Type I photosynthetic reaction center in an anoxygenic phototrophic member of the *Chloroflexota*

Tsuji JM[1†]*, Shaw NA[1], Nagashima S[2‡], Venkiteswaran JJ[1,3], Schiff SL[1], Hanada S[2§], Tank M[2,4], Neufeld JD[1]*

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

[2]Tokyo Metropolitan University, 1-1 Minami-osawa, Hachioji, Tokyo, Japan, 192-0397

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

[4]Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures GmbH, Inhoffenstrasse 7B, 38124 Braunschweig, Germany

[†]Current address: Institute of Low Temperature Science, Hokkaido University, Kita-19, Nishi-8, Kita-ku, Sapporo, Japan, 060-0819

[‡]Current address: Kanagawa University, 2946 Tsuchiya, Hiratsuka, Kanagawa, Japan, 259-1293

[§]Current address: Bioproduction Research Institute, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki, Japan, 305-8566

*Corresponding authors: jackson.tsuji@uwaterloo.ca; jneufeld@uwaterloo.ca

Keywords: *Chloroflexota*; *Chloroflexi*; filamentous anoxygenic phototroph; boreal lakes; anoxygenic photoautotrophy; evolution of photosynthesis; enrichment cultivation; metatranscriptomics

Summary

Phototrophic members of the *Chloroflexota* phylum are enigmas in the evolution of phototrophy. Although all previously characterized phototrophic *Chloroflexota* members use a Type II reaction center (RCII) to perform light energy conversion[1,2], many members host chlorosomes, a light-harvesting apparatus only otherwise reported among Type I reaction center (RCI)-containing phototrophs[3]. Here we report the discovery and cultivation of "*Candidatus* Chlorohelix allophototropha", the first known member of the *Chloroflexota* to use RCI, rather than RCII, for phototrophy. Cultivated from an iron-rich Boreal Shield lake, "*Ca.* Chx. allophototropha" cells contain chlorosomes and encode a novel fifth clade of RCI protein. Phylogenomic data demonstrate that "*Ca.* Chx. allophototropha" shares closest ancestry to RCII-utilizing relatives for multiple photosynthesis accessory genes, including chlorosomes. Thus, we propose that the enigmatic distribution of chlorosomes in the *Chloroflexota* phylum resulted from genetic interaction of RCI- and RCII-utilizing *Chloroflexota* members. Although undetected previously, metatranscriptome data demonstrate that RCI-utilizing *Chloroflexota* members can be among the most active populations in illuminated anoxic waters of Boreal Shield lakes, which number in the millions globally[4]. Our results establish the *Chloroflexota* as the only bacterial phylum containing RCI- and RCII-utilizing anoxygenic phototrophs, providing new context for understanding the origins of phototrophic life on Earth.

Main

Chlorophyll-based phototrophy drives the majority of global carbon fixation[5] and relies on a Type I (RCI) and/or Type II (RCII) photosynthetic reaction center for light energy conversion[2]. Although oxygenic phototrophy is the dominant chlorophototrophic process today[6], most phototrophic genetic diversity is distributed among anoxygenic phototrophs affiliated with seven bacterial phyla[1]. Unlike oxygenic phototrophs, which use RCII and RCI in tandem as Photosystem II (PSII) and Photosystem I (PSI), respectively, anoxygenic phototrophs consistently use either RCI or RCII for phototrophy[2]. The photosynthetic reaction centers of both oxygenic and anoxygenic phototrophs have been extensively characterized to understand their physiological and evolutionary properties (reviewed in [2,7,8]). Although it appears that RCI and RCII diverged early in the evolution of life[9], it remains unclear how they became incorporated in a single organism, in the case of oxygenic phototrophs, yet appear to be well separated otherwise within the tree of life[10,11]. Outside of members of the *Cyanobacteria* phylum (and, by extension, chloroplasts in eukaryotes), RCI and RCII-utilizing phototrophs appear distantly related, belonging to separate bacterial phyla and generally using divergent photosynthesis accessory genes[1]. Although some have speculated that genetic interactions of RCI and RCII ought to have occurred among anoxygenic phototrophs over evolutionary time[11,12], no direct published evidence has existed to support this hypothesis.

We sampled the anoxic water column of Lake 227, an iron-rich Boreal Shield lake[4], and cultivated the first known member of the *Chloroflexota* (formerly *Chloroflexi*) phylum to use RCI for phototrophy (Fig. 1; Supplementary Note 1). Given that only RCII-utilizing *Chloroflexota* members were known previously[1,13,14], this

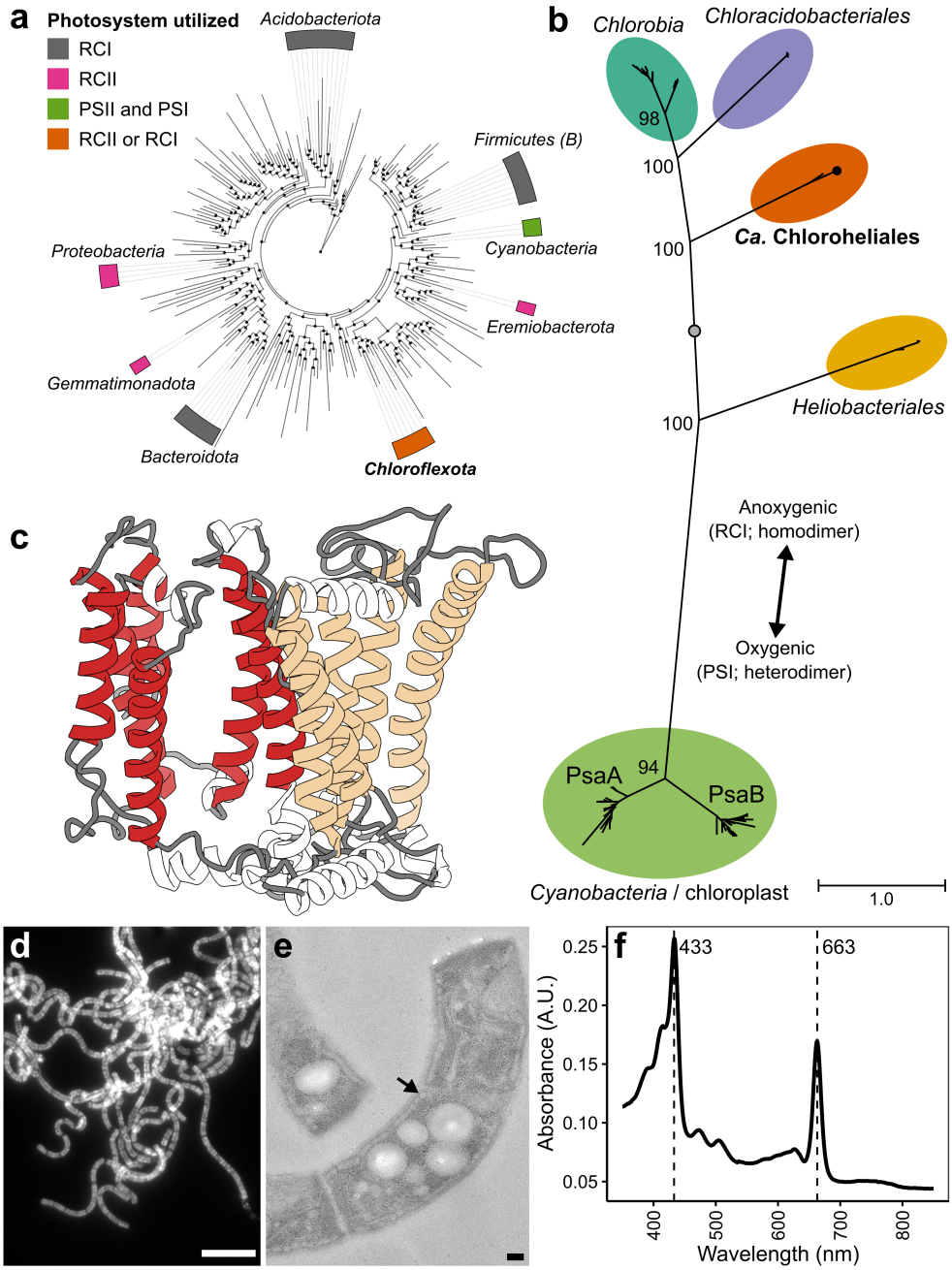


Fig. 1 | Type I reaction center-based phototrophy in "*Ca.* Chlorohelix allophototropha". **a**, Overview of the eight bacterial phyla containing known chlorophototrophs. A class-level summary of the Genome Taxonomy Database (GTDB) bacterial reference tree (release 89) is shown, with chlorophototrophic phyla highlighted based on the photosynthetic reaction center utilized. **b**, Maximum likelihood phylogeny of Type I reaction center primary sequences. Ultrafast bootstrap values for major branch points are shown out of 100 based on 1000 replicates, and the scale bar shows the expected proportion of amino acid change across the 548-residue masked sequence alignment (Extended Data Table 2). Chlorophototrophic lineages are summarized by order name. The placement of the PscA-like sequence of "*Ca*. Chx. allophototropha" is indicated by a black dot. All clades above they grey dot utilize chlorosomes. **c**, Predicted tertiary structure of the novel "*Ca*. Chlorohelix allophototropha" PscA-like primary sequence based on homology modelling. The six N-terminal and five C-terminal transmembrane helices expected for RCI are coloured in red and tan, respectively. **d,** Epifluorescence microscopy image of "*Ca.* Chx. allophototropha" cells. Cell autofluorescence is shown with light excitation and detection wavelengths optimized for bacteriochlorophyll *c*. **e**, Transmission electron microscopy image showing a longitudinal section of "*Ca.* Chx. allophototropha" cells. An example chlorosome is marked with an arrow. **f**, Absorption spectrum (350-850 nm) of an acetone:methanol extract of pigments from the "*Ca.* Chx. allophototropha" culture. Major absorption peaks are marked by dashed lines. Scale bars in panels **d** and **e** represent 10 µm and 0.1 µm, respectively.

novel phototroph breaks established paradigms that each bacterial phylum of anoxygenic phototrophs is associated with a single type of photosynthetic reaction center (Fig. 1a)[1,2,7]. We did not detect the *pufLM* genes
75 associated with RCII in the highly complete genome of the organism or in the enrichment culture metagenome from which that genome was derived (Supplementary Note 2; Extended Data Fig. 1). Instead, we detected a remote homolog of the *pscA* gene, encoding the RCI protein, on a 3.29 Mb scaffold that was flanked by other genes classified to the *Chloroflexota* phylum (Supplementary Data 1). Based on a maximum-likelihood amino acid sequence phylogeny (Fig. 1b; Extended Data Fig. 2), the PscA-like predicted protein of the novel
80 phototroph represents a distinct fifth clade of RCI/PSI protein and thus was not acquired by recent lateral gene transfer from known phototrophic clades[15]. Homology modelling indicated that the novel phototroph's PscA-like protein contained the expected six N-terminal transmembrane helices for coordinating antenna pigments and five C-terminal transmembrane helices involved in core electron transfer[16] (Fig. 1c), supporting its functional role. We did not detect a cyanobacterial *psaB*-like paralog of the RCI gene, implying that the novel phototroph uses a
85 homodimeric RCI complex like other known anoxygenic phototrophs[2].

In line with a phototrophic growth mode, cells of the novel strain grew as pigmented filaments (Fig. 1d) and contained light-harvesting chlorosomes (Fig. 1e; Extended Data Fig. 3), with the dominant pigment being bacteriochlorophyll *c* based on absorption spectrum data (Fig. 1f; see Supplementary Note 3)[17]. We also enriched a second phototrophic *Chloroflexota* member whose genome encoded a *pscA*-like gene belonging to the same
90 fifth RCI/PSI clade (Fig. 1b), supporting that the novel RCI gene is associated with *Chloroflexota* members (Supplementary Notes 1-2). Combined with metatranscriptome data (see below), our findings provide robust evidence that the novel *Chloroflexota* member relies on the detected RCI to perform phototrophy. Thus, we provisionally classify the novel phototroph as, "*Candidatus* Chlorohelix allophototropha" strain L227-S17 (Chlo.ro.he'lix. Gr. adj. *chloros*, green; Gr. fem. n. *helix*, spiral; N.L. fem. n. *Chlorohelix*, a green spiral.
95 al.lo.pho.to.tro'pha. Gr. adj. *allos*, other, different; Gr. n. *phos*, *-otos*, light; N.L. suff. *-tropha*, feeding; N.L. fem. adj. *allophototropha*, phototrophic in a different way), the first cultivated member of the provisional order "*Candidatus* Chloroheliales" within the *Chloroflexia* class (Supplementary Note 4).

