exist in microbial communities between lake hypolimnia implies that nutrient cycling will function differently in anoxic zones among different lakes. These findings add to a growing paradigm shift in limnology that complex microbial processes play an important role in lake function. The metagenome and metatranscriptome datasets presented in this work will allow for additional microbial processes, such as the potential for mercury methylation linked to iron reduction (Fleming *et al.* 2006), or the potential for modes of phototrophy coupled to other electron donors such as manganese(II) (Daye *et al.* 2019), to be explored prior to experimental work in the field. Such microbial data could enhance the ability to predict how abundant Boreal Shield lakes will respond to global change (Anas, Scott and Wissel 2015).

The findings presented in this thesis of the importance of cryptic sulfur cycling in the anoxic water columns of Boreal Shield lakes are relevant to early Earth research and to modern limnology. Of all studied ferruginous lakes to date, only Kabuno Bay of Lake Kivu appears to be dominated by photoferrotrophic activity in its anoxic water column (Llirós *et al.* 2015). Other iron-rich lakes, including Boreal Shield lakes, appear to have substantial activity of other phototrophs that complements or surpasses photoferrotroph activity (Crowe *et al.* 2014a; Walter *et al.* 2014). Cryptic sulfur cycling has been shown to play an important role in nutrient cycling in many modern environments (Johnston *et al.* 2014; Kappler and Bryce 2017), and given the potentially ancient evolutionary origin of microbial sulfur cycling (Lyons, Fike and Zerkle 2015), it is possible that the

cryptic sulfur cycle may have also become increasingly important in the Archaean oceans as additional organic carbon became available to fuel sulfate-reducing bacteria. The role of cryptic sulfur cycling and its impact on the deposition of banded iron formations could be further examined in future laboratory incubations or field experiments exploring the interplay of iron and sulfur-cycling microorganisms.

Key to understanding the potential functional role of *cyc2* in photoferrotrophy will be laboratory-based molecular studies of the gene's properties and activity. Newly cultivated photoferrotrophs such as *Chlorobium* sp. strain N1, which are known to be able to switch between the phototrophic oxidation of ferrous iron and sulfide (Laufer *et al.* 2017), as well as "*Ca.* Chl. canadense" cultivated in this work and *Chl. luteolum*, could be monitored for gene expression changes when different phototrophic electron donors are provided. Gene knock-out or knock-in studies may be required to determine whether *cyc2* plays a role in photoferrotrophy as well as to identify other genes potentially involved in the process. Such study would prove highly valuable for establishing molecular markers for iron oxidation generally among environmental samples. Furthermore, additional rate measurements of Boreal Shield lakes such as Lakes 304 and 221, including the potential to measure cryptic iron cycling, could help to validate the link between *cyc2* and environmental iron oxidation.

The spatial and temporal dynamics of the observed microbial communities in Boreal Shield lake anoxic zones remains poorly understood. Based on the mono- or dimictic nature of the studied lakes, anoxic zone microbial communities must re-establish themselves after each mixing event and likely undergo a period of dynamic change. The research presented in this thesis represents a first glimpse into the functional potential of these anoxic zone microbial communities. Future work at increased temporal sampling resolution, focusing in detail on a few selected lake systems (e.g., Lake 221 versus the neighbouring Lake 222), could yield important insights into potential successional changes in microbial community structure and gene expression over time. Such research could yield insights into the relative roles of the iron and sulfur cycles in fuelling phototrophic microbial communities over time or even the roles of alternative photosynthetic electron donors such as manganese(II) at different points during the season. In addition, increased spatial resolution sampling in the upper anoxic water columns of Boreal Shield lakes could show how dominant forms of phototrophy and their corresponding microbial communities shift across geochemical gradients, and such study could yield important insight into geochemical drivers of microbial community function.

In light of the findings presented in this thesis that the particulate methane monooxygenase gene is highly expressed in the Boreal Shield lake anoxic zones, further exploration of methane metabolism in Boreal Shield lakes is warranted and could yield ground-breaking discoveries about the biochemistry and ecology of methane oxidation. Other bacteria, such as "*Ca.* Methylomirabilis oxyfera", oxidize methane in anoxic environments by generating molecular oxygen internally (Ettwig *et al.* 2010). It is possible that intracellular or syntrophic generation of molecular oxygen could occur in Boreal Shield lake water columns or that some type of EET could occur to support methane oxidation. Confirming methanotrophic activity by these organisms in the Lake 227 water column would substantially broaden the potential niche of "aerobic" methanotrophs in environments globally, with implications for our understanding of the global methane cycle. Measurements of methane oxidation rates in the Lake 227 anoxic zone, or measures of per-cell methane uptake via nanoSIMS (Oswald *et al.* 2016b), could prove highly valuable as a next research step in this exciting work.

Cultivation of "*Ca.* Chx. allophototropha" has substantial future implications for photosynthesis research. Further physiological and biochemical characterization of the novel strain could yield key insights into the evolution of photosynthesis. For example, purification of the strain followed by confirmation of photoferrotrophic activity would broaden the known diversity of photoferrotrophy and iron-cycling genes by introducing a third phylum containing known photoferrotrophs (Melton *et al.* 2014; Bryce *et al.* 2018). Biochemical characterization of the "*Ca.* Chx. allophototropha" chlorosome, whose subunits appear to lack the H25 residue thought to be key for bacteriochlorophyll *a* binding, would further understanding of protein-pigment interactions. Longer-term, ancestral state reconstruction work and exploration of the global diversity of "*Ca.* Chloroheliales" members could allow for past evolutionary interactions between RCI- and RCII-utilizing *Chloroflexota* members to be better untangled. One exciting possibility is that a single common ancestor of both RCI- and RCII-utilizing *Chloroflexota* members encoded both photosynthesis systems simultaneously, with major implications for understanding of the evolution of oxygenic photosynthesis (Hohmann-Marriott and Blankenship 2011).

5.3 Research significance

In the current "golden age" of iron geomicrobiology (Melton *et al.* 2014), it is recognized that increasingly interdisciplinary efforts are needed to untangle the complexities of iron biogeochemical cycling (Raiswell and Canfield 2012). The research presented in this thesis represents the first application of genome-resolved metagenomics, metatranscriptomics, and biogeographic approaches to study of the microbial ecology of ferruginous water columns. This work summarizes the key microbial players and processes involved in phototrophy in Boreal Shield lakes, the potential interplay of phototrophs with the iron/sulfur cycles, and the underexplored microbial diversity of these lake systems. Work from Chapter 2 extends limited knowledge of the diversity of *cyc2* among *Chlorobia* members and suggests substantial metabolic potential for microbial iron- and sulfur-cycling in Boreal Shield lake anoxic zones. Ecological work in Chapter 3 shows that taxonomically and potentially functionally novel phototrophs are detectable among multiple Boreal Shield lakes and highlights the potential importance of cryptic sulfur cycling in iron-dominated systems. Lastly, cultivation of "*Ca.* Chx. allophototropha" reported in Chapter 4 showcases the novelty of Boreal Shield lake microbial communities and sheds light on basic processes governing the evolution of photosynthesis. The sequencing data presented in this thesis can continue to be mined and explored for additional functional novelty, and lakes at the IISD-ELA are readily accessible for follow-up observational and manipulative experimental work. Millions of additional Boreal Shield lakes could harbor distinct anoxic zone microbial communities related to those presented in this work and thus serve as valuable platforms for exploration of iron geomicrobiology, microbial ecology, and questions related to early Earth research and global change.

References

Allen JWA, Sawyer EB, Ginger ML *et al.* Variant c-type cytochromes as probes of the substrate specificity of the E. coli cytochrome c maturation (Ccm) apparatus. *Biochem J* 2009;**419**:177–86.

Altschul SF, Gish W, Miller W *et al.* Basic local alignment search tool. *J Mol Biol* 1990;**215**:403–10.

Amenabar MJ, Shock EL, Roden EE *et al.* Microbial substrate preference dictated by energy demand rather than supply. *Nat Geosci* 2017;**10**:577–81.

Anas MUM, Scott KA, Wissel B. Carbon budgets of boreal lakes: state of knowledge, challenges, and implications. *Environ Rev* 2015;**23**:275–87.

Anbar AD, Duan Y, Lyons TW *et al.* A whiff of oxygen before the Great Oxidation Event? *Science* 2007;**317**:1903–6.

Anderson MJ. A new method for non-parametric multivariate analysis of variance. *Austral Ecol* 2001;**26**:32–46.

Armstrong FAJ, Schindler DW. Preliminary chemical characterization of waters in the Experimental Lakes Area, northwestern Ontario. *J Fish Res Bd Can* 1971;**28**:171–87.

Barco RA, Emerson D, Sylvan JB *et al.* New insight into microbial iron oxidation as revealed by the proteomic profile of an obligate iron-oxidizing chemolithoautotroph. *Appl Environ Microbiol* 2015;**81**:5927–37.

Bartram AK, Lynch MD, Stearns JC *et al.* Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Appl Environ Microbiol* 2011;**77**:3846–52.

Beal EJ, Claire MW, House CH. High rates of anaerobic methanotrophy at low sulfate concentrations with implications for past and present methane levels. *Geobiology* 2011;**9**:131–9.

Beard BL, Johnson CM, Skulan JL *et al.* Application of Fe isotopes to tracing the geochemical and biological cycling of Fe. *Chem Geol* 2003;**195**:87–117.

Beaumont V, Robert F. Nitrogen isotope ratios of kerogens in Precambrian cherts: a record of the evolution of atmosphere chemistry? *Precambrian Res* 1999;**96**:63–82.

Berg JS, Jézéquel D, Duverger A *et al.* Microbial diversity involved in iron and cryptic sulfur cycling in the ferruginous, low-sulfate waters of Lake Pavin. *PLOS ONE* 2019;**14**:e0212787.

Berg JS, Michellod D, Pjevac P *et al.* Intensive cryptic microbial iron cycling in the low iron water column of the meromictic Lake Cadagno. *Environ Microbiol* 2016;**18**:5288–302.

Bergstrand LH, Cardenas E, Holert J *et al.* Delineation of steroid-degrading microorganisms through comparative genomic analysis. *mBio* 2016;**7**:e00166-16.

Biderre-Petit C, Boucher D, Kuever J *et al.* Identification of sulfur-cycle prokaryotes in a low-sulfate lake (Lake Pavin) using *aprA* and 16S rRNA gene markers. *Microb Ecol* 2011;**61**:313–27.

Biswas KC, Woodards NA, Xu H *et al.* Reduction of molybdate by sulfate-reducing bacteria. *Biometals* 2009;**22**:131–9.

Blankenship RE. *Molecular Mechanisms of Photosynthesis*. West Sussex, UK: John Wiley & Sons, 2014.

Boiteau RM, Mende DR, Hawco NJ *et al.* Siderophore-based microbial adaptations to iron scarcity across the eastern Pacific Ocean. *Proc Natl Acad Sci USA* 2016;**113**:14237–42.

Bokulich NA, Kaehler BD, Rideout JR *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 2018;**6**:90.

Bolyen E, Rideout JR, Dillon MR *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 2019;**37**:852–857.

Boyd PW, Ellwood MJ. The biogeochemical cycle of iron in the ocean. *Nat Geosci* 2010;**3**:675–82.

Bravo AG, Peura S, Buck M *et al.* Methanogens and iron-reducing bacteria: the overlooked members of mercury methylating microbial communities in boreal lakes. *Appl Environ Microbiol* 2018;**84**:e01774-18.

Brunskill GJ, Schindler DW. Geography and bathymetry of selected lake basins, Experimental Lakes Area, northwestern Ontario. *J Fish Res Bd Can* 1971;**28**:139–55.

Bryant DA, Costas AMG, Maresca JA *et al. Candidatus* Chloracidobacterium thermophilum: an aerobic phototrophic Acidobacterium. *Science* 2007;**317**:523–6.

Bryant DA, Liu Z, Li T *et al.* Comparative and functional genomics of anoxygenic green bacteria from the taxa *Chlorobi*, *Chloroflexi*, and *Acidobacteria*. In: Burnap R, Vermaas W (eds.). *Functional Genomics and Evolution of Photosynthetic Systems*. Dordrecht: Springer Netherlands, 2012, 47–102.

Bryce C, Blackwell N, Schmidt C *et al.* Microbial anaerobic Fe(II) oxidation – ecology, mechanisms and environmental implications. *Environ Microbiol* 2018;**20**:3462–83.

Bryce C, Blackwell N, Straub D *et al.* Draft genome sequence of *Chlorobium* sp. strain N1, a marine Fe(II)-oxidizing green sulfur bacterium. *Microbiol Resour Announc* 2019;**8**:e00080-19.

Busigny V, Planavsky NJ, Jézéquel D *et al.* Iron isotopes in an Archean ocean analogue. *Geochim Cosmochim Acta* 2014;**133**:443–62.

Cai C, Leu AO, Xie G-J *et al.* A methanotrophic archaeon couples anaerobic oxidation of methane to Fe(III) reduction. *ISME J* 2018;**12**:1929–39.

Callahan BJ, McMurdie PJ, Rosen MJ *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* 2016;**13**:581–3.

Camacho A, Walter XA, Picazo A *et al.* Photoferrotrophy: remains of an ancient photosynthesis in modern environments. *Front Microbiol* 2017;**8**:323.

Campbell P. Phosphorus budgets and stoichiometry during the open-water season in two unmanipulated lakes in the Experimental Lakes Area, northwestern Ontario. *Can J Fish Aquat Sci* 1994;**51**:2739–55.

Campbell P, Torgersen T. Maintenance of iron meromixis by iron redeposition in a rapidly flushed monimolimnion. *Can J Fish Aquat Sci* 1980;**37**:1303–13.

Caporaso JG, Kuczynski J, Stombaugh J *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* 2010;**7**:335–6.

Cardona T. Early Archean origin of heterodimeric Photosystem I. *Heliyon* 2018;**4**:e00548.

Cardona T. Thinking twice about the evolution of photosynthesis. *Open Biol* 2019;**9**:180246.

Cardona T, Rutherford AW. Evolution of photochemical reaction centres: more twists? *Trends Plant Sci* 2019;**24**:1008–21.

Cardona T, Sánchez-Baracaldo P, Rutherford AW *et al.* Early Archean origin of Photosystem II. *Geobiology* 2018;**17**:127–50.

Castelle C, Guiral M, Malarte G *et al.* A new iron-oxidizing/O_2-reducing supercomplex spanning both inner and outer membranes, isolated from the extreme acidophile *Acidithiobacillus ferrooxidans*. *J Biol Chem* 2008;**283**:25803–11.

Cavaco MA, St. Louis VL, Engel K *et al.* Freshwater microbial community diversity in a rapidly changing High Arctic watershed. *FEMS Microbiol Ecol* 2019;**95**:fiz161.

Chan C, McAllister SM, Garber A *et al.* Fe oxidation by a fused cytochrome-porin common to diverse Fe-oxidizing bacteria. *bioRxiv* 2018:228056.

Chan CS, Emerson D, Luther GW. The role of microaerophilic Fe-oxidizing micro-organisms in producing banded iron formations. *Geobiology* 2016;**14**:509–28.

Chan CS, Fakra SC, Emerson D *et al.* Lithotrophic iron-oxidizing bacteria produce organic stalks to control mineral growth: implications for biosignature formation. *ISME J* 2011;**5**:717–27.

Chaumeil P-A, Mussig AJ, Hugenholtz P *et al.* GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 2020;**36**:1925–7.

Chen Z, Pham L, Wu T-C *et al.* Ultra-low input single tube linked-read library method enables short-read second-generation sequencing systems to generate highly accurate and economical long-range sequencing information routinely. *Genome Res* 2020;**30**:898–909.

Cheng J, Pinnell L, Engel K *et al.* Versatile broad-host-range cosmids for construction of high quality metagenomic libraries. *J Microbiol Methods* 2014;**99**:27–34.

Clavier CGJ, Boucher G. The use of photosynthesis inhibitor (DCMU) for in situ metabolic and primary production studies on soft bottom benthos. *Hydrobiologia* 1992;**246**:141–5.

Clément J-C, Shrestha J, Ehrenfeld JG *et al.* Ammonium oxidation coupled to dissimilatory reduction of iron under anaerobic conditions in wetland soils. *Soil Biol Biochem* 2005;**37**:2323–8.

Cline JD. Spectrophotometric determination of hydrogen sulfide in natural waters. *Limnol Oceanogr* 1969;**14**:454–8.

Coates JD, Ellis DJ, Gaw CV *et al. Geothrix fermentans* gen. nov., sp. nov., a novel Fe(III)-reducing bacterium from a hydrocarbon-contaminated aquifer. *Int J Syst Bacteriol* 1999;**49**:1615–22.

Colbeau A, Vignais PM. The membrane-bound hydrogenase of *Rhodopseudomonas capsulata*: stability and catalytic properties. *Biochim Biophys Acta* 1981;**662**:271–84.

Cole JR, Wang Q, Fish JA *et al.* Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucl Acids Res* 2014;**42**:D633–42.

Cook RB. Distributions of ferrous iron and sulfide in an anoxic hypolimnion. *Can J Fish Aquat Sci* 1984;**41**:286–93.

Craddock PR, Dauphas N. Iron and carbon isotope evidence for microbial iron respiration throughout the Archean. *Earth Planet Sci Lett* 2011;**303**:121–32.

Crevecoeur S, Ruiz-González C, Prairie YT *et al.* Large-scale biogeography and environmental regulation of methanotrophic bacteria across boreal inland waters. *Mol Ecol* 2019;**28**:4181–96.

Croal LR, Jiao Y, Kappler A *et al.* Phototrophic Fe(II) oxidation in an atmosphere of H_2: implications for Archean banded iron formations. *Geobiology* 2009;**7**:21–4.

