Freshwater *Chlorobia* Exhibit Metabolic Specialization among Cosmopolitan and Endemic Populations

Sarahi L. Garcia,[a,b] Maliheh Mehrshad,[a,c] Moritz Buck,[a,c] Jackson M. Tsuji,[d*] Josh D. Neufeld,[d] Katherine D. McMahon,[e,f] Stefan Bertilsson,[a,c] Chris Greening,[g] Sari Peura[h]

[a]Department of Ecology and Genetics, Limnology, Uppsala University, Uppsala, Sweden
[b]Department of Ecology, Environment, and Plant Sciences, Science for Life Laboratory, Stockholm University, Stockholm, Sweden
[c]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Uppsala, Uppsala, Sweden
[d]Department of Biology, University of Waterloo, Waterloo, Ontario, Canada
[e]Department of Civil and Environmental Engineering, University of Wisconsin, Madison, Madison, Wisconsin, USA
[f]Department of Bacteriology, University of Wisconsin, Madison, Madison, Wisconsin, USA
[g]Department of Microbiology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia
[h]Department of Forest Mycology and Plant Pathology, Science for Life Laboratory, Swedish University of Agricultural Sciences, Uppsala, Uppsala, Sweden

Sarahi L. Garcia, Maliheh Mehrshad, and Moritz Buck contributed equally to this article. The order of the authors was designated by time commitment in the writing process.

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 the functional ecology and local adaptations of *Chlorobia* members came from isolates and merely 26 sequenced genomes that may not adequately represent 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-amplified 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 within the genus *Chlorobium*, and these mOTUs represented 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 lacked genes for the oxidation of specific sulfur and iron substrates. Surprisingly, one globally distributed Scandinavian clade encoded the ability to oxidize hydrogen, sulfur, and iron, suggesting that metabolic versatility facilitated such widespread colonization. Overall, these findings provide new insight 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 helped explore the ecology and evolution of environmental microbiota. We applied this 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 show they vary in their capacities to use different electron donors. Having photoautotrophic potential, these *Chlorobia* members could serve as carbon sources that support metalimnetic and hypolimnetic food webs.

KEYWORDS *Chlorobia*, freshwater, photosynthetic bacteria, planktonic

Although oxygenic phototrophs dominate contemporary carbon fixation, anoxygenic phototrophs have been important over planetary time scales and continue to occupy important niches in a wide range of ecosystems (1). Among the anoxygenic phototrophs, known members of the class *Chlorobia* (green sulfur bacteria, formerly phylum *Chlorobi*) are photolithoautotrophic anaerobes (2, 3) that leverage bacterio-chlorophyll-rich organelles called chlorosomes to harvest light at extremely low intensities (i.e., in the range of 1 to 10 nmol photons m^{-2} s^{-1}) (4). 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 (5, 6). Studies of isolates obtained from aquatic systems and microbial mats have revealed that members of the class *Chlorobia* use electrons derived from reduced sulfur compounds and/or hydrogen to fix carbon dioxide via the reverse tricarboxylic acid (rTCA) cycle (7–9). Although several strains can oxidize thiosulfate, they almost universally use sulfide as an 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 (10). Such photoferrotrophs include *Chlorobium ferrooxidans*, which obtains sulfur via assimilatory sulfate reduction, along with three other recently characterized strains (11–14). 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 (15–17). Together, this metabolic capacity and versatility enable members of the class *Chlorobia* to colonize dimly lit and anoxic aquatic environments.

In stratified freshwater ecosystems, *Chlorobia* microorganisms have particularly important ecological and biogeochemical roles (18–21). They accumulate to form dense populations at certain strata, manifesting as deep chlorophyll maxima (22), and contribute to carbon budgets of such lakes (23, 24), while concomitantly contributing to sulfur, iron, hydrogen, and/or nitrogen cycling (17, 25, 26). Despite progress, knowledge of the geographical distribution and metabolic capabilities of *Chlorobia* members within natural planktonic communities remains incomplete. Most understanding of the metabolic capabilities and ecophysiological strategies of this class come from studies of approximately 100 isolates (5), with genomes available for only 26 of these representatives (2, 27, 28). Especially given that cultures are often poor representatives of natural populations, the diversity and distribution of planktonic *Chlorobia* members are largely unknown, and almost nothing is understood about how genes conferring distinct ecophysiological traits are distributed among *Chlorobia*-associated genomes from geographically distinct water bodies.

Here, 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 (Fig. 1). We assembled and analyzed over 500 genomes of *Chlorobia* members, including 465 metagenome-assembled genomes (MAGs) and 19 single-amplified genomes (SAGs). These were further clustered into 71 metagenome operational taxonomic units (mOTUs). Each mOTU is a cluster of genomes that share 95% average nucleotide identity (ANI). We observed that some of these mOTUs exhibited cosmopolitan distributions, whereas others were locally constrained and possibly endemic to specific lakes or regions. In addition, clades varied in their encoded metabolic capabilities, and we found several mOTUs that share substrate specialization profiles. These findings reveal further metabolic versatility and niche specialization within abundant *Chlorobium*-associated bacteria.

RESULTS AND DISCUSSION

Metagenome assembly and characterization reveals that most planktonic members of the class *Chlorobia* affiliate with the genus *Chlorobium*. We investigated the planktonic *Chlorobia*-affiliated populations of 30 lakes and 12 ponds

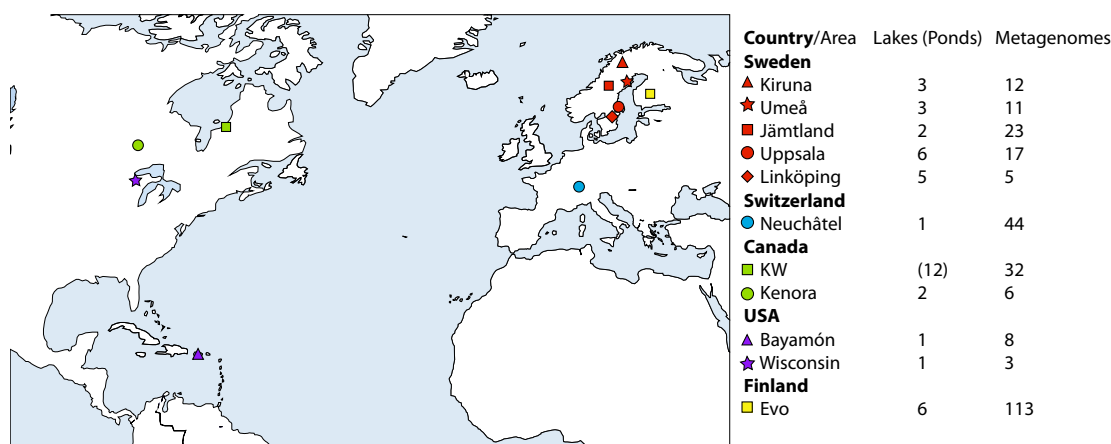| **Country**/Area	Lakes (Ponds)	Metagenomes
Sweden		
▲ Kiruna	3	12
★ Umeå	3	11
■ Jämtland	2	23
● Uppsala	6	17
◆ Linköping	5	5
Switzerland		
● Neuchâtel	1	44
Canada		
■ KW	(12)	32
● Kenora	2	6
USA		
▲ Bayamón	1	8
★ Wisconsin	1	3
Finland		
■ Evo	6	113

FIG 1 Map illustrating the lake and pond metagenome sampling sites for mapping and recovery of metagenome-assembled genomes (MAGs) and single-amplified genomes (SAGs). Symbols are colored by country, and a unique shape is used for each area within the same country. "KW" represents Kujjuarapik-Whapmagoostui.

sampled from Europe and North America (Fig. 1). Most of these samples were collected with the objective of investigating depth-associated changes in taxonomic diversity in different water masses that capture oxic epilimnion, metalimnion, and anoxic hypolimnion conditions (29). Following assembly and binning of individual metagenomes, we compiled 454 new 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 SAGs, we further compiled 25 genomes available from the Genome Taxonomy Database (GTDB) and 11 MAGs available from previous studies (27, 30–32). In total, the data set included 509 genomes, including the MAGs, SAGs, and complete genomes of isolates (see Table S1 in the supplemental material). 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% in all cases.