Phylogenomic analyses of the "*Ca.* Chx. allophototropha" genome revealed several phototrophy-associated genes that were highly divergent from those of known phototrophs (Fig. 2; Extended Data Table 1;
100 Supplementary Data 2). We detected a homolog of *fmoA*, encoding the Fenna-Matthews-Olson (FMO) protein involved in energy transfer from chlorosomes to RCI[18], with 21-26% predicted amino acid identity to known FmoA sequences (Extended Data Fig. 4). This finding makes the *Chloroflexota* the third known phylum to use the FMO protein in phototrophy[19]. Intriguingly, the FmoA predicted primary sequence lacked the Cys49 and Cys353 residues thought to be crucial for quenching energy transfer[20]. Matching physiological observations, we
105 detected a homolog of the key chlorosome structural gene *csmA*[12,21] that had ~33% identity at the predicted amino acid level to the CsmA primary sequence of *Chloroflexus aurantiacus* J-10-fl (Fig. 2), along with homologs of *csmM* and *csmY*[12]. The potential CsmA homolog notably lacked the His25 thought to be required for bacteriochlorophyll *a* binding[22] (Extended Data Fig. 5). In place of Alternative Complex III used by

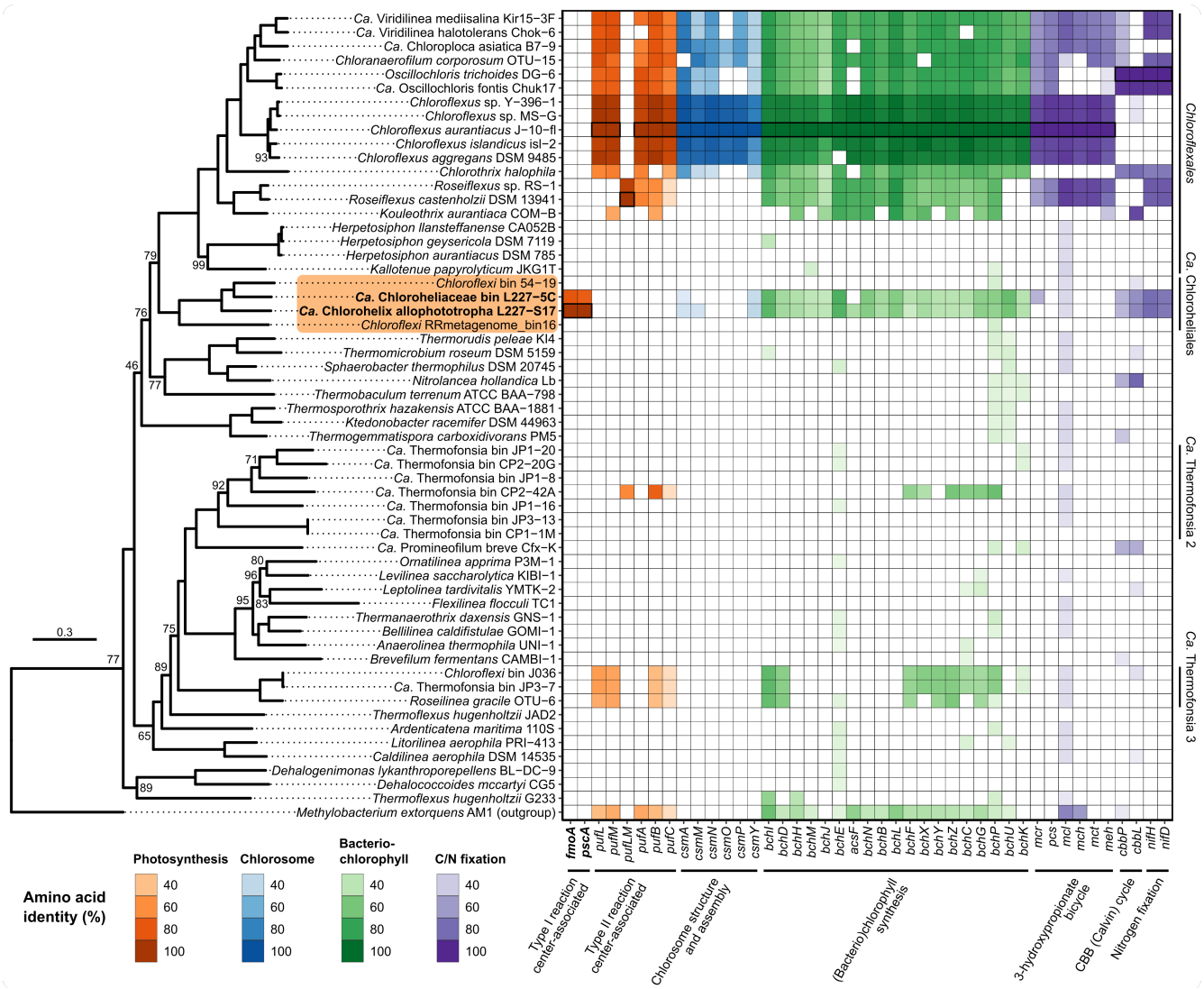


Fig. 2 | Genomic potential for phototrophy among members of the *Chloroflexota* phylum. A maximum likelihood phylogeny of key members of the phylum is shown on the left based on a set of 74 concatenated core bacterial proteins. Ultrafast bootstrap values are shown out of 100 based on 1000 replicates, and the scale bar shows the expected proportion of amino acid change across the 11 613-residue masked sequence alignment (Extended Data Table 2). The "*Ca.* Chloroheliales" clade is highlighted in orange, and genomes recovered in this study are in bold. On the right, a heatmap shows the presence/absence of genes involved in photosynthesis or related processes based on bidirectional BLASTP. The intensity of the fill colour of each tile corresponds to the percent identity of the BLASTP hit. Query sequences were derived from the genomes of "*Ca.* Chlorohelix allophototropha", *Chloroflexus aurantiacus*, *Roseiflexus castenholzii*, or *Oscillochloris trichoides*, and the query organism used for each gene is marked by bolded outlines on the heatmap. GenBank accession numbers of sequences associated with the novel "*Ca.* Chloroheliales" members enriched in this work are summarized in Extended Data Table 1, and raw results of bidirectional BLASTP are summarized in Supplementary Data 2. Orders containing potential phototrophs are labelled on the right of the heatmap. All genomes in the phylogeny encoded at least 50% of the core 74 protein set except for *Chloroflexi* RRmetagenome_bin16 and "*Ca*. Thermofonsia bin CP2-42A", which encoded 25 and 34 of the proteins, respectively.

Chloroflexales members for photosynthetic electron transport[23], we found a cytochrome b_6f-like gene cluster including a putative di-heme cytochrome *c* as described for *Heliobacterium modesticaldum*[24] (Supplementary Note 2). Moreover, we detected genes for the Calvin-Benson-Bassham (CBB) cycle for carbon fixation in the "*Ca.* Chx. allophototropha" genome, including a deep-branching Class IC/ID "red type" *cbbL* gene[25]
130 representing the large subunit of RuBisCO (Extended Data Fig. 6) and the *cbbP* gene encoding phosphoribulokinase[26] (Fig. 2). We did not detect genomic potential for the 3-hydroxypropionate bicycle, which is used for carbon fixation by some RCII-utilizing *Chloroflexales* members[27]. "*Ca.* Chx. allophototropha" thus represents the first known RCI-utilizing phototroph with potential for carbon fixation via the CBB cycle, with metatranscriptome data supporting that RuBisCO is actively used for photosynthesis.

135 Although previously overlooked, environmental metagenome and metatranscriptome data demonstrate that relatives of "*Ca.* Chx. allophototropha" are important phototrophic community members in iron-rich Boreal Shield lakes (Fig. 3). Surveying the water columns of eight seasonally anoxic Boreal Shield lakes at the International Institute for Sustainable Development Experimental Lakes Area (IISD-ELA) using metagenome sequencing, we recovered 756 dereplicated metagenome-assembled genomes (MAGs) having an average
140 completeness and contamination of 82.7% and 2.1%, respectively. From these, we identified two RCI-encoding MAGs affiliated with the "*Ca.* Chloroheliales" order (Supplementary Note 5). Short-read mapping of metagenome sequences to these MAGs demonstrated their widespread distribution within the illuminated anoxic zones of four lakes (i.e., Lakes 221, 222, 227, and 304), with relative abundances of up to 1.32% (Supplementary Data 3), although the relative abundance and activity of these MAGs were low in Lake 227 (Supplementary Note
145 5). Importantly, metatranscriptome data available for the anoxic water columns of Lakes 221 and 304 showed that the "*Ca.* Chx. allophototropha" relatives were highly photosynthetically active. In Lake 221, the two MAGs recruited a cumulative 3.3% of mapped RNA reads, and in Lake 304, "*Ca.* Chloroheliales" bin ELA729 was among the top ten highest RNA read-recruiting MAGs in the dataset (Fig. 3a). Both MAGs were found to have high expression of the *pscA*-like RCI gene, the *fmoA* gene, and the *cbbLS* genes encoding RuBisCO, with "*Ca.*
150 Chlorohelix" bin ELA319 (Lake 221) having all of these phototrophy-associated genes within the top 30 most highly expressed genes in its genome (Fig. 3b; Extended Data Fig. 7). Moreover, the two MAGs co-occurred with highly active RCII-encoding MAGs associated with the *Chloroflexales* order (Fig. 3a), which suggests that RCI- and RCII-utilizing *Chloroflexota* members can occupy the same or closely related niche space. Given that Boreal Shield lakes number in the tens of millions globally, and that an estimated 15% develop seasonally anoxic
155 conditions[4], relatives of "*Ca*. Chx. allophototropha" could be responsible for substantial and previously unreported contributions to the freshwater biogeochemistry.

Based on phylogenomic data, the proposed "*Ca.* Chloroheliales" order containing "*Ca.* Chx. allophototropha" places directly sister to the *Chloroflexales* order containing the canonical RCII-utilizing phototrophic clade in the phylum (Fig. 2). The phylogenetic proximity of these RCI- and RCII-utilizing groups is
160 unprecedented and contrasts with the observations of current literature[1,2,7]. Overlapping environmental niches between RCII-utilizing *Chloroflexota* members, which are known to inhabit mesophilic aquatic

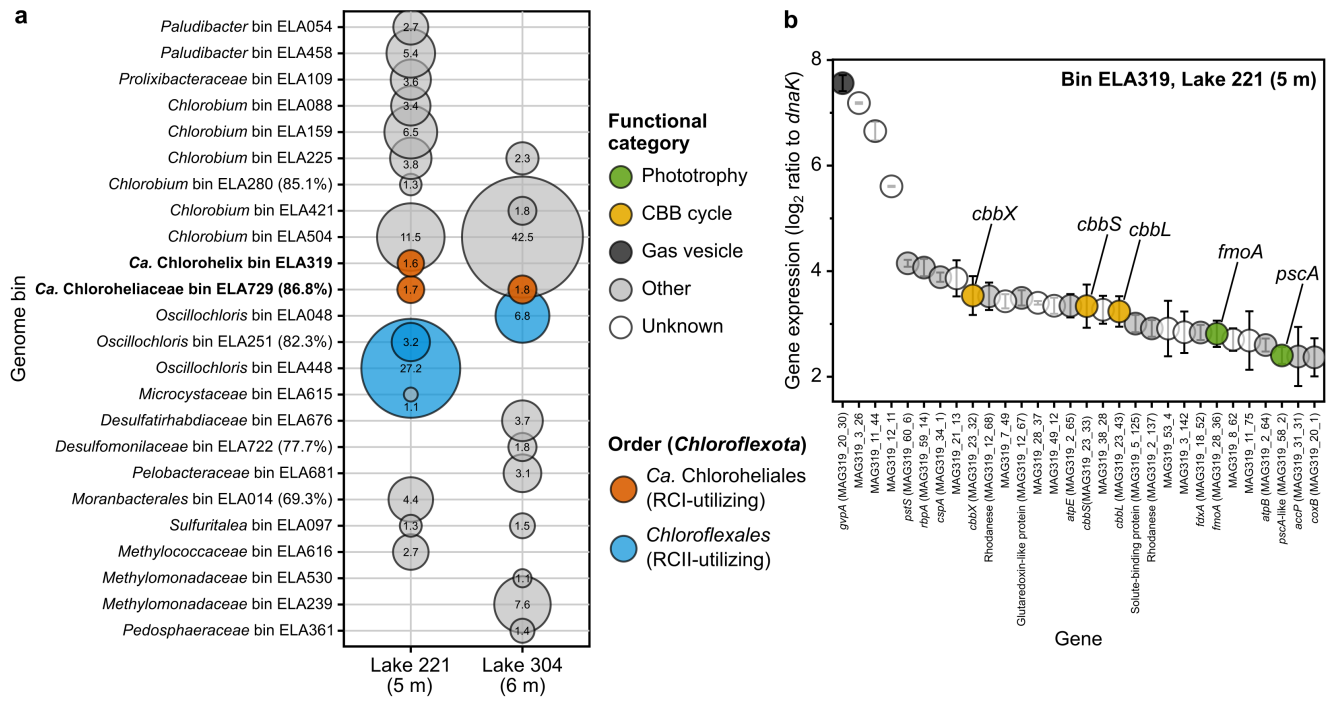


Fig. 3 | Environmental activity of "*Ca.* Chx. allophototropha" relatives in Boreal Shield lakes. a, Bubble plot showing active microbial populations in the water columns of two Boreal Shield lakes at the IISD-ELA based on read mapping of metatranscriptome data (from July 2018) to metagenome-assembled genomes (MAGs). The size of each bubble represents the mean "relative expression" (n=3) of each MAG, as described in the methods, and MAGs having >1% relative expression are shown. (See Supplementary Data 4 for standard deviations, which were on average 7.0% of the mean.) All MAGs shown had <5% estimated contamination, and for any MAGs with <90% estimated completion, the completion metric is shown beside the MAG ID. **b,** Normalized gene expression of "*Ca.* Chlorohelix bin ELA319" based on Lake 221 metatranscriptome data. The top 30 protein-coding genes with highest normalized expression values are shown. Error bars represent the standard deviation of the $\log_2$ expression ratio based on triplicate metatranscriptomes (n=3). Genes potentially involved in phototrophy, carbon fixation, and buoyancy are highlighted. Highly expressed genes based on Lake 304 metatranscriptome data and read mapping to bin ELA729 are shown in Extended Data Fig. 7, and normalized expression values for all genes, along with their sequences and annotations, are included in Supplementary Data 5.

environments[28,29], and RCI-utilizing "*Ca.* Chx. allophototropha" relatives could readily facilitate the genetic interaction of these chlorophototroph clades. We probed the phylogenetic history of several photosynthesis-associated genes of "*Ca.* Chx. allophototropha", in the context of known phototrophs, to explore whether such genetic interaction might have taken place (Fig. 4). In a maximum likelihood phylogeny of the chlorosome structural protein CsmA, "*Ca.* Chloroheliales" and RCII-utilizing *Chloroflexota* sequences grouped directly sister to one another (Fig. 4a; Extended Data Fig. 8). Similarly, using maximum likelihood phylogenies of the (bacterio)chlorophyll synthesis proteins BchIDH/ChlIDH (Fig. 4b; Extended Data Fig. 9) and BchLNB/ChlLNB/BchXYZ (Extended Data Fig. 10), we observed that "*Ca.* Chloroheliales" sequences placed immediately adjacent to the sister grouping of RCII-utilizing *Chloroflexota* and RCI-utilizing *Chlorobia* members. Although the long branch lengths in these phylogenies complicate inferring ancestral relationships, the consistent close placement of RCI- and RCII-utilizing *Chloroflexota* clades in these phylogenies provides first evidence that the two clades have a shared genetic history of phototrophy despite using different core reaction centers. Indeed, the presence of chlorosomes within phototrophic *Chloroflexota* members has been enigmatic

