Freshwater *Chlorobia* exhibit metabolic specialization among cosmopolitan and endemic populations

Authors: Sarahi L. Garcia [a,b,*] (sarahi.garcia@su.se), Maliheh Mehrshad [a,c,*] (maliheh.mehrshad@slu.se), Moritz Buck [a,c,*] (moritz.buck@slu.se), Jackson M. Tsuji [d] (jackson.tsuji@uwaterloo.ca), Josh D. Neufeld [d] (jneufeld@uwaterloo.ca), Katherine D. McMahon [e] (trina.mcmahon@wisc.edu), Stefan Bertilsson [a,c] (stefan.bertilsson@slu.se), Chris Greening [f] (chris.greening@monash.edu), Sari Peura [g] (sari.peura@slu.se)

* these authors contributed equally

Affiliations:

[a] Department of Ecology and Genetics, Limnology, Uppsala University, Norbyvägen 18 D, 75236 Uppsala, Sweden.

[b] Department of Ecology, Environment, and Plant Sciences, Science for Life Laboratory, Stockholm University, 10691 Stockholm, Sweden

[c] Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Lennart Hjelms väg 9, 75651 Uppsala, Sweden

[d] Department of Biology, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada, N2L 3G1

[e] Departments of Civil and Environmental Engineering, and Bacteriology, University of Wisconsin, Madison, 1550 Linden Drive, Madison, WI, 53706

[f] Department of Microbiology, Biomedicine Discovery Institute, Monash University, 23 Innovation Walk, Clayton, VIC 3800, Australia

[g] Department of Forest Mycology and Plant Pathology, Science for Life Laboratory, Swedish University of Agricultural Sciences, Almas Allé 5, 750 07 Uppsala, Sweden

Abstract

Photosynthetic bacteria from the class *Chlorobia* (formerly phylum *Chlorobi*) sustain carbon fixation in anoxic water columns. They harvest light at extremely low intensities and use various inorganic electron donors to fix carbon dioxide into biomass. Until now, most information on their functional ecology and local adaptations came from isolates and merely 26 sequenced genomes that are poor representatives of natural populations. To address these limitations, we analyzed global metagenomes to profile planktonic *Chlorobia* cells from the oxyclines of 42 freshwater bodies, spanning subarctic to tropical regions and encompassing all four seasons. We assembled and compiled over 500 genomes, including metagenome-assembled genomes (MAGs), single-cell genomes (SAGs), and reference genomes from cultures, clustering them into 71 metagenomic operational taxonomic units (mOTUs) or "species". Of the 71 mOTUs, 57 were classified as genus *Chlorobium* and these mOTUs varied in relative abundance up to ~60% of the microbial communities in the sampled anoxic waters. Several *Chlorobium*-associated mOTUs were globally distributed whereas others were endemic to individual lakes. Although most clades encoded the ability to oxidize hydrogen, many were lacking genes for the oxidation of specific sulfur and iron substrates. Surprisingly, one globally distributed Scandinavian *Chlorobium* clade encoded the ability to oxidize hydrogen, sulfur, and iron, suggesting that metabolic versatility facilitated such widespread colonization. Overall, these findings provide new insights into the biogeography of the *Chlorobia* and the metabolic traits that facilitate niche specialization within lake ecosystems.

Importance

The reconstruction of genomes from metagenomes has enabled unprecedented insights into the ecology and evolution of environmental microbiomes. We applied this powerful approach to 274 metagenomes collected from diverse freshwater habitats that spanned oxic and anoxic zones, sampling seasons, and latitudes. We demonstrate widespread and abundant distributions of planktonic *Chlorobia*-associated bacteria in hypolimnetic waters of stratified freshwater ecosystems and pinpoint nutrients that likely fuel their electron chains. Being photoautotrophs, these *Chlorobia* organisms also have the potential to serve as carbon sources that support metalimnetic and hypolimnetic food webs.

2

Introduction
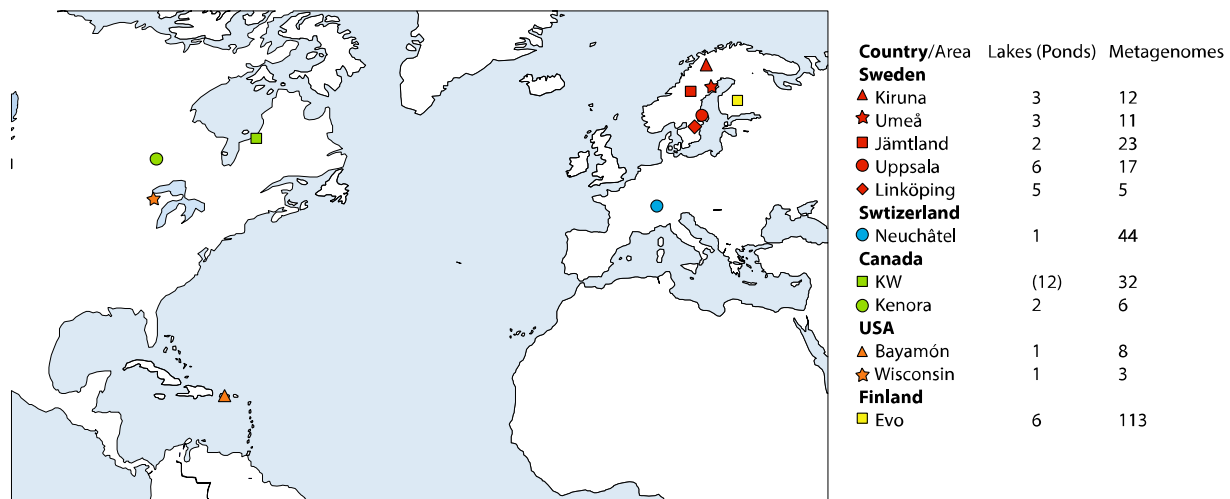
Although oxygenic phototrophs dominate contemporary carbon fixation, anoxygenic phototrophs have been important over planetary timescales and continue to occupy important niches in a wide range of ecosystems (1). Among the anoxygenic phototrophs, members of the class *Chlorobia* (green sulfur bacteria; formerly phylum *Chlorobi*) are physiologically remarkable. Known *Chlorobia* members are obligate photolithoautotrophic anaerobes (2) that leverage bacteriochlorophyll-rich organelles called chlorosomes to harvest light at extremely low intensities (i.e., in the range of 1–10 nmol photons $m^{-2} s^{-1}$) (3). Such efficient light harvesting determines the ecological niche of these bacteria at the lowermost strata of the photic zone in stratified water columns, where the least amount of light is available (4, 5). Studies on isolates obtained from different aquatic systems and microbial mats have revealed that members of class *Chlorobia* use electrons derived from reduced sulfur compounds and/or hydrogen to fix carbon dioxide via the reverse tricarboxylic acid (TCA) cycle (6-8). Although several strains can oxidize thiosulfate, they almost universally use sulfide as electron donor for CO_2 reduction and source of sulfur for assimilation (2). Several members of the class *Chlorobia* are additionally known to oxidize ferrous iron (Fe^{2+}) in a process called photoferrotrophy (9). Such photoferrotrophs include *Chlorobium ferrooxidans*, which obtains sulfur via assimilatory sulfate reduction, along with three other recently characterized strains (10-13). Although all characterized *Chlorobia* pure cultures can grow with carbon dioxide (CO_2) or bicarbonate as their sole carbon source, many isolates have been reported to assimilate simple organic acids, such as acetate and pyruvate, under photomixotrophic or photoheterotrophic conditions (2). Moreover, most *Chlorobia* cells also encode the potential for nitrogen fixation (14-16). Together, this metabolic capacity and versatility enables members of the class *Chlorobia* to colonize dimly lit anoxic environments.

In stratified freshwater ecosystems, *Chlorobia* microorganisms have particularly important ecological and biogeochemical roles (17-20). In such systems they accumulate to form dense populations at certain strata, manifesting as deep chlorophyll maxima (21), and they also play prominent roles in carbon budgets of such lakes (22, 23) while concomitantly contributing to various processes in sulfur, iron, hydrogen, and nitrogen cycling (16, 24, 25). However, our knowledge of the geographical distribution and metabolic capabilities of *Chlorobia* members within natural planktonic communities remains incomplete. This is because most of our understanding of the metabolic capabilities and ecophysiological strategies of this class come from studies of approximately 100 isolates (4), with genomes available for 26 of these representatives (2, 26, 27), and cultures are often poor representatives of natural populations. The diversity and distributions of planktonic *Chlorobia* members are largely unknown. Moreover, the question still remains: how are genes conferring distinct ecophysiological traits distributed among

Chlorobia-associated genomes from geographically distinct water bodies?

In this study, we used a cultivation-independent approach to gain a more comprehensive understanding of the natural distributions and metabolic capabilities of *Chlorobia* members within 42 freshwater bodies that are distributed across boreal, subarctic, and tropical regions in Europe and North America (Figure 1).