We clustered recovered genomes by 95% average nucleotide identity (ANI), which has previously been shown to unite classical species definitions and separate sequenced strains into consistent and distinct groups (33–35). Clustering into mOTUs provided a more comprehensive understanding of the genomic structure of genus *Chlorobium* members. We obtained 71 metagenomic operational taxonomic units (mOTUs) belonging to the class *Chlorobia* (see Table S2 in the supplemental material), of which 57 mOTUs were classified as members of the genus *Chlorobium* based on GTDB taxonomy (Fig. 2; Table S2) (27). These 57 mOTU genomes had an average completeness of approximately 90% and an estimated genome size of 2.6 Mbp (range, 2.1 to 3.7 Mbp), which are consistent with previous studies showing that *Chlorobium* genomes from isolates range from 1.9 to 3.3 Mbp (2). Among 57 *Chlorobium*-associated mOTUs, 13 were composed exclusively of different genomes from previously described nonidentical isolates (2), including *Chl. phaeobacteroides*, *Chl. limicola*, *Chl. luteolum*, *Chl. ferrooxidans*, and *Chl. phaeoclathratiforme* (Fig. 2; and Table S2). 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 suggests that the majority of natural diversity within freshwater planktonic members of the class *Chlorobia* is represented by uncultivated members of the genus *Chlorobium*.

The genus *Chlorobium* includes both cosmopolitan and endemic lineages. By identifying MAGs from different source locations that clustered within the same mOTUs, we were able to determine whether a "species" is widely distributed and putatively cosmopolitan or endemic to specific locations. To investigate the abundance
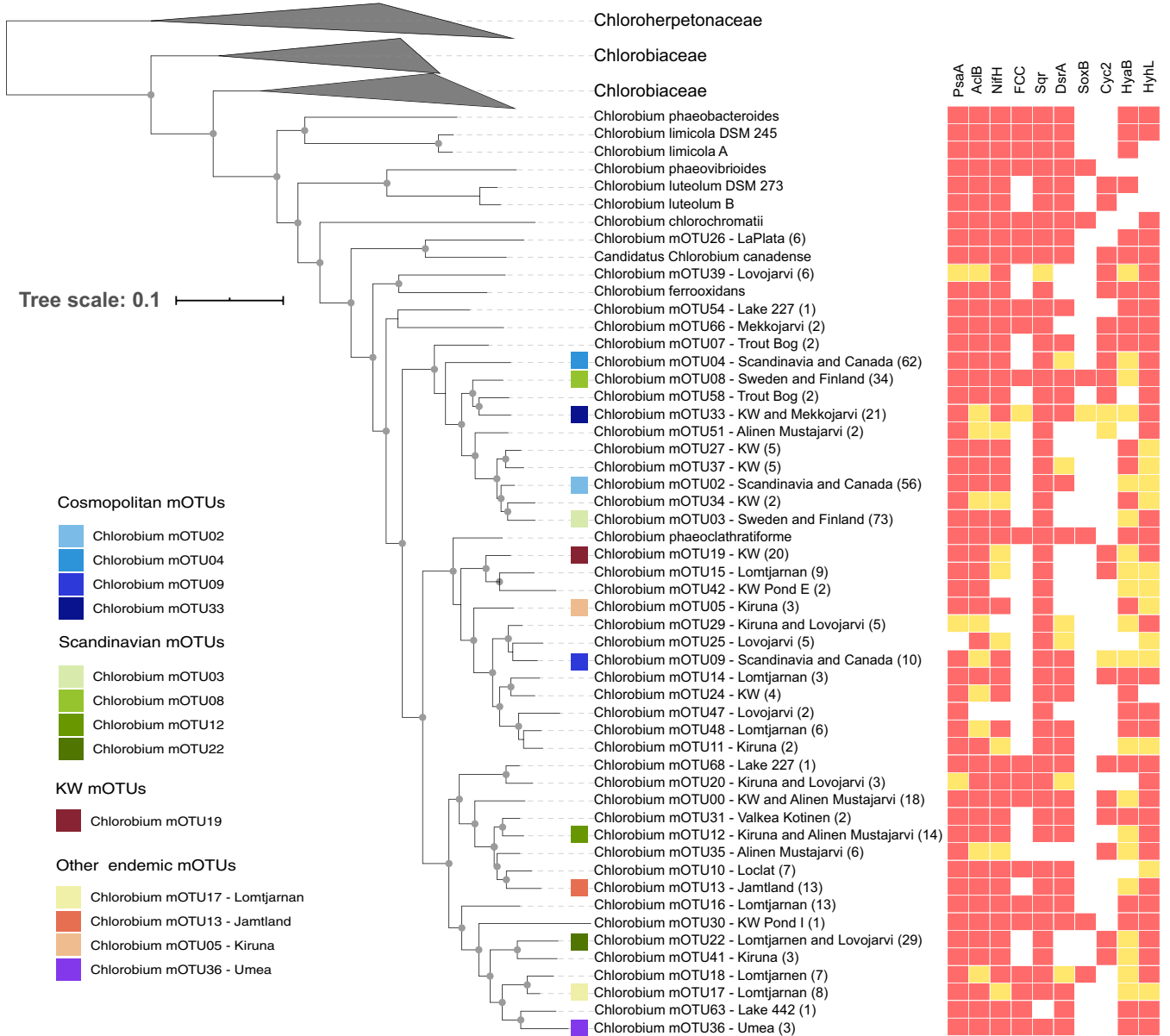
FIG 2 Phylogenetic relationships, metabolic capabilities, and ecological distributions of *Chlorobium* mOTUs. The tree was constructed using GTDB-Tk "de-novo" alignment and was annotated and curated in iTOL. Isolates are shown in the tree with species names instead of mOTU numbers. Other tips on the tree show the genus name and mOTU number followed by the location from which the genomes were assembled. If the MAGs were assembled from more than one lake in the same area, then the name of the area is written. If the MAGs were assembled from more than one area in the same country, then the name of the country is written. Lake La Plata is in Bayamon, Puerto Rico; 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, Finland. Lake Trout Bog is in Wisconsin, Lake Loclat is in Neuchatel, Switzerland, and Lake Lomtjarnen is in Jamtland, Sweden. Numbers in parentheses show 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 several genes, as indicated by their products: i.e., 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), and HyhL (group 3b [NiFe]-hydrogenase large subunit). In the heat map, red indicates that the gene is present in the core genome of the corresponding mOTU, and yellow indicates that it is present in an accessory genome of the mOTU. White indicates that the gene is absent in the corresponding mOTU. The 13 most abundant *Chlorobium* mOTUs in the investigated environments have a colored square before the name of the mOTU. Gray circles represent bootstrap values higher than 50%. Please see Table S2 for information on all 71 mOTUs.