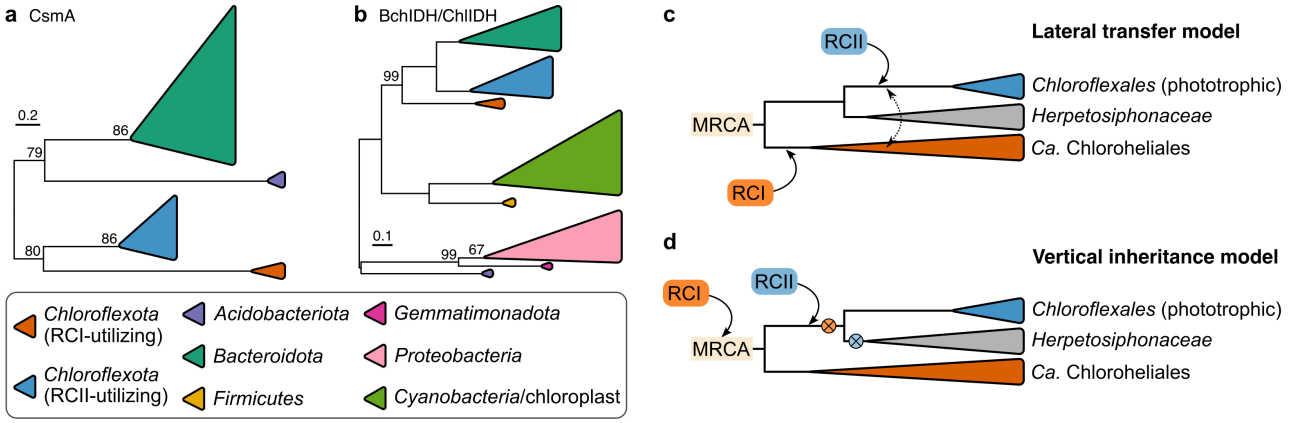


Fig. 4 | Evolution of phototrophy within the *Chloroflexota* phylum. a-b, Maximum likelihood phylogenies of the chlorosome structural protein CsmA (**a**) and (bacterio)chlorophyll synthesis proteins BchIDH/ChlIDH (**b**). The phylogenies are midpoint rooted, and ultrafast bootstrap values less than 100/100 are shown. Scale bars represent the expected proportion of amino acid change across the masked sequence alignments (see Extended Data Table 2). Detailed versions of the same phylogenies are shown in Extended Data Figs. 8-9. **c-d,** Proposed models of phototrophy evolution within the phylum. In the lateral transfer model (**c**), ancestors of the "*Ca.* Chloroheliales" and *Chloroflexales* orders independently obtained RCI and RCII, respectively. Subsequent lateral gene transfer between members of the two orders allowed for members of both orders to encode related photosynthesis accessory genes (e.g., for chlorosome structure). In the vertical inheritance model (**d**), the most recent common ancestor (MRCA) of the "*Ca.* Chloroheliales" and *Chloroflexales* orders encoded either RCI or RCII and several photosynthesis accessory genes (e.g., for chlorosome structure). Genes for the alternate reaction center were subsequently acquired by descendants of that ancestor, followed by loss of the original reaction center. Phototrophy genes were then lost in ancestors of certain clades, such as the non-phototrophic *Herpetosiphonaceae* family within the *Chloroflexales* order. The schematic depicts the MRCA having RCI, followed by acquisition of RCII, but alternative orders of photosynthesis gene gain/loss, or multiple gene gain/loss events, are also possible in the vertical inheritance model.

because these light-harvesting organelles are only otherwise found in RCI-utilizing phototrophs[3,12,21]. The discovery of "*Ca.* Chx. allophototropha" resolves this enigma if chlorosomes in RCII-utilizing *Chloroflexota* members were sourced from ancestors of the RCI-utilizing "*Ca.* Chx. allophototropha", or vice versa.

Our phylogenomic data leave open the possibility of two broad modes of genetic interaction between RCI- and RCII-utilizing *Chloroflexota* members. First, these two phototrophic *Chloroflexota* groups could have obtained their reaction centers independently and subsequently exchanged photosynthesis accessory genes, such as chlorosome-associated genes, via lateral gene transfer (Fig. 4c). Lateral gene transfer has been previously observed to play a role in the evolution of RCII-driven phototrophy in the phylum[30]. Given that chlorosomes are associated only with RCI-utilizing phototrophs outside the *Chloroflexota*, it would follow that chlorosomes were laterally transferred from RCI- to RCII-utilizing *Chloroflexota* members. Alternatively, RCI- and RCII-utilizing *Chloroflexota* members could share a common phototrophic ancestor and could have inherited chlorosomes and bacteriochlorophyll synthesis genes vertically from that ancestor (Fig. 4d). Consistent basal placement of "*Ca.* Chloroheliales"-associated photosynthesis accessory gene sequences compared to those of RCII-utilizing *Chloroflexota* members implies that the common ancestor contained RCI (Fig. 4a-b), although the presence of a *pscA-fmoA* gene cluster in the genome of one of our enriched "*Ca.* Chloroheliales" members could also imply more recent acquisition of RCI in this clade (Supplementary Note 2). The vertical inheritance scenario leaves open the intriguing possibility that a common ancestor of RCI- and RCII-utilizing *Chloroflexota* members used

RCI and RCII simultaneously, in a manner analogous to *Cyanobacteria* members, followed by a "genetic displacement" event of the other reaction center, as has been speculated previously[12]. Either the lateral gene transfer or vertical inheritance scenario potentially clarifies the origin of chlorosomes among RCII-utilizing *Chloroflexota* members and demonstrates the possibility for close genetic interaction of RCI and RCII within the *Chloroflexota*.

By cultivating the first-known RCI-utilizing phototroph within the *Chloroflexota* phylum, we demonstrate that RCI and RCII can be used by closely related anoxygenic phototrophs, in contrast to established paradigms. Discovering "*Ca.* Chx. allophototropha" demonstrates the possibility for massive photosynthesis gene movements over evolutionary time, which has been required by many evolutionary models previously but has been sparsely demonstrated in nature[8,15]. Cultivation of "*Ca.* Chx. allophototropha" substantially expands the known diversity of chlorophototrophy, with implications for the biochemistry, evolution, and ecology of photosynthesis.

References

1. Thiel, V., Tank, M. & Bryant, D. A. Diversity of chlorophototrophic bacteria revealed in the omics era. *Annu Rev Plant Biol* **69**, 21–49 (2018).
2. Fischer, W. W., Hemp, J. & Johnson, J. E. Evolution of oxygenic photosynthesis. *Annu Rev Earth Pl Sc* **44**, 647–683 (2016).
3. Hohmann-Marriott, M. F. & Blankenship, R. E. Hypothesis on chlorosome biogenesis in green photosynthetic bacteria. *FEBS Lett* **581**, 800–803 (2007).
4. Schiff, S. L. *et al.* Millions of Boreal Shield lakes can be used to probe Archaean Ocean biogeochemistry. *Sci Rep* **7**, 46708 (2017).
5. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
6. Raven, J. A. Contributions of anoxygenic and oxygenic phototrophy and chemolithotrophy to carbon and oxygen fluxes in aquatic environments. *Aquat Microb Ecol* **56**, 177–192 (2009).
7. Hohmann-Marriott, M. F. & Blankenship, R. E. Evolution of photosynthesis. *Annu Rev Plant Biol* **62**, 515–548 (2011).
8. Martin, W. F., Bryant, D. A. & Beatty, J. T. A physiological perspective on the origin and evolution of photosynthesis. *FEMS Microbiol Rev* **42**, 205–231 (2018).
9. Sadekar, S., Raymond, J. & Blankenship, R. E. Conservation of distantly related membrane proteins: photosynthetic reaction centers share a common structural core. *Mol Biol Evol* **23**, 2001–2007 (2006).
10. Olson, J. M. & Blankenship, R. E. Thinking about the evolution of photosynthesis. in *Discoveries in Photosynthesis* (eds. Govindjee, Beatty, J. T., Gest, H. & Allen, J. F.) 1073–1086 (Springer Netherlands, 2005). doi:10.1007/1-4020-3324-9_95.
11. Cardona, T. Thinking twice about the evolution of photosynthesis. *Open Biol* **9**, 180246 (2019).
12. Bryant, D. A. *et al.* Comparative and functional genomics of anoxygenic green bacteria from the taxa *Chlorobi*, *Chloroflexi*, and *Acidobacteria*. in *Functional Genomics and Evolution of Photosynthetic Systems* (eds. Burnap, R. & Vermaas, W.) 47–102 (Springer Netherlands, 2012). doi:10.1007/978-94-007-1533-2_3.
13. Pierson, B. K. & Castenholz, R. W. A phototrophic gliding filamentous bacterium of hot springs, *Chloroflexus aurantiacus*, gen. and sp. nov. *Arch Microbiol* **100**, 5–24 (1974).

14. Hanada, S. The phylum *Chloroflexi*, the family *Chloroflexaceae*, and the related phototrophic families *Oscillochloridaceae* and *Roseiflexaceae*. in *The Prokaryotes: Other Major Lineages of Bacteria and The Archaea* (eds. Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 515–532 (Springer Berlin Heidelberg, 2014). doi:10.1007/978-3-642-38954-2_165.
15. Zeng, Y., Feng, F., Medová, H., Dean, J. & Koblížek, M. Functional type 2 photosynthetic reaction centers found in the rare bacterial phylum *Gemmatimonadetes*. *Proc Natl Acad Sci USA* **111**, 7795–7800 (2014).
16. Gisriel, C. *et al.* Structure of a symmetric photosynthetic reaction center–photosystem. *Science* **357**, 1021–1025 (2017).
17. Overmann, J. & Garcia-Pichel, F. The phototrophic way of life. in *The Prokaryotes: Prokaryotic Communities and Ecophysiology* (eds. Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 203–257 (Springer Berlin Heidelberg, 2013). doi:10.1007/978-3-642-30123-0_51.
18. Olson, J. M. The FMO protein. in *Discoveries in Photosynthesis* (eds. Govindjee, Beatty, J. T., Gest, H. & Allen, J. F.) 421–427 (Springer Netherlands, 2005). doi:10.1007/1-4020-3324-9_40.
19. Bryant, D. A. *et al. Candidatus* Chloracidobacterium thermophilum: an aerobic phototrophic Acidobacterium. *Science* **317**, 523–526 (2007).
20. Orf, G. S. *et al.* Evidence for a cysteine-mediated mechanism of excitation energy regulation in a photosynthetic antenna complex. *Proc Natl Acad Sci USA* **113**, E4486–E4493 (2016).
21. Frigaard, N.-U. & Bryant, D. A. Seeing green bacteria in a new light: genomics-enabled studies of the photosynthetic apparatus in green sulfur bacteria and filamentous anoxygenic phototrophic bacteria. *Arch Microbiol* **182**, 265–276 (2004).
22. Pedersen, M. Ø., Linnanto, J., Frigaard, N.-U., Nielsen, N. Chr. & Miller, M. A model of the protein–pigment baseplate complex in chlorosomes of photosynthetic green bacteria. *Photosynth Res* **104**, 233–243 (2010).
23. Tang, K.-H. *et al.* Complete genome sequence of the filamentous anoxygenic phototrophic bacterium *Chloroflexus aurantiacus*. *BMC Genomics* **12**, 334 (2011).
24. Ducluzeau, A. L., Chenu, E., Capowiez, L. & Baymann, F. The Rieske/cytochrome *b* complex of Heliobacteria. *Biochim Biophys Acta, Bioenerg* **1777**, 1140–1146 (2008).
25. Tourova, T. P. *et al.* Phylogeny of anoxygenic filamentous phototrophic bacteria of the family *Oscillochloridaceae* as inferred from comparative analyses of the *rrs*, *cbbL*, and *nifH* genes. *Microbiology* **75**, 192–200 (2006).
26. Tabita, F. R. The biochemistry and molecular regulation of carbon dioxide metabolism in cyanobacteria. in *The Molecular Biology of Cyanobacteria* (ed. Bryant, D. A.) 437–467 (Springer, 1994). doi:10.1007/978-94-011-0227-8_14.
27. Shih, P. M., Ward, L. M. & Fischer, W. W. Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the *Chloroflexi*. *Proc Natl Acad Sci USA* **114**, 10749–10754 (2017).
28. Gorlenko, V. M. *et al. Candidatus* 'Chloroploca asiatica' gen. nov., sp. nov., a new mesophilic filamentous anoxygenic phototrophic bacterium. *Microbiology* **83**, 838–848 (2014).
29. Gaisin, V. A. *et al.* '*Candidatus* Viridilinea mediisalina', a novel phototrophic Chloroflexi bacterium from a Siberian soda lake. *FEMS Microbiol Lett* **366**, fnz043 (2019).
30. Ward, L. M., Hemp, J., Shih, P. M., McGlynn, S. E. & Fischer, W. W. Evolution of phototrophy in the *Chloroflexi* phylum driven by horizontal gene transfer. *Front Microbiol* **9**, 260 (2018).

Methods

Enrichment cultivation of "Ca. Chx. allophototropha"

To culture novel anoxygenic phototrophs, we sampled Lake 227 (49.69° N, 93.69° W), a seasonally anoxic and ferruginous Boreal Shield lake at the International Institute for Sustainable Development Experimental Lakes Area (IISD-ELA). The IISD-ELA sampling site (49.50-49.75° N, 93.50-94.00° W), located near Kenora, Canada, has been described in detail previously[4,31–34]. Lake 227 develops up to ~100 μM concentrations of dissolved ferrous iron in its anoxic water column[4], and anoxia is more pronounced than expected naturally due to the long-term experimental eutrophication of the lake[35]. We collected water from the illuminated portion of the upper anoxic zone of Lake 227 in September 2017, at 3.88 and 5.00 m depth, and transported this water to the laboratory under anoxic and chilled conditions in 120 mL glass serum bottles sealed with black rubber stoppers (Geo-Microbial Technology Company; Ochelata, Oklahoma, USA).

Water was supplemented with 2% v/v of a freshwater medium containing 8 mM ferrous chloride[36] and was distributed under anoxic conditions into 120 mL glass serum bottles, also sealed with black rubber stoppers (Geo-Microbial Technology Company), that had a headspace of dinitrogen gas at 1.5 atm final pressure. Bottles were spiked with a final concentration of 50 μM Diuron or 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU); Sigma-Aldrich; St. Louis, Missouri, USA) to block oxygenic phototrophic activity[37] and to eliminate oxygenic phototrophs. Spiking was performed either at the start of the experiment or as needed based on observations of oxygenic phototroph growth. Bottles were incubated at 10°C under white light (20-30 μmol photons m^{-2} s^{-1} at 400-700 nm wavelengths; blend of incandescent and fluorescent sources) or at 22°C under far red LED light (using a PARSource PowerPAR LED bulb; LED Grow Lights Depot; Portland, Oregon, U.S.A.). Cultures were monitored regularly for ferrous iron concentration using the ferrozine assay[38] and were amended with additional freshwater medium or ferrous chloride (e.g., in 0.1-1 mM increments) when ferrous iron levels dropped, presumably due to iron oxidation.

