Genomic potential for photoferrotrophy in a seasonally anoxic Boreal Shield lake

Tsuji JM[1], Tran N[1], Schiff SL[1], Venkiteswaran JJ[1,2], Molot LA[3], Neufeld JD[1]*

[1]University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada, N2L 3G1

[2]Wilfrid Laurier University, 75 University Avenue West, Waterloo Ontario, Canada, N2L 3C5

[3]York University, 4700 Keele Street, Toronto, Ontario, Canada, M3J 1P3

*Corresponding author: jneufeld@uwaterloo.ca

Running title: Photoferrotrophy in a Boreal Shield lake

Keywords: photoferrotrophy; *Chlorobia*; *cyc2*; boreal lakes; iron cycling; Archaean Eon; metagenomics; anoxygenic photoautotrophy

Abbreviations used throughout text: *Chl.* = Chlorobium; *Cba.* = Chlorobaculum; *Che.* = Chloroherpeton; *Ptc.* = Prosthecochloris

The authors declare no competing financial interests.

Abstract

Photoferrotrophy, the direct photosynthetic oxidation of ferrous iron, is thought to have contributed to primary production within Earth's early anoxic oceans yet is presumed to be of little modern environmental relevance. Here we use genome-resolved metagenomics and enrichment cultivation to provide first functional gene evidence that photoferrotrophic bacteria are important photoautotrophs in the anoxic water columns of globally abundant Boreal Shield lakes. We recovered six high-completeness, low-contamination draft genome bins belonging to the *Chlorobia* class. Half of these genome bins contained the *cyc2* candidate gene marker for iron oxidation, and one containing *cyc2* lacked the sulfur oxidation gene pathway altogether. Using a custom profile Hidden Markov Model, we recovered distinct *cyc2* homologues from our metagenome data and doubled the known diversity of *Chlorobia*-associated *cyc2* genes. Metagenome data also showed that potential photoferrotrophs co-occurred with sulfide-oxidizing populations of *Chlorobia* in one lake but that only sulfide-oxidizing populations were present in a neighboring lake, implying that strong ecological controls govern the favourability of photoferrotrophy in ferruginous lakes. Our results greatly expand knowledge of photoferrotroph ecology in natural aquatic systems, demonstrating modern environmental relevance for this iron-oxidizing photosynthetic mode and supporting future investigations into early Earth microbial communities.

Introduction

Anoxygenic photoautotrophs are common in oxygen-limited environments worldwide and serve as primary producers in sediments, hot springs, and anoxic water columns as deep as the Black Sea [1–3]. Microorganisms performing anoxygenic photosynthesis commonly use a reduced sulfur compound, such as sulfide or thiosulfate, as the photosynthetic electron donor [4, 5]. The discovery of alternative metabolic modes of photosynthesis challenge classical interpretations of how anoxic ecosystems function [6, 7]. Photoferrotrophy, an anoxygenic photosynthetic process involving the oxidation of ferrous iron in place of reduced sulfur, has been described in several laboratory cultures but remains poorly understood with respect to its ecological role [8]. Photoferrotrophs are of great interest for their potential involvement in early Earth microbial communities where they may have served as important primary producers in the ferruginous (i.e., iron-rich, sulfate-poor) and anoxic waters of Archaean Eon oceans (ca. 3.8-2.5 Ga) [9–11]. However, due to higher levels of sulfate in most modern aquatic systems, compared to Archaean oceans, photoferrotrophs are presumed to be of little modern environmental importance, acting as a remnant of a previously globally influential photosynthetic process [12].

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. However, recent work has begun to unravel the molecular basis for iron oxidation in photoferrotrophs and other neutrophilic iron-oxidizing bacteria (reviewed in [13, 14]). Candidate gene systems are emerging that can be used to probe for selected types of iron oxidation environmentally. In particular, genes encoding for porin-cytochrome *c* protein complexes (PCC), such as *pioAB* and *mtoAB*, and genes encoding for outer-membrane monoheme c-type cytochromes, such as *cyc2*, have been identified as potentially useful markers for iron oxidation [14–17]. The discovery of such genetic markers

expands capabilities to explore the role of photoferrotrophy and other forms of iron cycling in nature.

Additional photoferrotrophic bacteria are also being isolated in the laboratory. For example, among the *Chlorobia* class (i.e., "green sulfur bacteria"), the only known photoferrotroph was *Chlorobium ferrooxidans* strain KoFox, which was enriched from freshwater lake sediments [18]. Recently, two additional *Chlorobia*-associated photoferrotrophs were cultivated. *Chlorobium phaeoferrooxidans* strain KB01, a bacteriochlorophyll e-containing member of the *Chlorobia*, was isolated from the anoxic water column of the meromictic and ferruginous Kabuno Bay (see below; [19]). In addition, *Chlorobium* sp. strain N1 was isolated from marine sediments [20]. Both *Chl. ferrooxidans* and *Chl. phaeoferrooxidans* oxidize ferrous iron as their sole photosynthetic electron donor, assimilating sulfur as sulfate [18, 19]. *Chl.* sp. N1 is capable of oxidizing either ferrous iron or sulfide as the photosynthetic electron donor [20]. 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* with genomic potential for photoferrotrophy [19–21]. Draft genome sequences available for *Chl. ferrooxidans*, *Chl. phaeoferrooxidans*, and *Chl.* sp. N1 show that their genomes contain *cyc2* homologues, unlike all other sulfide-oxidizing *Chlorobia*, with the exception of *Chl. luteolum* [21–24]. Although photoferrotrophy has not yet been demonstrated in *Chl. luteolum*, the presence of *cyc2* in its genome and its relatedness to *Chl.* sp. N1 implies that it can perform photoferrotrophy, although the conditions needed to induce photoferrotrophy remain unclear [20]. The addition of *Chl. phaeoferrooxidans* and *Chl.* sp. N1 allow for first glimpses into the natural diversity of photoferrotrophy among *Chlorobia* and provide evidence that *cyc2* may be a valuable genomic marker for photoferrotrophy in the *Chlorobia* class.

Current knowledge of photoferrotroph ecology derives from studies of a select few modern environments. Among the best studied are meromictic ferruginous lakes that are used as analogues of Archaean ocean conditions (reviewed in [25]; see also [26]). Study of these Archaean ocean analogues has led to varying conclusions about the environmental importance of photoferrotrophy. In Lake Matano (Indonesia), bacteria classified to the *Chlorobium* genus were detected beneath the oxic-anoxic zone interface [27]. However, only genes implicated in the sulfide oxidation pathway among *Chlorobia* were detected in metagenome sequencing data, and anoxygenic photoautotrophy rates were below detection limits, leaving the role of photoferrotrophy unclear [28]. In Lake La Cruz (Spain), rates of photoferrotrophic iron oxidation were estimated at 0.2-1.4 μM Fe^{2+} day^{-1}, yet microscopy data and photoautotrophy rates suggested that photoferrotrophy was catalyzed by a small proportion of total phototrophic cells, and that photoferrotrophs co-existed with other phototrophic bacteria, such as sulfide oxidizers [29]. In Lake Kivu's Kabuno Bay (Democratic Republic of the Congo), anoxygenic phototrophs in the upper anoxic zone used ferrous iron for most or all photosynthetic activity at rates of up to 100 μM Fe^{2+} day^{-1} [19]. Anoxygenic photoautotrophs were estimated as being capable of supplying up to ~90% of total fixed carbon in the bay's anoxic zone [30]. Contrasting results from these lakes suggest that unknown factors in ferruginous systems modulate the relative importance of various anoxygenic photosynthetic modes. In addition, it appears as though sulfide-oxidizing photoautotrophs can play a relatively important role in primary production even in sulfide-limited systems (see also [31]). However, the limited number of such meromictic ferruginous lakes makes further speculation about photoferrotroph ecology and *in situ* activity challenging.