and prevalence of each reconstructed mOTU within resident *Chlorobia* assemblages, we mapped metagenomic reads from the lake and pond data sets (see Table S3 in the supplemental material) against the genome collection (Table S1). Reads of our lake/pond data set mapped to 45 (out of 71) *Chlorobia* mOTUs, of which 42 mOTUs were classified as members of the genus *Chlorobium* (Fig. 3). None of these mOTUs
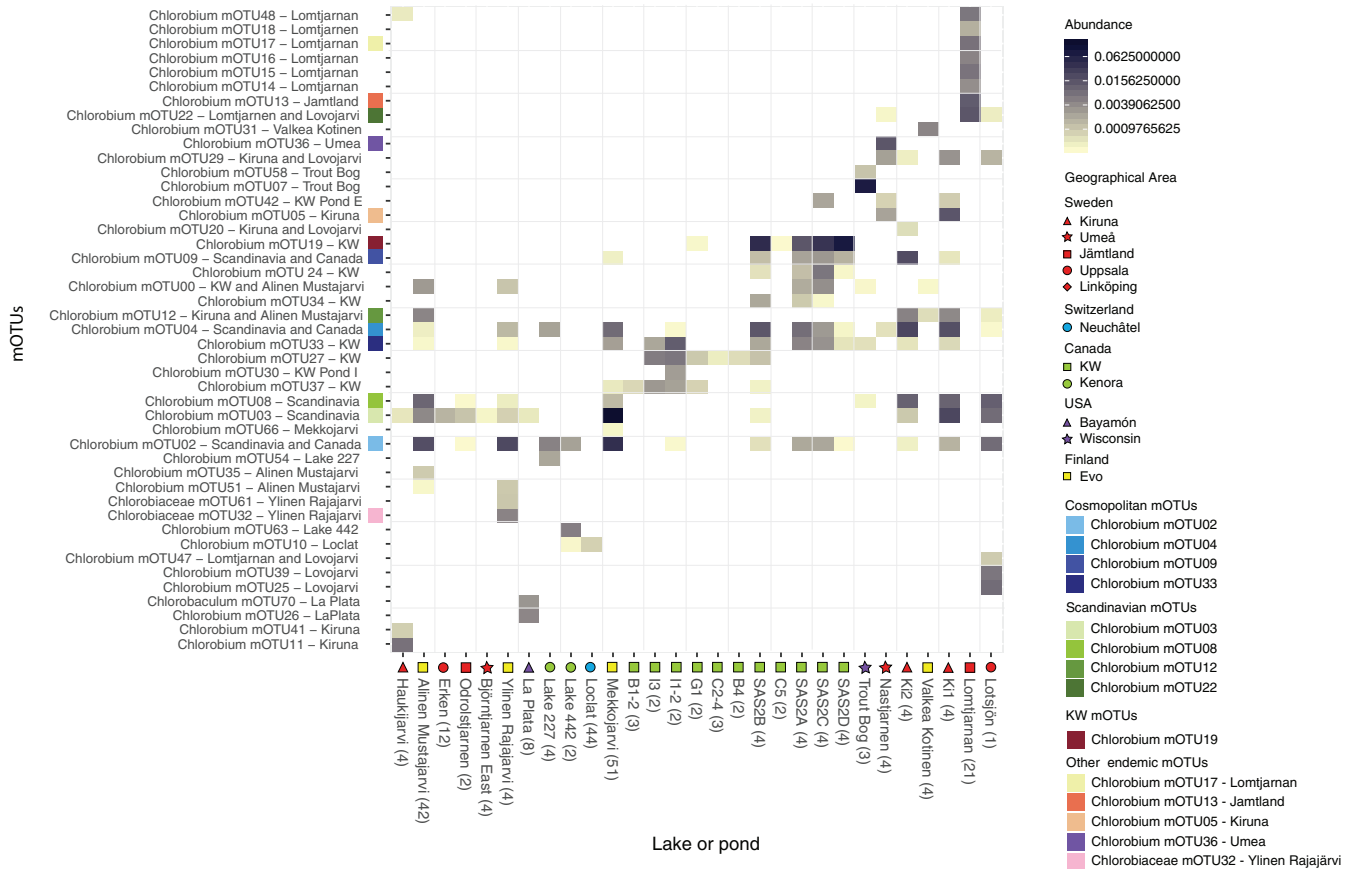
FIG 3 Relative abundance of metagenome reads for the 45 *Chlorobia* mOTUs present in lake and pond metagenomes. The top 14 *Chlorobia* mOTUs are color coded according to the location from which they were assembled. Metagenomes were subsampled to one million reads and then mapped competitively with 100% identity cutoff to all genomes in all mOTUs. The reads were normalized to the relative abundance of reads per metagenome. Relative abundances from all depths and time points were averaged, and the number of the samples averaged is shown in the parentheses next to the lake or pond name. The cutoff for presence of an mOTU to be included was 0.03% (i.e., 0.0003 in the figure) read abundance per lake/pond. The name of the mOTUs includes geographical information about the origin of the genomes. The lakes or ponds shown along the *x* axis include a symbol for the region where they are located.

corresponded to cultivated *Chlorobium* representatives. This provides further evidence that uncultivated *Chlorobium* members dominated planktonic *Chlorobia* present in sampled lakes and ponds.

Out of the 42 locations represented in this study, we found that metagenomes sequenced from 18 lakes (of 30 total) and 11 ponds (of 12 total) contained *Chlorobia*-affiliated reads (Fig. 3). However, for 10 of the lakes that did not show *Chlorobia* read abundance, we only have epilimnion samples with full oxygenation, and so it was expected to not find *Chlorobia* reads or genomes. We found four mOTUs composed of MAGs reconstructed from assemblies originating from both studied continents, i.e., mOTU02, mOTU04, mOTU09, and mOTU33 (in blue in Fig. 2 to 5), and we define these mOTUs as cosmopolitan because of their broad distributions. Metagenomic read recruitment confirmed that these cosmopolitan mOTUs were present in lakes and ponds from both studied continents (Fig. 3). In contrast, the other 41 mOTUs were less widely distributed and were found in just one lake, in one geographical area, or just 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* (Fig. 2).

Depth stratification of *Chlorobia* mOTUs was evident when mapped reads were visualized as relative abundances, based on total metagenomic reads, across lake water column samples (Fig. 4). As expected, our study showed that *Chlorobium* mOTUs were found mostly below the oxycline (Fig. 4), and several were highly abundant within their
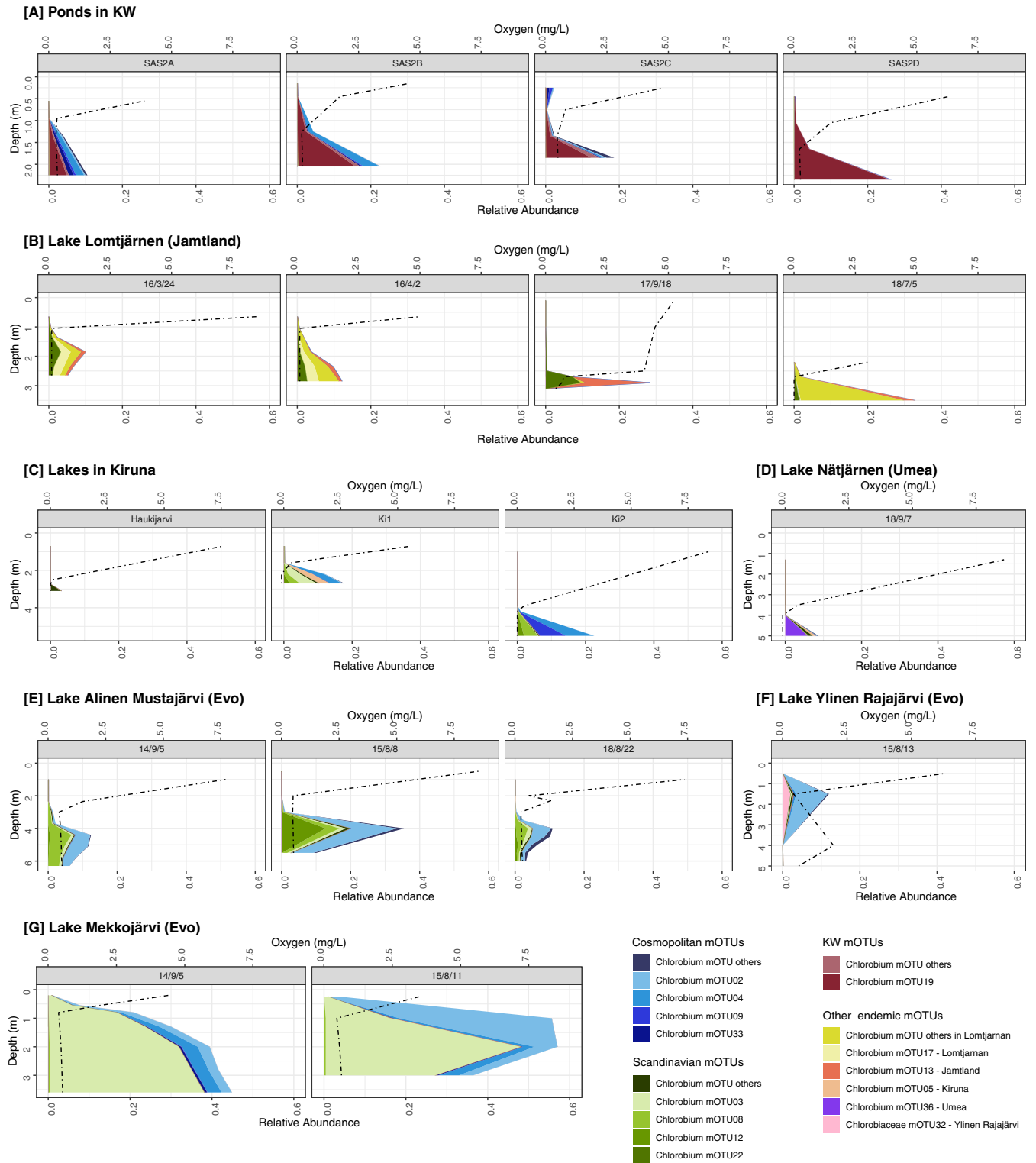
FIG 4 Distribution of environmentally abundant mOTUs across depth profiles in the lake and pond data sets with depth-discrete sampling. Note that the 14 most abundant mOTUs have individual color coding, and the remaining abundant mOTUs are combined in several categories labeled "other." Oxygen is represented by a dashed line. (A) Ponds in the KW area. (B) Time points for Lake Lomtjärnen in Sweden. (C) Different lakes in the Kiruna area in Sweden. (D) Lake Nästjärnen in Sweden. (E) Time points of Lake Alinen Mustajärvi in Finland. (F) Lake Ylinen Rajajärvi in Finland. (G) Time points for Lake Mekkojärvi in Finland.