90   Through this extensive study we assembled and analyzed over 500 genomes of *Chlorobia* members, including 465 metagenome-assembled genomes (MAGs) and 21 single-amplified genomes (SAGs). These were further clustered into 71 metagenome operational taxonomic units (mOTUs; where each mOTU is a cluster of genomes that share 95% average nucleotide identity). We observed that some of these mOTUs exhibited cosmopolitan distributions, whereas other lineages were locally constrained and

95   possibly adapted to specific lakes or regions. In addition, clades varied in their encoded metabolic capabilities and we found several mOTUs that appear consistent with substrate specialization. These findings reveal further metabolic versatility and niche specialization within the abundant members of *Chlorobium*-associated bacteria.

Country/Area	Lakes (Ponds)	Metagenomes
Sweden		
Kiruna	3	12
Umeå	3	11
Jämtland	2	23
Uppsala	6	17
Linköping	5	5
Swtizerland		
Neuchâtel	1	44
Canada		
KW	(12)	32
Kenora	2	6
USA		
Bayamón	1	8
Wisconsin	1	3
Finland		
Evo	6	113

100

Figure 1. Areas from which we gathered lake or pond metagenomes for mapping and recovery of metagenome-assembled genomes (MAGs) and single-cell genomes (SAGs). Symbols are colored by country, and a unique shape is used for each area within the same country. KW stands for Kujjuarapik-Whapmagoostui.

105   **Results and discussion**

Metagenome assembly and characterization reveals that most planktonic members of the class Chlorobia affiliate with the genus Chlorobium

In this study we investigated the planktonic *Chlorobia*-affiliated populations of 30 lakes and 12 ponds in different regions of Europe and North America (Figure 1). Most of these samples were collected with the

110   objective of investigating depth-associated changes in taxonomic diversity in different water masses that

4

capture oxic epilimnion, metalimnion, and anoxic hypolimnion conditions (28). Following assembly and binding of individual metagenomes, we compiled 454 new metagenome-assembled genomes (MAGs) that affiliated with the class *Chlorobia.* Moreover, we collected 19 SAGs belonging to the class *Chlorobia* from two of the sampled lakes. To characterize the phylogenetic distributions of these new MAGs and

115   SAGs, we further compiled 25 genomes available from the Genome Taxonomy Database (GTDB) and 11 MAGs available from previous studies (26, 29-31). In total, the dataset included 509 genomes including the MAGs, SAGs, and complete genomes of isolates (Table S1). Genome completeness varied from 50 to 100%, with an average of 89.0% and a median of 94.5% (Table S1). Contamination was below 5% for all genomes according to CheckM (32).

120

We clustered these genomes by 95% average nucleotide identity (ANI). This clustering into mOTUs enabled us to gain a more comprehensive understanding of the genomic structure of genus *Chlorobium* members. By clustering of >500 genomes at a 95% ANI threshold, previously shown to unite classical species definitions and separate sequenced strains into consistent and distinct groups (33-35), we obtained

125   71 metagenomic operational taxonomic units (mOTUs) or "species" belonging to the class *Chlorobia* (Table S2). Of those, 57 mOTUs were classified as members of the genus *Chlorobium* under the GTDB taxonomy (Figure 2 and table S2) (26). These 57 mOTU genomes had an average completeness of approximately 90%. Their mean estimated length was 2.6 Mbp, with the lowest estimated size being 2.1 Mbp and the highest 3.7 Mbp, which is consistent with previous studies showing that *Chlorobia* genomes

130   from isolates range in size from 1.9 to 3.3 Mbp (2). Among the 57 *Chlorobium*-associated mOTUs, 12 were composed exclusively of genomes from previously described isolates (2), including *Chlorobium (Chl.) phaeobacteroides, Chl. limicola, Chl. luteolum, Chl. ferrooxidans and Chl. phaeoclathratiforme* (Figure 2). None of the mOTUs included both isolate genomes and environmental MAGs. Whereas several of the cultured members of the class *Chlorobia* were isolated from lakes (36), our analysis

135   suggests that the majority of natural diversity within freshwater planktonic members of the class *Chlorobia* are found within uncultivated members of the genus *Chlorobium.*

*The genus Chlorobium includes both cosmopolitan and endemic lineages*
By identifying MAGs from different source locations that cluster within the same mOTUs, we were able

140   to determine whether a "species" is widely distributed and putatively cosmopolitan, or in contrast endemic to specific locations. To investigate the abundance and prevalence of each reconstructed mOTU within resident *Chlorobia* assemblages, we mapped metagenomic reads from the lake and pond datasets (Table S3) against the genome collection (Table S1). To reduce noise, we mapped reads with 100% identity to the reference and applied a cutoff for presence below 0.03% metagenomic read abundance.
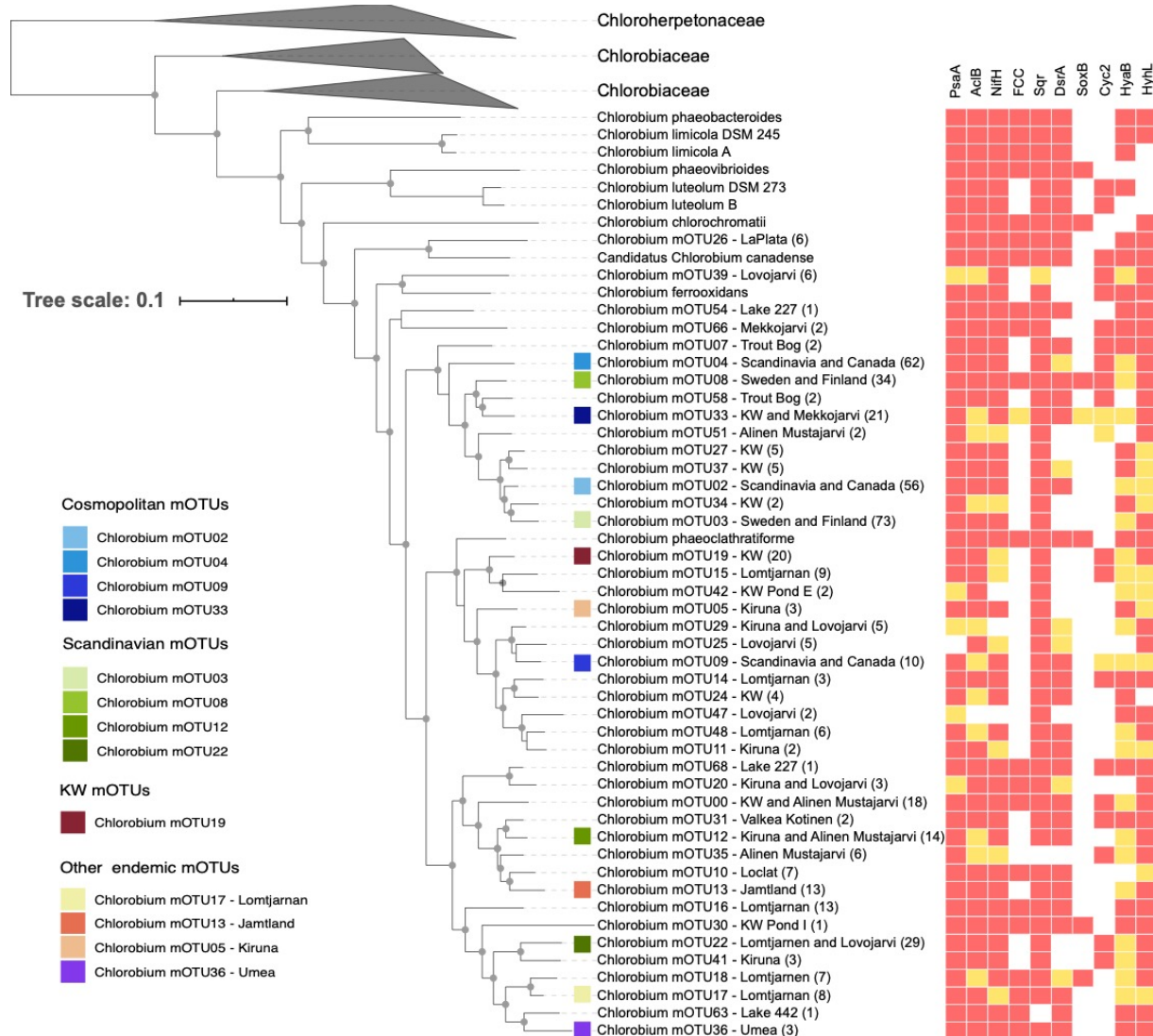
Figure 2. Phylogenetic relationship and metabolic capabilities of Chlorobium shows cosmopolitan and endemic mOTUs as non-monophyletic. The tree was constructed using GTDB-Tk "de-novo" alignment, and it was annotated and curated in iTOL. Isolates are shown in the tree with species names instead of mOTU number. Other tips on the tree show the genus name and mOTU number followed by location where the genomes were assembled from. If the MAGs where assembled from more than one lake in the same area, then the name of the area is written. If the MAGs where assembled from more than one area in the same country, then the name of the country is written. Lake La Plata is in Bayamon, Lakes 227 and 442 are in the IISD-ELA (near Kenora, Canada), Lake Lovojarvi, Lake Mekkojarvi, Lake Valkea Kotinen and Lake Alinen Mustajarvi are in Evo, Lake Trout Bog is in Wisconsin, Lake Loclat is in Neuchatel and Lake Lomtjarnen is in Jamtland. Numbers in parenthesis shows the number of genomes in each mOTU. The Genus *Chlorobium* is part of the Family *Chlorobiaceae*. Other genomes in different Genera within the family *Chlorobiaceae* are clustered. The tree also includes information about the presence/absence of some gene products (PsaA – Photosystem I P700 chlorophyll a apoprotein A1, AclB – ATP-citrate lyase beta-subunit, NifH – nitrogenase iron protein, FCC – flavocytochrome c sulfide dehydrogenase, Sqr – sulfide-quinone oxidoreductase, DsrA – reverse dissimilatory sulfite reductase, SoxB – thiosulfohydrolase, Cyc2 – iron-oxidizing outer-membrane c-type cytochrome, HyaB - group 1d [NiFe]-hydrogenase large subunit, HyhL - group 3b [NiFe]- hydrogenase). In the heatmap red means the gene is present in the core genome of the mOTU, and yellow means it is present in the accessory genome of the mOTU, and white means the gene is absent in the mOTU, see the methods for details. The 13 most abundant *Chlorobium* mOTUs in the investigated environments have a colored square before the name of the mOTU. Gray circles in the tree represent bootstrap values higher that 50%.