Recently, we combined 16S ribosomal RNA (16S rRNA) gene sequencing and iron isotope data to propose photoferrotrophy by populations of *Chlorobia* as a major photosynthetic process

in the seasonally anoxic water columns of two ferruginous Boreal Shield lakes at the International Institute for Sustainable Development Experimental Lakes Area (IISD-ELA; near Kenora, Canada) [32]. Boreal Shield lakes number in the tens of millions across northern regions globally, and we proposed that such lakes commonly contain photoferrotrophic bacteria and could serve as plentiful modern-day analogues of Archaean Ocean ecosystems. Here, we use genome-resolved metagenomics to recover high-completeness, low-contamination draft genomes of key populations of *Chlorobia* from IISD-ELA lakes. We then apply recent knowledge of iron functional gene markers to test whether these populations possess the genetic potential for photoferrotrophy or other phototrophic modes. We also report the enrichment of two novel species of *Chlorobia* from IISD-ELA lakes and analyze their genetic potential for sulfide and iron oxidation. Through this first usage of molecular tools to distinguish *Chlorobia*-driven photoferrotrophy from other modes of photosynthesis in an environmental context, we aim to clarify the diversity and ecology of photoferrotrophic bacteria with implications for understanding both Archaean microbial communities and modern ferruginous environments.

Materials and Methods

Lake sampling, metagenome sequencing, assembly, and binning

The IISD-ELA sampling site, lake geochemistry, and sample collection methods were described previously [32]. Based on preliminary 16S rRNA gene data, we selected six water column genomic DNA samples for re-sequencing via shotgun metagenomics. All six samples were collected at or beneath the oxic-anoxic zone boundary and had high relative abundances of anoxygenic phototrophs, dominated by populations of *Chlorobia*. 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, USA) and was sequenced on a single lane of a HiSeq 2500 (Illumina; San Diego, California, USA) 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 [33]. The configuration file with all settings for the pipeline is available in Supplementary File 1. 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, as described in the Supplementary Methods.

Identification of cyc2 genes and Chlorobia genome bins

Assembled contigs were assessed for the presence of *cyc2* gene homologues by building a custom profile Hidden Markov Model (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 iron-oxidizing bacteria as described by He and colleagues [13]. 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* [34]. Collected sequences were aligned using Clustal Omega, version 1.2.3 [35], and the alignment was used to build a profile HMM using the

hmmbuild function of HMMER3, version 3.2.1 [36]. Recovered *cyc2* genes over the course of the
155 study were added to the alignment and HMM; final HMMs are included in Supplementary File 2.

Recovered *cyc2* genes were compared to *cyc2* reference sequences by building a maximum-likelihood phylogeny. The alignment was masked using Gblocks, version 0.91b [37], as described in the Supplementary Methods. The phylogeny was then prepared from the masked sequence alignment via IQ-TREE, version 1.6.10 [38], using the LG+F+I+G4 sequence evolution model as
160 determined by the ModelFinder module of IQ-TREE [39]. 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 [40] with default settings. All genome bins from individual assembles and co-assemblies were pooled for dereplication to help
165 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 [41] were selected for further study. These genome bins were imported into Anvi'o version 4 [42] and were manually examined for contigs improperly binned based on read mapping, tetranucleotide frequencies, and contig taxonomic
170 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 [43] using the *Chl. luteolum* genome to guide ordering. To assess bin quality, tRNA genes were
175 predicted using prokka v1.13.3 [44].

Enrichment cultivation, sequencing, and assembly

Lake 227 was sampled again in July 2017, along with the nearby Lake 304, which has been described previously and also has seasonally anoxic bottom waters [45, 46]. Lake 304 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 [47]. Anoxic water was collected from both lakes at a depth of 6 m, where trace levels of light are generally detectable (i.e., in the range of 0.01-1 μmol photons/m^2/s). 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, USA), 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 media, prepared as described by Imhoff [4, 5], or ferrous iron-containing freshwater media, prepared as described by Hegler and colleagues [48]. Ferrous iron-containing media contained 8 mM $FeCl_2$, without filtration of precipitates, and used trace element solution SLA [4]. 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, USA). The remaining headspace was flushed with a 90:10 N_2:CO_2 gas mix at left at a pressure of 1.5 atm. Several bottles additionally had anoxic DCMU (i.e., diuron; Sigma-Aldrich; St. Louis, Missouri, USA) added to a concentration of 50 μM to block activity of oxygenic phototrophs [49, 50]. Bottles were incubated at 22°C and ~50 cm away from far red PARSource PowerPAR LED Bulbs (LED Grow Lights Depot; Portland, Oregon, USA) as the light source.

Two enrichment cultures survived repeated subculturing or dilution-to-extinction and
200 contained green sulfur bacteria based on pigmentation and marker gene sequencing. 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 lane of a HiSeq 2500 (Illumina) in
205 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 [42] via *atlas-to-anvi.sh* commit 99b85ac;
210 contigs of curated bins were subsequently ordered, and tRNA genes counted, as described above.

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 and photoferrotrophs from the family were downloaded from the NCBI
215 (Supplementary File 3). Average nucleotide identity (ANI) between genomes was calculated using FastANI version 1.1 [51]. 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 [52, 53]. IQ-TREE version 1.6.9 [38] was used to construct the phylogeny from the alignment. The LG+F+R4 model of sequence evolution was
220 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 or gene clusters implicated in iron or sulfur cycling metabolism in the collection of *Chlorobiaceae* genomes was assessed using reciprocal *blastp*. Genes were selected from the genomes of *Chl. ferrooxidans* and *Chl. clathratiforme* based on the genome comparison of Frigaard and Bryant [21]. Predicted primary sequences of these genes were used to identify putative homologues across other genomes of *Chlorobia* using the *BackBLAST.py* module of the BackBLAST pipeline (development commit cdaddc4) [54]. The e-value cutoff for BLAST 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 between reference genomes.

Metagenome taxonomic and functional profiling

Environmental relative abundances of microbial populations were estimated by read mapping to dereplicated genome bins (further details in Supplementary Methods). Taxonomy and functional gene information were then overlaid onto relative abundance data. Genome bin taxonomy was determined based on the Genome Taxonomy Database using GTDB-Tk, version v0.2.2, relying on GTDB release 86, version 3 (April 2019) [55]. 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 [56], respectively. Gene amino acid translations were predicted for all genome bins using prodigal version 2.6.3 [57], via the GTDB-Tk, and amino acid translations were searched via HMMs using the MetAnnotate pipeline development release version 0.9.2 [58]. An e-value cutoff of 10^{-40} was used to filter low-quality gene hits. In case of bias due to unassembled or incorrectly binned genes, genome bin-based environmental abundance metrics were cross-

245 compared to abundance metrics generated by directly scanning unassembled metagenome reads, as described in the Supplementary Methods.

Results

Recovery of Chlorobium genome bins

250 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 *Chlorobium* genus (Table 1). 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
255 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 [41]. Recovered genome bins were between 1.9 to 2.6 Mb in size, within the approximate length range of reference genomes of *Chlorobiaceae* (2.0-3.3 Mb; average 2.7 Mb) and were
260 represented by 13 to 208 contigs. Other lower-quality genome bins classified as belonging to the *Chlorobia* were also recovered (Supplementary File 4), 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 strain heterogeneity (25%).

265 The constructed HMM for the *cyc2* gene enabled the recovery of five potential *cyc2* homologues associated with *Chlorobia* from assembled contigs (Fig 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 [59] to contain a porin-like beta barrel structure for membrane binding (Fig. 1A; one example shown) [13]. Aside from the closely related genomes of *Chl. ferrooxidans* and *Chl. phaeoferrooxidans* (Supplementary Figure 1), all other genomes showed substantial genome re-arrangement around the *cyc2* homologue (Fig. 1B). However, detected *cyc2* homologues were always adjacent to a *c5* family cytochrome (Fig. 1B). These *c5* family cytochromes had low sequence identity to one another (i.e., as low as 30%) yet also contained a CXXCH motif near the N-terminal end (Supplementary File 5). 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 (Fig. 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.

Enrichment of sulfide-oxidizing cultures of Chlorobia

Enrichment cultivation was attempted on anoxic water using both sulfide- and ferrous iron-containing anoxic media. Growth was seen on both media tested. Enrichments grown on sulfide-containing media developed the expected 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 (using *Chlorobia*-targeted PCR primers 341f/GSB822r [60] and PGdsrAF/PGdsrAR [61], respectively). However, the appearance of cultures grown on ferrous iron-containing media differed from that of reference photoferrotrophic members of *Chlorobia*. Instead of forming a reddish-brown ferric iron precipitate, the enrichments blackened, potentially indicative of sulfate reduction and the formation of metal sulfides. After blackening, enrichments developed a green colour (Supplementary Figure 2). Suspecting that sulfate reduction was occurring in the ferrous iron enrichment bottles, one Lake 227 enrichment was transitioned to growth on sulfide-containing media and still developed green pigmentation, albeit slowly (Supplementary Figure 2). 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 these enrichments contained *cyc2* homologues. Although L227 enrichment S-6D ceased to grow in laboratory culture, the L304 enrichment S-6D continues to be maintained and purified.