respective aquatic microbial communities. In particular, *Chlorobium* sequences comprised 57% of the community at several depths of Lake Mekkojärvi, Finland (Fig. 4G). Such relative abundances are consistent with previous studies showing that this class can constitute 12 to 47% of reads from lake ecosystems (31, 32, 37, 38). Our data show one clear example of multiple *Chlorobium* populations typically coinhabiting the same lake with spatially separated niches. Specifically, mOTU02 and mOTU03 were differentially distributed by depth in August 2015 within Lake Mekkojärvi samples (Fig. 4G). Similar findings were previously reported in Trout Bog, WI, 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) (30, 32). Previous research suggests that such niche specialization may be explained by distinct pigment absorbance profiles (32, 39).

Several mOTUs appear to be endemic to specific geographical and seasonal niches. For example, we observed that mOTUs 22, 17, and 13 were temporally stable within Lake Lomtjärnen, Sweden (Fig. 4B). These mOTUs were observed across four time points: 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 time points. The mOTUs residing in this lake were mostly endemic to the lake or to the region, and they were consistently recovered at different times. In addition, the relative abundance of these mOTUs changed according to the sampling time. For example, mOTU13 increased in relative abundance at the September sampling point. We predict that some endemic mOTUs may occupy specific seasonal niches and suggest that a more temporally resolved sampling effort, together with cell counts to calculate absolute abundances, would help test this hypothesis.

The metabolic potential of the most abundant *Chlorobium* groups. We investigated the metabolic potential of all *Chlorobium*-associated mOTUs. Combining ANI clustering and statistical approaches allowed us to calculate the probability that specific genes are present in the core or accessory genomes of the "species," despite the incompleteness of several MAGs and SAGs. Overall, 42 out of the 71 *Chlorobia* mOTUs were composed of more than one genome, and hence we found power in replication (40). For example, mOTU08 included 34 MAGs with an average completeness of 95%, and the mOTU contained all surveyed functional genes (Fig. 2) in the core genome, except for HyaB, which was putatively within the accessory genome. Moreover, in order for a gene to be completely missing from the mOTU, the gene had to be absent in all of the MAGs and SAGs associated with that particular mOTU. The absence of a gene in our mOTU data is strong evidence for the absence of that gene from the corresponding population, unlike traditional MAG-based studies that select a single representative MAG for analysis.

We found an average of 710 annotated genes with designated KEGG functions per mOTU. For 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 (Fig. 2 and 5; see Fig. S1A in the supplemental material). These findings are consistent with previous studies showing very minor differences in genes encoding the photosynthetic type I reaction center unit in this class (41). Molybdenum- and iron-nitrogenases were also widely encoded by the representatives of the class *Chlorobia* (Fig. S1C). Thus, light harvesting, carbon fixation, and nitrogen fixation appear to be conserved core metabolic features of freshwater *Chlorobia* members.

In contrast to functions that were widely encoded, the products of *Chlorobia*-associated mOTUs varied significantly in their encoded 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
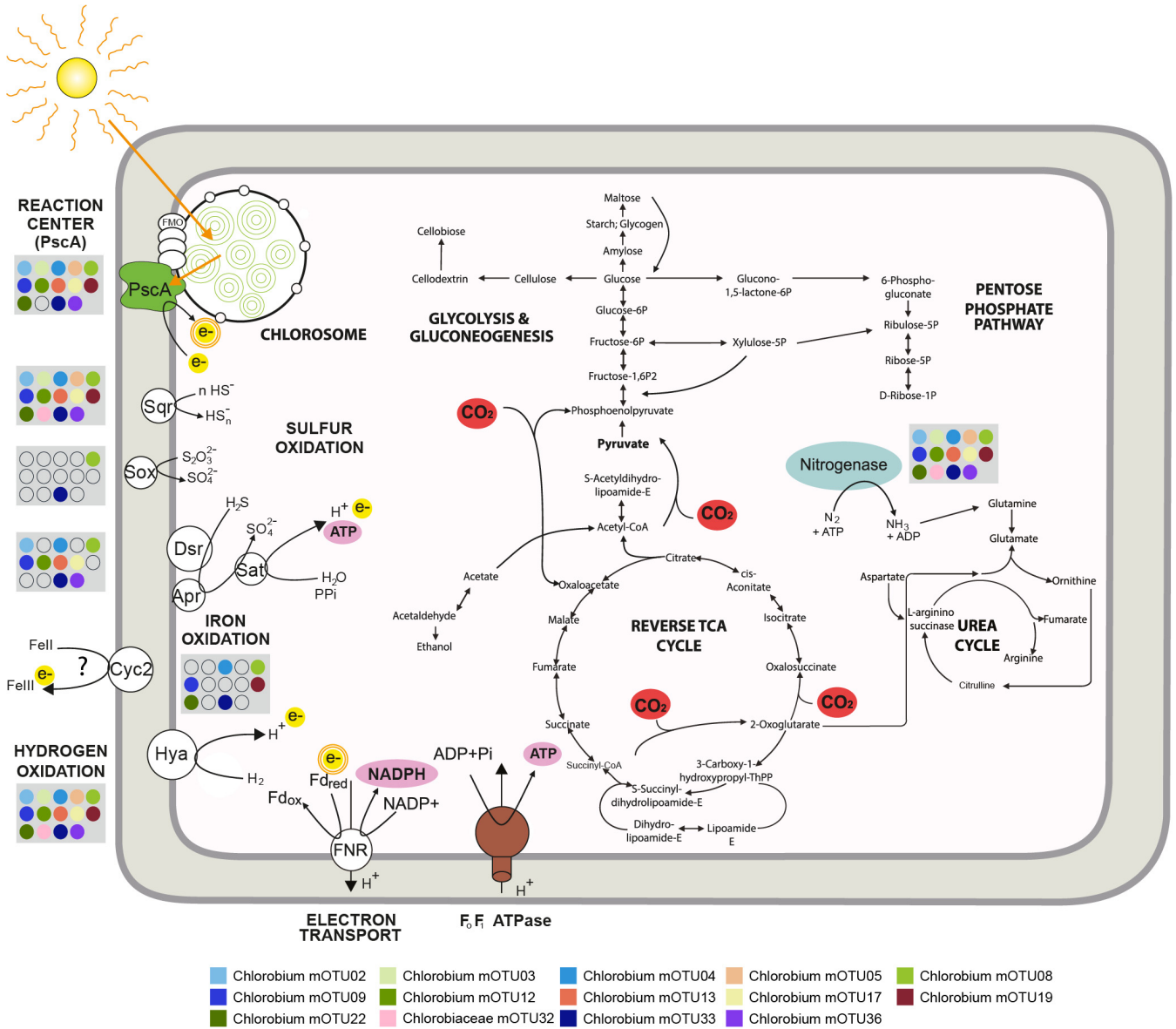
FIG 5 Metabolic potential of the 14 most abundant *Chlorobia* mOTUs detected in the studied freshwater ecosystems, with a focus on photosynthesis, electron transport, and carbon fixation. Blue indicates nitrogenase, green indicates chlorosome and photochemical reaction center PscA, yellow indicates electrons being donated through oxidation reactions and photosynthesis, pink indicates reductant (NADPH) and chemical energy (ATP) produced during oxidation reactions and photosynthesis, and red indicates 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, and iron oxidation and activates them using light energy, which allows proton pumping and ferredoxin reduction. Ferredoxin reduction is linked to photosystem activity, depicted with the double orange circle in electron. In the gray boxes, circles show whether marker genes are present or absent in the respective mOTUs in the legend. Genes are indicated by the following gene product abbreviations: 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 (ferredoxin-NADP oxidoreductase), and Fd (ferredoxin).