165   Considering these cutoffs, reads of our lake/pond dataset mapped to 45 (out of 71) *Chlorobia* mOTUs, of

which 42 mOTUs classified as members of the genus *Chlorobium* (Figure 3). None of these mOTUs

contained any of the cultivated *Chlorobium* representatives. This provides further evidence that

uncultivated *Chlorobium* members are the dominant members of the planktonic *Chlorobia* present in

sampled lakes and ponds.
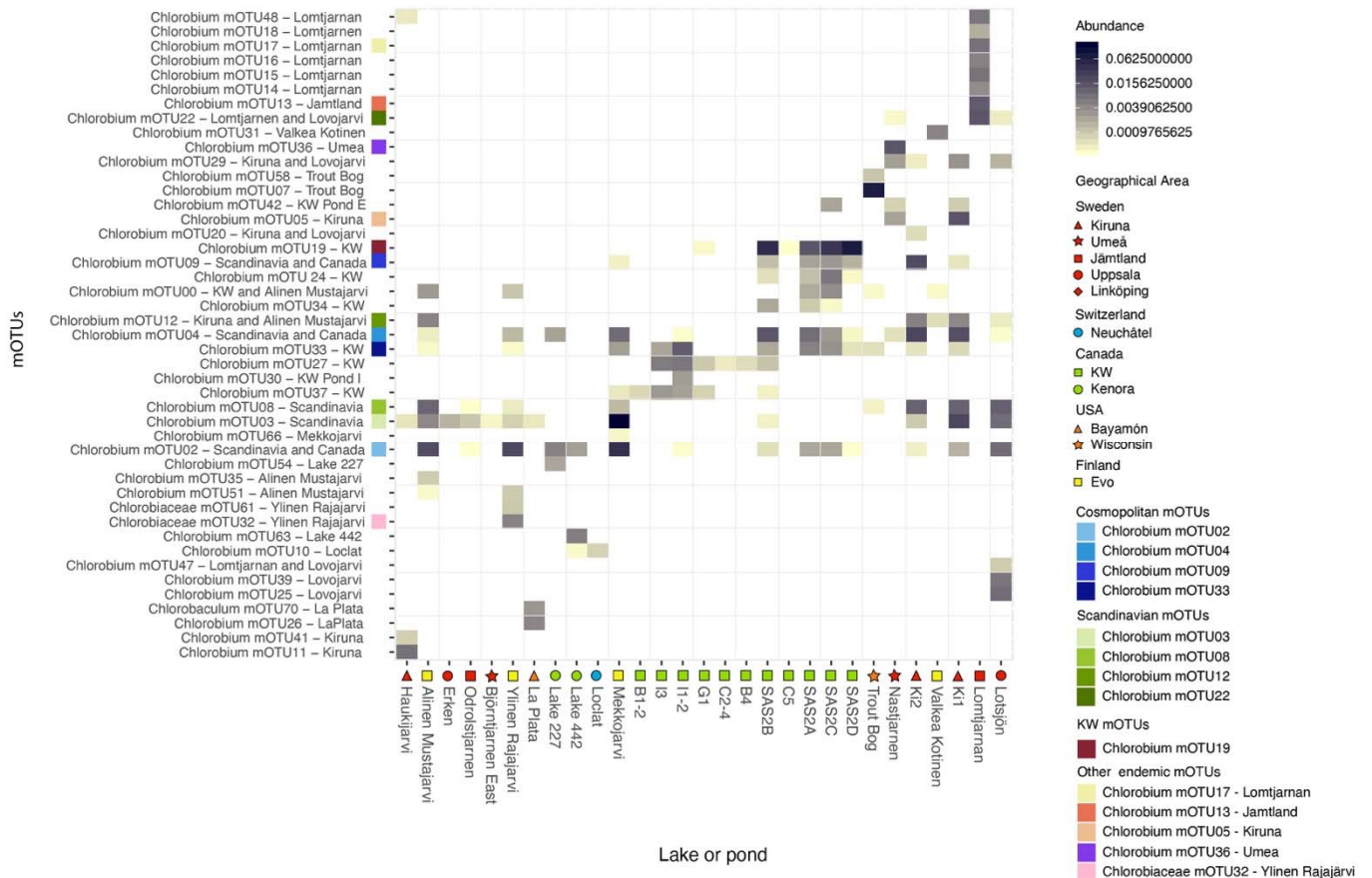
170

Figure 3. Relative abundance of the metagenome reads of the 45 abundant *Chlorobia* mOTUs in lake and pond metagenomes shows some mOTUs are truly endemic while others are more widespread or cosmopolitan. Top 14 *Chlorobia* mOTUs are color coded according to the location they were assembled from. Metagenomes were
175   subsampled to one million reads. Reads were then mapped competitively with 100% identity cutoff to all genomes in all mOTUs using BBmap. The reads were normalized to relative abundance of reads per metagenome. Relative abundances from all depths and timepoints in a given lake/pond were averaged. The cutoff for presence of a mOTU to be included in the figure is 0.03% read abundance per lake/pond. The name of the mOTUs includes geographical information about the origin of the genomes as in Figure 2. The lakes or ponds in the x axis include a
180   symbol for the region where they are located.

Out of the 42 locations represented by this study, we found that metagenomes sequenced from 18 lakes

(of 30 total) and 11 ponds (of 12 total) contained *Chlorobia* reads (Figure 3). However, for 10 of the lakes

that did not show *Chlorobia* read abundance, we only have epilimnion samples with full oxygenation, and

185   so it was expected to not find *Chlorobia* reads or genomes. If we deduct these 10 lakes, we found

7

Chlorobia reads in most of the metagenomes from anoxic freshwaters. In our data, we found that four mOTUs with MAGs reconstructed from assemblies originating on both studied continents (i.e., mOTU02, mOTU04, mOTU09, and mOTU33; in blue, Figures 2, 3, 4, and 5) and so define them as cosmopolitan because of their distribution pattern. Moreover, metagenomic read recruitment confirmed that these

190   cosmopolitan mOTUs were present in lakes and ponds from both studied continents (Figure 3). In contrast, the other 41 mOTUs were less widely distributed and were found either in just one lake, in one geographical area, or in Scandinavia. The apparent cosmopolitan or locally constrained clades did not appear to be monophyletic, but rather distributed across the phylogenetic tree of the class *Chlorobia* (Figure 2).

195

Depth stratification of *Chlorobia* mOTUs was evident when mapped reads were visualized as relative abundance of the total metagenomic reads per depth (Figure 4). As expected, our study showed that *Chlorobium* populations were mostly found below the oxycline (Figure 4) and some mOTUs were highly abundant. Most notably, for Lake Mekkojärvi (Finland), *Chlorobium* sequences comprised 57% of the

200   community by relative abundance at some depths (Figure 4G). These high levels are in line with previous studies showing that this class can constitute 12 - 47% of reads in lake ecosystems (30, 31, 37, 38). Our data shows one clear case in which multiple *Chlorobium* populations typically co-inhabit the same lake with spatially separated niches. Specifically, mOTU02 and mOTU03 demonstrated differential depth preferences in August 2015 based on Lake Mekkojärvi samples (Figure 4G). Similar findings were

205   previously reported in Trout Bog, where one of the populations was recovered only in the lowermost water layers (i.e., mOTU07, GSB-A/Chlorobium-111), whereas the other populations had broader distributions (e.g., mOTU58, GSB-B) (29, 31). It has been previously postulated that such niche specialization may reflect distinct pigment absorbance profiles (31, 39).