Croal LR, Johnson CM, Beard BL *et al.* Iron isotope fractionation by Fe(II)-oxidizing photoautotrophic bacteria. *Geochim Cosmochim Acta* 2004;**68**:1227–42.

Crowe SA, Døssing LN, Beukes NJ *et al.* Atmospheric oxygenation three billion years ago. *Nature* 2013;**501**:535–8.

Crowe SA, Hahn AS, Morgan-Lang C *et al.* Draft genome sequence of the pelagic photoferrotroph *Chlorobium phaeoferrooxidans*. *Genome Announc* 2017;**5**:e01584-16.

Crowe SA, Jones C, Katsev S *et al.* Photoferrotrophs thrive in an Archean Ocean analogue. *Proc Natl Acad Sci USA* 2008;**105**:15938–43.

Crowe SA, Katsev S, Leslie K *et al.* The methane cycle in ferruginous Lake Matano. *Geobiology* 2011;**9**:61–78.

Crowe SA, Maresca JA, Jones C *et al.* Deep-water anoxygenic photosythesis in a ferruginous chemocline. *Geobiology* 2014a;**12**:322–39.

Crowe SA, Paris G, Katsev S *et al.* Sulfate was a trace constituent of Archean seawater. *Science* 2014b;**346**:735–9.

Crusius J, Anderson RF. Evaluating the mobility of ^{137}Cs, $^{239+240}$Pu and ^{210}Pb from their distributions in laminated lake sediments. *J Paleolimnol* 1995;**13**:119–41.

Curtis PJ, Schindler DW. Hydrologic control of dissolved organic matter in low-order Precambrian Shield lakes. *Biogeochemistry* 1997;**36**:125–38.

Czaja AD, Johnson CM, Beard BL *et al.* Biological Fe oxidation controlled deposition of banded iron formation in the ca. 3770 Ma Isua Supracrustal Belt (West Greenland). *Earth Planet Sc Lett* 2013;**363**:192–203.

Daye M, Klepac-Ceraj V, Pajusalu M *et al.* Light-driven anaerobic microbial oxidation of manganese. *Nature* 2019;**576**:311–4.

DeSantis TZ, Hugenholtz P, Larsen N *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* 2006;**72**:5069–72.

Diao M, Sinnige R, Kalbitz K *et al.* Succession of bacterial communities in a seasonally stratified lake with an anoxic and sulfidic hypolimnion. *Front Microbiol* 2017;**8**:2511.

Eddy SR. Accelerated profile HMM searches. *PLOS Comput Biol* 2011;**7**:e1002195.

Edgar RC. UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat Methods* 2013;**10**:996–8.

Ehrenreich A, Widdel F. Anaerobic oxidation of ferrous iron by purple bacteria, a new type of phototrophic metabolism. *Appl Environ Microbiol* 1994;**60**:4517–26.

Emerson D, Field E, Chertkov O *et al*. Comparative genomics of freshwater Fe-oxidizing bacteria: implications for physiology, ecology, and systematics. *Front Microbiol* 2013;**4**:254.

Emerson D, Moyer C. Isolation and characterization of novel iron-oxidizing bacteria that grow at circumneutral pH. *Appl Environ Microbiol* 1997;**63**:4784–92.

Emerson S, Hesslein R. Distribution and uptake of artificially introduced radium-226 in a small lake. *J Fish Res Bd Can* 1973;**30**:1485–90.

Eren AM, Esen ÖC, Quince C *et al.* Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 2015;**3**:e1319.

Ettwig KF, Butler MK, Le Paslier D *et al.* Nitrite-driven anaerobic methane oxidation by oxygenic bacteria. *Nature* 2010;**464**:543–8.

Ettwig KF, Zhu B, Speth D *et al.* Archaea catalyze iron-dependent anaerobic oxidation of methane. *Proc Natl Acad Sci USA* 2016;**113**:12792–6.

Fakhraee M, Hancisse O, Canfield DE *et al.* Proterozoic seawater sulfate scarcity and the evolution of ocean–atmosphere chemistry. *Nat Geosci* 2019;**12**:375.

Falagán C, Johnson DB. *Acidibacter ferrireducens* gen. nov., sp. nov.: an acidophilic ferric iron-reducing gammaproteobacterium. *Extremophiles* 2014;**18**:1067–73.

Field CB, Behrenfeld MJ, Randerson JT *et al.* Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* 1998;**281**:237–40.

Field EK, Kato S, Findlay AJ *et al.* Planktonic marine iron oxidizers drive iron mineralization under low-oxygen conditions. *Geobiology* 2016;**14**:499–508.

Findlay DL, Kasian SEM. Phytoplankton community responses to nutrient addition in Lake 226, Experimental Lakes Area, northwestern Ontario. *Can J Fish Aquat Sci* 1987;**44**:S35–46.

Finn RD, Coggill P, Eberhardt RY *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 2016;**44**:D279–85.

Finneran KT, Johnsen CV, Lovley DR. *Rhodoferax ferrireducens* sp. nov., a psychrotolerant, facultatively anaerobic bacterium that oxidizes acetate with the reduction of Fe(III). *Int J Syst Evol Micr* 2003;**53**:669–73.

Fischer WW, Hemp J, Johnson JE. Evolution of oxygenic photosynthesis. *Annu Rev Earth Pl Sc* 2016;**44**:647–83.

Fish JA, Chai B, Wang Q *et al.* FunGene: the functional gene pipeline and repository. *Front Microbiol* 2013;**4**:291.

Fleming EJ, Mack EE, Green PG *et al.* Mercury methylation from unexpected sources: molybdate-inhibited freshwater sediments and an iron-reducing bacterium. *Appl Environ Microbiol* 2006;**72**:457–64.

Frigaard N-U, Bryant DA. Seeing green bacteria in a new light: genomics-enabled studies of the photosynthetic apparatus in green sulfur bacteria and filamentous anoxygenic phototrophic bacteria. *Arch Microbiol* 2004;**182**:265–76.

Frigaard N-U, Bryant DA. Genomic insights into the sulfur metabolism of phototrophic green sulfur bacteria. In: Hell R, Dahl DC, Knaff D, et al. (eds.). *Sulfur Metabolism in Phototrophic Organisms*. Springer Netherlands, 2008, 337–55.

Gaisin VA, Burganskaya EI, Grouzdev DS *et al.* '*Candidatus* Viridilinea mediisalina', a novel phototrophic Chloroflexi bacterium from a Siberian soda lake. *FEMS Microbiol Lett* 2019;**366**:fnz043.

Garber AI, Nealson KH, Okamoto A *et al.* FeGenie: A comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. *Front Microbiol* 2020;**11**:37.

Gich F, Garcia-Gil J, Overmann J. Previously unknown and phylogenetically diverse members of the green nonsulfur bacteria are indigenous to freshwater lakes. *Arch Microbiol* 2001;**177**:1–10.

Gisriel C, Sarrou I, Ferlez B *et al.* Structure of a symmetric photosynthetic reaction center–photosystem. *Science* 2017;**357**:1021–5.

Glöckner FO, Yilmaz P, Quast C *et al.* 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J Biotechnol* 2017;**261**:169–76.

Grégoire DS, Poulain AJ. A physiological role for Hg^{II} during phototrophic growth. *Nat Geosci* 2016;**9**:121–5.

Griffin BM, Schott J, Schink B. Nitrite, an electron donor for anoxygenic photosynthesis. *Science* 2007;**316**:1870–1870.

van Grinsven S, Damsté JSS, Asbun AA *et al.* Methane oxidation in anoxic lake water stimulated by nitrate and sulfate addition. *Environ Microbiol* 2019;**22**:766–82.

van Grinsven S, Damsté JSS, Harrison J *et al.* Impact of electron acceptor availability on methane-influenced microorganisms in an enrichment culture obtained from a stratified lake. *Front Microbiol* 2020;**11**:715.

Gupta D, Sutherland MC, Rengasamy K *et al.* Photoferrotrophs produce a PioAB electron conduit for extracellular electron uptake. *mBio* 2019;**10**:e02668-19.

Guzman MS, Rengasamy K, Binkley MM *et al.* Phototrophic extracellular electron uptake is linked to carbon dioxide fixation in the bacterium *Rhodopseudomonas palustris*. *Nat Commun* 2019;**10**:1355.

Haft DH, Selengut JD, Richter RA *et al.* TIGRFAMs and Genome Properties in 2013. *Nucleic Acids Res* 2013;**41**:D387–95.

Hanada S. The phylum *Chloroflexi*, the family *Chloroflexaceae*, and the related phototrophic families *Oscillochloridaceae* and *Roseiflexaceae*. In: Rosenberg E, DeLong EF, Lory S, et al. (eds.). *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea*. Springer Berlin Heidelberg, 2014, 515–32.

Hanada S, Hiraishi A, Shimada K *et al. Chloroflexus aggregans* sp. nov., a filamentous phototrophic bacterium which forms dense cell aggregates by active gliding movement. *Int J Syst Evol Microbiol* 1995;**45**:676–81.

Hanada S, Takaichi S, Matsuura K *et al. Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. *Int J Syst Evol Microbiol* 2002;**52**:187–93.

Hanson TE, Luther GWI, Findlay A *et al.* Phototrophic sulfide oxidation: environmental insights and a method for kinetic analysis. *Front Microbiol* 2013;**4**:382.

Hashizume K, Pinti DL, Orberger B *et al.* A biological switch at the ocean surface as a cause of laminations in a Precambrian iron formation. *Earth Planet Sci Lett* 2016;**446**:27–36.

He S, Barco RA, Emerson D *et al.* Comparative genomic analysis of neutrophilic iron(II) oxidizer genomes for candidate genes in extracellular electron transfer. *Front Microbiol* 2017;**8**:1584.

He S, Lau MP, Linz AM *et al.* Extracellular electron transfer may be an overlooked contribution to pelagic respiration in humic-rich freshwater lakes. *mSphere* 2019;**4**:e00436-18.

Hegler F, Posth NR, Jiang J *et al.* Physiology of phototrophic iron(II)-oxidizing bacteria: implications for modern and ancient environments. *FEMS Microbiol Ecol* 2008;**66**:250–60.

Heising S, Richter L, Ludwig W *et al. Chlorobium ferrooxidans* sp. nov., a phototrophic green sulfur bacterium that oxidizes ferrous iron in coculture with a "*Geospirillum*" sp. strain. *Arch Microbiol* 1999;**172**:116–24.

Heyes A, Bell J. Sulfide analysis using ion specific electrode (with preservation in sulfide anti-oxidant buffer) Appendix D, *Standard operating procedures*. *Academy of Natural Sciences, St Leonard, MD* 1999.

Hoang DT, Chernomor O, von Haeseler A *et al.* UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;**35**:518–22.

Hohmann-Marriott MF, Blankenship RE. Hypothesis on chlorosome biogenesis in green photosynthetic bacteria. *FEBS Lett* 2007;**581**:800–3.

Hohmann-Marriott MF, Blankenship RE. Evolution of photosynthesis. *Annu Rev Plant Biol* 2011;**62**:515–48.

Huang S, Jaffé PR. Characterization of incubation experiments and development of an enrichment culture capable of ammonium oxidation under iron-reducing conditions. *Biogeosciences* 2015;**12**:769–79.

Huang S, Jaffé PR. Isolation and characterization of an ammonium-oxidizing iron reducer: *Acidimicrobiaceae* sp. A6. *PLOS ONE* 2018;**13**:e0194007.

Hug LA, Baker BJ, Anantharaman K *et al.* A new view of the tree of life. *Nat Microbiol* 2016;**1**:16048.

Huson DH, Scornavacca C. Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst Biol* 2012;**61**:1061–7.

Hyatt D, Chen G-L, LoCascio PF *et al.* Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform* 2010;**11**:1–11.

Imhoff JF. The family *Chromatiaceae*. In: Rosenberg E, DeLong EF, Lory S, et al. (eds.). *The Prokaryotes*. Springer Berlin Heidelberg, 2014a, 151–78.

Imhoff JF. The family *Chlorobiaceae*. In: Rosenberg E, DeLong EF, Lory S, et al. (eds.). *The Prokaryotes*. Springer Berlin Heidelberg, 2014b, 501–14.

Jain C, Rodriguez-R LM, Phillippy AM *et al.* High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;**9**:5114.

Jeans C, Singer SW, Chan CS *et al.* Cytochrome 572 is a conspicuous membrane protein with iron oxidation activity purified directly from a natural acidophilic microbial community. *ISME J* 2008;**2**:542–50.

Ji M, Greening C, Vanwonterghem I *et al.* Atmospheric trace gases support primary production in Antarctic desert surface soil. *Nature* 2017;**552**:400–3.

Jiao Y, Kappler A, Croal LR *et al.* Isolation and characterization of a genetically tractable photoautotrophic Fe(II)-oxidizing bacterium, *Rhodopseudomonas palustris* strain TIE-1. *Appl Environ Microbiol* 2005;**71**:4487–96.

Johnson CM, Beard BL, Klein C *et al.* Iron isotopes constrain biologic and abiologic processes in banded iron formation genesis. *Geochim Cosmochim Acta* 2008;**72**:151–69.

Johnson JE, Webb SM, Thomas K *et al.* Manganese-oxidizing photosynthesis before the rise of cyanobacteria. *Proc Natl Acad Sci USA* 2013;**110**:11238–43.

Johnston DT, Gill BC, Masterson A *et al.* Placing an upper limit on cryptic marine sulphur cycling. *Nature* 2014;**513**:530–3.

Jones C, Crowe SA. No evidence for manganese-oxidizing photosynthesis. *Proc Natl Acad Sci USA* 2013;**110**:E4118.

Kadnikov VV, Savvichev AS, Mardanov AV *et al.* Microbial communities involved in the methane cycle in the near-bottom water layer and sediments of the meromictic subarctic Lake Svetloe. *Antonie van Leeuwenhoek* 2019;**112**:1801–1814.

Kalyaanamoorthy S, Minh BQ, Wong TKF *et al.* ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;**14**:587–9.

Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* 2016;**428**:726–31.

Kang DD, Li F, Kirton E *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019;**7**:e7359.

Kantor RS, Zyl AW van, Hille RP van *et al.* Bioreactor microbial ecosystems for thiocyanate and cyanide degradation unravelled with genome-resolved metagenomics. *Environ Microbiol* 2015;**17**:4929–41.

Kappler A, Bryce C. Cryptic biogeochemical cycles: unravelling hidden redox reactions. *Environ Microbiol* 2017;**19**:842–6.

Kappler A, Pasquero C, Konhauser KO *et al.* Deposition of banded iron formations by anoxygenic phototrophic Fe(II)-oxidizing bacteria. *Geology* 2005;**33**:865–8.

Karhunen J, Arvola L, Peura S *et al.* Green sulphur bacteria as a component of the photosynthetic plankton community in small dimictic humic lakes with an anoxic hypolimnion. *Aquat Microb Ecol* 2013;**68**:267–72.

Kasalický V, Zeng Y, Piwosz K *et al.* Aerobic anoxygenic photosynthesis is commonly present within the genus *Limnohabitans*. *Appl Environ Microbiol* 2018;**84**:e02116-17.

Kato S, Ohkuma M, Powell DH *et al.* Comparative genomic insights into ecophysiology of neutrophilic, microaerophilic iron oxidizing bacteria. *Front Microbiol* 2015;**6**:1265.

Keller W. Implications of climate warming for Boreal Shield lakes: a review and synthesis. *Environ Rev* 2007;**15**:99–112.

Kendall B, Creaser RA, Reinhard CT *et al.* Transient episodes of mild environmental oxygenation and oxidative continental weathering during the late Archean. *Sci Adv* 2015;**1**:e1500777.

Kennedy K, Hall MW, Lynch MDJ *et al.* Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Appl Environ Microbiol* 2014;**80**:5717–22.

Keppen OI, Baulina OI, Kondratieva EN. *Oscillochloris trichoides* neotype strain DG-6. *Photosynth Res* 1994;**41**:29–33.

Kerin EJ, Gilmour CC, Roden E *et al.* Mercury methylation by dissimilatory iron-reducing bacteria. *Appl Environ Microbiol* 2006;**72**:7919–21.

Kieser S, Brown J, Zdobnov EM *et al.* ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinform* 2020;**21**:257.

Koblížek M. Ecology of aerobic anoxygenic phototrophs in aquatic environments. *FEMS Microbiol Rev* 2015;**39**:854–70.

Koeksoy E, Halama M, Konhauser KO *et al.* Using modern ferruginous habitats to interpret Precambrian banded iron formation deposition. *Int J Astrobiol* 2016;**15**:205–17.

Konhauser KO. *Introduction to Geomicrobiology*. Malden, MA, USA: Blackwell Publishing, 2009.

Konhauser KO, Planavsky NJ, Hardisty DS *et al.* Iron formations: A global record of Neoarchaean to Palaeoproterozoic environmental history. *Earth-Science Reviews* 2017;**172**:140–77.

Kortelainen P, Pajunen H, Rantakari M *et al.* A large carbon pool and small sink in boreal Holocene lake sediments. *Glob Change Biol* 2004;**10**:1648–53.

Kuever J. The family *Desulfobacteraceae*. In: Rosenberg E, DeLong EF, Lory S, et al. (eds.). *The Prokaryotes: Deltaproteobacteria and Epsilonproteobacteria*. Springer Berlin Heidelberg, 2014, 45–73.

Kulp TR, Hoeft SE, Asao M *et al.* Arsenic(III) fuels anoxygenic photosynthesis in hot spring biofilms from Mono Lake, California. *Science* 2008;**321**:967–70.

Lambrecht N. Insights into early Earth ocean biogeochemistry from intensive monitoring of two ferruginous meromictic lakes. 2019. PhD thesis. Iowa State University, Iowa, USA, 2019.

Lambrecht N, Katsev S, Wittkop C *et al.* Biogeochemical and physical controls on methane fluxes from two ferruginous meromictic lakes. *Geobiology* 2019;**0**:1–16.