The "*Ca.* Chx. allophototropha" enrichment culture was gradually acclimated to higher concentrations of freshwater medium through repeated feeding and subculture. Cultures were moved to room temperature (20-22°C) growth and could be grown under white light (fluorescent, halogen, or incandescent; same light intensity as above) or under far red LED light (using a PARSource PowerPAR LED bulb). A deep agar dilution series[39] was used to eliminate contaminants from the culture. The same freshwater medium was used for this dilution series, amended with an acetate feeding solution[40]. Ferrous chloride-containing agar plugs (10 mM concentration) were initially used at the bottom of tubes to form a ferrous iron concentration gradient to qualitatively determine best conditions for growth, with 0.125 g/L sodium thioglycolate being used as a reducing agent in the medium. Over multiple rounds of optimization, the growth medium was adjusted to contain 2 mM ferrous chloride (without sodium thioglycolate), 1.2 mM sodium acetate, and 50 mg/L resazurin, and to be a 1:1 dilution of the original freshwater medium, kept at a pH of ~7.5. In addition, cultures were generally incubated

345 in soft agar (0.2-0.3%; w/v)[29] to promote faster growth of "*Ca.* Chx. allophototropha" cells compared to growth in standard concentration agar (~0.8%; w/v) or in liquid medium.

Agar shake tubes were incubated under 740 nm LED lights for a period of time to eliminate purple phototrophic bacteria, because 740 nm light is not readily used for photosynthesis by these bacteria[17]. Additionally, a higher medium pH of ~7.5-8.5 was used for a time to select against green phototrophic bacteria
350 belonging to the *Chlorobia* class, because these bacteria are known to grow poorly in moderately basic conditions[41]. Lastly, sodium molybdate was added to the medium at a final concentration of 100 μM, in a 1:10 concentration ratio relative to sulfate, to inhibit the activity of sulfate-reducing bacteria as optimized experimentally (Supplementary Methods).

355 *Physiological characterization*

Filaments of "*Ca.* Chx. allophototropha" cells were identified based on pigment autofluorescence using epifluorescence microscopy. An Eclipse 600 light microscope equipped for fluorescence microscopy (Nikon; Shinagawa, Tokyo, Japan) was used with a xenon light source. Autofluorescence in the wavelength range expected for bacteriochlorphyll *c* was detected using an excitation filter of 445(±45) nm, dichroic mirror of 520
360 nm, and emission filter of 715 (LP) nm (all from Semrock; Rochester, New York, USA), before light collection using an ORCA-Flash4.0 monochromatic camera (Hamamatsu Photonics; Hamamatsu City, Shizuoka, Japan). Examination of cells under epifluorescence microscopy enabled the growth of cells to be qualitatively assessed. In addition, growth of "*Ca.* Chx. allophototropha" was confirmed based on relative increases in cell pigmentation using a custom-built pigment fluorescence detection system (Supplementary Methods).

365 To perform transmission electron microscopy (TEM), cell biomass was picked from agar shake tubes, and residual agar surrounding cells was digested using agarase. One unit of β-agarase I (New England Biolabs; Ipswich, Massachusetts, USA) and 10 uL of 10x reaction buffer was added to 100 μL of cell suspension and incubated at 42°C for 1.5 h. Following cell pelleting and removal of supernatant, cells were then fixed for 2 h at 4°C in 4%/4% glutaraldehyde/paraformaldehyde (dissolved in 1x phosphate-buffered saline; 1x PBS) and stored
370 at 4°C. Sample preparation and imaging was performed at the Molecular and Cellular Imaging Facility of the Advanced Analysis Center (University of Guelph, Ontario, Canada). Washed cell pellets were enrobed in 4% Noble agar (ThermoFisher Scientific; Waltham, Massachusetts, USA), and enrobed samples were cut into 1 mm cubes. The cubes were fixed in 1% (w/v) osmium tetroxide for 45 min and subjected to a 25-100% ethanol dehydration series. Samples were then subjected to stepwise infiltration using 25-100% LR White Resin (Ted
375 Pella; Redding, California, USA), transferred to gelatin capsules in 100% LR White Resin, and allowed to polymerize overnight at 60°C. An Ultracut UCT ultramicrotome (Leica Microsystems; Wetzlar, Germany) equipped with a diamond knife (DiATOME; Hatfield, Pennsylvania, USA) was used to cut 50 nm thin sections, which were then floated on 100-mesh copper grids and stained with 2% uranyl acetate and Reynold's lead citrate to enhance contrast. Sections were imaged under standard operating conditions using a Tecnai G2 F20 TEM

380 (ThermoFisher Scientific; Waltham, Massachusetts, USA) that was running at 200 kV and was equipped with a Gatan 4k CCD camera (Gatan; Pleasanton, California, USA). For scanning electron microscopy (SEM), fixed cells were washed in 1x PBS and incubated in 1% (w/v) osmium tetroxide at room temperature for 30 min. Following incubation, cells were washed, deposited on an aluminum stub, dried, and then sputter coated with a gold:palladium mixture using a Desk V TSC Sample Preparation System (Denton Vacuum; Moorestown, New
385 Jersey, USA). Prepared samples were imaged using a Quanta FEG 250 SEM (ThermoFisher Scientific) with a high-voltage setting of 10 kV and working distance of 9.9 mm.

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

16S rRNA gene amplicon sequencing and analysis

395 The microbial community composition of early enrichment cultures was assessed using 16S ribosomal RNA (16S rRNA) gene amplicon sequencing with subsequent analysis of the obtained sequences. Genomic DNA was extracted from pelleted cell biomass using the DNeasy UltraClean Microbial Kit (Qiagen; Venlo, The Netherlands) according to the manufacturer protocol, with the addition of a 10 min treatment at 70°C after adding Solution SL for improved cell lysis. The V4-V5 hypervariable region of the 16S rRNA gene was then
400 amplified from extracted DNA using the universal prokaryotic PCR primers 515F-Y[42] and 926R[43], using primers with attached index sequences and sequencing adaptors as described previously[44–46]. Amplification was performed in singlicate. Library pooling, cleanup, and sequencing on a MiSeq (Illumina; San Diego, California, USA) was performed as described previously[44]. Sequence data analysis was performed using QIIME2[47] version 2019.10 via the AXIOME3[48] pipeline, commit 1ec1ea6 (https://github.com/neufeld/axiome3), with default
405 analysis parameters. Briefly, paired-end reads were trimmed, merged, and denoised using DADA2[49] to generate an amplicon sequence variant (ASV) table. Taxonomic classification of ASVs was performed using QIIME2's naive Bayes classifier[50] trained against the Silva SSU database[51,52], release 132. The classifier training file was prepared using QIIME2 version 2019.7.

410 *Metagenome sequencing of enrichment culture DNA*

The functional gene content of early liquid enrichment cultures was assessed via shotgun metagenome sequencing. Genomic DNA was extracted as above, and library preparation and sequencing was performed at The Centre for Applied Genomics (TCAG; The Hospital for Sick Children, Toronto, Canada). The Nextera DNA Flex Library Prep Kit (Illumina) kit was used for metagenome library preparation, and libraries were sequenced

415 using a fraction of a lane of a HiSeq 2500 (Illumina) with 2x125 base paired-end reads to obtain 5.0-7.3 million total reads per sample.

High molecular weight DNA was later extracted from the "*Ca.* Chx. allophototropha" enrichment culture grown in agar shake tubes. Colonies were picked from agar, and genomic DNA was extracted using a custom phenol:chloroform DNA extraction protocol followed by size selection and cleanup (Supplementary Methods).
420 Read cloud DNA sequencing library preparation was performed using the TELL-Seq WGS Library Prep Kit[53] (Universal Sequencing Technology; Canton, Massachusetts, USA) following ultralow input protocol recommendations for small genomes. Library amplification was performed using 10 µL TELL-beads and 16 amplification cycles. The resulting library was sequenced using a MiSeq Reagent Kit v2 (300-cycle; Illumina) with 2x150 bp read length on a MiSeq (Illumina) to a depth of 19.6 million total reads.
425

Genome assembly and binning

Read cloud metagenome sequencing data for the "*Ca.* Chx. allophototropha" enrichment culture were demultiplexed and quality control was performed on index reads using the Tell-Read pipeline, version 0.9.7 (Universal Sequencing Technology), with default settings. Demultiplexed reads were then assembled using Tell-
430 Link, version 1.0.0 (Universal Sequencing Technology), using global and local kmer lengths of 65 and 35, respectively. Metagenome data for this sample, together with metagenome data for an early enrichment of the same culture, were then partitioned into genome bins and analyzed using the ATLAS pipeline, version 2.2.0[54] (Supplementary Methods). The "*Ca.* Chx. allophototropha" genome bin was identified based on its taxonomic placement within the *Chloroflexota* phylum according to the Genome Taxonomy Database Toolkit (GTDB-Tk)
435 version 0.3.3, which relied on GTDB release 89[55,56]. This genome bin was manually curated to remove potentially incorrectly binned scaffolds (Supplementary Methods). Following curation, the genome bin was annotated using Prokka version 1.14.6[57], and selected portions of the genome were checked for functional genes using the BlastKOALA webserver version 2.2[58].

Short read metagenome sequencing data generated from a second early enrichment culture was also
440 analyzed. This second enrichment was lost early in the cultivation process but contained another member of the "*Ca.* Chloroheliales" (Supplementary Note 1). The single enrichment culture metagenome was processed using ATLAS version 2.2.0[54,59] for read quality control, assembly and genome binning. A single genome bin was classified using the GTDB-Tk[56] to the *Chloroflexota* phylum. This genome bin, named "*Ca.* Chloroheliaceae bin L227-5C", was manually curated (Supplementary Methods) and was then annotated using Prokka[57] as above. All
445 ATLAS settings for both runs are available in the code repository associated with this work.

To assess the community composition of the enrichment cultures, the relative abundances of each recovered genome bin from ATLAS were assessed by calculating the number of reads mapped to each bin divided by the total reads mapped to assembled scaffolds. To cross-validate these results, short protein fragments were predicted directly from unassembled read data (R1, or forward reads) using FragGeneScanPlusPlus[60]

450 commit 9a203d8. These protein fragments were searched using a profile Hidden Markov Model (HMM) for the single-copy housekeeping gene *rpoB*, and hits were assigned taxonomy using MetAnnotate[61] version 0.92 with default settings. The HMM for *rpoB* (2009 release) was downloaded from FunGene[62], and taxonomic classification was performed using the RefSeq database, release 80 (March 2017)[63]. Output of MetAnnotate was analyzed using the metannoviz R library, version 1.0.2 (https://github.com/metannotate/metannoviz) with an e-
455 value cutoff of 10^{-10}.

Identification of RCI-associated genes

To detect remote homologs of RCI-associated genes in the two "*Ca.* Chloroheliales"-associated genome bins, hmmsearch[64] version 3.1b2 was used to search predicted proteins from the bins using HMMs downloaded
460 from Pfam[65]. In particular, genes encoding the core Type I reaction center (*pscA*/*pshA*/*psaAB*; PF00223), a Type I reaction center-associated protein (*Chlorobia*-associated *pscD*; PF10657), chlorosomes structural units (*csmAC*; PF02043, PF11098), and a bacteriochlorophyll *a* binding protein (*fmoA*; PF02327), were queried. Genes encoding an RCI-associated iron-sulfur protein (*pscB*; PF12838) and RCI-associated *c*-type cytochromes (*Chlorobia*-associated *pscC* or *Heliobacterium*-associated PetJ; PF10643 or PF13442) were also queried,
465 although the HMMs used for these genes were non-specific. The genomes were also confirmed to lack the *pufLM* genes associated with RCII using the "Photo_RC" HMM (PF00124), which targets both the *pufL* and *pufM* RCII subunits. Short peptides predicted directly from unassembled read data (see above) were also queried using the "Photo_RC" HMM. Detected homologs of RCI-associated genes in the genome bins were assessed for their alignment to known RCI-associated genes encoded by phototrophic microorganisms belonging to other
470 phyla. The tertiary structures of the detected homologs were also predicted using I-TASSER[66], and phylogenetic placement was assessed (see below). Custom HMMs were also built for selected homologs. Primary sequences predicted for each gene of interest from the two genomes were aligned using Clustal Omega[67] version 1.2.3, and HMMs were generated using hmmbuild[64] version 3.1b2. Custom HMMs and homology models generated by I-TASSER are available in the code repository associated with this work.

475

Assessment of genomic potential for photosynthesis within the Chloroflexota phylum

Representative genomes for the phylum *Chloroflexota* were downloaded from NCBI based on information in the GTDB[55], release 89. All genomes that represented a type species of *Chloroflexota* according to NCBI and the GTDB were downloaded, as well as genomes representing members of known phototrophic clades (i.e., the
480 *Chloroflexaceae* family[14], the "*Ca.* Thermofonsia" order[30], and the "*Ca.* Roseiliniales" order[68]). Selected genomes of known phototrophs that were deposited in other genome databases were also downloaded (see details in the code repository associated with this work). In addition, any genome bins listed in the GTDB that belonged to the "54-19" order (which represented the name of the "*Ca.* Chloroheliales" order in this database) were downloaded. Non-phototrophic lineages were subsequently pruned to one representative per genus, with

485 the exception of non-phototrophic lineages closely related to phototrophic clades. In total, this left 58 genomes, including genomes of 28 known phototrophs. A concatenated core protein phylogeny was constructed for this genome collection using GToTree[69] version 1.4.11. The "Bacteria.hmm" collection of 74 single-copy marker genes was used, and a minimum of 30% of the marker genes were required in each genome for the genome to be kept in the final phylogeny. IQ-TREE[70] version 1.6.9 was used to build the phylogeny from the masked and
490 concatenated multiple sequence alignment. ModelFinder[71] was used to determine the optimal evolutionary model for phylogeny building. Branch supports were calculated using 1000 ultrafast bootstraps[72].

A collection of photosynthesis-associated genes, including genes for reaction centers, antenna proteins, chlorosome structure and attachment, bacteriochlorophyll synthesis, and carbon fixation, was selected based on the genome analyses of Tang and colleagues[23] and Bryant and colleagues[12]. Reference sequences for these genes
495 were selected from genomes of well-studied representatives of the phylum, *Chloroflexus aurantiacus*[13], *Oscillochloris trichoides*[73], and *Roseiflexus castenholzii*[74]. Bidirectional BLASTP[75] was performed on these reference sequences against the entire *Chloroflexota* genome collection (above) to detect potential orthologs. The BackBLAST pipeline[76], version 2.0.0-alpha3 (doi:10.5281/zenodo.3697265), was used for bidirectional BLASTP, and cutoffs for the *e* value, percent identity, and query coverage of hits were empirically optimized to
500 10^{-3}, 20%, and 50%, respectively. Genes involved in iron oxidation or reduction were also searched using FeGenie[77], commit 30174bb, with default settings.