210   Several of the putative endemic mOTUs were temporally stable within their respective lakes. For example, we sampled Lake Lomtjärnen (Figure 4B) at four timepoints: March 2016 (under ice and snow cover), April 2016 (ice cover only), September 2017 (ice free) and July 2018 (ice free). The oxycline is lower in the water column when the lake is ice free and *Chlorobium* populations were abundant below the oxycline at all timepoints. The mOTUs residing in this lake were mostly endemic to the lake or to the

215   region and they were consistently recovered at the different time points. It is also notable that the abundance of the different mOTUs changed according to the sampling time. For example, mOTU13 increased in relative abundance for the September sampling point. This allows us to speculate that, for endemic mOTUs, some may have specific seasonal niches. A more temporally resolved sampling effort would help test this hypothesis.
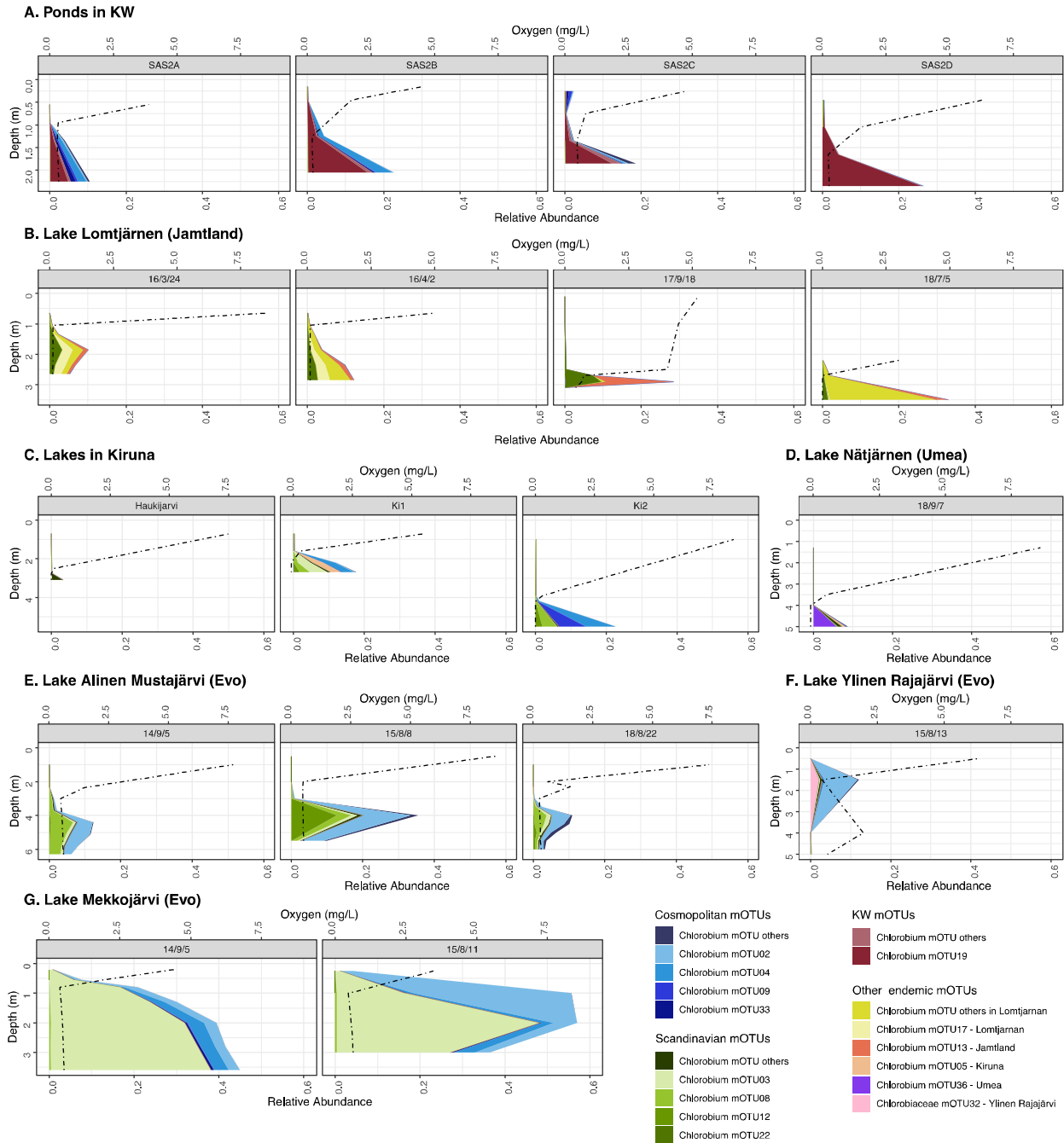
220

Figure 4. Distribution of environmentally abundant mOTUs across depth profiles in the lake and pond datasets with depth discrete sampling and high Chlorobium abundances shows abundance of *Chlorobia* below the oxycline. Note that the 14 most abundant *Chlorobia* mOTUs have individual color coding and while the remaining abundant mOTUs are combined in a category labeled "other". For cosmopolitans, this include mOTU 00 and 42 and for Scandinavians this include mOTU 35, 29, 11, 31, 51 and 66. Other mOTUs in Kuujjuarapik-Whapmagoostui include mOTU 37, 24, 27, 34 and 30. Finally other mOTUs in the category endemic include mOTU 14, 15, 16, 18 and 48. Oxygen is represented by a dashed line. (A) Ponds in the KW area. (B) Different time points for Lake Lomtjärnen in Sweden. (C) Different lakes in the Kiruna area. (D) Lake Nästjärnen in Sweden. (E) Different time points of Lake Alinen Mustajärvi in Finland. (F) Lake Ylinen Rajajärvi. (G) Different time points for Lake Mekkojärvi in Finland.

230

The metabolic potential of the most abundant Chlorobium groups

We investigated the metabolic potential of all *Chlorobium*-associated mOTUs by inspecting the functionally annotated genes via eggNOG-mapper assigned K0 numbers. Combining the power of ANI clustering and statistical approaches helps calculate the probability that some genes are present in the core
235   or accessory genomes of the "species", despite the incompleteness of some of the MAGs and SAGs. In our data, 42 out of the 71 *Chlorobia* mOTUs were composed of more than one genome, and hence we find power in replication (40). For example, mOTU08 includes 34 MAGs with a completeness average of 95%, and the mOTU encodes most genes displayed in the core (Figure 2), except for HyaB, which is putatively within the accessory genome. Moreover, in order for a gene to be completely missing from the
240   mOTU, the gene has to be absent in all of the MAGs and SAGs for the mOTU. Unlike traditional MAG-based studies, the absence of a gene in our mOTU data means that there is a high likelihood that the gene is really missing from the annotation or the population.

For individual mOTUs, we found an average of 710 annotated genes with designated KEGG functions.
245   With these mOTUs, we were able to investigate the core metabolic functions of freshwater members of the genus *Chlorobium.* As expected, given the photolithoautotrophic lifestyle of cultured *Chlorobia*, core features of the genomes included all genes related to glycolysis and gluconeogenesis, reverse TCA cycle, and chlorophyll and bacteriochlorophyll biosynthesis (Figure 5; Figure S1A). These findings are consistent with previous studies showing very minor differences in the genes encoding the photosynthetic
250   Type 1 reaction center unit in this class (41). We also observed molybdenum- and iron-nitrogenases widely encoded by the representatives of the class *Chlorobia* (Figure S1C). Thus, light harvesting, carbon fixation, and nitrogen fixation appear to be conserved core metabolic features of freshwater *Chlorobia* members.