Laufer K, Niemeyer A, Nikeleit V *et al.* Physiological characterization of a halotolerant anoxygenic phototrophic Fe(II)-oxidizing green-sulfur bacterium isolated from a marine sediment. *FEMS Microbiol Ecol* 2017;**93**:fix054.

Leang C, Coppi MV, Lovley DR. OmcB, a c-type polyheme cytochrome, involved in Fe(III) reduction in *Geobacter sulfurreducens*. *J Bacteriol* 2003;**185**:2096–103.

Lee MD. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 2019;**35**:4162–4164.

Li D, Liu C-M, Luo R *et al.* MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;**31**:1674–6.

Li X, Huang Y, Liu H *et al.* Simultaneous Fe(III) reduction and ammonia oxidation process in Anammox sludge. *J Environ Sci* 2018;**64**:42–50.

Liu K, Wu L, Couture R-M *et al.* Iron isotope fractionation in sediments of an oligotrophic freshwater lake. *Earth Planet Sc Lett* 2015;**423**:164–72.

Llirós M, García–Armisen T, Darchambeau F *et al.* Pelagic photoferrotrophy and iron cycling in a modern ferruginous basin. *Sci Rep* 2015;**5**:13803.

Louca S, Parfrey LW, Doebeli M. Decoupling function and taxonomy in the global ocean microbiome. *Science* 2016;**353**:1272–7.

Lovley DR. Dissimilatory Fe(III) and Mn(IV) reduction. *Microbiol Mol Biol Rev* 1991;**55**:259–87.

Lynch MD, Masella AP, Hall MW *et al.* AXIOME: automated exploration of microbial diversity. *GigaScience* 2013;**2**:3–3.

Lyons TW, Fike DA, Zerkle A. Emerging biogeochemical views of Earth's ancient microbial worlds. *Elements* 2015;**11**:415–21.

Martin WF, Bryant DA, Beatty JT. A physiological perspective on the origin and evolution of photosynthesis. *FEMS Microbiol Rev* 2018;**42**:205–31.

Martinez-Cruz K, Leewis M-C, Herriott IC *et al.* Anaerobic oxidation of methane by aerobic methanotrophs in sub-Arctic lake sediments. *Sci Total Environ* 2017;**607**:23–31.

Masella AP, Bartram AK, Truszkowski JM *et al.* PANDAseq: paired-end assembler for Illumina sequences. *BMC Bioinform* 2012;**13**:1–7.

Mayr MJ, Zimmermann M, Guggenheim C *et al.* Niche partitioning of methane-oxidizing bacteria along the oxygen–methane counter gradient of stratified lakes. *ISME J* 2019:1–14.

McAllister SM, Polson SW, Butterfield DA *et al.* Validating the Cyc2 neutrophilic iron oxidation pathway using meta-omics of Zetaproteobacteria iron mats at marine hydrothermal vents. *mSystems* 2020;**5**:e00553-19.

Melton ED, Schmidt C, Kappler A. Microbial iron(II) oxidation in littoral freshwater lake sediment: the potential for competition between phototrophic vs. nitrate-reducing iron(II)-oxidizers. *Front Microbiol* 2012;**3**:197.

Melton ED, Swanner ED, Behrens S *et al.* The interplay of microbially mediated and abiotic reactions in the biogeochemical Fe cycle. *Nat Rev Micro* 2014;**12**:797–808.

Mendler K, Chen H, Parks DH *et al.* AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res* 2019;**47**:4442–8.

Menzel P, Ng KL, Krogh A. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 2016;**7**:1–9.

Methé BA, Nelson KE, Eisen JA *et al.* Genome of *Geobacter sulfurreducens*: metal reduction in subsurface environments. *Science* 2003;**302**:1967–9.

Milucka J, Kirf M, Lu L *et al.* Methane oxidation coupled to oxygenic photosynthesis in anoxic waters. *ISME J* 2015;**9**:1991–2002.

Miot J, Jézéquel D, Benzerara K *et al.* Mineralogical diversity in Lake Pavin: connections with water column chemistry and biomineralization processes. *Minerals* 2016;**6**:24.

Molot LA, Watson SB, Creed IF *et al.* A novel model for cyanobacteria bloom formation: the critical role of anoxia and ferrous iron. *Freshw Biol* 2014;**59**:1323–40.

Morana C, Roland FAE, Crowe SA *et al.* Chemoautotrophy and anoxygenic photosynthesis within the water column of a large meromictic tropical lake (Lake Kivu, East Africa). *Limnol Oceanogr* 2016;**61**:1424–37.

Mori Y, Purdy KJ, Oakley BB *et al.* Comprehensive detection of phototrophic sulfur bacteria using PCR primers that target reverse dissimilatory sulfite reductase gene. *Microbes Environ* 2010;**25**:190–6.

Nguyen L-T, Schmidt HA, von Haeseler A *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**:268–74.

Nurk S, Meleshko D, Korobeynikov A *et al.* metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 2017;**27**:824–34.

O'Leary NA, Wright MW, Brister JR *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 2016;**44**:D733–45.

Olm MR, Brown CT, Brooks B *et al.* dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 2017;**11**:2864–8.

Olson JM. The FMO protein. In: Govindjee, Beatty JT, Gest H, et al. (eds.). *Discoveries in Photosynthesis*. Dordrecht: Springer Netherlands, 2005, 421–7.

Olson JM, Blankenship RE. Thinking about the evolution of photosynthesis. In: Govindjee, Beatty JT, Gest H, et al. (eds.). *Discoveries in Photosynthesis*. Dordrecht: Springer Netherlands, 2005, 1073–86.

Oswald K, Jegge C, Tischer J *et al.* Methanotrophy under versatile conditions in the water column of the ferruginous meromictic Lake La Cruz (Spain). *Front Microbiol* 2016a;**7**:1762.

Oswald K, Milucka J, Brand A *et al.* Light-dependent aerobic methane oxidation reduces methane emissions from seasonally stratified lakes. *PLoS ONE* 2015;**10**:e0132574.

Oswald K, Milucka J, Brand A *et al.* Aerobic gammaproteobacterial methanotrophs mitigate methane emissions from oxic and anoxic lake waters. *Limnol Oceanogr* 2016b;**61**:S101–18.

Overmann J, Coolen MJL, Tuschak C. Specific detection of different phylogenetic groups of chemocline bacteria based on PCR and denaturing gradient gel electrophoresis of 16S rRNA gene fragments. *Arch Microbiol* 1999;**172**:83–94.

Overmann J, Garcia-Pichel F. The phototrophic way of life. In: Rosenberg E, DeLong EF, Lory S, et al. (eds.). *The Prokaryotes: Prokaryotic Communities and Ecophysiology*. Springer Berlin Heidelberg, 2013, 203–57.

Ozaki K, Tajika E, Hong PK *et al.* Effects of primitive photosynthesis on Earth's early climate system. *Nat Geosci* 2018;**11**:55–9.

Ozaki K, Thompson KJ, Simister RL *et al.* Anoxygenic photosynthesis and the delayed oxygenation of Earth's atmosphere. *Nat Commun* 2019;**10**:3026.

Parada AE, Needham DM, Fuhrman JA. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 2016:1403–14.

Parks DH, Chuvochina M, Waite DW *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;**36**:996–1004.

Parks DH, Imelfort M, Skennerton CT *et al.* CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;**25**:1043–55.

Pedersen MØ, Linnanto J, Frigaard N-U *et al.* A model of the protein–pigment baseplate complex in chlorosomes of photosynthetic green bacteria. *Photosynth Res* 2010;**104**:233–43.

Pedersen MØ, Underhaug J, Dittmer J *et al.* The three-dimensional structure of CsmA: A small antenna protein from the green sulfur bacterium *Chlorobium tepidum*. *FEBS Lett* 2008;**582**:2869–74.

Pedregosa F, Varoquaux G, Gramfort A *et al.* Scikit-learn: machine learning in Python. *J Mach Learn Res* 2011;**12**:2825–30.

Peng C, Bryce C, Sundman A *et al.* Organic matter complexation promotes Fe(II) oxidation by the photoautotrophic Fe(II)-oxidizer *Rhodopseudomonas palustris* TIE-1. *ACS Earth Space Chem* 2019;**3**:531–6.

Petersen TN, Brunak S, von Heijne G *et al.* SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 2011;**8**:785–6.

Petrenko P, Lobb B, Kurtz DA *et al.* MetAnnotate: function-specific taxonomic profiling and comparison of metagenomes. *BMC Biol* 2015;**13**:1–8.

Pettersen EF, Goddard TD, Huang CC *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004;**25**:1605–12.

Peura S, Buck M, Aalto SL *et al.* Novel autotrophic organisms contribute significantly to the internal carbon cycling potential of a boreal lake. *mBio* 2018;**9**:e00916-18.

Pfennig N. *Rhodocyclus purpureus* gen. nov. and sp. nov., a ring-shaped, vitamin B12-requiring member of the family *Rhodospirillaceae*. *Int J Syst Evol Microbiol* 1978;**28**:283–8.

Pierson BK, Castenholz RW. A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. *Arch Microbiol* 1974;**100**:5–24.

Pinti DL, Hashizume K. Early life record from nitrogen isotopes. In: Golding SD, Glikson M (eds.). *Earliest Life on Earth: Habitats, Environments and Methods of Detection*. Dordrecht, The Netherlands: Springer, 2011, 183–205.

Planavsky N, Rouxel O, Bekker A *et al.* Iron-oxidizing microbial ecosystems thrived in late Paleoproterozoic redox-stratified oceans. *Earth Planet Sci Lett* 2009;**286**:230–42.

Posth NR, Bristow LA, Cox RP *et al.* Carbon isotope fractionation by anoxygenic phototrophic bacteria in euxinic Lake Cadagno. *Geobiology* 2017;**15**:798–816.

Posth NR, Konhauser KO, Kappler A. Microbiological processes in banded iron formation deposition. *Sedimentology* 2013;**60**:1733–54.

Poulton SW, Canfield DE. Ferruginous conditions: a dominant feature of the ocean through Earth's history. *Elements* 2011;**7**:107–12.

Pruesse E, Peplies J, Gloeckner FO. SINA: Accurate high-throughput multiple sequence alignment of ribosomal RNA genes. *Bioinformatics* 2012;**28**:1823–9.

Quast C, Pruesse E, Yilmaz P *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* 2013;**41**:D590–6.

Quay PD, Broecker WS, Hesslein RH *et al.* Vertical diffusion rates determined by tritium tracer experiments in the thermocline and hypolimnion of two lakes. *Limnol Oceanogr* 1980;**25**:201–18.

Quince C, Lanzen A, Davenport RJ *et al.* Removing noise from pyrosequenced amplicons. *BMC Bioinform* 2011;**12**:38.

Raiswell R, Canfield DE. The iron biogeochemical cycle past and present. *Geochem Perspect* 2012;**1**:1–2.

Ramírez-Flandes S, González B, Ulloa O. Redox traits characterize the organization of global microbial communities. *Proc Natl Acad Sci USA* 2019;**116**:3630–5.

Raven JA. Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat Microb Ecol* 2009;**56**:177–92.

Raymond J, Siefert JL, Staples CR *et al.* The natural history of nitrogen fixation. *Mol Biol Evol* 2004;**21**:541–54.

Rissman AI, Mau B, Biehl BS *et al.* Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* 2009;**25**:2071–3.

Roland FAE, Darchambeau F, Morana C *et al.* Anaerobic methane oxidation in an East African great lake (Lake Kivu). *Biogeosciences Discuss* 2016:1–27.

Sadekar S, Raymond J, Blankenship RE. Conservation of distantly related membrane proteins: photosynthetic reaction centers share a common structural core. *Mol Biol Evol* 2006;**23**:2001–7.

Savvichev AS, Babenko VV, Lunina ON *et al.* Sharp water column stratification with an extremely dense microbial population in a small meromictic lake, Trekhtzvetnoe. *Environ Microbiol* 2018;**20**:3784–97.

Savvichev AS, Kokryatskaya NM, Zabelina SA *et al.* Microbial processes of the carbon and sulfur cycles in an ice-covered, iron-rich meromictic lake Svetloe (Arkhangelsk region, Russia). *Environ Microbiol* 2017;**19**:659–72.

Schanz F. Vertical light attenuation and phytoplankton development in Lake Zürich. *Limnol Oceanogr* 1985;**30**:299-310.

Schiff SL, Tsuji JM, Wu L *et al.* Millions of Boreal Shield lakes can be used to probe Archaean Ocean biogeochemistry. *Sci Rep* 2017;**7**:46708.

Schindler D. The coupling of elemental cycles by organisms: evidence from whole-lake chemical perturbations. In: Stumm W (ed.). *Chemical Processes in Lakes*. New York, NY: John Wiley and Sons, 1985, 225–50.

Schindler DW. Light, temperature, and oxygen regimes of selected lakes in the Experimental Lakes Area, northwestern Ontario. *J Fish Res Bd Can* 1971;**28**:157–69.

Schindler DW. Eutrophication and recovery in experimental lakes: implications for lake management. *Science* 1974;**184**:897–9.

Schindler DW. The dilemma of controlling cultural eutrophication of lakes. *Proc R Soc B* 2012;**279**:4322–33.

Schindler DW, Armstrong FAJ, Holmgren SK *et al.* Eutrophication of Lake 227, Experimental Lakes Area, northwestern Ontario, by addition of phosphate and nitrate. *J Fish Res Bd Can* 1971;**28**:1763–82.

Schindler DW, Fee EJ. Experimental Lakes Area: whole-lake experiments in eutrophication. *J Fish Res Bd Can* 1974;**31**:937–53.

Schindler DW, Hecky RE, Findlay DL *et al.* Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. *Proc Natl Acad Sci USA* 2008;**105**:11254–8.

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;**30**:2068–9.

Shih PM, Ward LM, Fischer WW. Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the *Chloroflexi*. *Proc Natl Acad Sci USA* 2017;**114**:10749–54.

Sievers F, Wilm A, Dineen D *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;**7**:539.

Sinclair L, Peura S, Hernandez P *et al.* Novel chemolithotrophic and anoxygenic phototrophic genomes extracted from ice-covered boreal lakes. *bioRxiv* 2017:139212.

Singh RG, Tanca A, Palomba A *et al.* Unipept 4.0: functional analysis of metaproteome data. *J Proteome Res* 2018;**18**:606–15.

Soo RM, Hemp J, Hugenholtz P. Evolution of photosynthesis and aerobic respiration in the cyanobacteria. *Free Radical Bio Med* 2019;**140**:200–5.

Spasov E, Tsuji JM, Hug LA *et al.* High functional diversity among *Nitrospira* populations that dominate rotating biological contactor microbial communities in a municipal wastewater treatment plant. *ISME J* 2020;**14**:1857–72.

Spoelstra J, Murray M, Elgood R. *A simplified diffusion method for $\delta^{15}N$ analysis of dissolved ammonium*. National Water Research Institute, Report Number 11-038. Environment Canada. 2011.

Stainton MP, Capel MJ, Armstrong FA. *The Chemical Analysis of Freshwater*. 2nd Edition. Can Fish Mar Serv Misc Spec Publ, 1977.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;**30**:1312–3.

Stookey LL. Ferrozine - a new spectrophotometric reagent for iron. *Anal Chem* 1970;**42**:779–81.

Straub KL, Benz M, Schink B *et al.* Anaerobic, nitrate-dependent microbial oxidation of ferrous iron. *Appl Environ Microbiol* 1996;**62**:1458–60.

Tabita FR, Hanson TE, Satagopan S *et al.* Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Phil Trans R Soc B* 2008;**363**:2629–40.

Taipale S, Kankaala P, Hahn MW *et al.* Methane-oxidizing and photoautotrophic bacteria are major producers in a humic lake with a large anoxic hypolimnion. *Aquat Microb Ecol* 2011;**64**:81–95.

Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007;**56**:564–77.

Tang K-H, Barry K, Chertkov O *et al.* Complete genome sequence of the filamentous anoxygenic phototrophic bacterium *Chloroflexus aurantiacus*. *BMC Genomics* 2011;**12**:334.

Tang K-H, Tang YJ, Blankenship RE. Carbon metabolic pathways in phototrophic bacteria and their broader evolutionary implications. *Front Microbiol* 2011;**2**:165.

Tank M, Bryant DA. Nutrient requirements and growth physiology of the photoheterotrophic Acidobacterium, *Chloracidobacterium thermophilum*. *Front Microbiol* 2015;**6**:226.

Tank M, Thiel V, Ward DM *et al.* A panoply of phototrophs: an overview of the thermophilic chlorophototrophs of the microbial mats of alkaline siliceous hot springs in Yellowstone National Park, WY, USA. In: Hallenbeck PC (ed.). *Modern Topics in the Phototrophic Prokaryotes: Environmental and Applied Aspects*. Springer International Publishing Switzerland, 2017, 87–137.

Temple KL, Colmer AR. The autotrophic oxidation of iron by a new bacterium: *Thiobacillus ferrooxidans*. *J Bacteriol* 1951;**62**:605–11.

Teodoru CR, del Giorgio PA, Prairie YT *et al.* Patterns in pCO_2 in boreal streams and rivers of northern Quebec, Canada. *Global Biogeochem Cycles* 2009;**23**:GB2012.

Thiel V, Tank M, Bryant DA. Diversity of chlorophototrophic bacteria revealed in the omics era. *Annu Rev Plant Biol* 2018;**69**:21–49.

Thomazo C, Pinti DL, Busigny V *et al.* Biological activity and the Earth's surface evolution: Insights from carbon, sulfur, nitrogen and iron stable isotopes in the rock record. *C R Palevol* 2009;**8**:665–78.

Thompson KJ, Kenward PA, Bauer KW *et al.* Photoferrotrophy, deposition of banded iron formations, and methane production in Archean oceans. *Sci Adv* 2019;**5**:eaav2869.