Phylogenetic assessment of photosynthesis-associated genes

Selected photosynthesis-associated genes were tested against genes from other phyla to examine the
505 phylogenetic relationship of the novel "*Ca.* Chloroheliales" sequences to those of other phototrophic organisms. Genes for RCI/PSI (*pscA*/*pshA*/*psaAB*)[78], bacteriochlorophyll *a* binding (*fmoA*)[18], chlorosome structure (*csmA*)[22], (bacterio)chlorophyll synthesis (*bchIDH*/*chlIDH* and *bchLNB*/*chlLNB*/*bchXYZ*)[12], and RuBisCo (*cbbL*)[79], were examined. Genomes potentially containing these genes of interest were determined using a combination of automated detection via AnnoTree[80] and descriptions in the literature. Annotree was also used to summarize the
510 GTDB reference tree for downstream data visualization. Representative genomes from this initial genome set were selected from the GTDB, based on genome quality and taxonomic novelty, before being downloaded from NCBI. Potential orthologs of the genes of interest were identified in the downloaded genomes using bidirectional BLASTP, which was performed using the primary sequences of known reference genes as queries (Supplementary Data 2) via BackBLAST[76] version 2.0.0-alpha3. In addition, for some sequence sets (i.e., Type I
515 reaction centers; Bch proteins; CbbL), additional reference sequences were manually added based on literature references[12,78,79]. Sequence sets were then pruned to remove closely related sequences. Identified primary sequences were aligned using Clustal Omega[67] version 1.2.3, and the sequence sets were further verified by manually inspecting the sequence alignments for unusual variants. Alignments were then masked from non-informative regions using Gblocks[81] version 0.91b with relaxed settings (-t=p -b3=40 -b4=4 -b5=h) to preserve

regions with remote homology (see results in Extended Data Table 2). Maximum likelihood protein phylogenies were built from masked sequence alignments using IQ-TREE[70] version 1.6.9. Evolutionary rate model selection was performed using ModelFinder[71], and 1000 rapid bootstraps were used to calculate branch support values[72]. Rate models used are summarized in Extended Data Table 2.

Boreal Shield lake survey

To probe the environmental relevance of "*Ca.* Chloroheliales" members, we surveyed eight lakes (Lakes 221, 222, 224, 227, 304, 373, 442, and 626) within the IISD-ELA that were known or suspected to develop seasonal anoxia based on previous field data, along with a permanently oxic reference lake (Lake 239). These lakes were sampled for DNA in the summer or fall from 2016 until 2018 across four main sampling events. Sampling in June 2016, September 2016, and September 2017 involved multi-depth hypolimnion profiling of Lakes 227 and 442, for which metagenome data have been reported previously[82], along with mid-hypolimnion sampling of the other selected lakes. A full water column profile of DNA samples was collected for Lake 227 in September 2017, along with collection of RNA samples from a single depth of Lake 227 in the upper anoxic hypolimnion. Lastly, in July 2018, Lakes 221 and 304 were surveyed again. A single depth in the mid anoxic hypolimnion of both lakes was sampled for both DNA and RNA.

Samples for water column DNA were collected by pumping water, using a closed system gear pump and line, through sterile 0.22 µm Sterivex polyvinyl fluoride filters (Merck Millipore; Burlington, Massachusetts, U.S.A.). The sampling line was pre-flushed with water at the desired depth prior to sampling to allow sufficient flow time for water to be fully replaced twice through the line. When collecting water column RNA, the same filters were used as for DNA, but filters were immediately filled with 1.8 mL of DNA/RNA Shield (Zymo Research; Irvine, California, U.S.A.) once packed and purged of residual water. Filters were collected (and subsequently extracted and analyzed) in triplicate for RNA, with each filter replicate representing a distinct water column sample. Sterivex filters were kept chilled after collection until being frozen (at -20°C) upon return to the sampling camp later that day. Filters were then shipped chilled to the University of Waterloo and were kept frozen (at -20°C) until processing.

Environmental DNA/RNA extraction and sequencing

For each Sterivex filter used for DNA collection, the Sterivex filter case was opened, and the filter membrane was carefully removed using a flame-sterilized scalpel blade. Each filter membrane was cut in half lengthwise along the filter; one half was stored frozen to provide a backup sample, and DNA was extracted from the other half using the DNeasy PowerSoil or DNeasy PowerSoil HTP 96 Kit (Qiagen; Venlo, The Netherlands). Extractions were performed according to the kit protocol, and the optional 10 min incubation at 70°C after adding Solution C1 was performed to enhance cell lysis. Mechanical lysis was performed for samples in PowerBead Tubes using a FastPrep-24 instrument (MP Biomedicals; Santa Ana, California, U.S.A.) set at 5 m/s

555 for 45 s, and mechanical lysis was performed for samples in PowerBead DNA Plates using a mixer mill MM 400 (Retsch; Haan, Germany) set at 30 Hz for 10 min. Resulting DNA concentrations were then quantified using a Nanodrop spectrophotometer (Thermo Fisher Scientific; Waltham, Massachusetts, U.S.A.) or using the Qubit dsDNA HS Assay Kit with Qubit 2.0 fluorometer (Thermo Fisher Scientific).

RNA extraction was performed using the ZymoBIOMICS DNA/RNA Miniprep Kit (Zymo Research) with
560 initial steps of the protocol modified slightly to accommodate the volume of DNA/RNA Shield associated with each filter. Sterivex filters filled with DNA/RNA Shield were thawed, and the DNA/RNA Shield was pumped out of the filter cases and saved for downstream use. Filter cases were then opened and filters excised as described above. Each filter half was cut into small pieces. One half was stored in a clean 2 mL microfuge tube along with half of the collected DNA/RNA Shield solution and frozen as a backup sample. The other filter half,
565 along with the remaining DNA/RNA Shield, was added into a dry ZR BashingBead Lysis Tube for extraction. Mechanical lysis was performed using a FastPrep-24 instrument (MP Biomedicals). Lysis tubes were shaken at 6.5 m/s for 60 s twice, and tubes were allowed to cool on ice for at least 60 s between mechanical lysis rounds. After centrifugation, the entire supernatant volume was transferred to a new microfuge tube, and one volume of DNA/RNA Lysis Buffer was added to the tube and mixed. RNA extraction was then performed according to the
570 "DNA & RNA Parallel Purification" protocol in the kit manual using in-column DNase I treatment. Only RNA (and not DNA) extracts were saved because corresponding Sterivex filters for DNA were collected and processed for the same lake depths using the protocol described above. The resulting RNA extracts were quantified using a Nanodrop spectrophotometer (Thermo Fisher Scientific) and Qubit RNA Assay Kit with Qubit 2.0 fluorometer (Thermo Fisher Scientific). Extracts were also run on a 1% agarose gel stained with GelRed
575 (Biotium; Fremont, California, U.S.A.) to confirm that rRNA of the expected lengths was visible.

Metagenomes were sequenced in two separate batches. Lake hypolimnion DNA samples from the June and September 2016 sampling events were prepared into sequencing libraries and sequenced by the U.S. Department of Energy Joint Genome Institute (Lawrence Berkeley National Laboratory) . Metagenome library preparation was performed using the Nextera XT DNA Library Preparation Kit including library amplification
580 steps. Resulting libraries were sequenced on a HiSeq 2500 (Illumina) with 2x150 base paired end reads, generating 88.2 to 273.5 million reads per sample. Subsequently, DNA samples from 2017 and 2018 field sampling were prepared for metagenome sequencing and sequenced by the McMaster Genome Facility (McMaster University; Hamilton, Ontario, Canada). Sample DNA was sheared to the desired insert size using a Covaris ultrasonicator (Covaris; Woburn, Massachusetts, U.S.A.), and metagenome sequencing libraries were
585 prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England Biolabs) including library amplification steps. The resulting library was sequenced on a portion of a lane of a HiSeq 2500 with 2x200 base paired end reads. Because the reverse reads for that run were truncated at 114 bp due to a sequencer error, the library was run a second time on a single lane of a HiSeq 2500 with 2x250 base paired end reads. Data was combined between the two runs, generating a total of 41.2 to 128.2 million reads per sample.

590 Replicate RNA samples from 2017 and 2018 field sampling were also sequenced. Metatranscriptome library preparation and sequencing was performed by the McMaster Genome Facility (McMaster University). The Ribo-Zero rRNA Removal Kit (Bacteria; Illumina) was used to deplete rRNA genes in the samples prior to library preparation. Library preparation was then performed using the NEBNext Ultra II RNA Library Prep Kit for Illumina (New England Biolabs). Consistent RNA amounts for all samples, following rRNA depletion, were
595 used as input and without directional RNA selection. The resulting library was sequenced on a portion of a lane of a HiSeq 2500 in Rapid Run Mode with 2x200 bp paired end reads. This was the same sequencing run as for 2017-2018 metagenomes where the reverse reads were truncated at 114 bp, but because sufficient sequencing depth was obtained for the project's research goal, the library was not re-run. Sequencing generated 11.1 to 20.5 million reads per replicate. In total, 37 metagenomes were sequenced, along with nine metatranscriptomes
600 representing three samples (see Supplementary Data 3-4).

Metagenome and metatranscriptome analysis

All metagenome data was processed using the ATLAS pipeline, version 2.1.4[54]. Default settings were used for the run except that the minimum percent identity threshold for read mapping (via "contig_min_id") was set
605 to 99% and only MaxBin2 and MetaBAT2 were used as binning algorithms[83,84]. To enhance genome binning quality, six lake metagenomes that were previously sequenced from the water columns of Lakes 227 and 442[82] were included in the ATLAS run, along with a single metagenome from the aphotic zone of the nearby and meromictic Lake 111, which was also sampled in July 2018. The Lake 111 metagenome was not analyzed further in the context of this work. The entire ATLAS pipeline, including quality control on raw reads, metagenome
610 assembly of individual samples, metagenome binning, dereplication of bins from all samples and bin analysis, and gene clustering and annotation, was run end-to-end. For the genome binning step, all metagenome samples from the same lake were summarized in the same "BinGroup", allowing for read mapping information between each sample from the same lake to be used as differential abundance information to guide genome binning. After running ATLAS, the resulting dereplicated genome bins were taxonomically classified using the GTDB-Tk,
615 version 0.3.2[56], which relied on the GTDB, release 89[55]. All metagenome-assembled genomes (MAGs) had a minimum percent completeness of 50% and maximum percent contamination of 10% based on CheckM[85], according to default ATLAS settings. The relative abundance of each MAG in a sample was calculated by dividing the number of QC-processed reads mapped to the MAG by the total number of assembled reads (i.e., raw reads that mapped to assembled contigs) for that sample. A MAG was considered present in a sample if its
620 relative abundance was greater than 0.01%.

Metatranscriptome data were also processed using the ATLAS pipeline[54]. Using ATLAS commit 59da38f, the "qc" module was run to perform quality control of raw metatranscriptome read data. Then, a customized fork of ATLAS, commit 96e47df (available under the "maprna" branch at https://github.com/jmtsuji/atlas) was used to map RNA reads onto the set of dereplicated genome bins obtained from metagenome analyses (above) and to

625 summarize RNA read counts. Briefly, the dereplicated genome bins were used as input for the "genomes" module of ATLAS so that QC-processed metatranscriptome reads were mapped onto genome bins using bbmap. The minimum percent identity threshold for read mapping ("contig_min_id") was set to 99%. Following read mapping to genome bins, the counts of metatranscriptome read hits to genes within genome bins were summarized using featureCounts[86], in version 1.6.4 of the Subread package. Default settings were used, except
630 for the following flags: "-t CDS -g ID --donotsort". Raw analysis code is available in the Github repository associated with this work.

After generating RNA read mapping data via ATLAS, the relative expression of each dereplicated genome bin was calculated within each metatranscriptome. To perform this calculation, within each metatranscriptome, the number of RNA reads that mapped to each genome bin was divided by the total number of RNA reads that
635 mapped to all genome bins. The resulting relative expression values were averaged between replicate metatranscriptomes. In addition, the expression levels of individual genes within a genome were determined for "*Ca.* Chloroheliales"-associated genome bins, which were identified based on GTDB taxonomy. The single-copy taxonomic marker gene *dnaK* was identified in each genome bin based on eggNOG-mapper[87] annotations (version 1.0.3), and the annotation was confirmed based on a BLASTP[75] search of the predicted protein sequence
640 against the RefSeq[63] database. Relative expression levels of each gene within a genome bin were then calculated by dividing the gene length-normalized hit count of each gene by the gene-length normalized hit count of *dnaK*. Annotations of genes with high relative expression were determined by eggNOG-mapper as part of the ATLAS "Genecatalog" module, run on metagenome data, combined with manual searching for phototrophy-associated genes using methods described above.

645

Data availability

The three enrichment culture metagenomes, along with recovered genome bins, are available under the National Center for Biotechnology Information (NCBI) BioProject accession PRJNA640240. Amplicon sequencing data are available under the same NCBI BioProject accession. Metagenome data from 2016,
650 sequenced by the JGI, are available in the JGI Genome Portal under Proposal ID 502896. Environmental metagenome and metatranscriptome data from 2017-2018 are available under NCBI BioProject accession PRJNA664486. The full set of 756 metagenome-assembled genomes used for read mapping of metatranscriptome data are available in a Zenodo repository (doi:10.5281/zenodo.3930110) and will be made available on NCBI upon publication.

655

Code availability

Custom scripts and additional raw data files used to analyze the metagenome and genome data are available at https://github.com/jmtsuji/Ca-Chlorohelix-allophototropha-RCI (doi:10.5281/zenodo.3932366).