oxidoreductases (Sqr) were encoded by all but two *Chlorobium* mOTUs (Fig. 2; Fig. S1B), whereas reverse dissimilatory sulfite reductases (i.e., DsrA) and flavocytochrome *c* sulfide dehydrogenase (FCC) were encoded by 71% and 53% of the mOTUs, respectively (Fig. 2). A more restricted trait was the capacity to oxidize thiosulfate via thiosulfohydrolase (i.e., SoxB). The corresponding gene was present in only 14% of the *Chlorobium* mOTUs, although it was found in all *Chlorobaculum* mOTUs (Fig. 2; Table S2). Others have suggested that members of the class *Chlorobia* have adapted to different environments by acquiring distinct electron transfer complexes through horizontal gene transfer (41–43). Even though genes for the Dsr and Sox complexes may
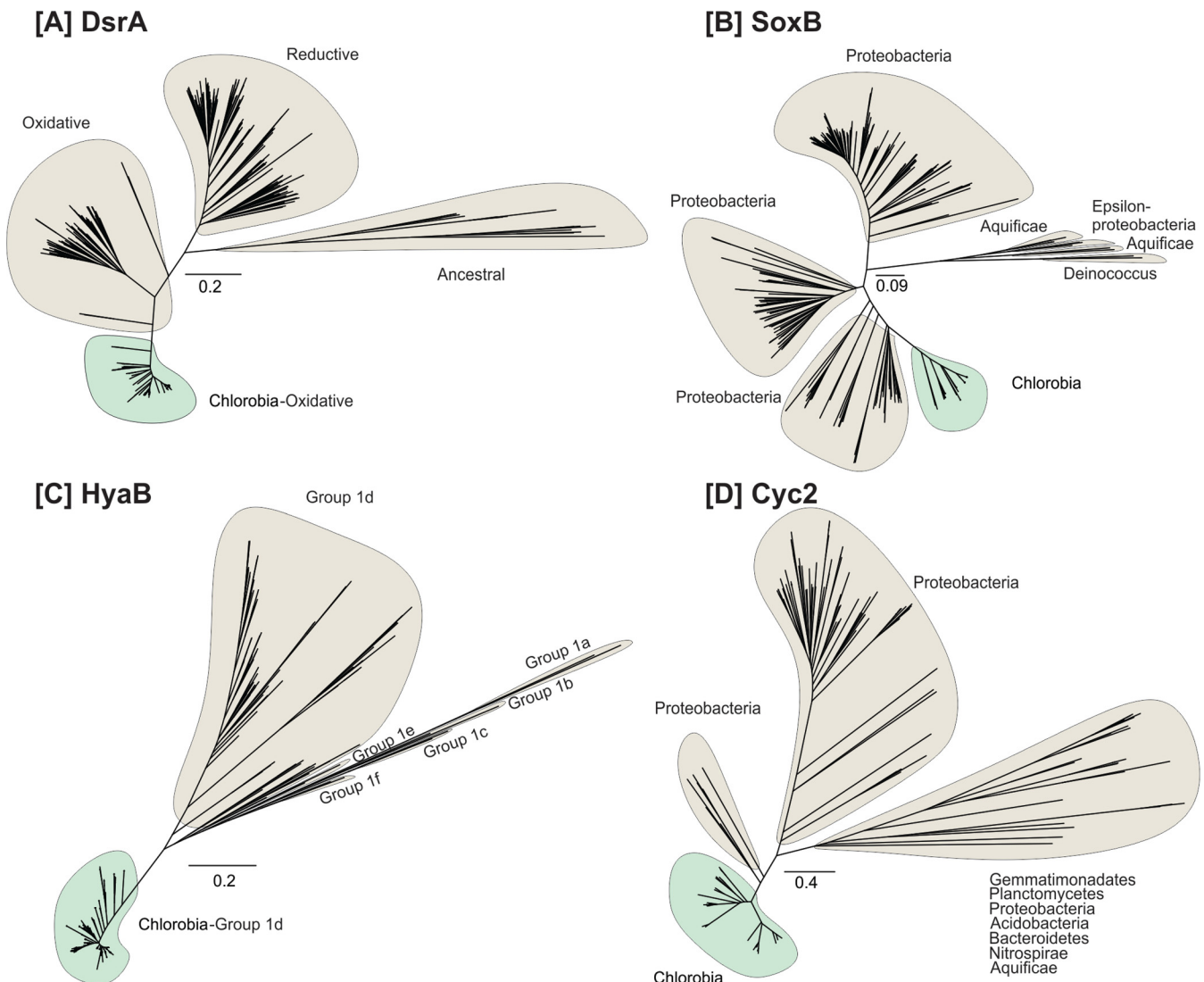
FIG 6 Maximum likelihood phylogeny of the recovered protein sequences from reconstructed *Chlorobia*-affiliated MAGs/SAGs (highlighted in green) together with the reference protein sequences recovered from GenBank, with the exception of hydrogenases, for which sequences were collected from HydDB (highlighted in gray). Panels A to D show the phylogeny of the following protein sequences: (A) dissimilatory sulfite reductase (DsrA; $n = 551$), (B) thiosulfohydrolase (SoxB; $n = 294$), (C) group 1d [NiFe]-hydrogenase (HyaB; $n = 579$), and (D) putative iron-oxidizing cytochrome (Cyc2; $n = 306$).