Thompson KJ, Simister RL, Hahn AS *et al.* Nutrient acquisition and the metabolic potential of photoferrotrophic *Chlorobi*. *Front Microbiol* 2017;**8**:1212.

Thottathil SD, Reis PCJ, Prairie YT. Methane oxidation kinetics in northern freshwater lakes. *Biogeochemistry* 2019;**143**:105–116.

Tourova TP, Kovaleva OL, Gorlenko VM *et al.* Use of genes of carbon metabolism enzymes as molecular markers of *Chlorobi* phylum representatives. *Microbiology* 2013;**82**:784–93.

Tourova TP, Spiridonova EM, Slobodova NV *et al.* Phylogeny of anoxygenic filamentous phototrophic bacteria of the family *Oscillochloridaceae* as inferred from comparative analyses of the *rrs*, *cbbL*, and *nifH* genes. *Microbiology* 2006;**75**:192–200.

Tran P, Ramachandran A, Khawasik O *et al.* Microbial life under ice: Metagenome diversity and *in situ* activity of Verrucomicrobia in seasonally ice-covered lakes. *Environ Microbiol* 2018;**20**:2568–84.

Vandermeulen JH, Davis ND, Muscatine L. The effect of inhibitors of photosynthesis on zooxanthellae in corals and other marine invertebrates. *Marine Biology* 1972;**16**:185–91.

Venkiteswaran JJ, Schiff SL, Wallin MB. Large carbon dioxide fluxes from headwater boreal and sub-boreal streams. *PLOS ONE* 2014;**9**:e101756.

Verpoorter C, Kutser T, Seekell DA *et al.* A global inventory of lakes based on high-resolution satellite imagery. *Geophys Res Lett* 2014;**41**:6396–402.

Verschoor MJ, Molot LA. A comparison of three colorimetric methods of ferrous and total reactive iron measurement in freshwaters. *Limnol Oceanogr Methods* 2013;**11**:113–25.

Viollier E, Inglett PW, Hunter K *et al.* The ferrozine method revisited: Fe(II)/Fe(III) determination in natural waters. *Appl Geochem* 2000;**15**:785–90.

Walter XA, Picazo A, Miracle MR *et al.* Phototrophic Fe(II)-oxidation in the chemocline of a ferruginous meromictic lake. *Front Microbiol* 2014;**5**:713.

Wang P, Benoit G. Modeling the biogeochemical role of photosynthetic sulfur bacteria in phosphorus cycling in a managed eutrophic lake. *Ecol Model* 2017;**361**:66–73.

Wang Q, Garrity GM, Tiedje JM *et al.* Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* 2007;**73**:5261–7.

Ward LM, Cardona T, Holland-Moritz H. Evolutionary implications of anoxygenic phototrophy in the bacterial phylum *Candidatus* Eremiobacterota (WPS-2). *Front Microbiol* 2019;**10**:1658.

Ward LM, Hemp J, Shih PM *et al.* Evolution of phototrophy in the *Chloroflexi* phylum driven by horizontal gene transfer. *Front Microbiol* 2018;**9**:260.

Watanabe T, Kojima H, Fukui M. Complete genomes of freshwater sulfur oxidizers *Sulfuricella denitrificans* skB26 and *Sulfuritalea hydrogenivorans* sk43H: genetic insights into the sulfur oxidation pathway of betaproteobacteria. *Syst Appl Microbiol* 2014;**37**:387–95.

Weber KA, Achenbach LA, Coates JD. Microorganisms pumping iron: anaerobic microbial iron oxidation and reduction. *Nat Rev Micro* 2006;**4**:752–64.

Welte CU. Revival of archaeal methane microbiology. *mSystems* 2018;**3**:e00181-17.

White III RA, Brown J, Colby S *et al.* ATLAS (Automatic Tool for Local Assembly Structures) - a comprehensive infrastructure for assembly, annotation, and genomic binning of metagenomic and metatranscriptomic data. *PeerJ Preprints* 2017;**5**:e2843v1.

Willems A. The family *Comamonadaceae*. In: Rosenberg E, DeLong EF, Lory S, et al. (eds.). *The Prokaryotes*. Springer Berlin Heidelberg, 2014, 777–851.

Wolfe B, Kling HJ, Brunskill GJ *et al.* Multiple dating of a freeze core from Lake 227, an experimentally fertilized lake with varved sediments. *Can J Fish Aquat Sci* 1994;**51**:2274–85.

Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 2016;**32**:605–7.

Yang J, Yan R, Roy A *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 2015;**12**:7–8.

Zeng Y, Feng F, Medová H *et al.* Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum *Gemmatimonadetes*. *Proc Natl Acad Sci USA* 2014;**111**:7795–800.

Zerkle AL, Claire MW, Domagal-Goldman SD *et al.* A bistable organic-rich atmosphere on the Neoarchaean Earth. *Nat Geosci* 2012;**5**:359–63.

Zhou J, Bruns MA, Tiedje JM. DNA recovery from soils of diverse composition. *Appl Environ Microbiol* 1996;**62**:316–22.

Appendix A

Millions of Boreal Shield lakes can be used to probe Archaean Ocean biogeochemistry[4]

Introduction

Ancient oceans on Earth were rich in iron, low in sulfur, and free of oxygen. Oxidation of Earth's early oceans and atmosphere, concurrent with the early evolution of life and onset of photosynthesis, has long fostered intense scientific debate. Bacterial photosynthesis remains important in modern systems but was key to early Earth oxidation. Only a few modern environments have been identified for whole ecosystem study because the physico-chemical conditions similar to those predicted for the Archaean Ocean are thought to be rare (Walter *et al.* 2014). Most evidence has been gleaned from the sedimentary rock record and laboratory simulations (reviewed by Koeksoy *et al.* 2016), spurred by controversy surrounding the origin of the globally ubiquitous and extensive banded iron formations (BIFs).

In a pioneering whole ecosystem study in Lake Matano, Indonesia, large populations of photosynthetic green sulfur bacteria (GSB) were found just below the permanent chemocline at 120 metres depth, including a close relative of *Chlorobium ferrooxidans*, a known photoferrotroph (Crowe *et al.* 2008). This bacterium uses light to fix inorganic carbon, with reduced iron (Fe^{2+}) as the electron donor, thereby producing oxidized iron:

$$4Fe^{2+} + HCO_3^- + 10H_2O \xrightarrow{hv} CH_2O + 4Fe(OH)_3 + H^+$$
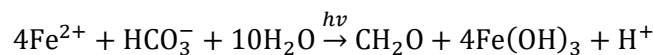
Photoferrotrophy has been proposed as the earliest photosynthetic process in Earth's history, predating oxygenic photosynthesis by cyanobacteria (Konhauser 2009). Photoferrotrophy could be responsible for a large part of early Earth oxidation leading to the mixed iron oxidation states that have proven difficult to explain in globally occurring BIFs, deposited when oxygen was still absent from the atmosphere (Kappler *et al.* 2005; Konhauser 2009). Since the initial discovery at Lake Matano, only two other ferruginous open water sites – Lac Cruz, Spain (Walter *et al.* 2014) and Kabuno Bay, a sub-basin of Lake Kivu, east Africa (Llirós *et al.* 2015) – have been identified that host microbial communities dominated by photoferrotrophs. Very recently, the first metabolic rate

---

measurements have been reported in the chemoclines of Lakes Kivu and La Cruz (Walter *et al.* 2014; Llirós *et al.* 2015; Morana *et al.* 2016), confirming photoferrotrophic activity and its biogeochemical importance by fueling microbial iron reduction and possibly co-existing pelagic heterotrophy. No other examples of such microbial consortia in natural aquatic systems have yet been reported. All of these systems are meromictic and have low sulfate and high iron due to their origins as volcanic craters or rift lakes. Such systems are naturally rare worldwide (Llirós *et al.* 2015), with expectations that only a handful of appropriate lakes will be found (Koeksoy *et al.* 2016). This limits progress toward understanding Archaean Ocean biogeochemistry and the origins of life.

Almost one half of the largest terrestrial biome on Earth, the boreal forest, is underlain by Precambrian Shield geology (Venkiteswaran, Schiff and Wallin 2014) of low sulfur content. Millions of lakes cover over 7% of the Boreal Shield areas of Canada, Fennoscandinavia, and Russia (Kortelainen *et al.* 2004). Bottom portions of stratified lakes and shallow ponds become anoxic periodically in summer and under ice in winter if the supply of terrestrial or aquatic organic matter is sufficient and/or the lake morphometry restricts mixing. After the onset of anoxia, the naturally low-sulfur bottom waters of such lakes become rich in dissolved ferrous iron, matching key parameters used in the selection of other Archaean Ocean analogues (Walter *et al.* 2014). Here we test the hypothesis that the bottom anoxic layers of seasonally stratified lakes on the Boreal Shield could provide modern *in situ* laboratories for advancing the scientific understanding of microbial metabolic pathways in both the Archaean Ocean and in modern lake environments.

Methods

Experimental design and site description

To assess the potential for the use of Boreal Shield lakes to yield insight into microbial iron cycling processes, two exceptionally well characterized lakes (L227 and L442) were selected for detailed study over several years. These lakes are located at a long term scientific research site where comprehensive data on lakes, streams, climate, and hydrology have been collected for over 47 years, providing support for studies of shorter duration. Here, a suite of geochemical measurements, novel stable isotopic analysis, and nucleic sequencing provided complementary information on the occurrence of distinctive microbial consortia in these lakes.

The Experimental Lakes Area (ELA) is located in northwestern Ontario, Canada at 49°40' N, 93°45' W. Information on geology, vegetation, and climate is available (Brunskill and Schindler 1971). Lake 227 is a small headwater lake of 5 ha with a mean depth of 4.4 meters and maximum depth of 10 m. Lake 442 is a small lake of 16 ha with a mean depth of 9.6 meters and maximum depth of 17.8 m.

Sample collection and analysis

Water and lake sediment samples from Lake 227 have been collected from the beginning of the nutrient addition experiment in 1969 and continue today. For this study, additional water, sediment trap, and sediment samples were collected over several years. Multiple water column profiles were collected in Lake 227 in 2010 and 2011 during the summer stratified period from May to October. Single profiles were collected around the time of the cyanobacterial bloom that occurs each year in June to July in 2012, 2013 and 2014. Exact sampling dates and parameters are given in Figures 1-2 and Supplementary Figure A2.

Water samples were collected using a gear pump in a closed system from the desired depth. Samples for NO_3^-, NH_4^+, DOC, SO_4^{2-}, and total dissolved iron (TDFe) were filtered shortly after collection with 0.45 µm filters and analyzed using conventional methods (Stainton, Capel and Armstrong 1977). Sulfide was measured with an ion selective electrode (ISE) on unfiltered samples following stabilization in a sulfide anti-oxidant buffer (Heyes and Bell 1999). However, it is recognized that "free sulfide" is overestimated by an order of magnitude using the commonly used methylene blue method for dissolved samples due to the presence of "multiple reduced diffusible sulfur species" (Crowe *et al.* 2008). The contribution of dissolved, colloidal and particulate sulfur species to the response by the ISE in our unfiltered samples is unknown but most likely causes a substantial overestimation of "free sulfide" species. Nitrate (NO_3^-) is very low in both the epilimnion and hypolimnion for these lakes with mean values of < 3.2 µM in the summer stratified period. The pH of these lakes varies between 6.2 to 6.8 throughout the water column Water column light levels are routinely measured at ELA in the ice-free seasons using a LICOR flat plate quantum sensor (LI-192) with a LI-250 light meter. The sensor measures PAR wavelengths (400-700 nm) with a limit of detection of 0.01 µmol photons m^{-2} s^{-1}. As a result of light scattering and absorption associated with high phytoplankton densities in L227, PAR is strongly attenuated over depth and only a very small proportion of surface irradiance reaches the hypolimnion. Throughout the May-October period, PAR

values at 6 m depth in L227 were < 1.5 μmol photons m^{-2} s^{-1}. Detailed light profiles from L227 and L442 in June and September 2016 are presented in supplemental data (Supplementary Figure A2).

Samples for concentrations and isotopic analysis of DIC and CH_4 were collected directly without headspace into small glass serum bottles with stoppers and preserved with injections of concentrated HCl. Particulate organic matter (POM) for isotopic analysis was collected on Whatman QMA quartz filters with a nominal pore size of ~1 μm. Samples for microbial sequencing were collected by pumping water directly onto sterilized 0.22 μm Sterivex polyvinylidene fluoride filters (EMD Millipore). Filters were frozen until shipped back to the University of Waterloo and subsequently kept at -20°C or -80°C until processing.

Sediments were collected by freeze-coring (Crusius and Anderson 1995) at 7 and 10 meters during the winter ice cover period at L227. The entire sediment profile was analyzed for $\delta^{13}C$ and $\delta^{15}N$ at 1 cm intervals. Selected samples were analyzed for $\delta^{56}Fe$. Additional sediment samples were also collected by subsampling an Eckman dredge at 1, 4, 8, and 10 meters in Lake 227 and at 1, 5 11, 13, 15, and 17 meters in Lake 442. The top 1 cm was collected from only those dredges where the surface was clearly visible and surface structures were undisturbed. Sediment traps constructed of acrylic tubes with a width to depth ratio of > 8 were suspended at 2.0, 5.5, and 8.5 meters depth in the water column at three locations in the central part of L227.

Samples for $\delta^{13}C$ of DIC and CH_4 were prepared by headspace equilibration after acidification. For analysis of $\delta^{15}N$ of NH_4^+, samples were prepared using a modified diffusion technique (Spoelstra, Murray and Elgood 2011) with a precision of ±0.3‰ in $\delta^{15}N$. Both $\delta^{13}C$ of DIC and CH_4 and $\delta^{15}N$ of NH_4^+ were then analyzed by GC-CF-IRMS using an Agilent 6890 GC coupled to an Isochrom isotope ratio mass spectrometer (IRMS: Micromass UK) with precision +/- 0.3‰. $\delta^{13}C$, $\delta^{15}N$, and C/N of POM on filters, freeze-dried DOM, lake sediments, sediment trap samples were analyzed by EA-CF-IRMS using a Carlo Erba Elemental Analyzer (CHNS-O EA1108) coupled with a Delta Plus (Thermo) isotope ratio mass spectrometer with a precision of 0.2‰ in $\delta^{13}C$ and 0.3‰ in $\delta^{15}N$.

Samples of POM, in water and sediment samples were analyzed for $\delta^{56}Fe$ of Fe by MC-ICP-MS (Micromass Isoprobe) after purification using ion-exchange chromatography (Beard et $al.$ 2003) and reported relative to the average of igneous rocks ($\delta^{56}Fe$ = 0.0±0.05‰) with a precision of 0.03‰ (2σ). Sediment samples were digested with concentrated HF and HNO_3 and then dried before loading onto the resin. The measured Fe isotope composition of the IRMM-019 Fe isotope standard was

-0.08±0.05‰, which lies within error of the long-term value used in the lab of -0.09‰ relative to average igneous rocks (Beard *et al.* 2003).

Microbial community analysis

Genomic DNA was extracted using the PowerWater Sterivex DNA Isolation Kit (MoBio) and quantified using agarose gel electrophoresis. The V3-V4 region of the bacterial 16S rRNA gene was amplified from each sample using triplicate PCR amplifications. Each reaction contained ≤10 ng of sample DNA and used reagent, volumes, and thermocycler conditions described previously (Kennedy *et al.* 2014). Modified forward and reverse PCR primers 341f-808r were used for Illumina sequencing in a previously described configuration (Bartram *et al.* 2011). After combining triplicate reaction products to reduce bias, products for each sample were pooled at a normalized concentration, and the pooled library was gel purified using the Wizard SV Gel and PCR Clean-Up System (Promega) and spiked with 8.5-10% PhiX prior to sequencing. Paired-end (2x250 base) high-throughput DNA sequencing was carried out using the MiSeq platform (Illumina), achieving a cluster density of 452-507 K mm^{-2} with 92.9-98.0% of clusters passing filter. An average of ~600,000 raw reads per sample were generated for downstream analysis. Raw demultiplexed sequencing reads were processed using the AXIOME2 software tool, version 1.5 (Lynch *et al.* 2013). Using AXIOME2, paired sequences were assembled using PANDAseq version 2.8 (Masella *et al.* 2012), were chimera checked and clustered at 97% using USEARCH version 7.0.1090 (Edgar 2013), and were rarefied using QIIME version 1.9.0 (Caporaso *et al.* 2010).

To predict the major functional roles of the lake bacterial communities, the top ten most abundant bacterial OTUs within each water column sample, based on rarefied data, were evaluated for their potential involvement in iron or sulfur oxidation or reduction, and methanotrophy. Representative sequences for each abundant OTU were assigned taxonomic ranks using the RDP Naïve Bayesian rRNA Classifier, version 2.10, with a confidence threshold of 50%, based on the RDP 16S rRNA training set 14 (Wang *et al.* 2007; Cole *et al.* 2014). If the classifier could not assign a rank at the genus level, OTU representative sequences were queried against the NCBI non-redundant nucleotide database using BLASTN (Altschul *et al.* 1990). Cultured strains identified through the query with ≥97% sequence identity were used to assign, where possible, a single genus to the OTU. Subsequently, genera matched to each abundant OTU were checked against the literature for whether they contained strains with documented involvement in targeted metabolic activities.

(OTUs classified as chloroplasts were excluded from further analysis.) Any OTU classified within a genus that was found to contain a strain involved in one of the selected metabolic activities was inferred to have the same potential metabolic activity. Potential metabolic functions of OTUs that could only be classified to the family level were inferred similarly but with greater scrutiny, requiring that OTUs placed phylogenetically within a genus that matched the above criteria or that most to all genera in the family were implicated in the same functional role. Any remaining OTUs not classified to the family or genus level were not assigned a metabolic role. To check the validity of this manually curated method, results were compared to those obtained from the automated functional classifier FAPROTAX, script version 1.1 and database version 1.0, with default settings used for the output functional table (Louca, Parfrey and Doebeli 2016). For compatibility reasons, OTUs were reclassified using the RDP classifier trained against the Greengenes database (May 2013 release) prior to the FAPROTAX analysis (DeSantis *et al.* 2006).