660 **References**

31. Armstrong, F. A. J. & Schindler, D. W. Preliminary chemical characterization of waters in the Experimental Lakes Area, northwestern Ontario. *J. Fish. Res. Bd. Can.* **28**, 171–187 (1971).
32. Brunskill, G. J. & Schindler, D. W. Geography and bathymetry of selected lake basins, Experimental Lakes Area, northwestern Ontario. *J Fish Res Bd Can* **28**, 139–155 (1971).
33. Schindler, D. W. Light, temperature, and oxygen regimes of selected lakes in the Experimental Lakes Area, northwestern Ontario. *J. Fish. Res. Bd. Can.* **28**, 157–169 (1971).
34. Schindler, D. W. & Fee, E. J. Experimental Lakes Area: whole-lake experiments in eutrophication. *J. Fish. Res. Bd. Can.* **31**, 937–953 (1974).
35. Schindler, D. W. *et al.* Eutrophication of lakes cannot be controlled by reducing nitrogen input: Results of a 37-year whole-ecosystem experiment. *Proc Natl Acad Sci USA* **105**, 11254–11258 (2008).
36. Hegler, F., Posth, N. R., Jiang, J. & Kappler, A. Physiology of phototrophic iron(II)-oxidizing bacteria: implications for modern and ancient environments. *FEMS Microbiol Ecol* **66**, 250–260 (2008).
37. Vandermeulen, J. H., Davis, N. D. & Muscatine, L. The effect of inhibitors of photosynthesis on zooxanthellae in corals and other marine invertebrates. *Marine Biology* **16**, 185–191 (1972).
38. Stookey, L. L. Ferrozine - a new spectrophotometric reagent for iron. *Anal Chem* **42**, 779–781 (1970).
39. Pfennig, N. *Rhodocyclus purpureus* gen. nov. and sp. nov., a ring-shaped, vitamin B12-requiring member of the family *Rhodospirillaceae*. *Int J Syst Evol Microbiol* **28**, 283–288 (1978).
40. Imhoff, J. F. The family *Chromatiaceae*. in *The Prokaryotes* (eds. Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 151–178 (Springer Berlin Heidelberg, 2014). doi:10.1007/978-3-642-38922-1_295.
41. Imhoff, J. F. The family *Chlorobiaceae*. in *The Prokaryotes* (eds. Rosenberg, E., DeLong, E. F., Lory, S., Stackebrandt, E. & Thompson, F.) 501–514 (Springer Berlin Heidelberg, 2014). doi:10.1007/978-3-642-38954-2_142.
42. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* 1403–1414 (2016) doi:10.1111/1462-2920.13023.
43. Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. Removing noise from pyrosequenced amplicons. *BMC Bioinform* **12**, 38 (2011).
44. Kennedy, K., Hall, M. W., Lynch, M. D. J., Moreno-Hagelsieb, G. & Neufeld, J. D. Evaluating bias of Illumina-based bacterial 16S rRNA gene profiles. *Appl Environ Microbiol* **80**, 5717–5722 (2014).
45. Bartram, A. K., Lynch, M. D., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Appl Environ Microbiol* **77**, 3846–3852 (2011).
46. Cavaco, M. A. *et al.* Freshwater microbial community diversity in a rapidly changing High Arctic watershed. *FEMS Microbiol Ecol* **95**, fiz161 (2019).
47. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852–857 (2019).
48. Min, D., Doxey, A. C. & Neufeld, J. D. AXIOME3: automation, extension, and integration of microbial ecology. *GigaScience* **10**, giab006 (2021).
49. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).
50. Bokulich, N. A. *et al.* Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* **6**, 90 (2018).
51. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**, D590–D596 (2013).

52. Glöckner, F. O. *et al.* 25 years of serving the community with ribosomal RNA gene reference databases and tools. *J Biotechnol* **261**, 169–176 (2017).
53. Chen, Z. *et al.* Ultra-low input single tube linked-read library method enables short-read second-generation sequencing systems to generate highly accurate and economical long-range sequencing information routinely. *Genome Res* **30**, 898–909 (2020).
54. Kieser, S., Brown, J., Zdobnov, E. M., Trajkovski, M. & McCue, L. A. ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinform* **21**, 257 (2020).
55. Parks, D. H. *et al.* A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* **36**, 996–1004 (2018).
56. Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P. & Parks, D. H. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**, 1925–1927 (2020).
57. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
58. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J Mol Biol* **428**, 726–731 (2016).
59. White III, R. A. *et al.* ATLAS (Automatic Tool for Local Assembly Structures) - a comprehensive infrastructure for assembly, annotation, and genomic binning of metagenomic and metatranscriptomic data. *PeerJ Preprints* **5**, e2843v1 (2017).
60. Singh, R. G. *et al.* Unipept 4.0: functional analysis of metaproteome data. *J Proteome Res* **18**, 606–615 (2018).
61. Petrenko, P., Lobb, B., Kurtz, D. A., Neufeld, J. D. & Doxey, A. C. MetAnnotate: function-specific taxonomic profiling and comparison of metagenomes. *BMC Biol* **13**, 1–8 (2015).
62. Fish, J. A. *et al.* FunGene: the functional gene pipeline and repository. *Front Microbiol* **4**, 291 (2013).
63. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–D745 (2016).
64. Eddy, S. R. Accelerated profile HMM searches. *PLOS Comput Biol* **7**, e1002195 (2011).
65. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* **44**, D279–D285 (2016).
66. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7–8 (2015).
67. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).
68. Tank, M., Thiel, V., Ward, D. M. & Bryant, D. A. A panoply of phototrophs: an overview of the thermophilic chlorophototrophs of the microbial mats of alkaline siliceous hot springs in Yellowstone National Park, WY, USA. in *Modern Topics in the Phototrophic Prokaryotes: Environmental and Applied Aspects* (ed. Hallenbeck, P. C.) 87–137 (Springer International Publishing Switzerland, 2017). doi:10.1007/978-3-319-46261-5_3.
69. Lee, M. D. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* **35**, 4162–4164 (2019).
70. Nguyen, L.-T., Schmidt, H. A., von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268–274 (2015).
71. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermiin, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**, 587–589 (2017).
72. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* **35**, 518–522 (2018).

73. Keppen, O. I., Baulina, O. I. & Kondratieva, E. N. *Oscillochloris trichoides* neotype strain DG-6. *Photosynth Res* **41**, 29–33 (1994).
74. Hanada, S., Takaichi, S., Matsuura, K. & Nakamura, K. *Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. *Int J Syst Evol Microbiol* **52**, 187–193 (2002).
75. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403–410 (1990).
76. Bergstrand, L. H., Cardenas, E., Holert, J., Hamme, J. D. V. & Mohn, W. W. Delineation of steroid-degrading microorganisms through comparative genomic analysis. *mBio* **7**, e00166-16 (2016).
77. Garber, A. I. *et al.* FeGenie: A comprehensive tool for the identification of iron genes and iron gene neighborhoods in genome and metagenome assemblies. *Front Microbiol* **11**, 37 (2020).
78. Cardona, T. Early Archean origin of heterodimeric Photosystem I. *Heliyon* **4**, e00548 (2018).
79. Tabita, F. R., Hanson, T. E., Satagopan, S., Witte, B. H. & Kreel, N. E. Phylogenetic and evolutionary relationships of RubisCO and the RubisCO-like proteins and the functional lessons provided by diverse molecular forms. *Phil Trans R Soc B* **363**, 2629–2640 (2008).
80. Mendler, K. *et al.* AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res* **47**, 4442–4448 (2019).
81. Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564–577 (2007).
82. Tsuji, J. M. *et al.* Anoxygenic photosynthesis and iron–sulfur metabolic potential of *Chlorobia* populations from seasonally anoxic Boreal Shield lakes. *ISME J* **14**, 2732–2747 (2020).
83. Wu, Y.-W., Simmons, B. A. & Singer, S. W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* **32**, 605–607 (2016).
84. Kang, D. D. *et al.* MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* **7**, e7359 (2019).
85. Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P. & Tyson, G. W. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* **25**, 1043–1055 (2015).
86. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
87. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol Biol Evol* **34**, 2115–2122 (2017).

Acknowledgments

With gratitude (chi-miigwech), we acknowledge that our samples were obtained from the traditional territory of the Anishinaabe Peoples. We also thank staff at the IISD-ELA and R. Elgood for providing baseline limnological data and sampling advice; J. Mead, E. Barber, J. Wolfe, K. Thompson, E. McQuay, and R. Henderson for assistance with lake sampling; X. Lu, E. Spasov, and L. Shakib, for assistance with DNA and RNA extraction; K. Engel for assistance with high-throughput DNA sequencing; Y. Shirotori and M. Saini for assistance with enrichment cultivation; V. Gaisin for advice for cultivating filamentous phototrophs; S. Hirose for assistance with Sanger sequencing and pigment analyses; R. Harris and E. Roach for assistance with electron microscopy; T. Chen for advice during read cloud sequencing library preparation; Y. Wang for assistance with read cloud sequence assembly; V. Thiel for advice about reference genomes of *Chloroflexota* members; B.

Schink for advice about binomial nomenclature; N. Tran, K. Shimada, and J. Hemp for discussion and advice about biochemistry; and D. Bryant for critical reading of this work. This work was supported by a Strategic Partnership Grant for Projects from the National Sciences and Engineering Research Council of Canada (NSERC), Discovery Grants from NSERC, and a research grant from the Institute for Fermentation, Osaka. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, was supported under Contract No. DE-AC02-05CH11231.

Author contributions

J.J.V., S.L.S., and J.D.N. conceived the study. J.M.T., N.A.S., and M.T. performed enrichment cultivation. J.M.T. collected epifluorescence microscopy images and ran pigment analyses. J.M.T. and N.A.S. performed electron microscopy, extracted culture genomic DNA, and performed read cloud metagenome library preparation and sequencing. J.M.T., J.J.V., and S.L.S. performed the lake survey, and J.M.T. and N.A.S. performed environmental DNA and RNA extraction. J.M.T. analyzed 16S rRNA gene, metagenome, and metatranscriptome sequencing data, conducted phylogenetic and statistical analyses, and visualized the data. J.M.T., S.N., S.H., M.T., and J.D.N. interpreted the data and its impacts on understanding the evolution of photosynthesis. S.H., M.T., and J.D.N. supervised research. J.M.T. wrote the paper with N.A.S., S.N., J.D.N., and comments from all other authors.

Ethics declarations

Competing interests

The authors declare no competing interests.

Supplementary Information | PDF file containing Supplementary Notes 1-5, Supplementary Methods, and Supplementary References.

820 **Supplementary Data 1 | Genomic context of novel Type I reaction center genes detected in this study.** The Excel file summarizes the top five BLASTP hits against the RefSeq database for predicted open reading frames (ORFs) of each of 20 genes up/downstream of the detected *pscA*-like genes.

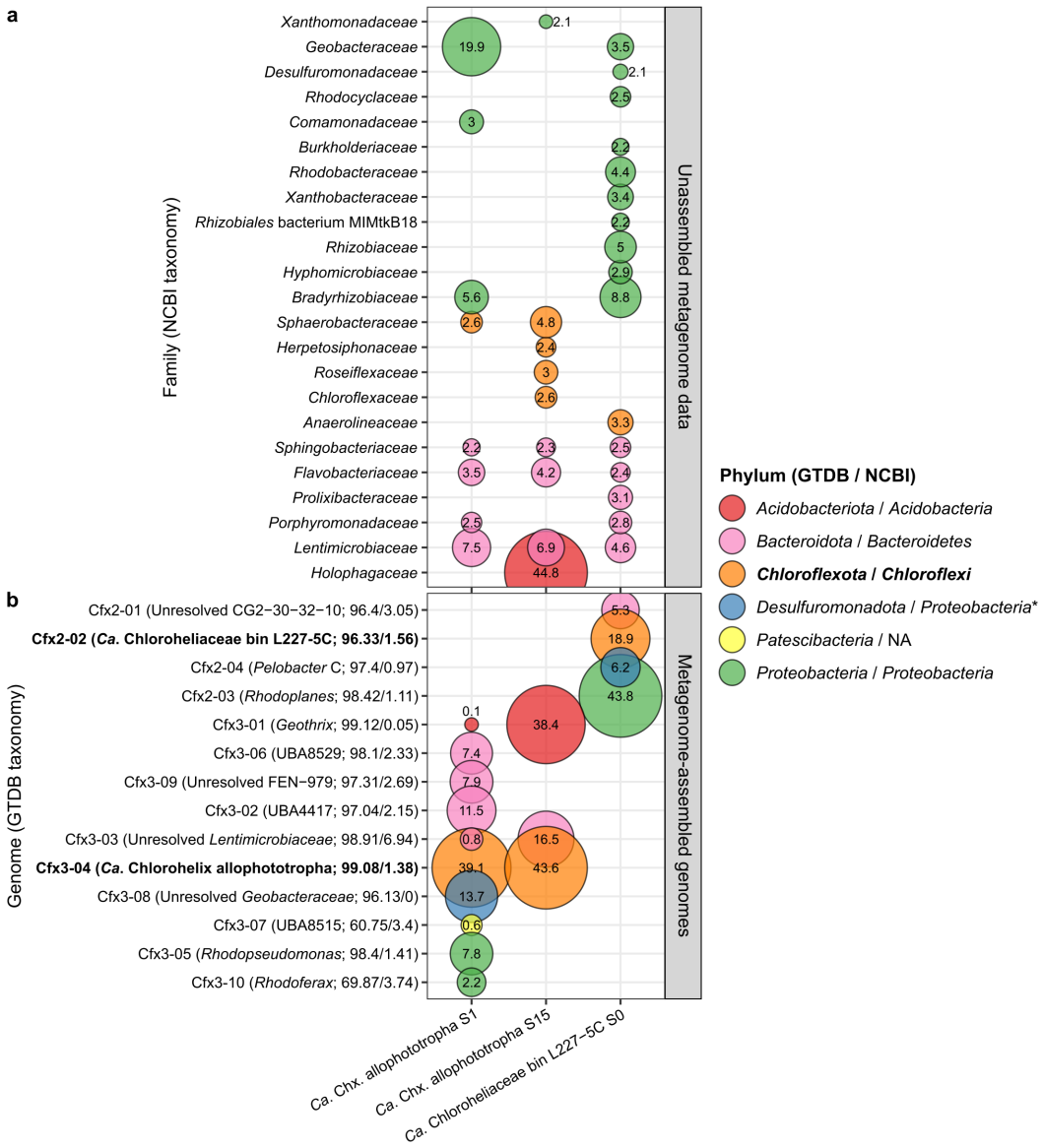
Supplementary Data 2 | Bidirectional BLASTP results for photosynthesis-related genes among the
825 ***Chloroflexota* phylum.** The Excel file summarizes the query sequences and results of bidirectional BLASTP to search for photosynthesis-related genes in genomes of *Chloroflexota* members. These data are visualized in Fig. 2.

Supplementary Data 3 | Mapping of metagenome reads to metagenome-assembled genomes. The Excel file
830 summarizes the percent mapped reads from each analyzed Boreal Shield lake metagenome to each of the 756 dereplicated MAG recovered in this study.