255   In contrast to functions that were widely encoded, *Chlorobia* mOTUs significantly varied in their capacity to use electron donors (i.e., hydrogen, sulfide, thiosulfate, and iron). With respect to sulfur compounds, most mOTUs were capable of oxidizing sulfide via one or more enzymes. Sulfide-quinone oxidoreductases (Sqr) were encoded by all but two *Chlorobium* mOTUs (Figure 5 and S1B), whereas reverse dissimilatory sulfite reductases (i.e., DsrA) and flavocytochrome *c* sulfide dehydrogenase (*FCC*)
260   were encoded by 71% and 53% of the mOTUs, respectively (Figure 2). A more restricted trait was the capacity to oxidize thiosulfate via thiosulfohydrolase (i.e., SoxB), which was only present in 14% of the *Chlorobium* mOTUs but in all *Chlorobaculum* mOTUs (Figure 2 and Table S2). Others have suggested that members of the Class *Chlorobia* have adapted to different environments by acquiring different electron transfer complexes through horizontal gene transfer (41-43). Even though genes for the Dsr and

10

265   Sox complexes may have been horizontally acquired in this class (41), our phylogenetic trees show that DsrA and SoxB form monophyletic clades, consistent with a model that they were each acquired on one occasion during the evolution of *Chlorobia* and have potentially been lost from certain clades (Figure 2, Figure 6A and B).
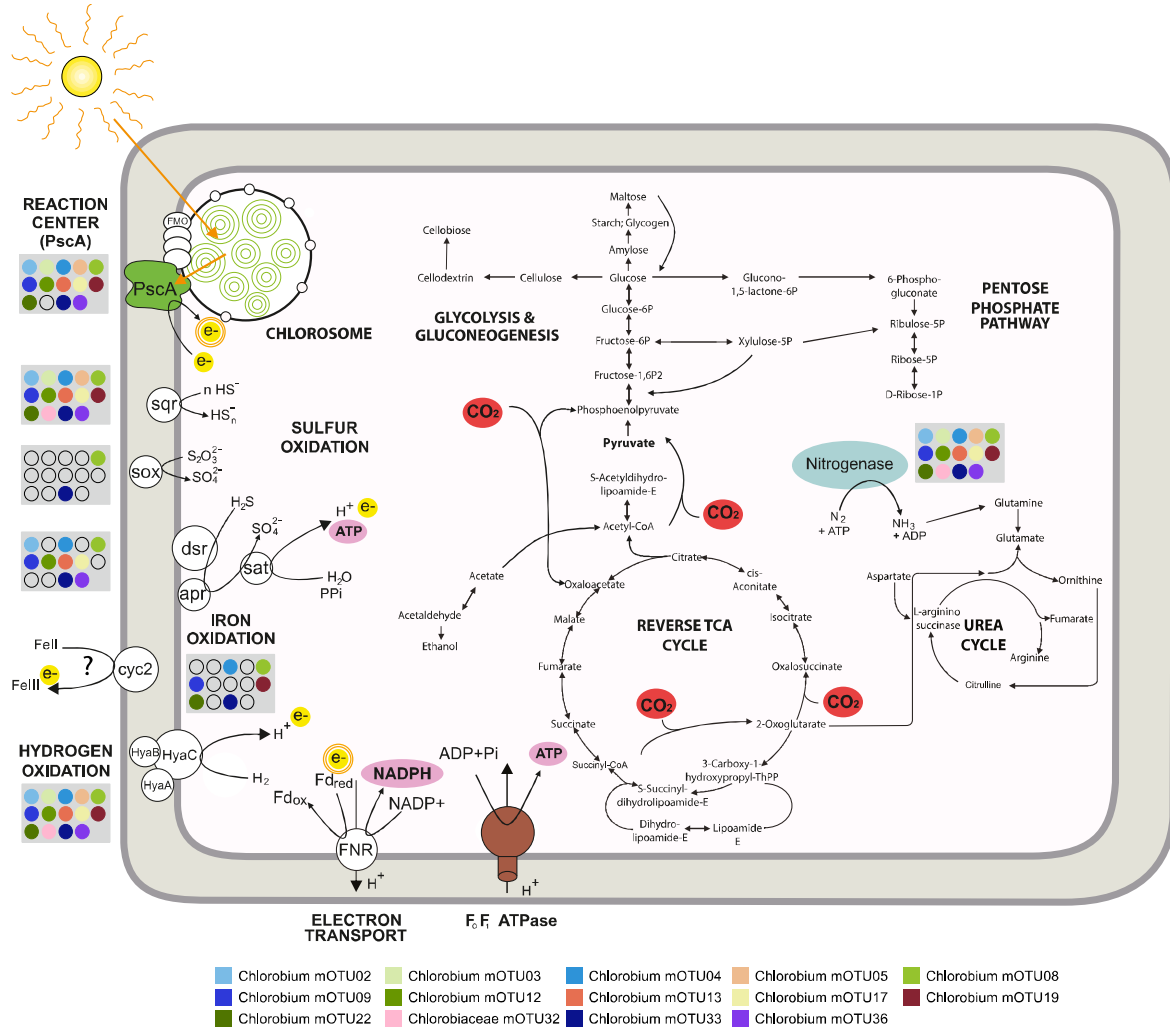
270

Figure 5. Metabolic potential of the 14 most abundant *Chlorobia* mOTUs found in the studied freshwaters, with focus on photosynthesis, electron donor, electron transport, and carbon fixation. Blue: the nitrogenase. Green: the chlorosome and photochemical reaction center PscA. Yellow: the electrons being donated through oxidation reactions and photosynthesis. Pink: reductant (NADPH) and chemical energy (ATP) produced during oxidation

275   reactions and photosynthesis. Red: Pathways of carbon assimilation through the reverse tricarboxylic acid cycle, as well as anaplerotic gluconeogenesis steps, using electrons derived from inorganic compound oxidation and energy derived from photosynthesis. The photosystem uses electrons derived from sulfide, hydrogen, thiosulfate, iron oxidation and use light to activate them. This allows proton pumping and ferredoxin reduction. Ferredoxin reduction is linked to photosystem activity depicted with the double orange circle in electron. In the grey boxes

280   there are circles where we colored according to presence/absence of the genes in the respective mOTUs in the legend. Gene product abbreviation, PscA – Photosystem I P700 chlorophyll a apoprotein A1, sqr – sulfide-quinone oxidoreductase, dsr – reverse dissimilatory sulfite reductase, apr - adenylylsulfate reductase, sat- sulfate adenylyltransferase, sox – thiosulfohydrolase, cyc2 – iron-oxidizing outer-membrane *c*-type cytochrome, Hya - group 1d [NiFe]-hydrogenase, FNR - fumarate and nitrate reductase, Fd - ferredoxin.

285

Genes encoding HyaB and HyhL (group 1d and group 3b [NiFe]-hydrogenases, respectively) were detected in 82% and 91% of the *Chlorobium* mOTUs, respectively (Figure 2, 6C and S1C). The group 1d enzymes are known to support hydrogenotrophic respiration and anoxygenic photosynthesis in diverse

290   bacterial lineages (44-46). It is likely that the electrons liberated by these enzymes are transferred to the photosynthetic reaction center to reduce ferredoxin and $NADP^+$ as a source of reductant for carbon fixation, nitrogen fixation, and other biosynthetic processes (47). Phylogenetic analysis confirmed that these subunits formed monophyletic radiations together with the hydrogenases retrieved from *Chlorobia* reference genomes (Figure 6C). The physiological role of group 3b hydrogenases (HyhL) in *Chlorobia*

295   genomes is less clear. These bidirectional cytosolic enzymes are likely to support either hydrogenotrophic carbon fixation or facultative hydrogenogenic fermentation in this class (Figure S1D) (48, 49). Although uptake hydrogenases have previously been reported for *Chlorobium* members (50), the ecological significance of H_2 metabolism has been overlooked in this class. Overall, the results suggest that H_2 metabolism is an ancestral and conserved trait of the class *Chlorobia*. This is also consistent with the

300   widespread distribution of [NiFe]-hydrogenases across bacterial phyla (51, 52) and the conservation of group 1d lineages in sister classes of the phylum *Bacteroidota*.
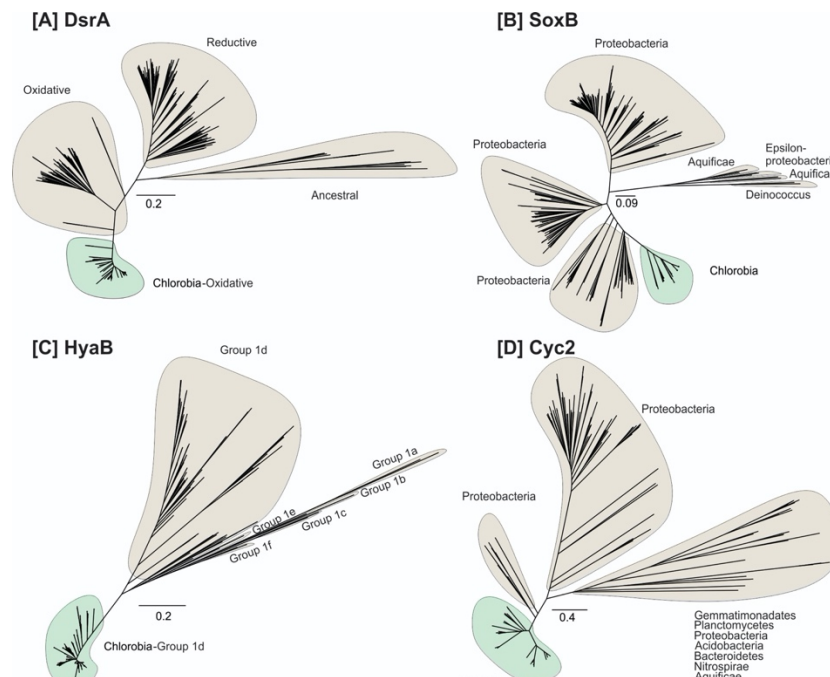
Figure 6. Maximum-likelihood phylogeny of the recovered protein sequences from reconstructed Chlorobia
305   MAGs/SAGs (highlighted in green) together with the reference protein sequences recovered from NCBI GeneBank, with the exception of hydrogenases for which sequences were collected from HydDB (highlighted in gray). Panels show the phylogeny of (A) dissimilatory sulfite reductase, DsrA (n=551), (B) Thiosulfohydrolase, SoxB (n=294), (C) Group 1d NiFe hydrogenase, HyaB (n=579), and (D) Iron oxidizing cytochrome, Cyc2 (n=306) protein sequences.