High abundance OTUs classified within the family *Chlorobiaceae* (i.e., green sulfur bacteria) or *Comamonadaceae* were also examined phylogenetically to evaluate their potential functional roles in photoferrotrophy and iron reduction, respectively. Reference 16S rRNA gene sequences of cultured strains, along with appropriate outgroup sequences, were obtained from the Silva SSU-RefNR database, release 122, for *Chlorobiaceae*, and the All-Species Living Tree Project database, release 128, for *Comamonadaceae*, supplemented with additional SSU-RefNR type strain sequences known to have the function of interest (Quast *et al.* 2013). Due to the high diversity of the family *Comamonadaceae*, reference sequences were obtained from a monophyletic subset of genera within the family according to Willems (Willems 2014). Obtained sequences were aligned to OTU representative sequences using SINA version 1.2.11 (Pruesse, Peplies and Gloeckner 2012), and the alignment was then truncated to the V3-V4 region of the 16S rRNA gene. Using this alignment, a phylogeny was built using RAxML version 8.1.17 (Stamatakis 2014), using 100 maximum likelihood searches and the GTRCAT sequence evolution model. Node support values were calculated using the Shimodaira-Hasegawa test. The resulting phylogeny was visualized using Dendroscope version 3 (Huson and Scornavacca 2012).

Results

Geochemical and stable isotopic evidence

Boreal Shield lakes have been studied intensely at the Experimental Lakes Area (ELA) in northwestern Ontario, Canada. One of these small lakes, Lake 227 (L227), is the site of the world's longest running nutrient addition experiment, with amendments of nitrogen and phosphorus, or phosphorus alone, for over 47 years (Schindler *et al.* 2008). As a result of fertilization, summer phytoplankton biomass is high, resulting in a hypolimnion that is devoid of oxygen. However, the presence of over 50 cm of continuous annually varved sediments in the deepest part of the lake (10 m; Wolfe *et al.* 1994) indicates that the bottom of the hypolimnion has been naturally anoxic during lake stratification for over 300 years under natural nutrient loading conditions. Nearby Lake 442 (L442) has not been manipulated experimentally but the lower portion of the hypolimnion develops anoxia. L442 has physico-chemical features typical of natural lakes on the Boreal Shield. The anoxic zones of both L227 and L442 have low sulfate concentrations (5 to 21 µM) and high total dissolved iron (TDFe) concentrations (115 to 162 µM) over the summer sampling season (Figure A1), comparable to the levels reported in Lake Matano, Lake La Cruz, and Lake Pavin (Walter *et al.* 2014). Light penetration into the upper anoxic zones of both lakes is low but detectable, in the range of 0.01-0.03 and 0.09-1.64 µmol photons m^{-2} s^{-1} in L227 and L442, respectively (Supplementary Figure A1).

Hypolimnetic waters in L227 have distinctive patterns in natural abundance stable isotopes (Figure A2). These waters are high in iron, ammonium (NH_4^+), dissolved inorganic carbon (DIC), and methane (CH_4), yet low in sulfate (Figure A1). Stable carbon isotopes ($\delta^{13}C$) of particulate organic matter (POM) in the hypolimnion were offset from the overlying epilimnion where phytoplankton biomass was high (Figure A2a). In contrast, hypolimnetic sediments and sediment trap samples were similar in $\delta^{13}C$ to epilimnetic POM (Figure A2a), consistent with the high flux of organic carbon from the surface and indicating that the offset in $\delta^{13}C$ was not an effect of POM diagenesis during transit through the short water column to underlying lake sediments. Further, $\delta^{13}C$ of hypolimnetic DIC increased with depth (Figure A2a). Hypolimnetic POM of lower $\delta^{13}C$ than either epilimnetic POM or hypolimnetic DIC can only be attributed to isotopic fractionation associated with photosynthesis or assimilation of C with very negative $\delta^{13}C$, such as that measured in CH_4 (Figure A2a). Similarly, nitrogen isotopic composition ($\delta^{15}N$) of hypolimnetic POM was offset from both epilimnetic POM and hypolimnetic NH_4^+ (Figure A2b), consistent with isotopic fractionation during biological uptake of NH_4^+ and not N_2 fixation. Finally, the $\delta^{56}Fe$ of POM in the epilimnion and hypolimnion (Figure

A2c) also differed and the isotopic fractionation between dissolved and particulate phase was reversed from the epilimnion to hypolimnion. Rates of vertical mixing in the hypolimnion, previously determined in L227 using additions of ^{226}Ra and ^{3}H as tracers, are very low (Emerson and Hesslein 1973; Quay *et al.* 1980), similar to rates of molecular diffusion. These low mixing rates and the persistence over the stratification period mean that microbiota contributing to hypolimnetic POM are suspended in the water column at the observed depth. Together, these data imply that the microbial consortia in the anoxic zone are metabolically active and sufficiently abundant to alter the isotopic POM signatures of δ^{13}C, δ^{15}N, and δ^{56}Fe in POM from values typical of the epilimnion. Furthermore, the offset of the carbon and nitrogen stable isotopes in POM from inorganic dissolved substrates is consistent with fractionation associated with photosynthesis, despite extremely low light levels in the hypolimnion (Supplementary Figure A1) due to high summer epilimnetic biomass.

Microbial community composition

High-throughput sequencing of bacterial 16S rRNA genes from L227 water column samples followed by manual functional inferences of the most abundant taxa suggested that the hypolimnion of L227 supports a metabolically unique bacterial community (Figure A3). Throughout the L227 hypolimnion, we detected microbial communities that were dominated by potential iron-cycling bacteria, along with populations of potential sulfur- and methane-cycling bacteria, consistent with microbial consortia reported in Lake Kivu, where photoferrotrophic activity was found (Llirós *et al.* 2015). In particular, the 16S rRNA gene data showed a dominance of *Chlorobi*, closely related to photoferrotrophic *C. ferrooxidans*, among the most abundant operational taxonomic units (OTUs) in the anoxic zone. Indeed, sequences classified within the genus *Chlorobium* comprised up to 8% of all reads just below the oxic-anoxic interface. One particular *Chlorobium* OTU (Figure A3, *Chlorobium* OTU 1), which was the most abundant across all L227 samples, had over 99.5% 16S rRNA gene sequence identity with known *C. ferrooxidans* strains (Supplementary Figure A3). Putative methanotrophic OTUs belonging to the family *Methylococcaceae* of the *Gammaproteobacteria* were abundant in the anoxic zone, and potential sulfur oxidizers and reducers, such as *Sulfuritalea* and *Desulfatirhabdium* spp., were detected at lower abundance, implying relatively low contributions to overall lake metabolism. The hypoxic and anoxic zones of L227 were rich in bacterial taxa implicated in iron reduction, including members of the genera *Rhodoferax*, *Albidiferax*, and *Geothrix*.

In addition to manual functional annotations of 16S rRNA gene data, functional predictions were performed in parallel using FAPROTAX (Louca, Parfrey and Doebeli 2016). These automated functional assignments took into account the entire microbial community of each sample, rather than the most abundant taxa alone. Results largely agreed with manual inferences: iron oxidation ("anoxygenic phototrophy"), methanotrophy, and sulfur respiration increased in the anoxic zone with similar relative abundance trends compared to the manual approach (Supplementary Figure A4). Although discrepancy existed in the case of iron reduction, additional phylogenetic analysis confirmed that dominant OTUs implicated in iron reduction from the manual approach were closely related to known iron-reducing strains despite the functional diversity of the *Comamonadaceae* family to which they were classified (Supplementary Figure A5; (Finneran, Johnsen and Lovley 2003). High iron levels in the lower L227 water column and the position of these OTUs near where potentially photoferrotrophic *Chlorobi* were identified are also suggestive of an iron-reducing role. However, it is difficult to infer the functional roles of these microorganisms without further cultivation-based or metagenomic analyses.

Similar populations of *Chlorobium*, *Rhodoferax*, and *Methylobacter* (family *Methylococcaceae*), along with other potential sulfur and iron cyclers, were reported in a sub-basin of Lake Kivu, the only Archaean Ocean analogues whose whole microbial community has been described to date (Llirós *et al.* 2015). Evidence from targeted sequencing studies in other Archaean Ocean analogues suggests that a similar microbial consortium may be present. In Lake La Cruz, which is known to host photoferrotrophic *Chlorobi*, populations of methanotrophs belonging to the *Alpha*- and *Gammaproteobacteria* were recently identified in the anoxic zone using quantitative polymerase chain reaction (qPCR; Oswald *et al.* 2016a). Evidence of anaerobic methane oxidation, perhaps mediated by archaeal methane oxidizers, was found in Lake Matano near the same depth where large *Chlorobi* populations were detected (Crowe *et al.* 2011). Although the presence of photoferrotrophy has not been evaluated in Lake Pavin, sulfur-cycling microorganisms have been identified in the water column, along with populations of bacterial methanotrophs belonging to the genus *Methylobacter*, which have been identified in the lake anoxic zone (Biderre-Petit *et al.* 2011). These pelagic microbial consortia of potential iron-cycling bacteria, sulfur-cycling bacteria, and methanotrophs may be a shared trait of Archaean Ocean analogue systems. To our knowledge, such consortia have not been reported outside of these systems.

Natural abundance stable isotopes of iron

Key evidence of potential photoferrotrophy is provided by stable iron isotopes, a tool only very recently applied to freshwater lake iron cycling (Busigny *et al.* 2014; Liu *et al.* 2015), and not yet in Lakes Matano or Kivu. In the anoxic hypolimnion of L227, δ^{56}Fe in suspended particulates does not match any other particulate or dissolved pool in the water column or sediments (Figure A2c). Further, observed differences in δ^{56}Fe between dissolved and particulate iron are consistent in magnitude and direction with those of other ferrotrophs in laboratory cultures (Croal *et al.* 2004). Although we cannot rule out the possibility of isotopic separation due to partial iron oxidation at the oxycline, the exceedingly low iron concentrations above the oxycline, the low water mixing rates between the metalimnion and hypolimnion and the high abundance of bacteria that are likely able to reduce ferric iron below and just above the oxycline suggests that most or all accessible and re-oxidized iron should be reduced rapidly. This allows the hypolimnetic iron isotope signatures to be interpreted as the result of biological rather than chemical processes.

Our dissolved iron isotope composition profiles in the anoxic layer of L227 (and Lake 442) are distinct from those observed in Lake Pavin (Busigny *et al.* 2014), which is a permanently stratified lake with an anoxic bottom water layer and hypothesized as an Archaean ocean analogue. The observed gradient of dissolved δ^{56}Fe in Lake Pavin has been interpreted as resulting from partial oxidation of ferrous iron by O_2 at the redox boundary (Busigny *et al.* 2014). In addition, iron phosphates have been shown to be the predominant phase of the particulate matter in the anoxic monimolimnion (Miot *et al.* 2016). The total dissolved phosphorus (TDP) concentrations in both L227 and L442, up to 2 μM in the hypolimnion, are more than two orders of magnitude lower than that in Lake Pavin (up to 400 μM in the monimolimnion, Busigny *et al.* 2014). Further, soluble reactive phosphorus concentrations are at or below detection limits (data not shown) implying that much of the TDP is in organic form. Thus, the formation of iron phosphates does not contribute to the observed water column δ^{56}Fe in these lakes. In addition, photoferrotrophic bacteria have not yet been identified in association with minerals in the Lake Pavin water column (Miot *et al.* 2016). A lack of iron isotope compositions for the suspended particulates in Lake Pavin prevents further comparison between the two lake systems.

The similarity of δ^{56}Fe in the hypolimnetic dissolved phase, epilimnetic POM and sediments (Figure A2c), and the offset between hypolimnetic and epilimnetic POM (concomitant with δ^{13}C and δ^{15}N evidence), indicate that the δ^{56}Fe signature in hypolimnetic POM results from microorganisms

residing at that depth. Together, the microbial and geochemical data indicate that iron isotopes could be diagnostic of photoferrotropic activity in softwater lakes.

Evidence from other Boreal Shield lakes

Photoferrotrophic activity is not confined to experimentally eutrophied L227. Similar geochemistry and microbiota were identified in unperturbed L442. Unlike L227, where the hypolimnion is entirely anoxic, L442 has both an anoxic zone and overlying oxic portion within the hypolimnion. Given low hypolimnetic mixing rates, this feature allows the stable isotopic signatures to be interpreted as a direct result of anoxia at similar temperature and organic matter supply. Natural abundance stable isotopes of C and N (Figure A2) and water column chemistry (high Fe, DIC and NH_4^+, and low SO_4^{2-}; Figure A1) in the anoxic hypolimnion of L442 are typical of anoxic hypolimnia in other ELA lakes. Patterns in particulate and dissolved phase $\delta^{13}C$, $\delta^{15}N$, and $\delta^{56}Fe$ in L442 are all consistent with the anoxic hypolimnion of L227, and distinctly different from the oxic hypolimnion (Figure A2). Molecular sequencing in the anoxic zone confirms high abundance of *Chlorobi* that are closely related to *C. ferrooxidans*, in addition to a similar microbial consortium of putative iron reducers, sulfur reducers and oxidizers, and methanotrophs (Figure A3). Thus, even though boreal shield lakes are high in terrestrially derived organic carbon, unlike the Archaean Ocean, processes similar to the other reported Archaean analogues (e.g., Crowe *et al.* 2008; Walter *et al.* 2014) likely occur in Boreal Shield lakes with seasonally anoxic hypolimnia because the geology naturally leads to low sulfate and high iron conditions. Observations of both *Chlorobi* and methanotrophs have also been reported in anoxic hypolimnia in numerous Boreal Shield lakes in Finland (Taipale *et al.* 2011; Hanson *et al.* 2013; Karhunen *et al.* 2013). Further, in a recent study in Sweden, a shotgun metagenomics approach to the characterization of total DNA extraction was applied to a suite of Boreal lakes and indicated the presence of organisms having the ability to perform photoferrotrophy and anaerobic oxidation of methane (Sinclair *et al.* 2017). These new studies yield added support that these processes are widespread in Boreal Shield lakes.

Discussion

Collectively, our results provide first evidence for potential photoferrotrophy in Boreal Shield lakes, showing how distinct isotopic and chemical indicators can be used to prospect for corresponding microbial consortia. Although conditions on present day Earth cannot totally mimic conditions in the Archaean Ocean, Boreal Shield lakes and ponds are globally abundant and have metabolically active processes relevant to Archaean Ocean microbial life. Moreover, new metabolic pathways associated with potential photoferrotrophy will also alter current paradigms of biogeochemistry in modern lakes and reservoirs.

Important and novel to the debate surrounding evolution of life in the Archaean Ocean is that the microbial consortia associated with potential photoferrotrophy in lakes are robust and establish rapidly. Both L227 and L442 have two mixing periods per year, with fall turnover being the most complete. In wind protected L227, the water column is well oxygenated every year in both spring and fall to at least 6 m where *Chlorobium* spp. are abundant (Figure A3 and Supplementary Figure A6). In L442, mixing is complete in both spring and fall (Supplementary Figure A7). Isotopic and molecular analyses over several years show that these characteristic microbial consortia are re-established following oxygenation (Figs. 2 and 3). Although *Chlorobi* and methanotrophs have been found to re-establish in other lake hypolimnia following turnover (Hanson *et al.* 2013; Oswald *et al.* 2015), our study is the first to show that the microbial consortia associated with photoferrotrophy are sufficiently resilient to rapidly recolonize after periods of oxygenation. The potential for anoxygenic photosynthesis in iron-rich systems that undergo periodic oxygenation may be relevant to understanding the formation of banded iron formations, especially immediately prior to the Great Oxidation Event, where the relative roles of oxygenic and anoxygenic photosynthesis in iron oxidation and deposition remain controversial and oxygen-rich conditions were intermittent (Kendall *et al.* 2015). Further, a strain of the filamentous cyanobacterium *Aphanizomenon schindlerii* was also observed at 6.5 m (Molot *et al.* 2014), leaving open the possibility that a hypothesized photoferrotrophic precursor to oxygenic photosynthesis may still be operating today under conditions similar to early anoxic oceans. That potential photoferrotrophs were detected in several small and nondescript Boreal Shield lakes in northern Canada, and likely also in Finland and Sweden, implies that these bacteria are more widespread than previously thought and are quite resilient. In addition, limnological research conducted on Boreal lakes can offer new information to the Archaean ocean debate. Questions about how these microbial consortia can re-establish quickly following lake

149

overturn and how they adapt to rapidly changing gradients of light, sulfate, nitrate, and other parameters are fodder for future scientific research. Boreal Shield lakes provide new opportunities for whole ecosystem study of iron cycling in ferruginous systems that may also provide constraints on early Earth processes.