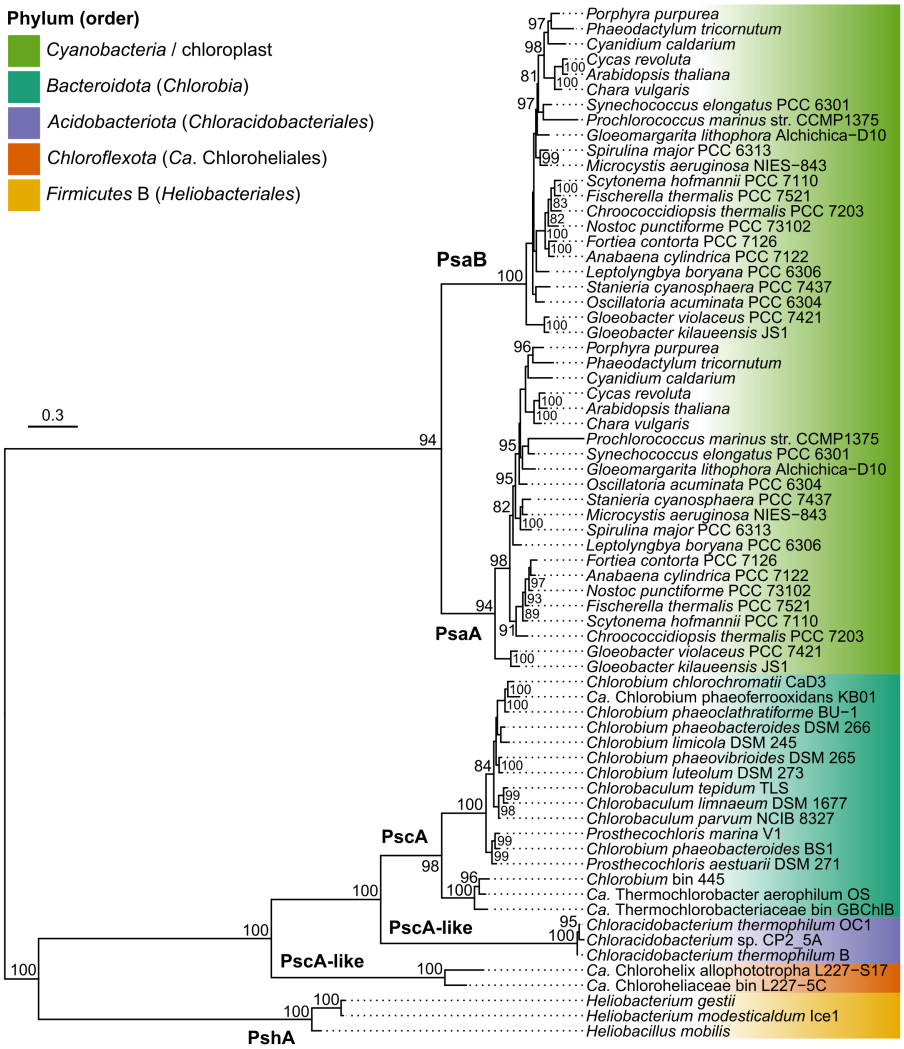
Supplementary Data 4 | Mapping of metatranscriptome reads to metagenome-assembled genomes. The Excel file summarizes the percent mapped reads from each analyzed Boreal Shield lake metatranscriptome to
835 each of the 756 dereplicated MAG recovered in this study. For each metatranscriptome, the percentage of total reads that mapped to the MAG set is also summarized. Only MAGs with at least one mapped read are shown. Data for Lakes 221 and 304 are visualized in Fig. 3a.

Supplementary Data 5 | Gene expression of "*Ca.* Chloroheliales"-associated genome bins. The Excel file
840 summarizes the normalized expression values of protein-coding genes in MAGs 319 and 729 based on environmental metatranscriptome data from Lakes 221 and 304. These data are visualized in Fig. 3b and Extended Data Fig. 7.

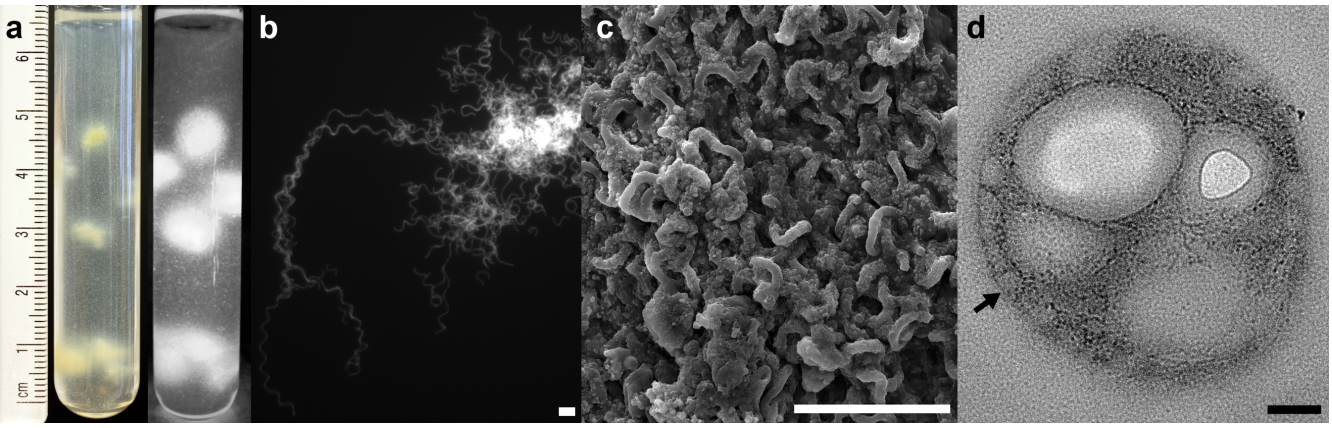
Supplementary Data 6 | Amplicon sequencing variant table of early phototroph enrichment cultures from
845 **this study.** The Excel file summarizes the percent relative abundances of amplicon sequencing variants (ASVs) detected in 16S ribosomal RNA gene amplicon sequencing data representing early enrichment cultures of "*Ca.* Chloroheliales" members.



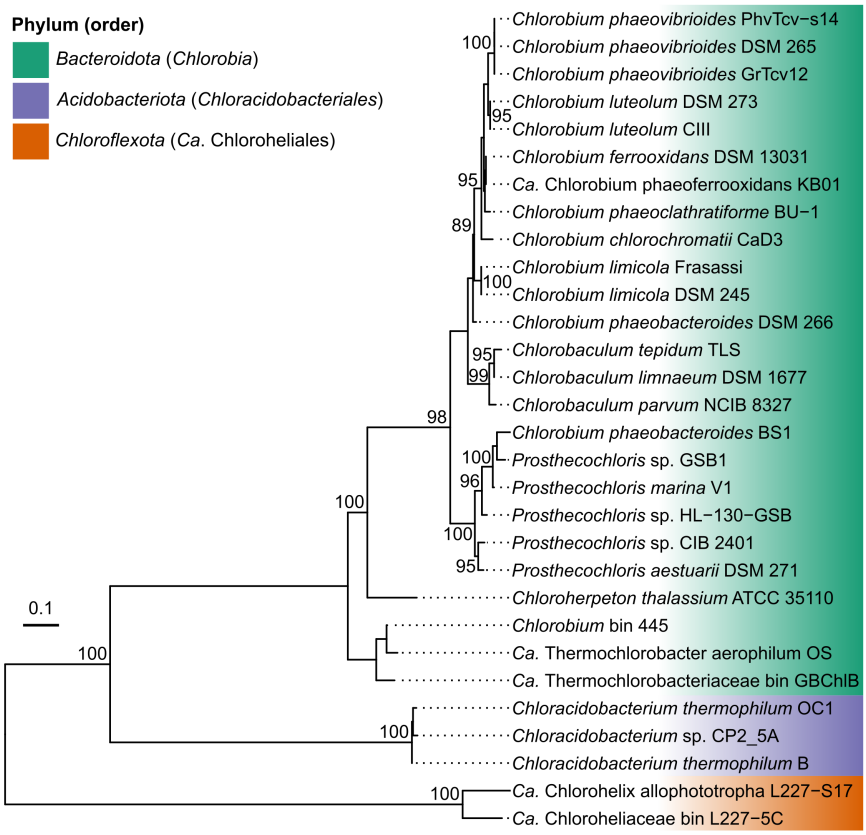
Extended Data Fig. 1 | Composition of the "*Ca.* Chloroheliales" enrichment cultures described in this study. a, Bubble plot showing the relative abundances of taxa within "*Ca.* Chloroheliales" enrichment culture metagenomes based on classification of partial RpoB fragments recovered from unassembled reads. Bubbles are coloured based on phylum according to NCBI taxonomy, because reads were classified against the RefSeq database. Microbial families with greater than 2% relative abundance are shown. Two metagenomes were derived from subcultures #1 and 15 of the "*Ca.* Chx. allophototropha" culture, and the third metagenome was derived from the initial enrichment of "*Ca.* Chloroheliaceae bin L227-5C" that was eventually lost. The subculture 15 metagenome of "*Ca.* Chx. allophototropha" was used to assemble the final genome of the strain. **b**, Bubble plot showing the relative abundances of uncurated genome bins recovered from the same metagenomes. Relative abundances are expressed as the percentage of reads mapped to the genome bin compared to the total number of assembled reads. Bubbles are coloured by phylum according to GTDB taxonomy and are shown for all entries greater than 0.05% relative abundance. Listed beside each genome bin name are the GTDB genus that the bin was classified to and the percent completeness and contamination of the bin according to CheckM. Statistics shown for the "*Ca.* Chx. allophototropha" bin differ slightly compared to Supplementary Note 2, which describes the statistics for the bin after manual bin curation.



Extended Data Fig. 2 | Maximum likelihood phylogeny of oxygenic and anoxygenic Type I reaction center predicted protein sequences. A simplified depiction of the same phylogeny is shown in Fig. 1b. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 548-residue masked sequence alignment (Extended Data Table 2).

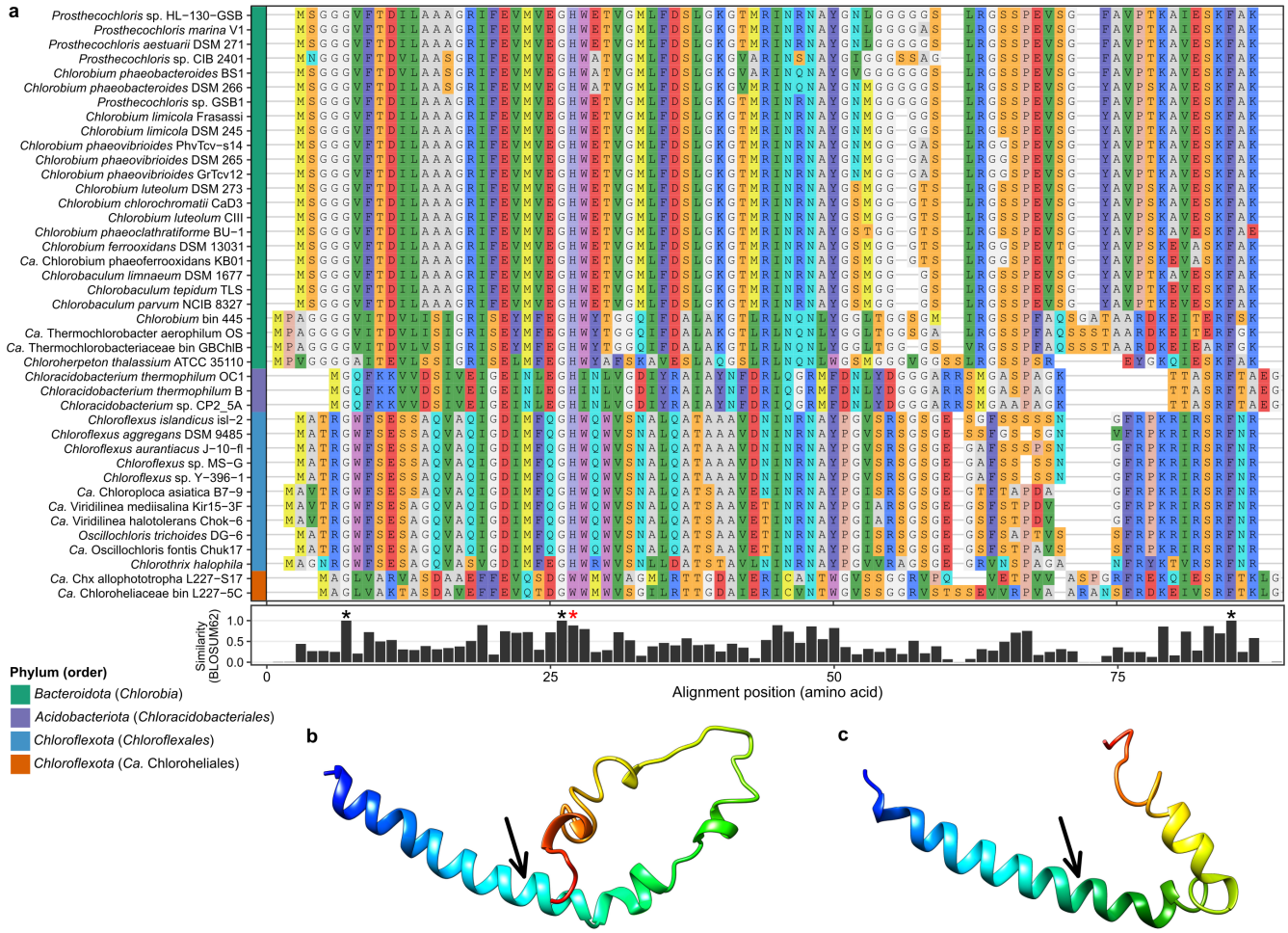


Extended Data Fig. 3 | Physiology of the "*Ca.* Chx. allophototropha" culture. a, Example growth of the "*Ca.* Chx. allophototropha" enrichment in soft agar. The left panel is a photograph of a test tube with a ruler shown for scale, and the right panel shows the same tube when viewed under a pigment fluorescence detection system optimized for bacteriochlorophyll *c* (Supplementary Methods; Extended Data Fig. 11). **b**, Epifluorescence microscopy image of "*Ca.* Chx. allophototropha" cells. Cell autofluorescence is shown, with light excitation and detection wavelengths optimized for bacteriochlorophyll *c*. **c**, Scanning electron microscopy image of an aggregate of "*Ca.* Chx. allophototropha" cells. **d**, Transmission electron microscopy image showing a cross sections of "*Ca.* Chx. allophototropha" cells. An example chlorosome is marked with an arrow. The scale bar in panel **b** represents 10 µm, and scale bars in panels **c-d** represent 0.1 µm. See Fig. 1 for additional microscopy images.