310   Our study also shows that 40% of the *Chlorobium* mOTUs harbor *cyc2* genes (Figure 2 and 6D), encoding a potential outer-membrane *c*-type cytochrome capable of iron oxidation (30, 53). This suggests that the distribution of the *cyc2* gene among *Chlorobium* genomes in lakes is higher than previously recognized. If the *cyc2* gene allows for extracellular electron transfer or ferrous iron oxidation, as speculated based on comparative genomics of cultured photoferrotrophs (30), this implies that *Chlorobia*

315   members could be profoundly important for iron biogeochemistry in lake systems globally.

We did not find any mOTU encoding all the putative oxidation genes in the core genome (Figure 2 and 5). Metabolic flexibility is a key factor governing distributions of taxa across ecosystems (54). However, in this study we did not find a correlation between how widespread an mOTU is and their capacity to use different electron acceptors. A broader global sampling could reveal if metabolic versatility governs

320   prevalence and abundance of *Chlorobium* bacteria globally.

Outlook of ecological roles of Chlorobium
At a first glance, *Chlorobia* members appear to populate a very specific niche in water columns of global lakes and ponds. As we confirm in our datasets, they thrive under the oxycline in the anaerobic layer where light is still available. They have evolved physiological adaptations to low light intensities (3) and

325   therefore seem to do particularly well in freshwater systems characterized by high concentrations of colored dissolved organic carbon (DOC) that absorb the incident solar radiation (55). In these lakes, solar radiation only penetrates to shallow depths in the water column, constraining populations of *Cyanobacteria* and most other photoautotrophs to the surface layer (56). *Chlorobium* members may contribute up to 83% of the total annual productivity in such humic waters (24) and also seem to play an

330   important role as diazotrophs (15, 16). Despite the very specific niche of *Chlorobium,* our work illustrates considerable metabolic versatility with regards to oxidation genes that can be used to obtain electrons for carbon fixation. By their variable capacity to recycle sulfide, hydrogen, and iron produced by dissimilatory sulfate reducers, hydrogenogenic fermenters, and dissimilatory iron reducers, *Chlorobium* may potentially influence the entire microbiome of anoxic freshwaters (57). Our analyses describe the

335   metabolic potential of environmentally relevant *Chlorobium* genomes along with their geographical distributions. Our results reveal both endemic and cosmopolitan clades but we were unable to link metabolic versatility to those distributions. Instead, the distributions of *Chlorobium* populations seem to distribute largely based on ecological factors. Our study also identified clades of *Chlorobium*, with abundant and cosmopolitan geographical distributions, that do not show a monophyletic clustering. In

340   order to pinpoint the exact physicochemical background providing them with this ecological advantage,

13

we need more comprehensive metadata and better geographical resolution. Nevertheless, our study highlights the ecological diversity of this enigmatic bacterial lineage.

Materials and methods

345 *Collection of lakes and pond metagenomes*

We obtained 265 metagenomes from 42 locations that spanned subarctic to tropical regions as a part of a project aiming to study microbial diversity in anoxic freshwater environments (Figure 1; Table S3) (28). Oxygen concentrations were measured with a YSI 55 oxygen probe (Yellow Springs Instruments, Yellow Springs, OH, USA). Moreover, we added six metagenomic libraries generated for a *Chlorobium* study in

350 Canada focusing on iron oxidation (30) and three more metagenomic libraries are part of a 15-year time-series study of Trout Bog (29).

Genome collection

Assembling and binning the 265 metagenomes resulted in ~12,000 metagenome-assembled genomes

355 (MAGs) from water bodies in 5 different countries (28). Of these, 454 belonged to the class *Chlorobia* and were used in this study. In brief, the raw data was trimmed using Trimmomatic (version 0.36) (58). The trimmed data were assembled using MEGAHIT (version 1.1.13)(59) with default settings. Two types of assemblies were performed: single sample assemblies for all samples individually and 53 co-assemblies that were performed mostly on lake-specific metagenome data. Detailed descriptions of the

360 sample collection, metadata, sample processing, metagenome generation and analyses methods are reported elsewhere (28).

The relevant quality controlled reads were mapped to all the assemblies using BBmap (60) with default settings and the mapping results were used to bin the contigs using MetaBAT (version 2.12.1, parameters

365 --maxP 93 --minS 50 -m 1500 -s 10000) (61). Moreover, we collected genomes of the class *Chlorobia* from the GTDB and several other published MAGs that fulfilled the medium quality threshold of ≥50% completeness and ≤5% contamination (62). In total, we compiled 509 genomes that included 454 MAGs and 19 SAGs that were from stratified water bodies in Sweden, Finland, Canada, Switzerland, and Puerto Rico (28), seven MAGs from enrichment culture or lake metagenomes of lakes at the International

370 Institute for Sustainable Development Experimental Lakes Area (IISD-ELA; near Kenora, Canada) (30), four MAGs from Trout Bog (29, 31), and 25 genomes (MAGs, SAGs or isolates) from the GTDB (2, 63). Completeness and contamination was assessed with CheckM (31).

375   *Single cell collection*

In brief, sorting was done in 2016 on a MoFlo Astrios EQ sorter (Beckman Coulter, USA) using a 488 and 532 nm laser for excitation, a 70-µm nozzle, a sheath pressure of 60 psi and 0.1 µm filtered 1 x PBS as sheath fluid. An ND filter ND=1 and the masks M1 and M2 were used. The trigger channel was set to forward scatter (FSC) at a threshold of 0.025% and sort regions were defined on autofluorescence using a

380   532 nm laser and band pass filters 710/45 and 664/22. Sorted plates were stored at -80°C. The whole genome amplification was run using the REPLI-g Single Cell Kit (QIAGEN) following the manual's instruction but with reduced total reaction volume to 12.5 µl. Amplified DNA was mixed thoroughly by pipetting up and down. The DNA was screened for bacterial 16S rRNA genes and DNA of confirmed *Chlorobia* members was sent for sequencing as previously described (28).

385

    mOTU analysis

Average nucleotide identity (ANI) for all genome pairs was computed with FastANI (version 1.3) (64) and the genomes were then clustered into metagenomic operational taxonomic units (mOTUs) with 70% completeness and 5% contamination thresholds. Genome pairs with ANI values above 95% were

390   clustered into connected components. Additionally, less complete genomes (above 50% yet below 70% completeness) were recruited to the mOTU if its ANI similarity was above 95%. The classification of the genomes was done using GTDB-Tk and sourmash (version 1.0) (64, 66). The most complete MAGs per mOTU were selected as representatives and a phylogenetic tree was calculated using GTDB-Tk (version 102 with database release 89) with *Chloroherpetonaceae* as an outgroup (63). The aligned genome tree

395   was loaded and curated in iTOL (version 5.5.1) (65). Moreover, functions (as in KO numbers) were classified as core if the probability of their presence-absence profile is higher under the assumption that it is in every genome, assuming their incompleteness, then the converse probability as computed by mOTUpan (https://github.com/moritzbuck/mOTUlizer).

400   *Abundance profiles*

To explore abundance profiles, metagenomic libraries were subsampled to 1,000,000 reads (Table S2). Metagenomic reads were then mapped to all collected *Chlorobia*-associated genomes (Table S1). Reads were mapped with 100% identity using BBMap (60). The results were then normalized to relative abundance of mOTU per metagenome. All relative abundances of all samples in each location were

405   averaged. The cutoff for presence of a mOTU was set to 0.03% read abundance per site. Heatmaps of abundance (Figure 3) were calculated and plotted using the R packages ggplot2 and phyloseq with parameters NMDS and Bray-Curtis for choosing the order of the x-axes (66-69). Depth-discrete abundance profiles were plotted using R packages ggplot2, phyloseq, and cowplot (67-69).