Metabolic diversity and phylogeny of Chlorobium genome bins

Recovered genome bins of *Chlorobia* were metabolically diverse and grouped into four general categories with respect to the presence of genes involved in iron and sulfur cycling (Fig. 2). One of the lake-recovered bins (L227 2013 Bin 56) lacked all 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. The final lake-recovered bin (L227 2013 Bin 22) lacked genes for both sulfur and iron oxidation, which is possibly due to incomplete genome binning. Lastly, the two enrichment culture bins contained genes involved in both sulfide and iron oxidation, similarly to *Chl.* sp. N1 and *Chl. luteolum*.

All recovered *Chlorobia* genome bins contained the *cys* gene cluster for assimilatory sulfate uptake, which is present only in the genomes of potentially photoferrotrophic reference species of *Chlorobia* [21]. The L227 2013 Bin 56 only contained half of the *cys* cluster, but the specific contig containing the cluster was checked and was found to have an assembly break point halfway through the gene cluster, making it likely that the complete *cys* cluster is present in the genome but failed to bin. However, the fact that *Chlorobia*-associated *cyc2* was completely undetectable in Lake 442 metagenome data (see below) implies that, although Lake 442 Bin 64 contains the *cys* cluster, it does not contain *cyc2*. Also, the genome of *Chl.* sp. N1 lacks the *cys* cluster, implying that the cluster is not absolutely necessary for photoferrotrophic growth. As such, the *cys* cluster may not be a reliable indicator of whether *Chlorobia* genomes contain *cyc2*. Two of the genome bins also contained putative homologues to a bacterioferritin, hypothesized by Frigaard and Bryant [21] to play a role in photoferrotrophy.

Phylogenetic analysis of the 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 (Fig. 2). Directly basal to this subgroup was a branch containing *Chl. ferrooxidans* and *Chl. phaeoferrooxidans*. By comparison, the Lake 304 enrichment bin 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* predicted primary sequence phylogeny reveals some congruency between

the two phylogenies for potentially photoferrotrophic members of the *Chlorobia*, with the exception of L227-S-6D and L304-S-6D (Supplementary Figure 3). Overall, all six recovered *Chlorobium* genome bins appear to represent novel species, having an ANI of < 79.6% compared to the genomes of all available type strains (Supplementary Figure 1).

Functional and taxonomic profiling

Read mapping to the genome bins of *Chlorobia* showed that populations with varying genomic potential for iron and sulfur-cycling were relevant to the lake environment (Fig. 3). 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 (Fig. 3A). 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 (Supplementary Figure 4). 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 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 MetAnnotate-based HMM searches (Fig. 3B). Genome bins containing potential *cyc2*

homologues grouped into the *Rhodoferax* or *Gallionella* genera containing known iron-cycling bacteria [17, 62] or into families not associated with iron oxidation (i.e., *Methylomonadaceae* and *Steroidobacteraceae*). The genome bins classified to the *Methylomonadaceae* family were at high relative abundance in the lake metagenomes, suggesting an abundance potentially as high as ~17% in Lake 227 (Fig. 3C). All *Methylomonadaceae*-affiliated bins, except for L442 2011 16.5m Bin 5, encoded particulate methane monooxygenase (*pmoA*) genes. Genome bins containing *dsrA* included known chemolithotrophic sulfide-oxidizing bacteria of the *Sulfuritalea* genus, heterotrophic sulfate-reducing bacteria of the *Desulfatirhabdiaceae* family, a bin classified to the *Smithellaceae* family of the *Syntrophales* order, and a bin classified to the *Rhodoferax* genus [63, 64]. These taxa, with the exception of *Rhodoferax* and *Smithellaceae*, also appeared in unassembled read-based marker gene assessments of the metagenomes (Supplementary Figure 4) and could play a role in the broader iron and sulfur cycles of the lakes.

Chlorobia 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 (Fig. 3C). Predicted relative abundances of populations of *Chlorobia* were 1.2-1.8 times higher when unassembled reads were assessed directly (Supplementary Figure 4), with a predicted maximum relative abundance of 12.3%, showing that high relative abundance is not a result of assembly bias (Fig. 3D). The only other potential chlorophototroph detected in the dataset at >1 % relative abundance was a single genome bin classified to the *Chloroflexaceae* family (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 (Fig. 3C) and contained the *bchL* and *pufM* photosynthesis genes (data not shown). This genome bin was undetected in all other samples.

Discussion

We present genome-resolved functional gene evidence that photoferrotrophs are present and can even be among the most abundant photoautotrophs in the anoxic water columns of seasonally anoxic Boreal Shield lakes. These results represent the first application of genome-resolved metagenomics to an Archaean ocean analogue system and the first use of this approach to distinguish metabolic modes of coexisting, closely related anoxygenic photoautotrophs. Within the context of photoferrotroph ecology, our findings provide genomic context to previous 16S rRNA gene data supporting the existence of photoferrotrophy in Boreal Shield lakes [32]. We also expand sparse knowledge of the diversity of *cyc2* and how *cyc2* is distributed within the *Chlorobia* class.

Our data strongly imply that photoferrotrophy is an active process in Boreal Shield lakes. The highly complete L227 2013 Bin 56 classified to the *Chlorobia* possessed the *cyc2* gene and yet had no genes involved in sulfur oxidation, indicating that the population represented by this genome bin could rely solely on ferrous iron oxidation mediated by *cyc2* for photosynthesis. This population represented as much as 2.8% of the anoxic zone microbial community in Lake 227 based on metagenome read recruitment to the genome bin and occurred at 6 m depth, where the lake generally reoxygenates each year [32]. This implies that the population must actively grow to recolonize that portion of the water column regularly and persists despite periodic reoxygenation events during bi-annual lake turnovers.

We enriched species of *Chlorobia* from Lakes 227 and 304 using standard techniques and a classical sulfide-containing medium, and yet we still recovered novel *cyc2* gene variants and doubled the number of known members of *Chlorobia* that oxidize sulfide but have putative iron-oxidizing genomic potential. Given that this medium has been used to enrich and isolate species of *Chlorobia* from numerous environments globally, and yet only a handful of *Chlorobia* species

containing *cyc2* have been reported to date, our enrichment of two *cyc2*-containing species of *Chlorobia* implies that the lake inoculum source itself is unique, not the cultivation method. This further supports that iron-rich Boreal Shield lake waters are conducive to the robust growth of photoferrotrophs. That *cyc2*-containing species of *Chlorobia* were enriched from both Lakes 227 and 304 also implies that photoferrotrophic potential is distributed more broadly among Boreal Shield lakes beyond Lake 227. Although we have not yet been able to demonstrate iron oxidation in Boreal Shield lake enrichment cultures of *Chlorobia*, it is possible that more carefully tuned cultivation conditions are needed to induce a switch in their metabolic modes, as has been hypothesized for *Chl. luteolum* [20]. Altogether, active photoferrotroph-based photosynthesis in these lakes could partially decouple carbon and sulfur cycles, hinting at broader iron-based metabolisms in Boreal Shield lake anoxic zones that rely on gene pathways and transformations whose function and significance remain poorly understood [65].

Our results indicate the potential coexistence of ferrous iron- and sulfide-oxidizing *Chlorobia* in Boreal Shield lakes. Genome bins of *Chlorobia* with photoferrotrophic genomic potential co-occurred consistently with bins of *Chlorobia* associated with sulfide oxidation at similar relative abundances across all four analyzed Lake 227 metagenomes (Fig. 3A). Such coexistence aligns well with observations from Lake La Cruz [29] and agrees with existing evidence that phototrophic sulfide oxidation can occur in ferruginous systems [28]. Observations from enrichment culturing on ferrous iron-containing media suggest that sulfide oxidation among Boreal Shield lake *Chlorobia* is possible even when sulfide is at trace levels, generated directly from sulfate-reducing bacteria (Supplementary Figure 2). Such “cryptic sulfur cycling” has been reported in various other aquatic systems and is likely an ecologically common phenomenon beyond Boreal Shield lakes [65], possibly favouring phototrophic sulfide oxidation over photoferrotrophy for as yet poorly understood reasons. That the genomes of two enriched species

of *Chlorobia* contained both *cyc2* and the sulfide oxidation gene system also implies that some *Chlorobia* in Boreal Shield lakes, like *Chl.* sp. N1, may be able to switch between these two metabolic pathways via unknown gene regulatory mechanisms.