Exploring the physicochemical conditions and isotopic fractionations associated with Boreal Shield lake anoxic zone microbial consortia will catalyze further scientific advances in both modern and ancient systems. We report first evidence of distinctive microbial fractionation of iron isotopes *in situ* in a setting with likely photoferrotrophic activity. Probing conditions in modern systems will facilitate new interpretations of the wide range of iron isotopic values observed throughout the Archaean (Craddock and Dauphas 2011). Similarly, our carbon isotope results have implications for both modern and ancient carbon cycling. Isotopic fractionation inherent in the reductive citric acid pathway (4 to 13 ‰; Thomazo *et al.* 2009), hypothesized as the photosynthetic pathway for GSB (Konhauser 2009; Llirós *et al.* 2015) is too small to offset the much higher $\delta^{13}C$ of the high flux of transitory organic matter and the associated heterotrophic activity in order to match the observed $\delta^{13}C$ of the POM in the L227 hypolimnion given the measured $\delta^{13}C$ of the DIC. Consumption of dissolved CH_4 with very low $\delta^{13}C$, perhaps by the abundant methanotrophs in the anoxic zone, must contribute to L227 POM. Use of CH_4 as a carbon substrate within such microbial consortia could alter interpretation of the $\delta^{13}C$ signature for photosynthesis in ancient rocks and in modern settings, such as that published for Lake Kivu (Llirós *et al.* 2015). In addition, biogenic CH_4 is hypothesized to be an important contributor to the Archaean atmosphere (Zerkle *et al.* 2012). Thus, anaerobic consumption of biogenic CH_4 also has implications for ancient carbon cycling. Finally, analyses of $\delta^{15}N$ show evidence of isotopic fractionation when large amounts of NH_4^+ are present. Little is known about the expected ranges for $\delta^{15}N$ in the Archaean Ocean. Although the observed shift in $\delta^{15}N$ from negative values in the early Archean to more positive values in various geologic materials has been linked to the Great Oxidation Event, the interpretation remains controversial (Pinti and Hashizume 2011). Negative values in $\delta^{15}N$ observed in early Archaean kerogens (Beaumont and Robert 1999) could be attributed to metabolic fractionation of NH_4^+. However, the use of these kerogens as biosignatures has been challenged (Pinti and Hashizume 2011) and only laboratory experiments have been invoked to constrain the isotopic effects (Pinti and Hashizume 2011). Finally, large shifts in C, N, and Fe isotopic compositions have been recorded between ~2.8 and ~2.5 Ga ago and attributed to environmental or metabolic changes in the nascent period of early Earth oxidation (Thomazo *et al.*

2009; Hashizume *et al.* 2016). Interpretation of the wide isotopic ranges preserved in different proxies in the geologic record would be aided by the modern study of these distinctive microbial consortia.

Our finding of active microbial iron oxidation in anoxic hypolimnia of Boreal Shield lakes also has novel implications for limnology, water management, and microbial ecology. Although bacterial iron cycling has been observed in the water columns of meromictic lakes (Berg *et al.* 2016), the overall importance of internal iron reduction-oxidation to metabolism within seasonally anoxic hypolimnetic water columns, including the presence of a putative iron oxidizer and high relative abundance and diversity of iron reducers, has not yet been recognized. Metabolism of gammaproteobacterial methanotrophs that are present in high abundance in the anoxic zone has been suggested to be coupled to oxygen produced by cyanobacteria at the same depth (Kendall *et al.* 2015; Milucka *et al.* 2015), but it is also possible that this process may be coupled to the reduction of alternative electron acceptors, such as newly oxidized iron by photoferrotrophs. Very recently, methanotrophs of the same taxonomic group as we found in L227 and L442 were found thriving in the anoxic zones of Lake La Cruz and meromictic Lake Zug, Switzerland and able to oxidize methane without oxygen in the absence of light (Oswald *et al.* 2016a, 2016b). Further, these methanotrophs were stimulated by the addition of oxidized iron and manganese (Oswald *et al.* 2016a, 2016b). Iron reduction coupled to anaerobic methane oxidation has also been reported in an archaeal enrichment culture, showing the validity of this redox process in the natural environment (Ettwig *et al.* 2016). The importance of methanotrophy in anoxic hypolimnia with respect to either metabolism or CH_4 emissions to the atmosphere is only starting to receive scrutiny (Milucka *et al.* 2015; Oswald *et al.* 2015; Roland *et al.* 2016), and not yet in the context of the metabolic pathways of the entire microbial consortium. Metagenomic data may shed light on the nature of anaerobic methane oxidation processes in Boreal Shield lakes and how these processes relate to those reported in other low-sulfate environments (Beal, Claire and House 2011). Further, availability of reduced iron has recently been implicated as a factor controlling the dominance of cyanobacteria in both non-eutrophic systems and in the formation of hazardous algal blooms that have increased in frequency and severity in recent years (Molot *et al.* 2014); further understanding of iron redox cycling is urgently needed. Also unknown is the role of oxidized iron production in anoxic hypolimnia with respect to sequestration of phosphorous, the limiting nutrient in most Boreal Shield lakes. Lake 227 and similar boreal lakes have atypical low internal phosphorous release (Findlay and Kasian 1987). Lastly, some iron-cycling organisms are known to play a role in mercury (Hg) cycling and formation of methylated

Hg (Grégoire and Poulain 2016), a toxic and bioaccumulating contaminant of fish in Boreal Shield lakes.

Using real systems is a powerful approach for promoting new discoveries. In contrast to simplified and hard to maintain laboratory cultures, whole thriving microbial communities can be studied under *in situ* environmental conditions, with additional layers of complexity that cannot be simulated in a laboratory. Boreal lakes and ponds number in the tens of millions (Verpoorter *et al.* 2014) and, of these, some 15% could have a portion of the water column that is seasonally anoxic, opening new avenues of exploration. Water column chemistry and light penetration differ considerably among boreal lakes depending on geology and contribution of dissolved organic carbon from the catchment. Thus, broad gradients of physico-chemical conditions such as flux and quality of organic matter, sulfate and sulfide concentrations and light penetration can be exploited to better understand these unique microbial communities. Furthermore, small lakes, such as those at ELA, can be manipulated experimentally so that a specific range of conditions can be targeted or purposeful additions of carbon or iron isotope tracers can be used to probe isotopic fractionation *in situ*. Because hypolimnia become isolated during the stratified period and mixing is at rates of diffusion, accumulation or loss of constituents and mass balances can be used to infer metabolic pathways, products, and rates. Coupling these data to parallel metagenomic and metatranscriptomic analyses will facilitate reconstruction of genomes and active metabolic processes associated with photoferrotrophy. In addition, isotopic fractionations of δ^{13}C, δ^{15}N, and δ^{56}Fe associated with anoxygenic photosynthesis and iron cycling that may be preserved in the global rock record can be studied *in situ* and during early stages of diagenesis in lake sediments deposited over the last 10,000 years since deglaciation. Boreal Shield lakes provide natural and accessible incubators for the study of processes relevant to the biogeochemistry and evolution of life in early Earth history.

Data availability

Raw amplicon sequencing data is available in the NCBI sequence read archive under BioProject PRJNA354806. Representative sequences for each OTU, determined using AXIOME2, were deposited in the NCBI under accession numbers KY515586-KY522665. Counts of OTUs in each lake water sample are available in Supplementary Data File A1 as an OTU table.

Acknowledgements

Author contributions

SLS, JJV, LM, MP were involved in study design of the larger project that supported data collected for this paper. SLS, JJV, RJE collected samples and analyzed δ^{13}C and δ^{15}N. LW analyzed samples for δ^{56}Fe. JMT and JDN performed molecular analysis. SLS wrote the paper with JDN, JMT, LW and comments from all other authors.
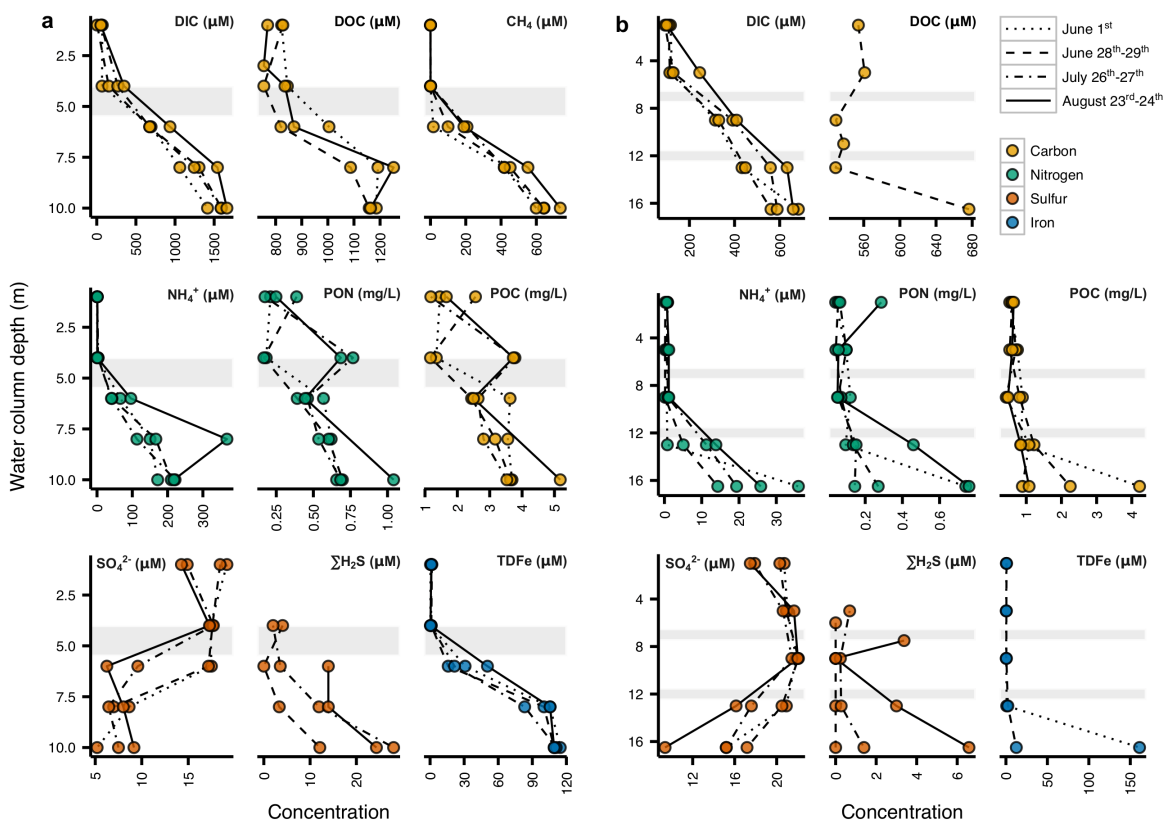
Figure A1. Water chemistry of (a) L227 and (b) L442 water columns. All samples were collected in June-August 2010. Measurements from the same sampling dates are connected by lines for visual clarity. Sampling dates include those from before the onset of the annual cyanobacterial bloom (June 1st), during the bloom (June 28th-29th), and after the bloom (July 26th-27th, August 23rd-24th), in L227. In each sub-panel for L227, the surface mixed layer is separated from the seasonally anoxic hypolimnion by a grey transition zone. Each sub-panel for L442 is divided by grey transition zones into the surfaced mixed layer (top), the cool, oxic hypolimnion (middle), and the seasonally anoxic hypolimnion (bottom). The transition zones dividing each lake layer may vary seasonally and annually in both thickness and water column location (depth) due to differing climate (see Supplementary Figure A2).
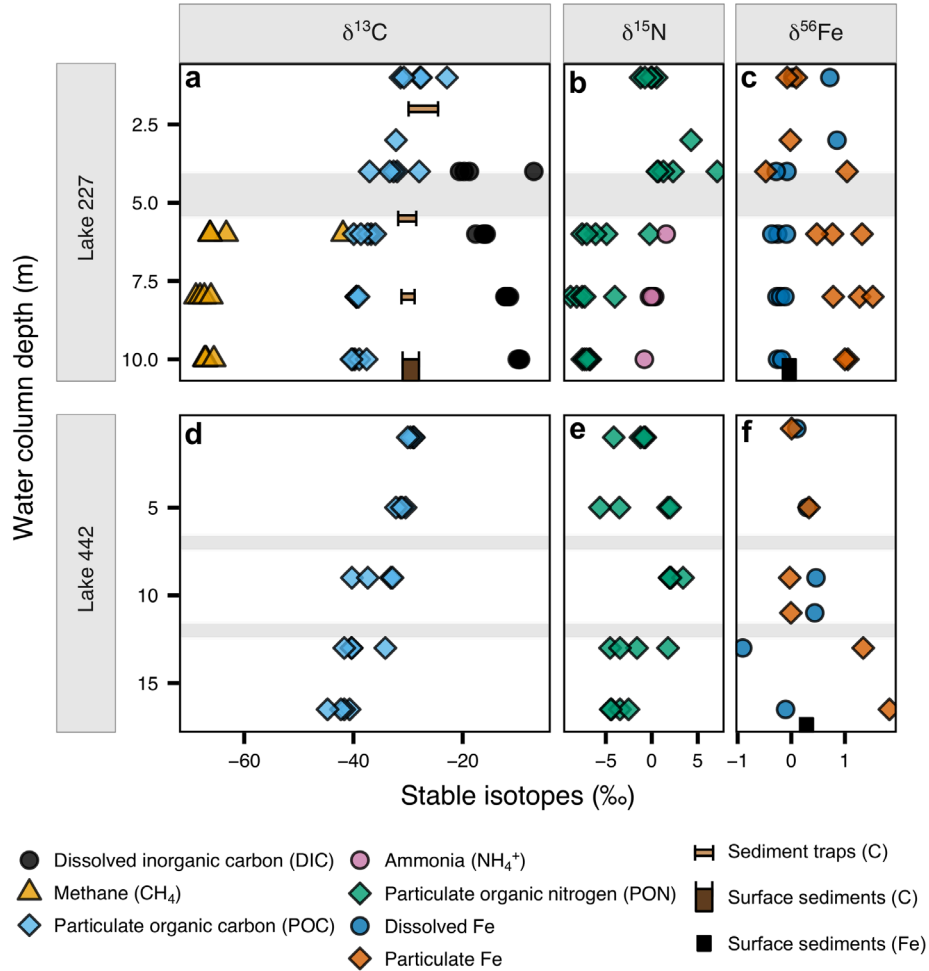
Figure A2. Stable isotopic values in the water columns of (a-c) L227 and (d-f) L442. (a, d) $\delta^{13}C$ in DIC, CH_4, and POC, sampled in June-August 2010 and July 2014. $\delta^{13}C$ in sediment traps and surficial sediments (1-2 cm) is also shown. The width of plotted sediment markers represents the range of $\delta^{13}C$ values measured across the summer of 2011. (b, e) $\delta^{15}N$ in NH_4^+ and PON, sampled in June-August 2010 and July 2014. (c, f) $\delta^{56}Fe$ in dissolved and particulate Fe sampled in June-August 2011 and July 2014. $\delta^{56}Fe$ in surficial sediments from July 2010 is also shown. In each L227 panel, the surface mixed layer is separated from the seasonally anoxic hypolimnion by a grey transition zone. Each panel for L442 is divided by grey transition zones into the surfaced mixed layer (top), the cool, oxic hypolimnion (middle), and the seasonally anoxic hypolimnion (bottom). The transition zones dividing each lake layer may vary seasonally and annually in both thickness and water column location (depth) due to differing climate. The water column of L227 at 4 m depth varied substantially in oxygen status over the sampling dates (see Supplementary Figure A2), which may be responsible for the wide range of isotopic values measured at this depth.
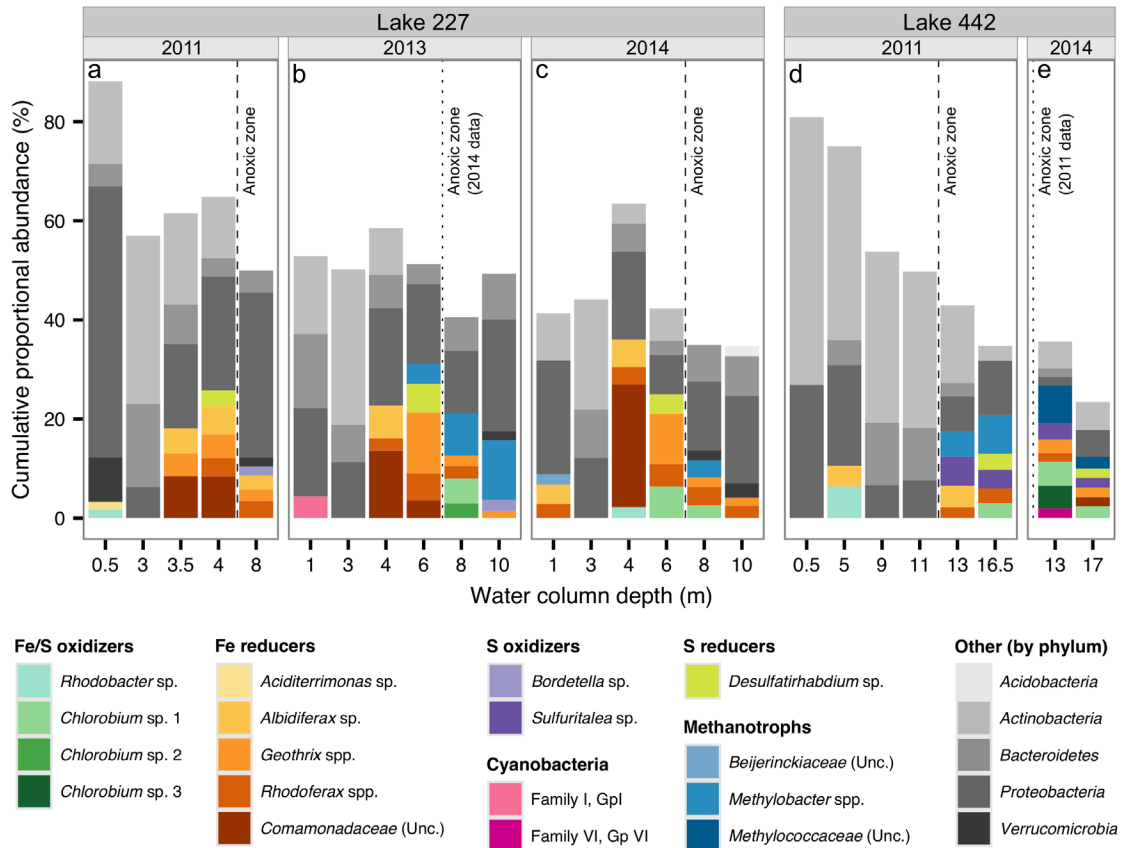
155