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

Genes encoding the catalytic subunits of group 1d and group 3b [NiFe]-hydrogenases (HyaB and HyhL, respectively) were detected in 82 and 91% of the *Chlorobium* mOTUs, respectively (Fig. 2 and 6C; Fig. S1C). The group 1d enzymes are known to support hydrogenotrophic respiration and anoxygenic photosynthesis in diverse 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 (Fig. 6C). The physiological role of group 3b hydrogenases (HyhL) in *Chlorobia* genomes is less clear. These bidirectional cytosolic enzymes are likely to support either hydrogenotrophic carbon fixation or facultative hydrogenogenic fermentation in this class (Fig. 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 widespread distribution of [NiFe]-hydrogenases across bacterial phyla (51, 52) and the conservation of group 1d lineages in sister classes within the phylum *Bacteroidota*.

Our study also shows that 40% of the *Chlorobium* mOTUs harbor *cyc2* gene homologs (Fig. 2 and 6D), encoding a potential outer membrane *c*-type cytochrome capable of ferrous iron oxidation (31, 53). This finding 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 (31), this implies that *Chlorobia* members could play an important role in iron biogeochemistry in lake systems globally.

Metabolic flexibility has been found to be a key factor governing distributions of taxa across ecosystems with disturbances (54). However, we did not find any mOTU containing all the putative oxidation genes in the core genome (Fig. 2 and 5), nor did we find any correlation between how widespread an mOTU is and its capacity to use different electron donors. A broader sampling across different temporal and spatial scales could reveal whether metabolic versatility governs the prevalence and abundance of *Chlorobium* members on a global scale.

Outlook of ecological roles of *Chlorobium* members. At first glance, members of the *Chlorobia* appear to populate a very specific niche in water columns of global lakes and ponds. As we confirm in our data sets, they thrive under the oxycline in the anaerobic layer where light is still available. They have evolved physiological adaptations to low light intensities (4) and therefore are particularly well adapted to freshwater systems characterized by high concentrations of colored dissolved organic carbon (DOC) that absorb 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 (25) and also seem to play an important role as diazotrophs (16, 17). Despite the specific niches occupied by *Chlorobium* populations, our work illustrates considerable metabolic versatility with regard to electron donors that may be used for carbon fixation. Through their variable capacity to recycle sulfide, hydrogen, and iron produced by dissimilatory sulfate reducers, hydrogenogenic fermenters, and dissimilatory iron reducers, *Chlorobium* members potentially influence the entire microbiome of anoxic freshwaters (57).

Our analyses describe the metabolic potential of environmentally relevant *Chlorobium* genomes, along with their geographical distributions, revealing both endemic and cosmopolitan clades and suggesting complex metabolic implications for those distributions. Distributions of *Chlorobia* populations appear governed by ecological factors beyond overall known metabolic potential. Moreover, our study identified clades of *Chlorobium* with abundant and cosmopolitan geographical distributions that do not show monophyletic clustering. Although comprehensive metadata and better geographical and temporal resolution would help identify the specific physicochemical mechanisms that lead to specific ecological advantages, our data nonetheless highlight previously overlooked ecological diversity for this globally distributed bacterial lineage.

MATERIALS AND METHODS

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 (Fig. 1; Table S3) (29). Briefly, samples were collected between 2009 and 2018, with most obtained by biomass collection on 0.22-μm-pore Sterivex filters (Millipore). Oxygen concentrations were measured with a YSI 55 oxygen probe (Yellow Springs Instruments, Yellow Springs, OH, USA). For all samples, DNA was extracted using the PowerSoil DNA extraction kit (MoBio, Carlsbad, CA, USA) following the manufacturer's instructions. Sequencing libraries were prepared from 10 to 20 ng of purified DNA using the ThruPLEX DNA-seq Prep kit according to the manufacturer's preparation guide.

The quality of the libraries was evaluated using the Agilent Fragment Analyzer system and the DNF-910-kit. The adapter-ligated fragments were quantified by quantitative PCR (qPCR) using the Library quantification kit for Illumina (KAPA Biosystems/Roche) on a CFX384 Touch real-time PCR detection system (Bio-Rad) prior to cluster generation and sequencing. The sequencing libraries were pooled and subjected to cluster generation and paired-end sequencing with 150-base read-length S2/S4 flow cells and the NovaSeq6000 (Illumina Inc.) using the v1 chemistry according to the manufacturer's protocols. Sequencing was performed by the SNP&SEQ Technology platform (Uppsala, Sweden). Each sample generated an average of 10.5 Gb of sequence data. Moreover, we added six metagenomic libraries generated for a *Chlorobium* study in Canada, focusing on iron oxidation (31) and three additional metagenomic libraries as part of a 15-year time series analysis of Trout Bog, WI (30).

Genome collection. Assembling and binning the 265 metagenomes resulted in ~12,000 metagenome-assembled genomes (MAGs) from water bodies spanning five different countries (29). Of these, 454 belonged to the class *Chlorobia* and were used in this study. In brief, raw data were trimmed using Trimmomatic version 0.36 (58) and were then assembled with MEGAHIT version 1.1.13 (59) with default settings. Two assemblies were performed: single assemblies for all samples individually and 53 coassemblies that were performed mostly with lake-specific metagenome data (see Table S4 in the supplemental material). More detailed methods for sample collection, metadata, processing, metagenome generation, and analyses are reported elsewhere (29).

The relevant quality-controlled reads were mapped to all assemblies using BBMap (60) with default settings. Mapping results were used to bin contigs using MetaBAT version 2.12.1 (parameters –maxP 93 –minS 50 -m 1500 -s 10000) (61). Moreover, we collected genomes of the class *Chlorobia* from the GTDB and collected several other published MAGs that fulfilled the medium-quality threshold of ≥50% completeness and ≤5% contamination (62). In total, we compiled 509 genomes that comprised the following: 454 MAGs and 19 SAGs that were from stratified water bodies in Sweden, Finland, Canada, Switzerland, and Puerto Rico (29); 7 MAGs from enrichment culture or lake metagenomes of lakes at the International Institute for Sustainable Development Experimental Lakes Area (IISD-ELA; near Kenora, Canada) (31); 4 MAGs from Trout Bog, WI (30, 32); and 25 genomes (MAGs, SAGs, or isolates) from the GTDB (2, 63). Completeness and contamination were assessed with CheckM (64).

Single-cell collection. Cell sorting was done in 2016 on a MoFlo Astrios EQ sorter (Beckman Coulter, USA) using 488- and 532-nm lasers for excitation, a 70-μm nozzle, a sheath pressure of 60 lb/in^2, and 0.1-μm-pore-filtered 1$\times$ phosphate-buffered saline (PBS) as sheath fluid. An ND (neutral-density) filter (ND = 1) and 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 by autofluorescence using a 532-nm laser and band-pass filters 710/45 and 664/22. Sorted plates were stored at −80°C. Whole-genome amplification was conducted using the REPLI-g single-cell kit (Qiagen) following the manual's instruction, but with a reduced total reaction volume (12.5 μl). Amplified DNA was mixed thoroughly by pipetting up and down. After screening for bacterial 16S rRNA genes, DNA from confirmed *Chlorobia* members was sequenced on an Illumina HiSeqX v2.5 PE at 2 × 150 bp (29).

mOTU analysis. Average nucleotide identity (ANI) for all genome pairs was computed with FastANI version 1.3 (65), 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 clustered into connected components. Additionally, less complete genomes (completeness above 50% yet below 70%) were recruited to the mOTU if the ANI similarity was above 95%. Classification of genomes was done with GTDB-Tk and sourmash version 1.0 (63, 66). The most complete MAG for each mOTU was selected as a representative, and a phylogenetic tree was calculated using GTDB-Tk version 102, with database release 89, using *Chloroherpetonaceae* as an outgroup (63). The aligned genome tree was loaded and curated in iTOL version 5.5.1 (67). Moreover, KO (KEGG Orthology) number functions were classified as core if the probabilities of their presence-absence profiles were higher under the assumption that they were in every genome, considering their incompleteness, than the converse probability (as computed by mOTUpan in mOTUlizer version 0.1.3 [https://github.com/moritzbuck/mOTUlizer]).