The lack of detectable photoferrotrophic genomic potential in Lake 442 metagenomes has important implications for the ecology of photoferrotrophy. Although populations of *Chlorobia* were estimated to occur at high relative abundance in Lake 442 based on metagenome data (i.e., 6.5% at 15 m depth in 2014), they only showed the potential for sulfide oxidation (Fig. 3A). *Chlorobia*-affiliated *cyc2* genes were undetectable in L442 metagenomes even among unassembled reads (Supplementary Figure 4). Lake 442 has similar underlying geology to Lake 227 and is nearby (~13.5 km distance). This suggests that photoferrotroph ecology may be heavily influenced by physicochemical drivers within the natural range possible among Boreal Shield lakes. Generally, levels of sulfate in the water column are higher in Lake 442 than Lake 227 (i.e., ~15 μM in the anoxic zone compared to ~7.5 μM [32]), and total dissolved iron concentrations in Lake 442 tend to be lower at the depths where populations of *Chlorobia* are present (i.e., ~5-15 μM compared to ~10-100 μM in Lake 227). It is possible that these differences, or other factors such as pH, light quality, or dissolved organic carbon, could shift the geochemical balance in favour of phototrophic sulfide oxidation. Sulfate and ferrous iron concentrations are near the approximate range predicted for the Archaean ocean (~0.06-0.08 μM sulfate and ~40-120 μM ferrous iron; summarized in [25]), implying that any physicochemical controls determined from Boreal Shield lakes are relevant for estimating photoferrotrophic activity in an Archaean ocean setting. Overall, our data supports that ferruginous conditions alone are insufficient evidence for photoferrotrophy, adding further complexity to the prediction of primary production modes that would have operated in Archaean oceans [66].

Beyond *Chlorobia*, we also demonstrate the potential utility of *cyc2* as an environmental gene marker. Using the same HMM-based search method, we detected *cyc2* homologues in other genome bins, such as bins representing potential methanotrophs in the *Methylomonadaceae* family. These genome bins of *Methylomonadaceae* represent high relative abundance members of the Lake 227 and Lake 442 microbial communities (Fig. 3), and the *cyc2* genes they contain, despite low homology to references, still include the N-terminal CXXCH motif (Supplementary File 5). Another recent study suggested the presence of *cyc2* genes in members of the *Methylomonadaceae* family [13]. It is intriguing that bacteria in this family, thought to be obligately aerobic methane oxidizers, are found at high relative abundance > 2 m below the oxic-anoxic zone boundaries of the lakes. Several authors have reported populations of *Methylomonadaceae* apparently linked to methane oxidation in anoxic environments [67–69]. The possibility of some form of iron cycling or extracellular electron transfer mediated by *cyc2* to supplement aerobic methanotrophy by *Methylomonadaceae* members, as recently suggested for other cytochromes in *Methylococcus capsulatus* Bath [70], would be intriguing to test in an effort to delineate the functional space of the *cyc2* gene product.

Our results expand the known diversity of both *cyc2* and the *Chlorobia* class. We add four novel *cyc2* gene homologues to those known among the *Chlorobia*, and we double the number of *cyc2* homologue-containing genomes of *Chlorobia* available for study. The genomes added include a potentially obligate photoferrotroph lacking a sulfide oxidation pathway and two potentially facultative photoferrotrophs with gene patterns analogous to *Chl. luteolum* or *Chl.* sp. N1(Fig. 2). The addition of these genomes to the reference ribosomal protein tree shows that photoferrotrophic potential does not group monophyletically among *Chlorobia*. Instead, photoferrotrophic potential appears polyphyletic, interspersed with genomes having various forms of sulfur oxidation potential. The thiosulfate oxidation (*sox*) gene system among *Chlorobia*

480 is known to not follow an inheritance pattern matching taxonomic marker genes, and lateral gene transfer within the *Chlorobia* may have been an important factor in its acquisition [21, 71]. However, the *cyc2* predicted primary sequence phylogeny among *Chlorobia* is somewhat congruent to the *Chlorobia* ribosomal protein phylogeny, suggesting that loss from a single common ancestor, rather than extensive lateral gene transfer within the class, may account for the 485 majority of how *cyc2* genes are distributed within the *Chlorobia* class (Supplementary Figure S3). Recovery of additional *cyc2*-containing *Chlorobia* genomes is required to confirm this hypothesis, but it is an intriguing example of the types of evolutionary analyses possible as photoferrotophy among the *Chlorobia* is further studied at the genomic level.

Our study is not the first to implicate photoferrotrophy in freshwater lakes using 490 metagenomic tools. Sinclair and colleagues suggested that two metagenome-assembled genomes classified to *Chlorobi* and *Chloroflexi* recovered from boreal lakes could potentially be photoferrotrophic based on the absence of genes for sulfur cycling, but iron oxidation genes could not be detected in these genomes [72]. In addition, He and colleagues recently hypothesized that *cyc2* may be more broadly involved in extracellular electron transfer, for example to organic 495 compounds, in humic lakes [73]. Although they reported a *cyc2* homologue associated with *Chlorobi* in their metagenome data, the potential for *Chlorobi* populations to use other electron donors, such as reduced sulfur, was not discussed, nor was the context of the detected *Chlorobi* among other photoautotrophs. Our study is unique in comparing coexisting and closely related populations of anoxygenic photoautotrophs and identifying gene clusters involved in distinct 500 photosynthetic modes.

Given incomplete knowledge about the *cyc2*-mediated electron transfer pathway [15], it is possible that alternative electron donors aside from ferrous iron could supply the electrons required for photosynthesis for the populations of *Chlorobia* detected in this study. Such a

hypothesis would be aligned with the proposal of He and colleagues [73] that *cyc2* mediates electron exchange with humic substances. Alternatively, other metals, such as manganese, could potentially be oxidized, although anoxygenic phototrophic manganese oxidation has not yet been demonstrated to occur in nature [74, 75]. Given that *cyc2* has thus far only been associated with iron oxidation in laboratory cultures, and given the high ferrous iron content of these lakes, we argue that ferrous iron electron transfer is the most plausible function for Cyc2 encoded by populations of *Chlorobia* in Boreal Shield lakes.

The potential presence of photoferrotrophs in Boreal Shield lakes has substantial implications for understanding Boreal Shield lake biogeochemistry, as well as for the evolution and ecology of anoxygenic photoautotrophy and the iron cycle. Boreal Shield lake biogeochemistry may be more tightly linked to the iron cycle than previously assumed, making iron a potentially understudied factor affecting the function and management of boreal ecosystems (e.g., [76, 77]). In addition, the new diversity of *cyc2* among *Chlorobia* reported here implies that substantial unmined environmental *cyc2* diversity exists. Thus, recovering genomes of additional photoferrotrophic members of *Chlorobia* from other boreal lake metagenome datasets [78] could help characterize the evolutionary history of iron oxidation and elucidate the full iron oxidation pathway utilized by the *Chlorobia*. Lastly, initial evidence of strong controls on photoferrotrophy in Boreal Shield lakes implies that further lake surveys could be used to build a conceptual framework for how photoferrotrophy may be modulated by environmental factors. Because this framework could be applied to Archaean contexts to postulate the role of photoferrotrophy in primary production over geological timeframes, such future work could help elucidate dynamics of photoferrotrophy in both ancient and modern ecosystems.

By recovering highly complete and low-contamination genome bins representing key anoxygenic phototroph populations from ferruginous Boreal Shield lakes, we show that these

lakes differ ecologically and functionally with respect to anoxygenic photosynthesis compared to classical lake models. We provide gene pathway-level evidence that some Boreal Shield lake populations of *Chlorobia* rely on *cyc2*-driven photoferrotrophy as their photosynthetic mode, and we propose that strong ecological controls may exist governing photoferrotrophy in Boreal Shield lakes. We also provide an expanded knowledge base for *cyc2*, including a new *cyc2*-containing enrichment of *Chlorobia*, as genetic studies continue to progress in linking *cyc2* to its specific cellular function [79]. Together, these findings serve as an important basis for future work probing the rates, gene expression, and enrichment culture activity of potentially photoferrotrophic *Chlorobia* in Boreal Shield lakes. Beyond being a process relevant to prehistoric environments, photoferrotrophy appears to be a common metabolic process in Boreal Shield lakes globally, with unexplored metabolic diversity and ecological controls relevant to understanding early life on Earth.