Figure A3. Microbial community in the water columns of (a, b, c) L227 and (d, e) L442. In each depth sample, the ten most abundant bacterial operational taxonomic units (OTUs) are shown as stacked bars in terms of their proportional abundance within rarefied molecular sequencing data. Bacterial OTUs are grouped according to their potential involvement in iron cycling, sulfur cycling, and methanotrophy (see Methods). Cyanobacterial OTUs are also shown. For clarity, OTUs not associated with these metabolic roles are displayed at the phylum level. Potentially photoferrotrophic *Chlorobium* OTUs are shown with a naming and colour scheme matching that of Supplementary Figure A3. Although not shown, *Chlorobium* OTU 1 is also present in the anoxic water column in L227 at 6 m depth in 2013 (rank 13, 2.9%) and in L442 at 13 m depth in 2011 (rank 31, 0.67%). All samples were collected between June 25th and July 9th in their respective sampling year.
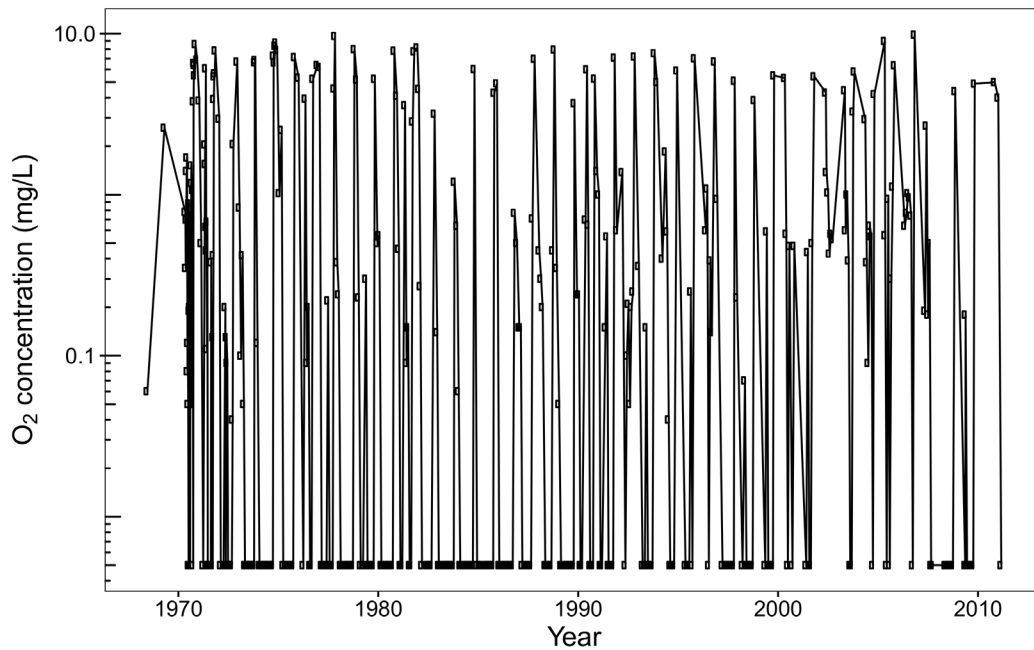
Figure A4. Dissolved oxygen at 6 m depth in the L227 water column from 1969 to 2011. The detection limit for O_2 is 0.005 mg O_2 L^{-1}. *Chlorobium* sequences were detected at high abundance at 6 m depth in both 2013 and 2014 (Figure A3, please also see legend). Dissolved oxygen samples were collected typically at least every two weeks in summer but were collected at most twice during the winter. Following this sampling schedule, the full extent of the typical spring and fall re-oxygenation events (overturns) in L227 may not have been measured in some years. Dissolved oxygen is typically, but not always, measured after fall overturn. Spring overturn measurements can be missed following ice-off due to logistical reasons and especially in years when temperatures warm rapidly after ice-off. Thus, the oxygen record at 6 m reflects the minimum number of re-oxygenation events at this depth.
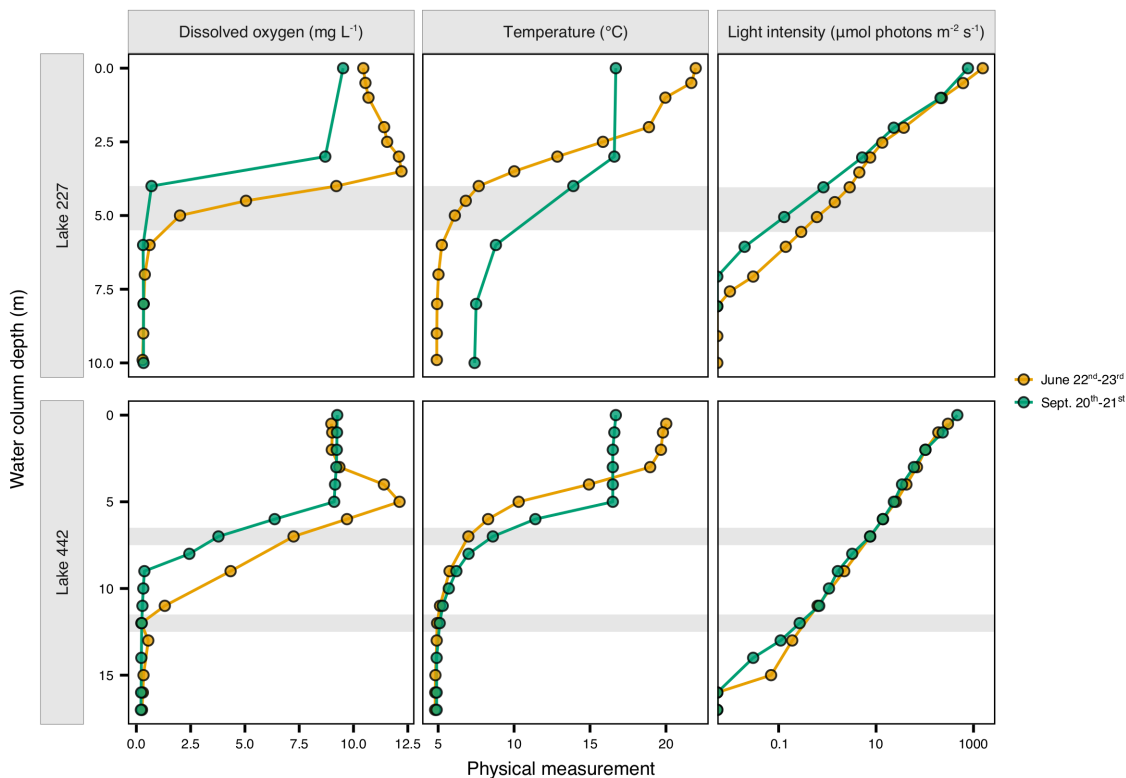
Supplementary Figure A1. Dissolved oxygen, temperature and light profiles of the L227 and L442 water columns in 2016. Grey bars represent approximate transition zones between lake layers and are identical to those shown in Figs. 1-2 (see figure captions). Exact layer boundaries at the sampling times may be estimated using the temperature and dissolved oxygen profiles shown. In both lakes, low but detectable light levels reached the top of the anoxic zone in both June and September. Light measurements below the detection limit (0.01 µmol m^{-2} s^{-1}) are plotted along the left edge of the y-axis for reference.
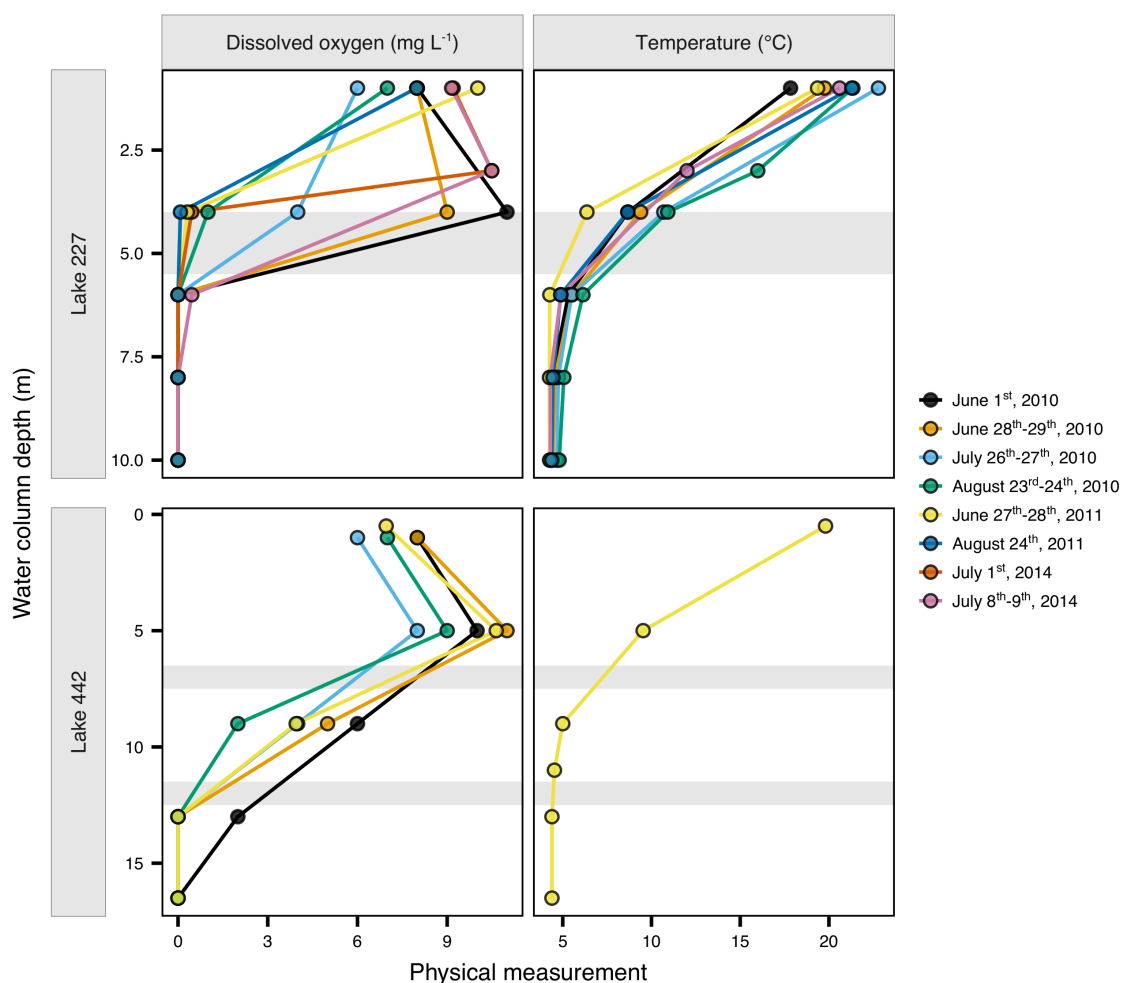
Supplementary Figure A2. Physical profiles of the L227 and L442 water columns during isotopic and molecular sampling. Grey bars represent approximate transition zones between lake layers and are identical to those shown in Figs. 1-2 (see figure captions). For comparison, actual lake zone boundaries at each time of measurement may be determined visually from temperature and dissolved oxygen data. Seasonal variations can be noted in the temperature and oxygen status of the water columns at depths near the approximate transition zones, for example for L227 at 4 m depth.
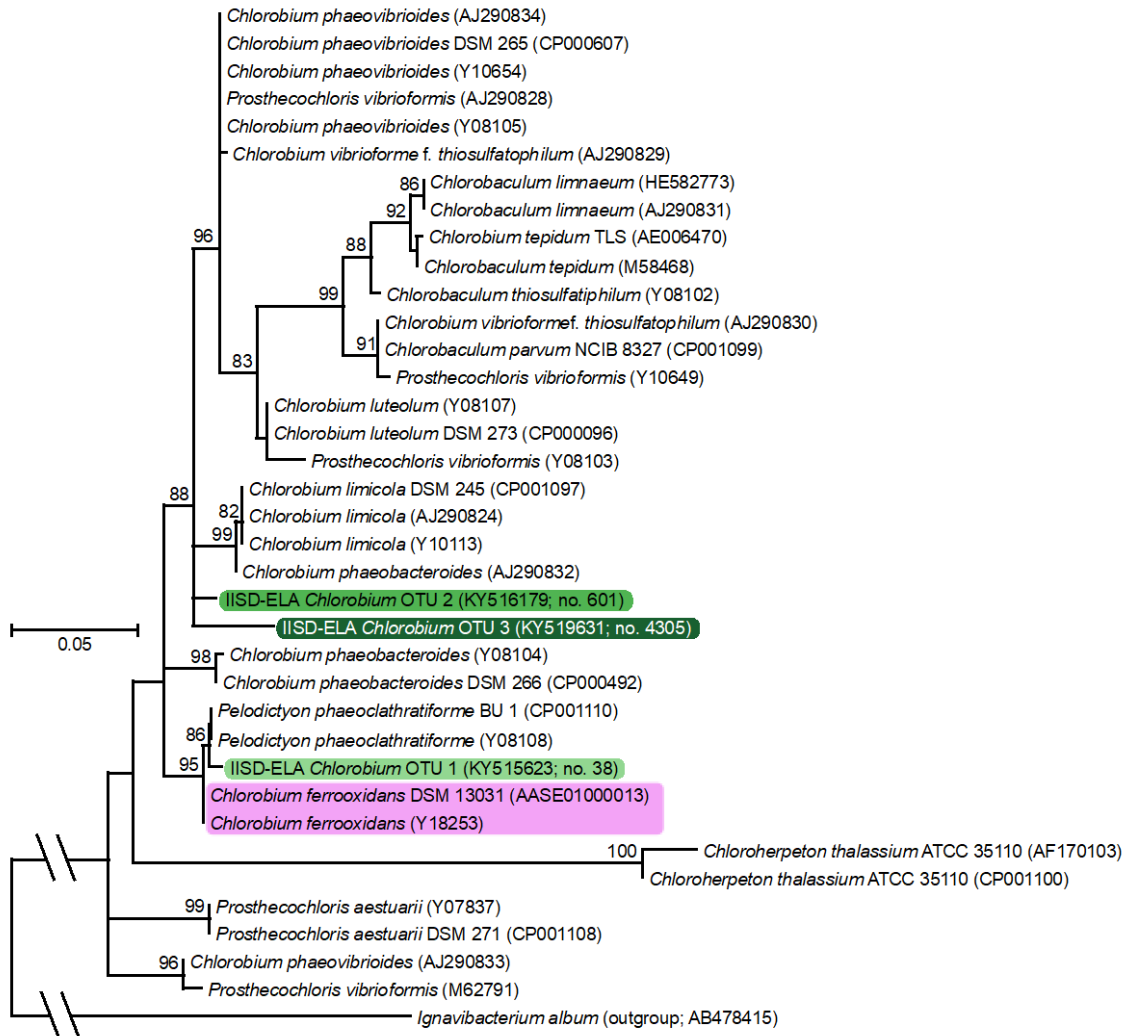
Supplementary Figure A3. Phylogenetic placement of potential photoferrotrophs within the family *Chlorobiaceae*. Reference *Chlorobiaceae* sequences represent cultured strains. Node support values, calculated using the Shimodaira-Hasegawa test, are shown where 80% or higher. Sequences highlighted in pink correspond to known photoferrotrophs, and those highlighted in green represent OTUs identified at high abundance in the water columns of L227 and L442 (Figure A3). Importantly, IISD-ELA *Chlorobium* OTU 1 was identified at high abundance in the water columns of both lakes.
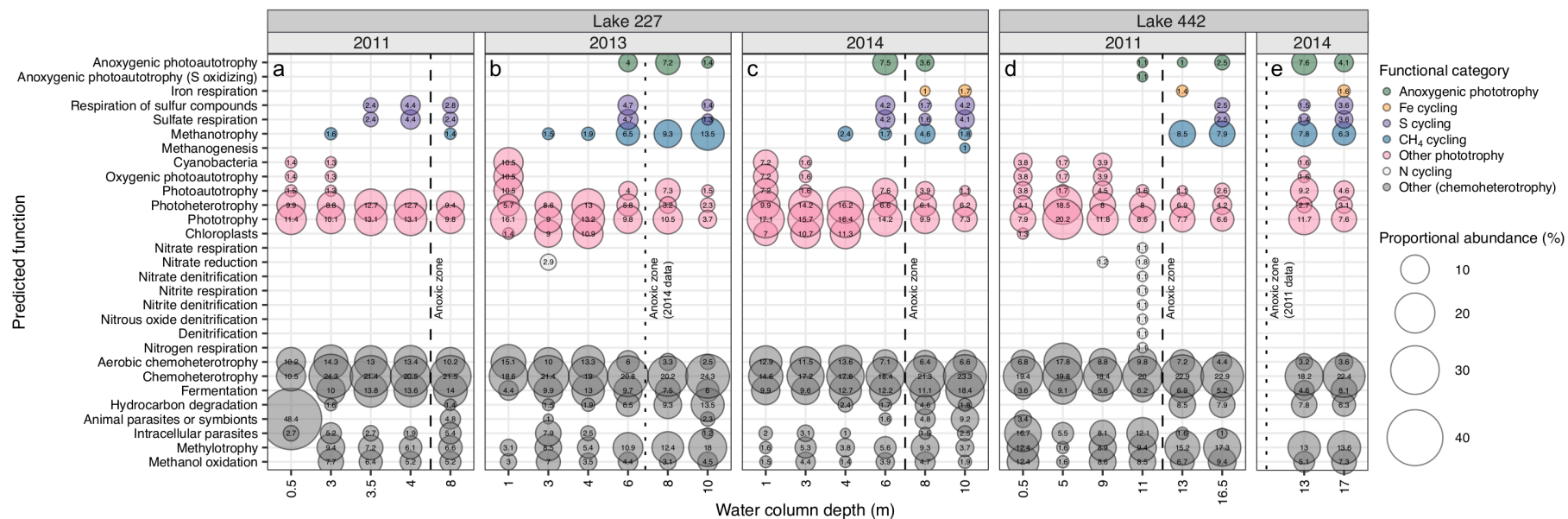
Supplementary Figure A4. Computational prediction of bacterial community function within the water columns of (a, b, c) L227 and (d, e) L442. Predicted functional roles were assigned to OTUs using FAPROTAX (*14*) for comparison to the manually curated method used for functional assignment in Figure A3 (see Methods). For each depth sample, functional roles making up more than 1% of all predicted roles are shown as bubbles according to their proportional abundance. Importantly, the "anoxygenic phototrophy" group includes green sulfur bacteria OTUs, which may perform sulfide oxidation or photoferrotrophy (see Supplementary Fig. S3). Bacteria from the family *Comamonadaceae* (e.g. *Rhodoferax* spp., *Albidiferax* spp.) were classified as potential chemoheterotrophs or photoheterotrophs, in contrast to their classification as potential iron reducing bacteria using the manually curated approach (Fig. 3). See Supplementary Figure A5 for a phylogeny of this family showing the placement of detected OTUs compared to known iron reducing strains. In total, 1123 of 6671 OTUs (i.e., 16.8%) could be assigned a functional role using FAPROTAX, including 48 of 74 (i.e., 64.9%) of the top ten most abundant OTUs in each sample.