Abundance profiles. To explore abundance profiles, metagenomic libraries were subsampled to 1,000,000 reads. 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 the relative abundance of each mOTU per metagenome. All relative abundances for all samples in each location were averaged. The cutoff for the presence of an mOTU was set to 0.03% per site (0.0003 in Fig. 3). Heat maps of abundance were calculated and plotted using the R packages ggplot2 and phyloseq with parameters NMDS (nonmetric multidimensional scaling) and Bray-Curtis for choosing the order of the *x* axes (68–71). Depth-discrete abundance profiles were plotted using R packages ggplot2, phyloseq, and cowplot (69–71).

Metabolic genes. The metabolic potentials of the *Chlorobia* MAGs, SAGs, and reference genomes were reconstructed based on eggNOG-mapper (72, 73) annotations. Ten of the 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 (dissimilatory sulfite reductase A subunit), SoxB (thiosulfohydrolase), Cyc2 (outer membrane monoheme *c*-type cytochrome), HyaB (group 1d [NiFe]-hydrogenase large subunit), and HyhL (group 3b [NiFe]-hydrogenase large subunit)—were further annotated in reconstructed MAGs, SAGs, and reference genomes using Diamond version 0.9.31 (74) against custom-built databases for each marker protein with an 80% query coverage threshold. A 50% identity threshold was used for all marker gene products, except for an 80% threshold used for the PsaA protein. Annotations were further validated by constructing phylogenetic trees.

Single-gene phylogeny. Maximum likelihood trees were constructed using the amino acid sequences for six metabolic marker proteins: PsaA (photosystem I P700 chlorophyll *a* apoprotein A1), NifH (nitrogenase iron protein), Sqr (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), and HyhL (group 3b [NiFe]-hydrogenase large subunit). Reference genes for each marker were collected from NCBI GenBank, with the exception of hydrogenases, for which sequences were collected from HydDB (46). For each metabolic marker protein, sequences retrieved from the *Chlorobia* mOTUs were aligned against reference protein sequences using ClustalW in MEGA7 (75, 76). Evolutionary relationships were visualized by constructing a maximum likelihood phylogenetic tree. Specifically, initial trees for the heuristic search were obtained automatically by applying Neighbor-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 50 replicates.

Data availability. All data sets are available in public repositories under NCBI project accession numbers PRJEB38681, PRJNA518727, and PRJNA534305. Accession numbers of the genomes are available (Table S1), as are the accession numbers of metagenomes (Table S3). Assembling and binning of the original data set used scripts available at https://github.com/moritzbuck/metasssnake, and general processing scripts for this project are available at https://github.com/moritzbuck/0026_Chlorobi.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.
FIG S1, JPG file, 2.5 MB.
TABLE S1, CSV file, 0.1 MB.
TABLE S2, CSV file, 0.02 MB.
TABLE S3, CSV file, 0.03 MB.
TABLE S4, CSV file, 0.03 MB.

REFERENCES

1. Bryant DA, Frigaard NU. 2006. Prokaryotic photosynthesis and phototrophy illuminated. Trends Microbiol 14:488–496. https://doi.org/10.1016/j.tim.2006.09.001.
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,

p 47–102. *In* Burnap RL, Vermaas W (ed), Functional genomics and evolution of photosynthetic systems, vol 33. Springer, Dordrecht, Netherlands.
3. Bhatnagar S, Cowley ES, Kopf SH, Pérez Castro S, Kearney S, Dawson SC, Hanselmann K, Ruff SE. 2020. Microbial community dynamics and coexistence in a sulfide-driven phototrophic bloom. Environ Microbiome 15:3. https://doi.org/10.1186/s40793-019-0348-0.

4. 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. Proc Natl Acad Sci U S A 102:9306–9310. https://doi.org/10.1073/pnas.0503674102.

5. Imhoff JF, Thiel V. 2010. Phylogeny and taxonomy of *Chlorobiaceae*. Photosynth Res 104:123–136. https://doi.org/10.1007/s11120-009-9510-7.

6. 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:e00626-18. https://doi.org/10.1128/mSphere.00626-18.

7. Overmann J. 2008. Ecology of phototrophic sulfur bacteria, p 375–396. *In* Hell R, Dahl C, Knaff DB, Leustek KT (ed), Advances in photosynthesis and respiration sulfur metabolism in phototrophic organisms, vol 27. Springer, New York, NY.

8. Sirevåg R. 1995. Carbon metabolism in green bacteria., p 871–883. *In* Blankenship RE, Madigan MT, Bauer CE (ed), Anoxygenic photosynthetic bacteria. Advances in photosynthesis and respiration, vol 2. Springer, Dordrecht, Netherlands.

9. Buchanan BB, Arnon DI. 1990. A reverse Krebs cycle in photosynthesis—consensus at last. Photosynth Res 24:47–53. https://doi.org/10.1007/BF00032643.

10. Camacho A, Walter XA, Picazo A, Zopfi J. 2017. Photoferrotrophy: remains of an ancient photosynthesis in modern environments. Front Microbiol 8:323. https://doi.org/10.3389/fmicb.2017.00323.

11. Heising S, Richter L, Ludwig W, Schink B. 1999. *Chlorobium ferrooxidans* sp. nov., a phototrophic green sulfur bacterium that oxidizes ferrous iron in coculture with a "*Geospirillum*" sp strain. Arch Microbiol 172:116–124. https://doi.org/10.1007/s002030050748.

12. Llirós M, García-Armisen T, Darchambeau F, Morana C, Triadó-Margarit X, Inceoğlu Ö, Borrego CM, Bouillon S, Servais P, Borges AV, Descy J-P, Canfield DE, Crowe SA. 2015. Pelagic photoferrotrophy and iron cycling in a modern ferruginous basin. Sci Rep 5:13803. https://doi.org/10.1038/srep13803.

13. 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 Microbiol Ecol 93:fix054. https://doi.org/10.1093/femsec/fix054.

14. Lambrecht N. 2019. Insights into early Earth ocean biogeochemistry from intensive monitoring of two ferruginous meromictic lakes. PhD dissertation. Iowa State University, Ames, IA.

15. Heda GD, Madigan MT. 1986. Aspects of nitrogen-fixation in *Chlorobium*. Arch Microbiol 143:330–336. https://doi.org/10.1007/BF00412798.

16. 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 11:1500. https://doi.org/10.3389/fmicb.2020.01500.

17. Fernandez L, Bertilsson S, Peura S. 2020. Non-cyanobacterial diazotrophs dominate nitrogen-fixing communities in permafrost thaw ponds. Limnol Oceanogr 65:S180–S193. https://doi.org/10.1002/lno.11243.

18. 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. ISME J 4:1002–1019. https://doi.org/10.1038/ismej.2010.28.

19. Peura S, Sinclair L, Bertilsson S, Eiler A. 2015. Metagenomic insights into strategies of aerobic and anaerobic carbon and nitrogen transformation in boreal lakes. Sci Rep 5:12102. https://doi.org/10.1038/srep12102.

20. 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. https://doi.org/10.1111/j.1574-6941.2009.00737.x.

21. 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. https://doi.org/10.1186/s40168-020-00889-8.

22. Rimmer A, Ostrovsky I, Yacobi YZ. 2008. Light availability for *Chlorobium phaeobacteroides* development in Lake Kinneret. J Plankton Res 30:765–776. https://doi.org/10.1093/plankt/fbn037.

23. Fry B. 1986. Sources of carbon and sulfur nutrition for consumers in three meromictic lakes of New York State. Limnol Oceanogr 31:79–88. https://doi.org/10.4319/lo.1986.31.1.0079.

24. 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. J Plankton Res 30:1189–1202. https://doi.org/10.1093/plankt/fbn070.