Data availability [to be provided before publication]

Raw metagenome reads for the freshwater lake and enrichment culture metagenomes are available in the NCBI sequence read archive (SRA) under BioProjects PRJNA518727 and PRJNA534305, respectively. Six curated genome bins of *Chlorobia* are available under the same BioProjects, along with all assembled contigs from freshwater 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 is available at https://github.com/jmtsuji/Chlorobia-cyc2-genomics (doi:10.5281/zenodo.3228523).

Acknowledgements

We thank R. Henderson and staff at the IISD-ELA for assistance in field sampling and interpretation of the dynamics of ELA lakes. In addition, we thank S. A. Crowe and M. Tank for valuable feedback on the manuscript, K. J. Thompson for helpful discussion related to the work, L. H. Bergstrand for assistance with bioinformatics, and K. E. Engel, R. C. Beaver, and N. A. Shaw for assistance with laboratory tasks. JMT thanks M. Jetten, C. Welte, and staff at the Soehngen Institute of Anaerobic Microbiology (SIAM) for providing helpful training for anaerobic microbiology methods. This research was supported by Discovery and Strategic Projects Grants from the National Sciences and Engineering Research Council of Canada (NSERC).

Author contributions

SLS, JJV, and LM were involved in the overall lake sampling project that led to this work and provided input and assistance with sample collection and data interpretation. JMT and JDN designed the experimental work in this study. JMT and NT performed enrichment cultivation. JMT analyzed the sequence data and drafted the manuscript with the edits and feedback of all authors.

Competing interests

The authors declare no competing financial interests.

Supplementary information is available at The ISME Journal's website.

Supplementary information

Supplementary Methods

Supplementary Figure 1 – Average nucleotide identity between recovered genome bins and reference genomes of *Chlorobia*. The same ribosomal protein tree as in Figure 2 is shown to the left of the heatmap.

Supplementary Figure 2 – Photographs enrichment cultures of *Chlorobia* in ferrous iron- and sulfide-containing media.

Supplementary Figure 3 – Tanglegram comparing the concatenated ribosomal protein phylogeny (Fig. 2) and Cyc2 phylogeny (Fig. 1C) among *Chlorobia*. The two tips whose placements differ between the two phylogenies are highlighted in blue or green.

Supplementary Figure 4 – Bubble plot showing taxonomic and functional profiling of unassembled metagenome data, in the same general layout as Figure 3. 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% 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.

Supplementary File 1 – ZIP file containing the YAML config files for the ATLAS metagenome assemblies performed for this publication.

Supplementary File 2 – ZIP file of two alternative profile Hidden Markov Models for the *cyc2* gene (general or *Chlorobia*-specific).

Supplementary File 3 – Excel file containing accessions of reference genomes of *Chlorobia* and statistics from the reciprocal BLAST search of Fe/S gene pathways.

Supplementary File 4 – CSV file containing genome statistics for all dereplicated MAGs from this study.

Supplementary File 5 – ZIP file of amino acid multiple sequence alignments: alignment of *Chlorobia*-associated Cyc2 predicted primary sequences from this study with reference sequences (unmasked and masked), alignment of c5 family cytochromes presented in this study, and alignment of additional low-homology Cyc2 hits from metagenomes.

References

1. Van Gemerden H, Mas J. Ecology of phototrophic sulfur bacteria. In: Blankenship RE, Madigan MT, Bauer CE (eds). *Anoxygenic Photosynthetic Bacteria*. 1995. Springer Netherlands, Dordrecht, pp 49–85.

2. Tank M, Thiel V, Ward DM, Bryant DA. 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*. 2017. Springer International Publishing Switzerland, pp 87–137.

3. Manske AK, Glaeser J, Kuypers MMM, Overmann J. Physiology and phylogeny of green sulfur bacteria forming a monospecific phototrophic assemblage at a depth of 100 meters in the Black Sea. *Appl Environ Microbiol* 2005; **71**: 8049–8060.

4. Imhoff JF. The family *Chromatiaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds). *The Prokaryotes*. 2014. Springer Berlin Heidelberg, pp 151–178.

5. Imhoff JF. The family *Chlorobiaceae*. In: Rosenberg E, DeLong EF, Lory S, Stackebrandt E, Thompson F (eds). *The Prokaryotes*. 2014. Springer Berlin Heidelberg, pp 501–514.

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

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

8. Ehrenreich A, Widdel F. Anaerobic oxidation of ferrous iron by purple bacteria, a new type of phototrophic metabolism. *Appl Environ Microbiol* 1994; **60**: 4517–4526.

9. Poulton SW, Canfield DE. Ferruginous conditions: a dominant feature of the ocean through Earth's history. *Elements* 2011; **7**: 107–112.

10. Kappler A, Pasquero C, Konhauser KO, Newman DK. Deposition of banded iron formations by anoxygenic phototrophic Fe(II)-oxidizing bacteria. *Geology* 2005; **33**: 865–868.

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

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

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

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

15. Castelle C, Guiral M, Malarte G, Ledgham F, Leroy G, Brugna M, 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**: 25803–25811.

16. Barco RA, Emerson D, Sylvan JB, Orcutt BN, Meyers MEJ, Ramírez GA, 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**: 5927–5937.

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

660 18. Heising S, Richter L, Ludwig W, Schink B. *Chlorobium ferrooxidans* sp. nov., a phototrophic green sulfur bacterium that oxidizes ferrous iron in coculture with a "*Geospirillum*" sp. strain. *Arch Microbiol* 1999; **172**: 116–124.

19. Llirós M, García–Armisen T, Darchambeau F, Morana C, Triadó–Margarit X, Inceoğlu Ö, et al. Pelagic photoferrotrophy and iron cycling in a modern ferruginous basin. *Sci Rep* 2015; **5**:
665 13803.

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

21. Frigaard N-U, Bryant DA. Genomic insights into the sulfur metabolism of phototrophic
670 green sulfur bacteria. In: Hell R, Dahl DC, Knaff D, Leustek T (eds). *Sulfur Metabolism in Phototrophic Organisms*. 2008. Springer Netherlands, pp 337–355.

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

675 23. Thompson KJ, Simister RL, Hahn AS, Hallam SJ, Crowe SA. Nutrient acquisition and the metabolic potential of photoferrotrophic *Chlorobi*. *Front Microbiol* 2017; **8**: 1212.

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

680 25. Koeksoy E, Halama M, Konhauser KO, Kappler A. Using modern ferruginous habitats to interpret Precambrian banded iron formation deposition. *Int J Astrobiol* 2016; **15**: 205–217.

26. Savvichev AS, Kokryatskaya NM, Zabelina SA, Rusanov II, Zakharova EE, Veslopolova EF, 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**: 659–672.

27. Crowe SA, Jones C, Katsev S, Magen C, O'Neill AH, Sturm A, et al. Photoferrotrophs thrive in an Archean Ocean analogue. *Proc Nat Acad Sci USA* 2008; **105**: 15938–15943.

28. Crowe SA, Maresca JA, Jones C, Sturm A, Henny C, Fowle DA, et al. Deep-water anoxygenic photosythesis in a ferruginous chemocline. *Geobiology* 2014; **12**: 322–339.

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

30. Morana C, Roland FAE, Crowe SA, Llirós M, Borges AV, Darchambeau F, et al. Chemoautotrophy and anoxygenic photosynthesis within the water column of a large meromictic tropical lake (Lake Kivu, East Africa). *Limnol Oceanogr* 2016; **61**: 1424–1437.

31. Berg JS, Jézéquel D, Duverger A, Lamy D, Laberty-Robert C, Miot J. Microbial diversity involved in iron and cryptic sulfur cycling in the ferruginous, low-sulfate waters of Lake Pavin. *PLOS ONE* 2019; **14**: e0212787.

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

33. White III RA, Brown J, Colby S, Overall CC, Lee J-Y, Zucker J, 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.

705 34. Jeans C, Singer SW, Chan CS, VerBerkmoes NC, Shah M, Hettich RL, 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**: 542–550.

35. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 2011;
710 **7**: 539.

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

37. Talavera G, Castresana J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* 2007; **56**: 564–577.

38. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective
715 stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015; **32**: 268–274.

39. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017; **14**: 587–589.

40. Olm MR, Brown CT, Brooks B, Banfield JF. dRep: a tool for fast and accurate genomic
720 comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 2017; **11**: 2864–2868.

41. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015; **25**: 1043–1055.

725 42. Eren AM, Esen ÖC, Quince C, Vineis JH, Morrison HG, Sogin ML, et al. Anvi'o: an advanced analysis and visualization platform for 'omics data. *PeerJ* 2015; **3**: e1319.