161
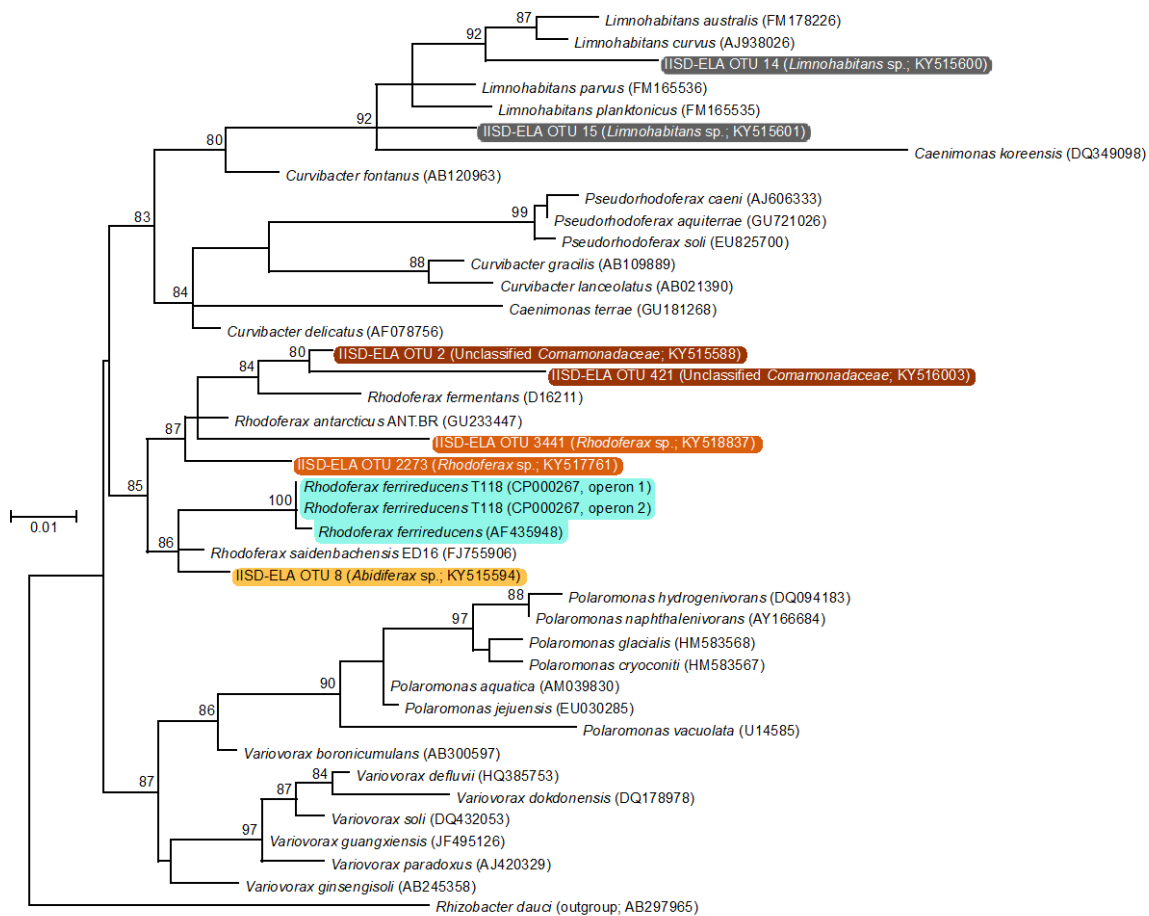
Supplementary Figure A5. Phylogenetic placement of potential iron reducing bacteria within the family *Comamonadaceae*. Reference *Comamonadaceae* sequences represent cultured strains from a monophyletic subset of the family (see Methods). Node support values, calculated using the Shimodaira-Hasegawa test, are shown where 80% or higher. Sequences highlighted in blue correspond to known iron reducing bacteria, and those highlighted in yellow or orange represent OTUs identified at high abundance in the water columns of L227 and L442 that were assigned a potential role in iron reduction using the manual annotation approach (Figure A3; colours match those in figure). Two *Comamonadaceae* OTUs at high abundance that were manually classified as likely chemoheterotrophs are highlighted in grey for comparison.
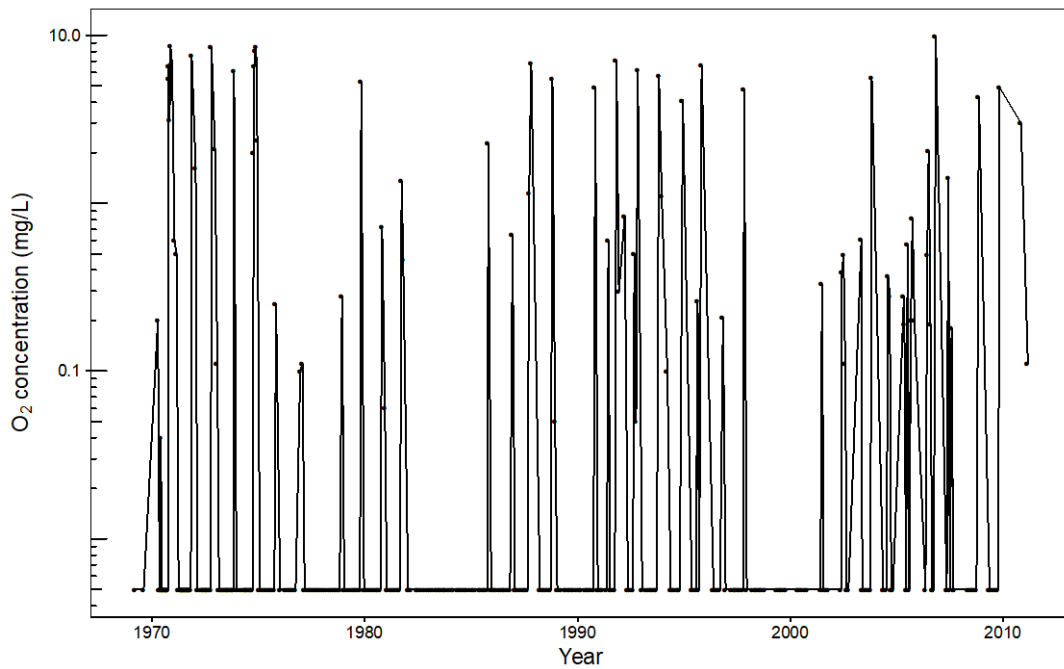
Supplementary Figure A6. Dissolved oxygen at 8 m depth in the L227 water column from 1969 to 2011. The detection limit for O_2 is 0.005 mg O_2 L^{-1}. *Chlorobium* sequences were detected at high abundance in both 2013 and 2014 at this depth (Fig. A3). Dissolved oxygen samples were collected typically at least every two weeks in summer but were collected at most twice during the winter. Following this sampling schedule, the full extent of the typical spring and fall re-oxygenation events (overturns) in L227 may not have been measured in some years. Dissolved oxygen is typically, but not always, measured after fall overturn. Spring overturn measurements can be missed following ice-off due to logistic reasons and especially in years when temperatures warm rapidly after ice-off. Thus, the oxygen record at 8 m reflects the minimum number of re-oxygenation events at this depth.
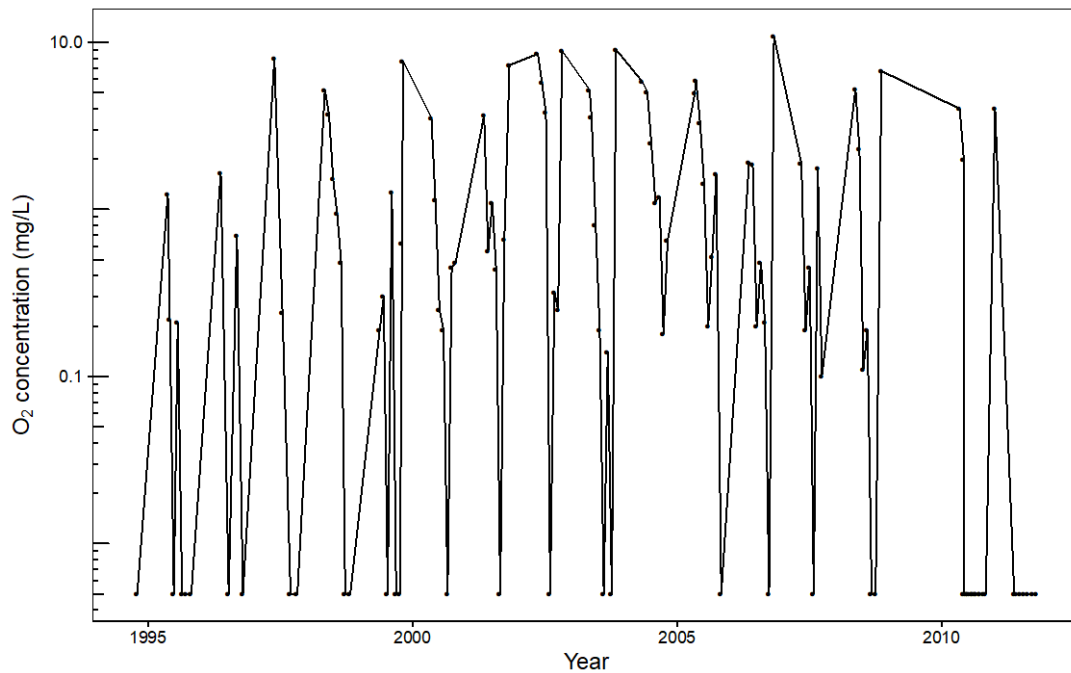
Supplementary Figure A7. Dissolved oxygen at 13 m depth in the L442 water column from 1994 to 2012. The detection limit for O_2 is 0.005 mg O_2 L^{-1}. *Chlorobium* sequences were detected at high abundance at this depth in both 2011 and 2014 (Figure A3). Dissolved oxygen samples were collected typically once per month in summer but were collected at most once during the winter. Following this sampling schedule, the full extent of the typical spring and fall re-oxygenation events (overturns) in L442 may not have been measured in some years. Dissolved oxygen is typically, but not always, measured after fall overturn. Spring overturn measurements can be missed following ice-off due to logistic reasons and especially in years when temperatures warm rapidly after ice-off. Thus, the oxygen record at 13 m reflects the minimum number of re-oxygenation events at this depth.

Supplementary Data File A1. Rarefied hit counts of operational taxonomic units (OTU) within water column samples of L227 and L442. The OTU table (CSV format) was prepared using the software tool AXIOME2 (see Materials and Methods). Representative sequences of each OTU are included. Classification of each OTU is shown up to the taxonomic rank where the RDP classifier's confidence value fell below 50%. The file is available online at doi:10.1038/srep46708.

Appendix B

Enrichment cultivation notes for "*Ca.* Chlorohelix allophototropha"

Cultivation of "*Candidatus* Chlorohelix allophototropha" began from an experiment attempting to culture photoferrotrophic *Chlorobia* members from Lake 227. Once DNA sequencing data became available showing that a phylogenetically distinct *Chloroflexota* member was growing in selected bottles from that experiment, incubation conditions had to be adjusted iteratively to promote the growth of the novel strain compared to other organisms. Ultimately, after several cultivation attempts, one lineage of enrichment culture could successfully be cultivated and survive multiple subculture transfers. It was from this lineage that "*Candidatus* Chlorohelix allophototropha" was derived. A log specifically for that lineage of enrichment culture is provided in Table B1 for reference.

Supplementary Table B1 Log of the enrichment cultivation process for "*Ca.* Chx. allophototropha". All subcultures are shown spanning from initial Lake 227 water to the culture from which DNA was extracted for read cloud DNA sequencing. Unless otherwise indicated, "freshwater medium" refers to a ferrous iron-containing medium described previously (Hegler *et al*. 2008) that contained 8 mM ferrous chloride.

Cultivation phase	Sub-culture	Start date	Notes
	0	2017-09-23	Setup: 98% lake water (L227, 3.88 m), 2% freshwater medium, resulting in ~160 µM ferrous iron in the bottle; N_2 headspace. Incubated at 22°C under far-red LED lights as described in Chapter 2.
			Monitored iron redox state every ~2 days via the ferrozine assay. Fed with ferrous chloride to 160 µM concentration, then with freshwater medium to 2% concentration, from Sept. 25th to 27th.
Acclimatization			Hypothesized that rapid ferrous iron oxidation activity was due to oxygenic phototrophs.
	1	2017-09-28	Due to rapid ferrous iron oxidation, a subculture with 10% inoculum was performed into 4% freshwater medium (containing 320 µM ferrous iron); N_2 headspace. Diuron was spiked into the bottle to a final concentration of 50 µM. Same incubation conditions as above.

Table B1, cont.

Cultivation phase	Sub-culture	Start date	Notes
Acclimatization	1, cont.	2017-09-28	Monitored iron redox status regularly via the ferrozine assay. Depletion of ferrous iron began to be observed after Dec. 14th.
			Bottle was incrementally fed with iron-containing freshwater medium whenever ferrous iron depletion was measured. Five feedings occurred from Dec. 14th, 2017 to Feb. 8th, 2018, increasing total iron levels in the bottle to ~2.7 mM and total medium concentration to ~34% of full concentration.
			Genomic DNA was extracted on Jan. 16th, 2018 was later used for 16S rRNA gene amplicon and metagenome sequencing.
			Subcultured into 20% and 100% freshwater medium; only the latter treatment is followed below.
Liquid enrichment	2	2018-02-14	10% inoculum into full-strength freshwater medium containing 8 mM ferrous iron; used a 90:10 N_2:CO_2 headspace. Same incubation conditions as above.
			Noted visual signs of iron oxidation on May 18th.
	3	2018-05-18	10% inoculum using the same medium and conditions as above.
			Noted visual signs of iron oxidation on Jun. 15th.
			Stored as a long-term reference sample at 4°C in the dark on Jun. 15th.
Iron gradient agar tubes	4	2018-11-09	10% inoculum into 0.6% agar shake tubes in dilution series having a 10 mM ferrous chloride plug and 50% concentration freshwater medium, lacking ferrous iron. Sodium thioglycolate (0.125 g/L) was used as a reducing agent, acetate (~6 mM) was added as an organic carbon source, and medium was supplemented with 50 uL/L of a 13-vitamin solution (Tank and Bryant 2015). Tubes were incubated at ~18°C under halogen lights (5-35 µmol PAR photons m^{-2} s^{-1}).
			Performed epifluorescence microscopy on Nov. 23rd; noticed growth of fluorescent filaments but also many potential purple phototrophic bacteria.
			Observed blackening of tubes on Jan. 17th, 2019. When tubes were opened, found that a mixture of golden and green colonies could be found in the lower halves of the tubes.

Table B1, cont.

Cultivation phase	Sub-culture	Start date	Notes
Iron gradient agar tubes	5	2019-01-17	Same setup as above, but used 0.8% agar and increased the pH of the medium to ~7.8, after addition of ferrous chloride, to inhibit growth of green phototrophic bacteria. Added 0.25 mg/L resazurin and omitted the 13-vitamin mix.
	6	2019-02-16	10% inoculum into deep agar dilution series, without plugs of ferrous iron.
			Used 50% concentration freshwater medium containing 2 mM ferrous iron, ~3 mM acetate, 0.8% agar, and 0.25 mg/L resazurin.
			Moved tubes to growth at room temperature (~20-25°C) in a 740 nm light incubator to block growth of purple phototrophic bacteria
			Observed blackening of tubes on Feb. 28th.
	7	2019-03-02	Same conditions as above (subculture 6), except that soft agar (0.2-0.3%) was used and the pH was adjusted ~7.5-7.8 before ferrous chloride addition.
Media refinement and culture purification	8	2019-03-22	Same as subculture 7. Moved to a 22°C white light incubator using a mixture of fluorescent and incandescent lights; purple bacteria seemed to have been eliminated.
	9	2019-04-02	Same as subculture 8 except that acetate levels were lowered to ~1.2 mM concentration.
	10	2019-05-30	No change from subculture 9.
	11	2019-06-18	No change from subculture 9.
	12	2019-07-12	No change from subculture 9.
	13	2019-09-05	Same as subculture 9, except that resazurin concentration was increased to 0.5 mg/L.
	14	2019-09-18	No change from subculture 13.
	15	2019-10-17	Same as subculture 13, except 100 μM molybdate was added to inhibit sulfate-reducing bacteria.
			Genomic DNA extracted on Nov. 28th.