25. Culver DA, Brunskill GJ. 1969. Fayetteville Green Lake, New York. V. Studies of primary production and zooplankton in a meromictic marl lake. Limnol Oceanogr 14:862–873. https://doi.org/10.4319/lo.1969.14.6.0862.

26. Comeau AM, Harding T, Galand PE, Vincent WF, Lovejoy C. 2012. Vertical distribution of microbial communities in a perennially stratified Arctic lake with saline, anoxic bottom waters. Sci Rep 2:604. https://doi.org/10.1038/srep00604.

27. 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. Nat Biotechnol 36:996–1004. https://doi.org/10.1038/nbt.4229.

28. 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. Proc Natl Acad Sci U S A 99:9509–9514. https://doi.org/10.1073/pnas.132181499.

29. Buck M, Garcia SL, Fernandez Vidal L, Martin G, Martinez Rodriguez GA, Saarenheimo J, Zopfi J, Bertilsson S, Peura S. 2020. Comprehensive dataset of shotgun metagenomes from stratified freshwater lakes and ponds. bioRxiv https://doi.org/10.1101/2020.11.12.379446.

30. 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. https://doi.org/10.1038/ismej.2015.241.

31. 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 14:2732–2747. https://doi.org/10.1038/s41396-020-0725-0.

32. Berg M, Goudeau D, Olmsted C, McMahon KD, Yitbarek S, Thweatt JL, Bryant DA, Eloe-Fadrosh EA, Malmstrom RR, Roux S. 2021. Host population diversity as a driver of viral infection cycle in wild populations of green sulfur bacteria with long standing virus-host interactions. ISME J https://doi.org/10.1038/s41396-020-00870-1.

33. Konstantinidis KT, Tiedje JM. 2005. Genomic insights that advance the species definition for prokaryotes. Proc Natl Acad Sci U S A 102:2567–2572. https://doi.org/10.1073/pnas.0409727102.

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 Res 43:6761–6771. https://doi.org/10.1093/nar/gkv657.

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. https://doi.org/10.1038/s41396-017-0001-0.

36. Overmann J, Pfennig N. 1989. *Pelodictyon phaeoclathratiforme* sp. nov., a new brown-colored member of the Chlorobiaceae forming net-like colonies. Arch Microbiol 152:401–406. https://doi.org/10.1007/BF00425181.

37. Taipale S, Jones RI, Tiirola M. 2009. Vertical diversity of bacteria in an oxygen-stratified humic lake, evaluated using DNA and phospholipid analyses. Aquat Microb Ecol 55:1–16. https://doi.org/10.3354/ame01277.

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. Aquat Microb Ecol 68:267–272. https://doi.org/10.3354/ame01620.

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. https://doi.org/10.1128/JB.186.9.2558-2566.2004.

40. Evans JT, Denef VJ. 2020. To dereplicate or not to dereplicate? mSphere 5:e00971-19. https://doi.org/10.1128/mSphere.00971-19.

41. Bryant DA, Liu ZF. 2013. Green bacteria: insights into green bacterial evolution through genomic analyses, p 99–150. *In* Beatty JT (ed), Genome evolution of photosynthetic bacteria, vol 66. Elsevier, Philadelphia, PA.

42. Gregersen LH, Bryant DA, Frigaard NU. 2011. Mechanisms and evolution of oxidative sulfur metabolism in green sulfur bacteria. Front Microbiol 2. https://doi.org/10.3389/fmicb.2011.00116.

43. Kelly DP. 2008. Stable sulfur isotope fractionation by the green bacterium *Chlorobaculum parvum* during photolithoautotrophic growth on sulfide. Pol J Microbiol 57:275–279.

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 479:253–256. https://doi.org/10.1038/nature10504.

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. https://doi.org/10.1038/nature10505.

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 10:761–777. https://doi.org/10.1038/ismej.2015.153.

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 complete genome sequence. Photosynthesis Res 78:93–117. https://doi.org/10.1023/B:PRES.0000004310.96189.b4.

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. Proc Natl Acad Sci U S A 111:11479–11484. https://doi.org/10.1073/pnas.1407034111.

49. Silva PJ, van den Ban EC, Wassink H, Haaker H, de Castro B, Robb FT, Hagen WR. 2000. Enzymes of hydrogen metabolism in *Pyrococcus furiosus*. Eur J Biochem 267:6541–6551. https://doi.org/10.1046/j.1432-1327.2000.01745.x.

50. Zorin NA. 1986. Redox properties and active-center of phototrophic bacteria hydrogenases. Biochimie 68:97–101. https://doi.org/10.1016/S0300-9084(86)81074-4.

51. Greening C, Boyd E. 2020. Editorial: microbial hydrogen metabolism. Front Microbiol 11:56. https://doi.org/10.3389/fmicb.2020.00056.

52. Lane N, Allen JF, Martin W. 2010. How did LUCA make a living? Chemiosmosis in the origin of life. Bioessays 32:271–280. https://doi.org/10.1002/bies.200900131.

53. McAllister SM, Polson SW, Butterfield DA, Glazer BT, Sylvan JB, Chan CS. 2020. Validating the Cyc2 neutrophilic iron oxidation pathway using metaomics of *Zetaproteobacteria* iron mats at marine hydrothermal vents. mSystems 5:e00553-19. https://doi.org/10.1128/mSystems.00553-19.

54. Chen Y-J, Leung PM, Bay SK, Hugenholtz P, Kessler AJ, Shelley G, Waite DW, Cook PLM, Greening C. 2020. Metabolic flexibility allows generalist bacteria to become dominant in a frequently disturbed ecosystem. bioRxiv https://doi.org/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. Limnol Oceanogr 52:1208–1219. https://doi.org/10.4319/lo.2007.52.3.1208.

56. Roberts BJ, Howarth RW. 2006. Nutrient and light availability regulate the relative contribution of autotrophs and heterotrophs to respiration in freshwater pelagic ecosystems. Limnol Oceanogr 51:288–298. https://doi.org/10.4319/lo.2006.51.1.0288.

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:e00916-18. https://doi.org/10.1128/mBio.00916-18.

58. Bolger A, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

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. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.

60. Bushnell B. 2016. BBMap short read aligner. University of California, Berkeley, CA. 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:e1165. https://doi.org/10.7717/peerj.1165.

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, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Murat Eren A, 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. https://doi.org/10.1038/nbt.3893.

63. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

64. 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 25:1043–1055. https://doi.org/10.1101/gr.186072.114.

65. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.

66. Brown CT, Irber L. 2016. sourmash: a library for MinHash sketching of DNA. J Open Source Software https://doi.org/10.21105/joss.00027.

67. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W256–W259. https://doi.org/10.1093/nar/gkz239.

68. Rajaram S, Oono Y. 2010. NeatMap—non-clustering heat map alternatives in R. BMC Bioinformatics 11:45. https://doi.org/10.1186/1471-2105-11-45.

69. R Core Team. 2020. R: a language and environment for statistical computing. http://www.R-project.org/.

70. McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217. https://doi.org/10.1371/journal.pone.0061217.

71. Ginestet C. 2011. ggplot2: elegant graphics for data analysis. J R Stat Soc Ser A Stat Soc 174:245–246. https://doi.org/10.1111/j.1467-985X.2010.00676_9.x.

72. Huerta-Cepas J, Szklarczyk D, Heller D, Hernandez-Plaza A, Forslund SK, Cook H, Mende DR, 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. https://doi.org/10.1093/nar/gky1085.

73. Kanehisa M, Sato Y. 2020. KEGG Mapper for inferring cellular functions from protein sequences. Protein Sci 29:28–35. https://doi.org/10.1002/pro.3711.

74. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.

75. Sondergaard D, Pedersen CNS, Greening C. 2016. HydDB: a web tool for hydrogenase classification and analysis. Sci Rep 6:34212. https://doi.org/10.1038/srep34212.

76. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol 33:1870–1874. https://doi.org/10.1093/molbev/msw054.