43. Rissman AI, Mau B, Biehl BS, Darling AE, Glasner JD, Perna NT. Reordering contigs of draft genomes using the Mauve Aligner. *Bioinformatics* 2009; **25**: 2071–2073.

44. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014; **30**: 2068–2069.

45. Armstrong FAJ, Schindler DW. Preliminary chemical characterization of waters in the Experimental Lakes Area, northwestern Ontario. *J Fish Res Bd Can* 1971; **28**: 171–187.

46. Schindler DW. Eutrophication and recovery in experimental lakes: implications for lake management. *Science* 1974; **184**: 897–899.

47. Schindler D. The coupling of elemental cycles by organisms: evidence from whole-lake chemical perturbations. In: Stumm W (ed). *Chemical Processes in Lakes*. 1985. John Wiley and Sons, New York, NY, pp 225–250.

48. Hegler F, Posth NR, Jiang J, Kappler A. Physiology of phototrophic iron(II)-oxidizing bacteria: implications for modern and ancient environments. *FEMS Microbiol Ecol* 2008; **66**: 250–260.

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

50. 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**: 141–145.

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

52. Lee MD. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* 2019; btz188.

53. Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, et al. A new view of the tree of life. *Nat Microbiol* 2016; **1**: 16048.

54. Bergstrand LH, Cardenas E, Holert J, Hamme JDV, Mohn WW. Delineation of Steroid-Degrading Microorganisms through Comparative Genomic Analysis. *mBio* 2016; **7**: e00166-16.

755 55. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018.

56. Fish JA, Chai B, Wang Q, Sun Y, Brown CT, Tiedje JM, et al. FunGene: the functional gene pipeline and repository. *Front Microbiol* 2013; **4**: 291.

760 57. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010; **11**: 1–11.

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

765 59. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nature Methods* 2015; **12**: 7–8.

60. 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.

770 61. Mori Y, Purdy KJ, Oakley BB, Kondo R. Comprehensive detection of phototrophic sulfur bacteria using PCR primers that target reverse dissimilatory sulfite reductase gene. *Microbes Environ* 2010; **25**: 190–196.

62. Finneran KT, Johnsen CV, Lovley DR. *Rhodoferax ferrireducens* sp. nov., a psychrotolerant, facultatively anaerobic bacterium that oxidizes acetate with the reduction of

775 Fe(III). *Int J Syst Evol Micr* 2003; **53**: 669–673.

63. 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**: 387–395.

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

65. Kappler A, Bryce C. Cryptic biogeochemical cycles: unravelling hidden redox reactions. *Environ Microbiol* 2017; **19**: 842–846.

66. Posth NR, Konhauser KO, Kappler A. Microbiological processes in banded iron formation deposition. *Sedimentology* 2013; **60**: 1733–1754.

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

68. Rissanen AJ, Saarenheimo J, Tiirola M, Peura S, Aalto SL, Karvinen A, et al. Gammaproteobacterial methanotrophs dominate methanotrophy in aerobic and anaerobic layers of boreal lake waters. *Aquat Microb Ecol* 2018; **81**: 257–276.

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

70. Tanaka K, Yokoe S, Igarashi K, Takashino M, Ishikawa M, Hori K, et al. Extracellular electron transfer via outer membrane cytochromes in a methanotrophic bacterium *Methylococcus capsulatus* (Bath). *Front Microbiol* 2018; **9**: 2905.

71. Meyer B, Imhoff JF, Kuever J. Molecular analysis of the distribution and phylogeny of the *soxB* gene among sulfur-oxidizing bacteria – evolution of the Sox sulfur oxidation enzyme system. *Environ Microbiol* 2007; **9**: 2957–2977.

72. Sinclair L, Peura S, Hernandez P, Schattenhofer M, Eiler A. Novel chemolithotrophic and anoxygenic phototrophic genomes extracted from ice-covered boreal lakes. *bioRxiv* 2017; 139212.

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

74. Johnson JE, Webb SM, Thomas K, Ono S, Kirschvink JL, Fischer WW. Manganese-oxidizing photosynthesis before the rise of cyanobacteria. *Proc Nat Acad Sci USA* 2013; **110**: 11238–11243.

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

76. Molot LA, Watson SB, Creed IF, Trick CG, McCabe SK, Verschoor MJ, et al. A novel model for cyanobacteria bloom formation: the critical role of anoxia and ferrous iron. *Freshw Biol* 2014; **59**: 1323–1340.

77. Bravo AG, Peura S, Buck M, Ahmed O, Mateos-Rivera A, Ortega SH, et al. Methanogens and iron-reducing bacteria: the overlooked members of mercury methylating microbial communities in boreal lakes. *Appl Environ Microbiol* 2018; AEM.01774-18.

78. Linz AM, He S, Stevens SLR, Anantharaman K, Rohwer RR, Malmstrom RR, et al. Freshwater carbon and nutrient cycles revealed through reconstructed population genomes. *PeerJ* 2018; **6**: e6075.

79. Chan C, McAllister SM, Garber A, Hallahan BJ, Rozovsky S. Fe oxidation by a fused cytochrome-porin common to diverse Fe-oxidizing bacteria. *bioRxiv* 2018; 228056.

80. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods* 2011; **8**: 785–786.

825 81. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004; **25**: 1605–1612.

830

Table and figure captions

Table 1. 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. A summary of bin statistics for lower-quality genome bins is included in Supplementary File 4.

Figure 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 [80]) and visualized using UCSF Chimera [81]. 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

855 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. The full sequence alignment is available in Supplementary File 5. Note that
860 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.

Figure 2. 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
865 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
870 presence/absence of genes implicated in iron and sulfur metabolism among *Chlorobia* based on reciprocal blastp. Heatmap tiles are shaded based on the percent amino acid identity of an identified gene compared to the reference sequence (*Chl. ferrooxidans* for iron-related genes, and *Chl. phaeoclathratiforme* for sulfur-related genes). Although the cytochrome c5 family gene upstream of *cyc2* was not searched for directly due to poor sequence homology, the gene was
875 verified manually to be adjacent to all hits of *cyc2*.

Figure 3. 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
880 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 1 and Fig. 2. In parentheses beside each name are bin quality statistics reported by CheckM – the predicted % completion, % contamination, and % strain heterogeneity, respectively. The displayed bins all
885 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%
890 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)
895 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. Supplementary Figure 4 shows relative abundances based on predictions from unassembled metagenome reads for comparison.

Table 1

Bin ID	Contigs	Length (Mb)	N50 (kb)	L50	GC content (%)	Genes	tRNA genes	Complete-ness (%)	Contam-ination (%)
L227 2013 Bin 55	208	2.38	16.7	42	48.5	2397	39	97.39	0.11
L227 2013 Bin 56	140	2.09	27.6	21	43.9	2009	38	96.7	0.57
L227 2013 Bin 22	57	2.55	79.5	12	45.9	2471	37	98.5	2.75
L442 Bin 64	128	1.9	28.6	22	48.3	1871	35	90.11	0
L227 enr. S-6D Bin 1	13	2.5	306.7	3	46.8	2385	44	99.45	1.01
L304 enr. S-6D Bin 1	15	2.61	271.5	3	48.7	2504	45	99.43	0