15

410 *Metabolic genes*

The metabolic potential of the *Chlorobia* MAGs, SAGs, and reference genomes were reconstructed based on eggNOG-mapper (70, 71) annotations. Ten metabolic marker gene products (PsaA – Photosystem I P700 chlorophyll a apoprotein A1, AclB – ATP-citrate lyase beta-subunit, NifH – nitrogenase iron protein, FCC – flavocytochrome *c* sulfide dehydrogenase, Sqr – sulfide-quinone oxidoreductase, DsrA –

415 dissimilatory sulfite reductase A subunit, SoxB – thiosulfohydrolase, Cyc2 – outer-membrane monoheme *c*-type cytochrome, HyaB - group 1d [NiFe]-hydrogenase large subunit, HyhL - group 3b [NiFe]-hydrogenase) were annotated in reconstructed MAGs, SAGs, and reference genomes using Diamond (version 0.9.31) (72) against custom build databases for each marker protein (query coverage threshold of 80%). A 50% identity threshold was used for all marker gene products except for an 80% threshold used

420 for the PsaA protein. Annotations were further validated by reconstructing phylogeny.

Single Gene phylogeny

Maximum-likelihood trees were constructed using the amino acid sequences of six metabolic marker proteins (PsaA – Photosystem I P700 chlorophyll a apoprotein A1, NifH – nitrogenase iron protein, Sqr –

425 sulfide-quinone oxidoreductase, DsrA – sulfite reductase A subunit, SoxB – thiosulfohydrolase, Cyc2 – outer-membrane monoheme c-type cytochrome, HyaB/HyaA - group 1d [NiFe]-hydrogenase large subunit and small subunit, HyhL - group 3b [NiFe]-hydrogenase large subunit). Reference genes for each marker were collected from NCBI GeneBank, with the exception of hydrogenases for which sequences were collected from HydDB (46). For each metabolic marker protein, sequences retrieved from the

430 *Chlorobia* mOTUs were aligned against reference protein sequences using ClustalW in MEGA7 (73, 74). Evolutionary relationships were visualized by constructing a maximum-likelihood phylogenetic tree. Specifically, initial trees for the heuristic search were obtained automatically by applying Neighbour-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. All residues were used and trees were bootstrapped with

435 50 replicates.

Data availability

All datasets are available in public repositories with NCBI project accession numbers PRJEB3868, PRJNA518727, and PRJNA534305. Accession numbers of the genomes are available in Table S1, and

440 the accession numbers of the metagenomes are available in Table S3. Assembling and binning of the original dataset has been done with the scripts available at https://github.com/moritzbuck/metasssnake General processing scripts for this project are available at https://github.com/moritzbuck/0026_Chlorobi

Acknowledgements

SLG and SP conceptualized the research idea. All authors provided data. MB, MM, SLG, JMT, CG and

460 KDM performed data analysis. SLG, MM, CG and SP did literature searches. SLG, MM, CG and SP drafted the manuscript and all authors contributed to writing and editing of the manuscript.

References

465 1. Bryant DA, Frigaard NU. 2006. Prokaryotic photosynthesis and phototrophy illuminated. Trends Microbiol 14:488-96.

2. Bryant DA, Liu Z, Li T, Zhao F, Garcia Costas AM, Klatt CG, Ward DM, Frigaard N-U, Overmann J. 2012. Comparative and functional genomics of anoxygenic green bacteria from the taxa *Chlorobi*, *Chloroflexi*, and *Acidobacteria*, Functional Genomics and Evolution of

470 Photosynthetic Systems, vol 33. Springer, Dordrecht.

3. Beatty JT, Overmann J, Lince MT, Manske AK, Lang AS, Blankenship RE, Van Dover CL, Martinson TA, Plumley FG. 2005. An obligately photosynthetic bacterial anaerobe from a deep-sea hydrothermal vent. Proceedings of the National Academy of Sciences of the United States of America 102:9306-9310.

475 4. Imhoff JF, Thiel V. 2010. Phylogeny and taxonomy of *Chlorobiaceae*. Photosynth Res 104:123-136.

5. Garcia SL, Szekely AJ, Bergvall C, Schattenhofer M, Peura S. 2019. Decreased snow cover stimulates under-ice primary producers but impairs methanotrophic capacity. mSphere 4.

6. Overmann J. 2008. Ecology of phototrophic sulfur bacteria, p 375 - 396. *In* R. Hell CD, D. B.,

480 Leustek KT (ed), Advances in photosynthesis and respiration Sulfur metabolism in phototrophic organisms, vol 27. New York: Springer.

7.   Sirevåg R. 1995. Carbon metabolism in green bacteria. *In* Blankenship RE, Madigan MT, C.E. B (ed), Anoxygenic Photosynthetic Bacteria Advances in Photosynthesis and Respiration, vol 2. Springer, Dordrecht.

485   8.   Buchanan BB, Arnon DI. 1990. A Reverse Krebs Cycle in Photosynthesis - Consensus at Last. Photosynth Res 24:47-53.

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

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

11.   Lliros M, Garcia-Armisen T, Darchambeau F, Morana C, Triado-Margarit X, Inceoglu O, Borrego CM, Bouillon S, Servais P, Borges AV, Descy JP, Canfield DE, Crowe SA. 2015. Pelagic photoferrotrophy and iron cycling in a modern ferruginous basin. Scientific Reports 5.

495   12.   Laufer K, Niemeyer A, Nikeleit V, Halama M, Byrne JM, Kappler A. 2017. Physiological characterization of a halotolerant anoxygenic phototrophic Fe(II)-oxidizing green-sulfur bacterium isolated from a marine sediment. Fems Microbiology Ecology 93.

13.   Lambrecht N. 2019. Insights into early Earth ocean biogeochemistry from intensive monitoring of two ferruginous meromictic lakesIowa State University, Ann Arbor.

500   14.   Heda GD, Madigan MT. 1986. Aspects of nitrogen-fixation in *Chlorobium*. Archives of Microbiology 143:330-336.

15.   Fernandez L, Peura S, Eiler A, Linz AM, McMahon KD, Bertilsson S. 2020. Diazotroph genomes and their seasonal dynamics in a stratified humic bog lake. Front Microbiol 11.

16.   Fernandez L, Bertilsson S, Peura S. 2020. Non□cyanobacterial diazotrophs dominate
505   nitrogen□fixing communities in permafrost thaw ponds. Limnology and Oceanography 65:S180-S193.

17.   Ng C, DeMaere MZ, Williams TJ, Lauro FM, Raftery M, Gibson JA, Andrews-Pfannkoch C, Lewis M, Hoffman JM, Thomas T, Cavicchioli R. 2010. Metaproteogenomic analysis of a dominant green sulfur bacterium from Ace Lake, Antarctica. ISME J 4:1002-19.

510   18.   Peura S, Sinclair L, Bertilsson S, Eiler A. 2015. Metagenomic insights into strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes. Scientific Reports 5:12102.

19.   Gregersen LH, Habicht KS, Peduzzi S, Tonolla M, Canfield DE, Miller M, Cox RP, Frigaard NU. 2009. Dominance of a clonal green sulfur bacterial population in a stratified lake. FEMS Microbiol Ecol 70:30-41.

515   20.   Panwar P, Allen MA, Williams TJ, Hancock AM, Brazendale S, Bevington J, Roux S, Paez-Espino D, Nayfach S, Berg M, Schulz F, Chen IA, Huntemann M, Shapiro N, Kyrpides NC, Woyke T, Eloe-Fadrosh EA, Cavicchioli R. 2020. Influence of the polar light cycle on seasonal dynamics of an Antarctic lake microbial community. Microbiome 8:116.

21.   Rimmer A, Ostrovsky I, Yacobi YZ. 2008. Light availability for *Chlorobium phaeobacteroides*
520   development in Lake Kinneret. Journal of Plankton Research 30:765-776.

22.   Fry B. 1986. Sources of carbon and sulfur nutrition for consumers in three meromictic lakes of New York State. Limnol Oceanogr 31:79-88.

23.   Yacobi YZ, Ostrovsky I. 2008. Downward flux of organic matter and pigments in Lake Kinneret (Israel): relationships between phytoplankton and the material collected in sediment traps. Journal
525   of Plankton Research 30:1189-1202.

24.   Culver DA, Brunskill GJ. 1969. Fayetteville Green Lake, New York. V. Studies of primary production and zooplankton in a meromictic marl lake. Limnology and Oceanography 14:862-873.

25.   Comeau AM, Harding T, Galand PE, Vincent WF, Lovejoy C. 2012. Vertical distribution of
530   microbial communities in a perennially stratified Arctic lake with saline, anoxic bottom waters. Sci Rep 2:604.

26. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nature Biotechnology 36:996-1004.

27. Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, Dodson RJ, Deboy R, Gwinn ML, Nelson WC, Haft DH, Hickey EK, Peterson JD, Durkin AS, Kolonay JL, Yang F, Holt I, Umayam LA, Mason T, Brenner M, Shea TP, Parksey D, Nierman WC, Feldblyum TV, Hansen CL, Craven MB, Radune D, Vamathevan J, Khouri H, White O, Gruber TM, Ketchum KA, Venter JC, Tettelin H, Bryant DA, Fraser CM. 2002. The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. PNAS 99:9509-9514.

28. Buck M, Garcia SL, Bertilsson S, Peura S. 2020. Comprehensive dataset of shotgun metagenomes from stratified freshwater lakes and ponds. *In preparation*.

29. Bendall ML, Stevens SL, Chan LK, Malfatti S, Schwientek P, Tremblay J, Schackwitz W, Martin J, Pati A, Bushnell B, Froula J, Kang D, Tringe SG, Bertilsson S, Moran MA, Shade A, Newton RJ, McMahon KD, Malmstrom RR. 2016. Genome-wide selective sweeps and gene-specific sweeps in natural bacterial populations. ISME J 10:1589–1601.

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

31. Berg M, Goudeau D, Olmsted C, McMahon KD, Thweatt J, Bryant D, Eloe-Fadrosh EA, Malmstrom RR, Roux S. 2020. Host population diversity as a driver of viral infection cycle in wild populations of green sulfur bacteria with long standing virus-host interactions. bioRxiv doi:https://doi.org/10.1101/2020.03.05.979559.

32. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res doi:10.1101/gr.186072.114.

33. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. P Natl Acad Sci USA 102:2567-2572.

34. Varghese NJ, Mukherjee S, Ivanova N, Konstantinidis KT, Mavrommatis K, Kyrpides NC, Pati A. 2015. Microbial species delineation using whole genome sequences. Nucleic Acids Research 43:6761-6771.

35. Garcia SL, Stevens SLR, Crary B, Martinez-Garcia M, Stepanauskas R, Woyke T, Tringe SG, Andersson SGE, Bertilsson S, Malmstrom RR, McMahon KD. 2018. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. ISME J 12:742-755.

36. Overmann J, Pfennig N. 1989. *Pelodictyon phaeoclathratiforme* sp. nov., a new brown-colored member of the Chlorobiaceae forming net-like colonies. Archives of Microbiology 152:401-406.

37. Taipale S, Jones RI, Tiirola M. 2009. Vertical diversity of bacteria in an oxygen-stratified humic lake, evaluated using DNA and phospholipid analyses. Aquatic Microbial Ecology 55:1-16.

38. Karhunen J, Arvola L, Peura S, Tiirola M. 2013. Green sulphur bacteria as a component of the photosynthetic plankton community in small dimictic humic lakes with an anoxic hypolimnion. Aquatic Microbial Ecology 68:267-272.

39. Maresca JA, Gomez Maqueo Chew A, Ponsati MR, Frigaard NU, Ormerod JG, Bryant DA. 2004. The bchU gene of *Chlorobium tepidum* encodes the c-20 methyltransferase in bacteriochlorophyll c biosynthesis. J Bacteriol 186:2558-2566.

40. Evans JT, Denef VJ. 2020. To dereplicate or not to dereplicate? mSphere 5:e00971-19.

41. Bryant DA, Liu ZF. 2013. Chapter four - Green bacteria: Insights into green bacterial evolution through genomic analyses, p 99-150. *In* Beatty JT (ed), Genome Evolution of Photosynthetic Bacteria, vol 66.

42. Gregersen LH, Bryant DA, Frigaard NU. 2011. Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. Frontiers in Microbiology 2.

43. Kelly DP. 2008. Stable sulfur isotope fractionation by the green bacterium *Chlorobaculum parvum* during photolithoautotrophic growth on sulfide. Polish Journal of Microbiology 57:275-279.

585 44. Shomura Y, Yoon KS, Nishihara H, Higuchi Y. 2011. Structural basis for a [4Fe-3S] cluster in the oxygen-tolerant membrane-bound [NiFe]-hydrogenase. Nature 479:253-U143.

45. Fritsch J, Scheerer P, Frielingsdorf S, Kroschinsky S, Friedrich B, Lenz O, Spahn CMT. 2011. The crystal structure of an oxygen-tolerant hydrogenase uncovers a novel iron-sulphur centre. Nature 479:249-252.

590 46. Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, Cook GM, Morales SE. 2016. Genomic and metagenomic surveys of hydrogenase distribution indicate H2 is a widely utilised energy source for microbial growth and survival. ISME J 10:761-77.

47. Frigaard NU, Chew AGM, Li H, Maresca JA, Bryant DA. 2003. Chlorobium tepidum: insights into the structure, physiology, and metabolism of a green sulfur bacterium derived from the

595 complete genome sequence. Photosynthesis Research 78:93-117.

48. Berney M, Greening C, Conrad R, Jacobs WR, Jr., Cook GM. 2014. An obligately aerobic soil bacterium activates fermentative hydrogen production to survive reductive stress during hypoxia. PNAS 111:11479-11484.

49. Silva PJ, van den Ban EC, Wassink H, Haaker H, de Castro B, Robb FT, Hagen WR. 2001.

600 Enzymes of hydrogen metabolism in *Pyrococcus furiosus*. Eur J Biochem 267:6541-51.

50. Zorin NA. 1986. Redox properties and active-center of phototrophic bacteria hydrogenases. Biochimie 68:97-101.

51. Greening C, Boyd E. 2020. Editorial: Microbial hydrogen metabolism. Front Microbiol 11:56.

52. Lane N, Allen JF, Martin W. 2010. How did LUCA make a living? Chemiosmosis in the origin of

605 life. Bioessays 32:271-280.

53. McAllister SM, Polson SW, Butterfield DA, Glazer BT, Sylvan JB, Chan CS. 2020. Validating the Cyc2 neutrophilic iron oxidation pathway using meta-omics of *Zetaproteobacteria* iron mats at marine hydrothermal vents. mSystems 5:e00553-19.

54. Chen Y-J, Leung PM, Bay SK, Hugenholtz P, Kessler AJ, Shelley G, Waite DW, Cook PLM,

610 Greening C. 2020. Metabolic flexibility allows generalist bacteria to become dominant in a frequently disturbed ecosystem. bioRxiv doi:10.1101/2020.02.12.945220.

55. Sobek S, Tranvik LJ, Prairie YT, Kortelainen P, Cole JJ. 2007. Patterns and regulation of dissolved organic carbon: An analysis of 7,500 widely distributed lakes. Limnology and Oceanography 52:1208-1219.

615 56. Roberts BJ, Howarth RW. 2006. Nutrient and light availability regulate the relative contribution of autotrophs and heterotrophs to respiration in freshwater pelagic ecosystems. Limnology and Oceanography 51:288-298.

57. Peura S, Buck M, Aalto SL, Morales SE, Nykanen H, Eiler A. 2018. Novel autotrophic organisms contribute significantly to the internal carbon cycling potential of a boreal lake. MBio 9:e00916-

620 18.

58. Bolger A, Lohse M, Usadel B. 2014. Trimmomatic: A flexible trimmer for Illumina Sequence Data. Bioinformatics 30:2114-2120.

59. Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph.

625 Bioinformatics 31:1674-6.

60. Bushnell B. 2016. BBMap short read aligner, University of California, Berkeley. https://sourceforge.net/projects/bbmap/.

61. Kang DWD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. Peerj 3:15.

630 62. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA,

Yooseph S, Sutton G, Glockner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu WT, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon

635   KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Genome Standards C, Lapidus A, Meyer F, Yilmaz P, Parks DH, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725-731.

63.   Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify
640   genomes with the Genome Taxonomy Database. Bioinformatics 36:1925-1927.

64.   Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nature Communications 9:5114.

65.   Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
645   developments. Nucleic Acids Res 47:W256-W259.

66.   Rajaram S, Oono Y. 2010. NeatMap--non-clustering heat map alternatives in R. BMC Bioinformatics 11:45.

67.   Team RC. 2014. R: A language and environment for statistical computing. http://www.R-project.org/. Accessed

650   68.   McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. Plos One 8:e61217.

69.   Ginestet C. 2011. ggplot2: Elegant graphics for data analysis. Journal of the Royal Statistical Society Series a-Statistics in Society 174:245-245.

70.   Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR,
655   Letunic I, Rattei T, Jensen LJ, von Mering C, Bork P. 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res 47:D309-D314.

71.   Kanehisa M, Sato Y. 2020. KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci 29:28-35.

660   72.   Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59-60.

73.   Sondergaard D, Pedersen CNS, Greening C. 2016. HydDB: A web tool for hydrogenase classification and analysis. Scientific Reports 6:34212

74.   Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular evolutionary genetics analysis version
665   7.0 for bigger datasets. Mol Biol Evol 33:1870-1874.