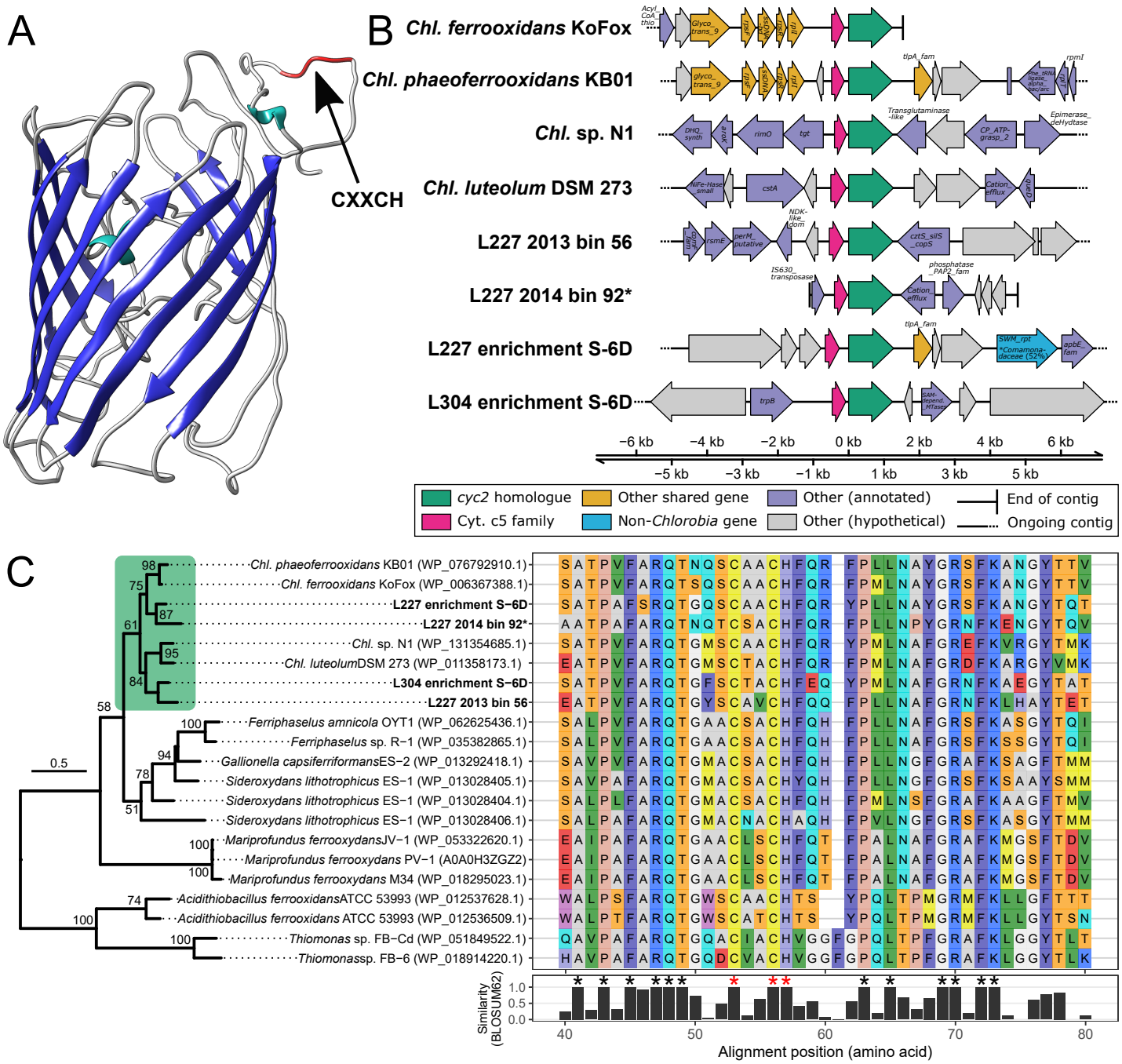
A
CXXCH
B
Chl. ferrooxidans KoFox
Chl. phaeoferrooxidans KB01
Chl. sp. N1
Chl. luteolum DSM 273
L227 2013 bin 56
L227 2014 bin 92*
L227 enrichment S-6D
L304 enrichment S-6D
−6 kb −4 kb −2 kb 0 kb 2 kb 4 kb 6 kb
−5 kb −3 kb −1 kb 1 kb 3 kb 5 kb
cyc2 homologue
Cyt. c5 family
Other shared gene
Non-Chlorobia gene
Other (annotated)
Other (hypothetical)
End of contig
Ongoing contig
C
Chl. phaeoferrooxidans KB01 (WP_076792910.1)
Chl. ferrooxidans KoFox (WP_006367388.1)
L227 enrichment S−6D
L227 2014 bin 92*
Chl. sp. N1 (WP_131354685.1)
Chl. luteolumDSM 273 (WP_011358173.1)
L304 enrichment S−6D
L227 2013 bin 56
Ferriphaselus amnicola OYT1 (WP_062625436.1)
Ferriphaselus sp. R−1 (WP_035382865.1)
Gallionella capsiferriformansES−2 (WP_013292418.1)
Sideroxydans lithotrophicus ES−1 (WP_013028405.1)
Sideroxydans lithotrophicus ES−1 (WP_013028404.1)
Sideroxydans lithotrophicus ES−1 (WP_013028406.1)
Mariprofundus ferrooxydansJV−1 (WP_053322620.1)
Mariprofundus ferrooxydans PV−1 (A0A0H3ZGZ2)
Mariprofundus ferrooxydans M34 (WP_018295023.1)
Acidithiobacillus ferrooxidansATCC 53993 (WP_012537628.1)
Acidithiobacillus ferrooxidans ATCC 53993 (WP_012536509.1)
Thiomonas sp. FB−Cd (WP_051849522.1)
Thiomonassp. FB−6 (WP_018914220.1)
0.5
Similarity (BLOSUM62)
Alignment position (amino acid)

Fig. 1

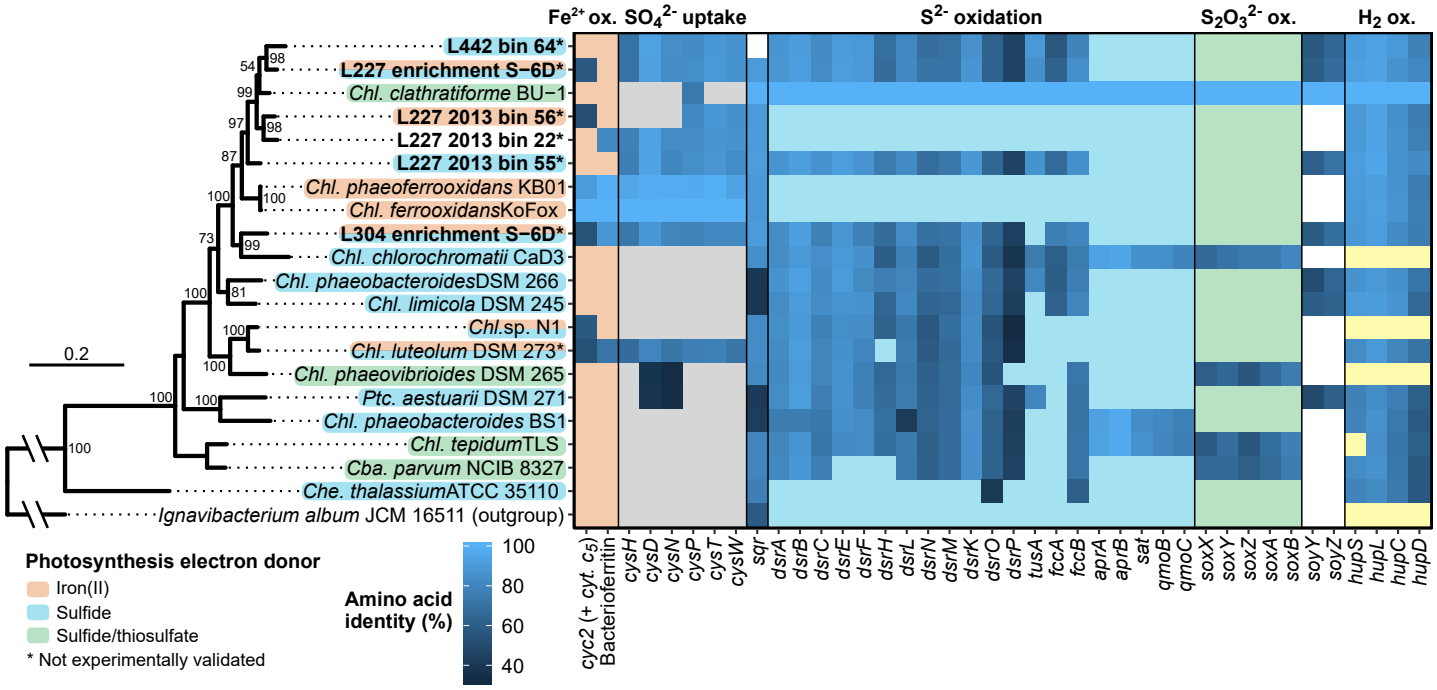
Fe2+ ox.
SO42- uptake
S2- oxidation
S2O32- ox.
H2 ox.
L442 bin 64*
L227 enrichment S−6D*
Chl. clathratiforme BU−1
L227 2013 bin 56*
L227 2013 bin 22*
L227 2013 bin 55*
Chl. phaeoferrooxidans KB01
Chl. ferrooxidansKoFox
L304 enrichment S−6D*
Chl. chlorochromatii CaD3
Chl. phaeobacteroidesDSM 266
Chl. limicola DSM 245
Chl.sp. N1
Chl. luteolum DSM 273*
Chl. phaeovibrioides DSM 265
Ptc. aestuarii DSM 271
Chl. phaeobacteroides BS1
Chl. tepidumTLS
Cba. parvum NCIB 8327
Che. thalassiumATCC 35110
Ignavibacterium album JCM 16511 (outgroup)
0.2
cyc2 (+ cyt. c5)
Bacterioferritin
cysH
cysD
cysN
cysP
cysT
cysW
sqr
dsrA
dsrB
dsrC
dsrE
dsrF
dsrH
dsrL
dsrN
dsrM
dsrK
dsrO
dsrP
tusA
fccA
fccB
aprA
aprB
sat
qmoB
qmoC
soxX
soxY
soxZ
soxA
soxB
soyY
soyZ
hupS
hupL
hupC
hupD
Photosynthesis electron donor
Iron(II)
Sulfide
Sulfide/thiosulfate
* Not experimentally validated
Amino acid identity (%)
100
80
60
40

Fig. 2

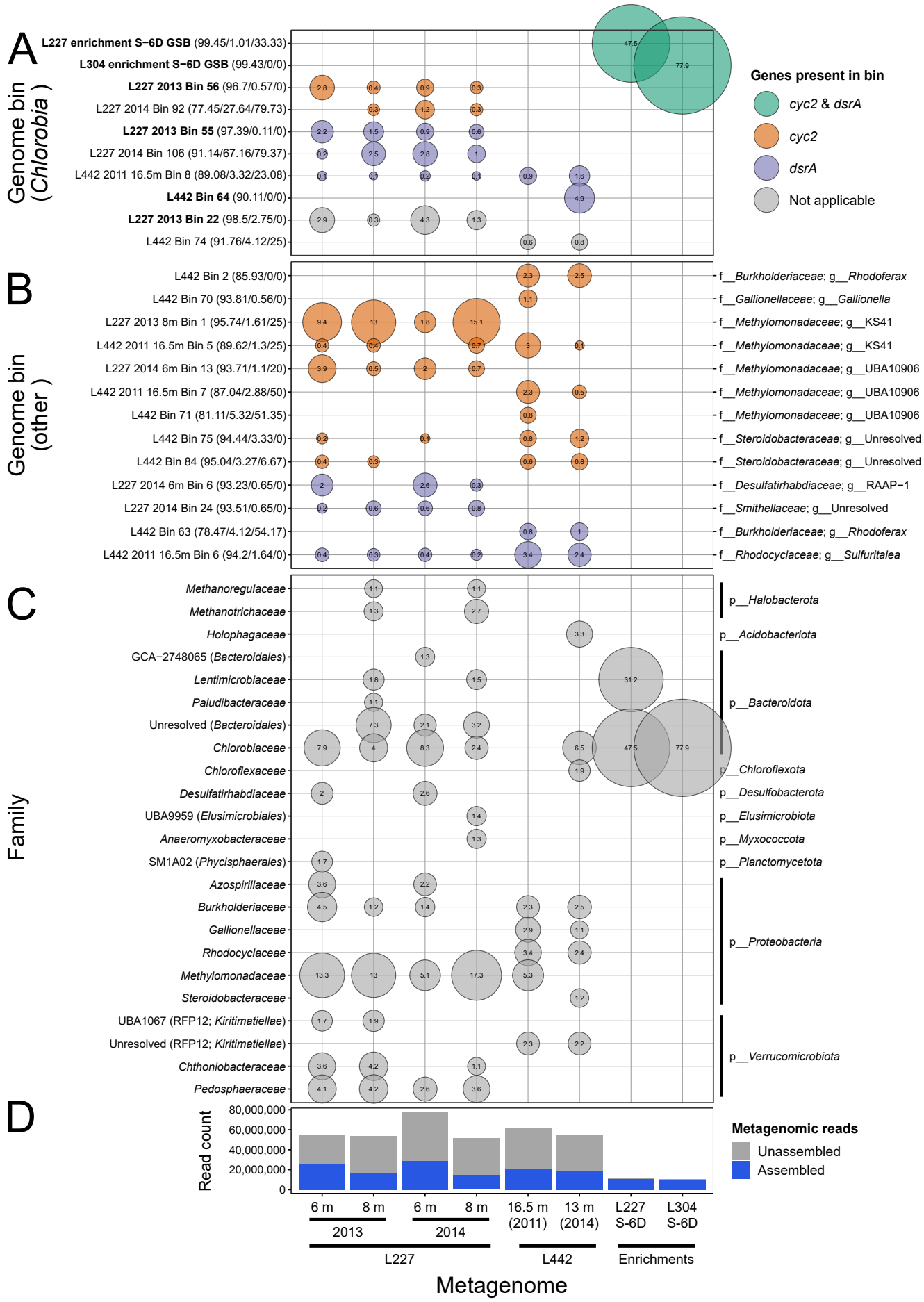
A
Genome bin (Chlorobia)
L227 enrichment S−6D GSB (99.45/1.01/33.33)
L304 enrichment S−6D GSB (99.43/0/0)
L227 2013 Bin 56 (96.7/0.57/0)
L227 2014 Bin 92 (77.45/27.64/79.73)
L227 2013 Bin 55 (97.39/0.11/0)
L227 2014 Bin 106 (91.14/67.16/79.37)
L442 2011 16.5m Bin 8 (89.08/3.32/23.08)
L442 Bin 64 (90.11/0/0)
L227 2013 Bin 22 (98.5/2.75/0)
L442 Bin 74 (91.76/4.12/25)
Genes present in bin
cyc2 & dsrA
cyc2
dsrA
Not applicable
B
Genome bin (other)
L442 Bin 2 (85.93/0/0)
L442 Bin 70 (93.81/0.56/0)
L227 2013 8m Bin 1 (95.74/1.61/25)
L442 2011 16.5m Bin 5 (89.62/1.3/25)
L227 2014 6m Bin 13 (93.71/1.1/20)
L442 2011 16.5m Bin 7 (87.04/2.88/50)
L442 Bin 71 (81.11/5.32/51.35)
L442 Bin 75 (94.44/3.33/0)
L442 Bin 84 (95.04/3.27/6.67)
L227 2014 6m Bin 6 (93.23/0.65/0)
L227 2014 Bin 24 (93.51/0.65/0)
L442 Bin 63 (78.47/4.12/54.17)
L442 2011 16.5m Bin 6 (94.2/1.64/0)
f__Burkholderiaceae; g__Rhodoferax
f__Gallionellaceae; g__Gallionella
f__Methylomonadaceae; g__KS41
f__Methylomonadaceae; g__KS41
f__Methylomonadaceae; g__UBA10906
f__Methylomonadaceae; g__UBA10906
f__Methylomonadaceae; g__UBA10906
f__Steroidobacteraceae; g__Unresolved
f__Steroidobacteraceae; g__Unresolved
f__Desulfatirhabdiaceae; g__RAAP−1
f__Smithellaceae; g__Unresolved
f__Burkholderiaceae; g__Rhodoferax
f__Rhodocyclaceae; g__Sulfuritalea
C
Family
Methanoregulaceae
Methanotrichaceae
Holophagaceae
GCA−2748065 (Bacteroidales)
Lentimicrobiaceae
Paludibacteraceae
Unresolved (Bacteroidales)
Chlorobiaceae
Chloroflexaceae
Desulfatirhabdiaceae
UBA9959 (Elusimicrobiales)
Anaeromyxobacteraceae
SM1A02 (Phycisphaerales)
Azospirillaceae
Burkholderiaceae
Gallionellaceae
Rhodocyclaceae
Methylomonadaceae
Steroidobacteraceae
UBA1067 (RFP12; Kiritimatiellae)
Unresolved (RFP12; Kiritimatiellae)
Chthoniobacteraceae
Pedosphaeraceae
p__Halobacterota
p__Acidobacteriota
p__Bacteroidota
p__Chloroflexota
p__Desulfobacterota
p__Elusimicrobiota
p__Myxococcota
p__Planctomycetota
p__Proteobacteria
p__Verrucomicrobiota
D
Read count
80,000,000
60,000,000
40,000,000
20,000,000
0
Metagenomic reads
Unassembled
Assembled
6 m
8 m
6 m
8 m
16.5 m (2011)
13 m (2014)
L227 S-6D
L304 S-6D
2013
2014
L227
L442
Enrichments
Metagenome

Fig. 3