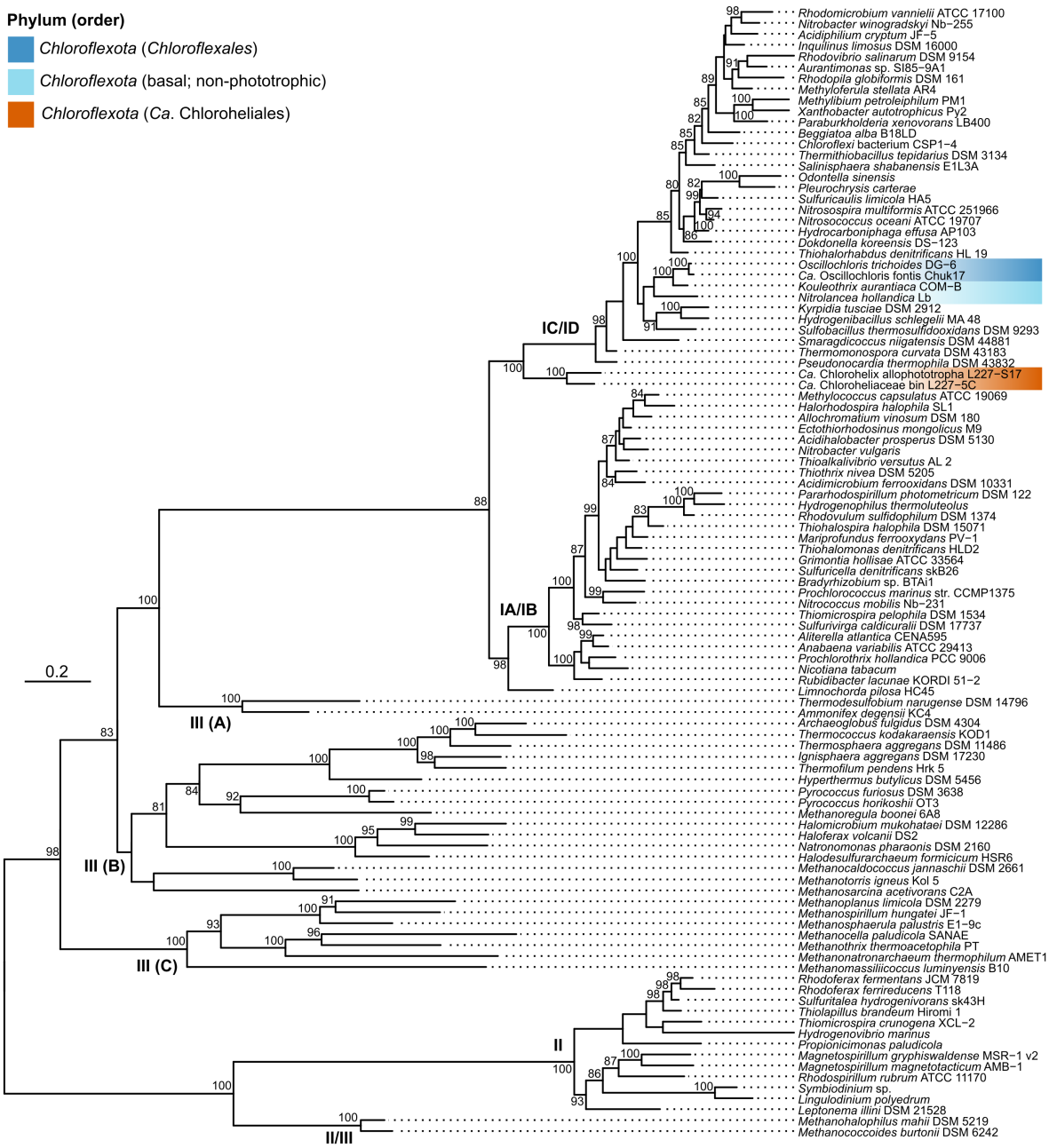


885

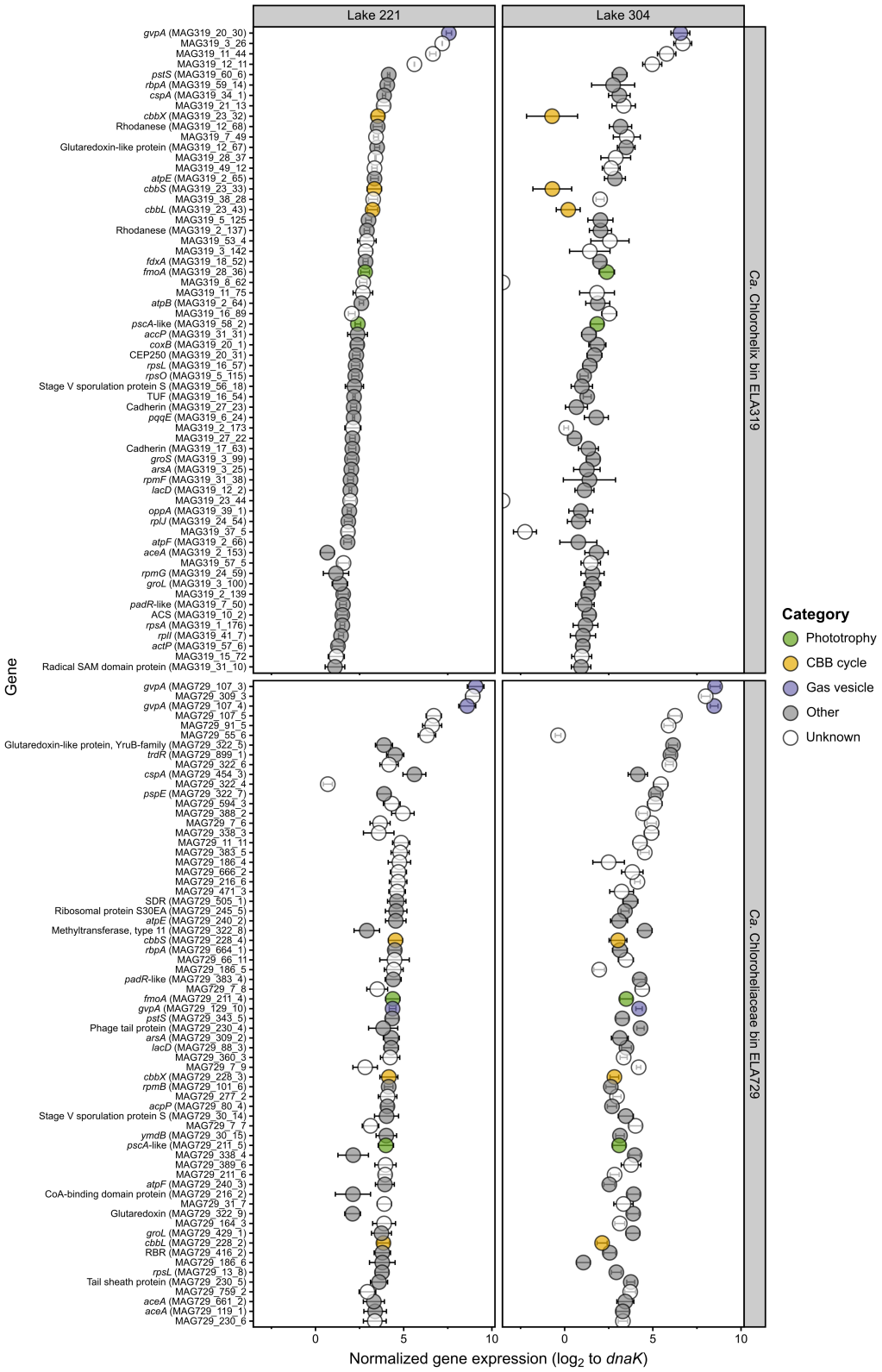
Extended Data Fig. 4 | Maximum likelihood phylogeny of Fenna-Matthews-Olson (FMO) protein (FmoA) sequences. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 356-residue masked sequence alignment (Extended Data Table 2).



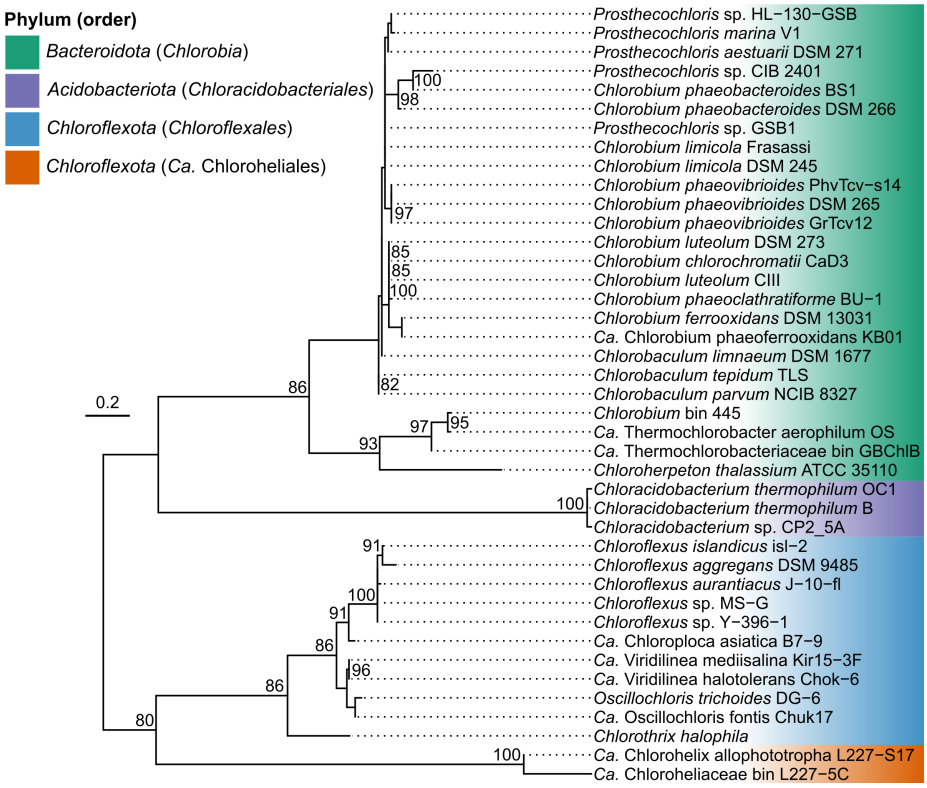
Extended Data Fig. 5 | Structural properties of the predicted chlorosome protein CsmA in "*Ca.* Chx. allophototropha". a, Multiple sequence alignment of the CsmA primary sequence including representatives from known chlorosome-containing phyla. The entire CsmA primary sequence is shown, and a colour bar on the left of the alignment indicates the phylum and order associated with each sequence. A corresponding maximum likelihood phylogeny is shown in Extended Data Fig. 8. The BLOSUM62 sequence similarity score is shown underneath the alignment. Residues that are conserved across all sequences are marked with a black asterisk, and the His25 residue predicted to be involved in bacteriochlorophyll *a* binding is marked with a red asterisk. **b**, Homology model showing the possible tertiary structure of the "*Ca.* Chx. allophototropha" CsmA protein as predicted by I-TASSER. The tryptophan residue corresponding to the typical His25 residue is indicated by an arrow. **c**, Tertiary structure of the CsmA protein from *Chlorobium tepidum* (PDB accession 2K37); the His25 site is marked with an arrow.



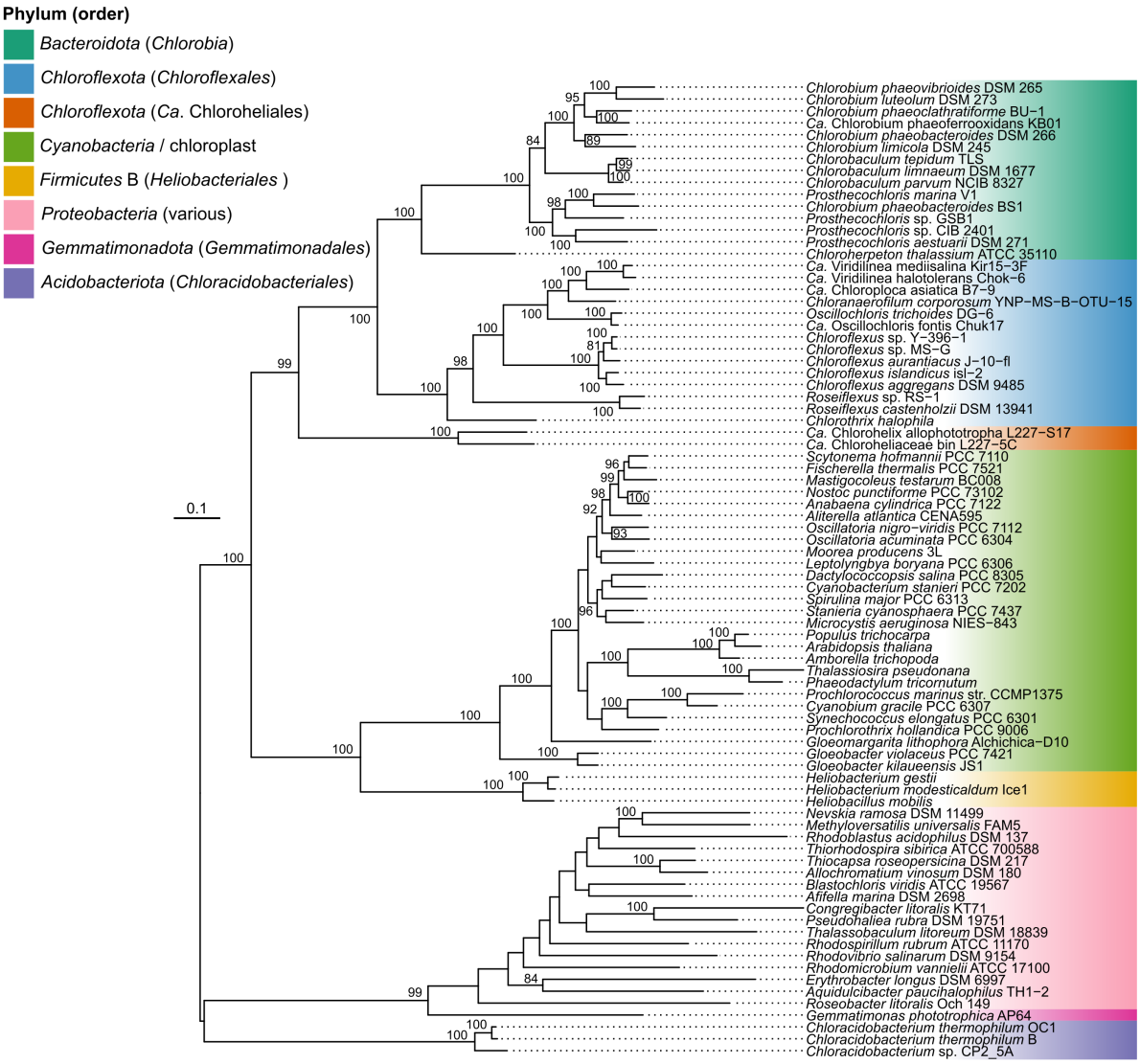
Extended Data Fig. 6 | Maximum likelihood phylogeny of RuBisCO large subunit (CbbL) predicted protein sequences. Group I to III CbbL sequences are shown, and group IV sequences, which do not form proteins involved in carbon fixation, are omitted for conciseness. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 412-residue masked sequence alignment (Extended Data Table 2).



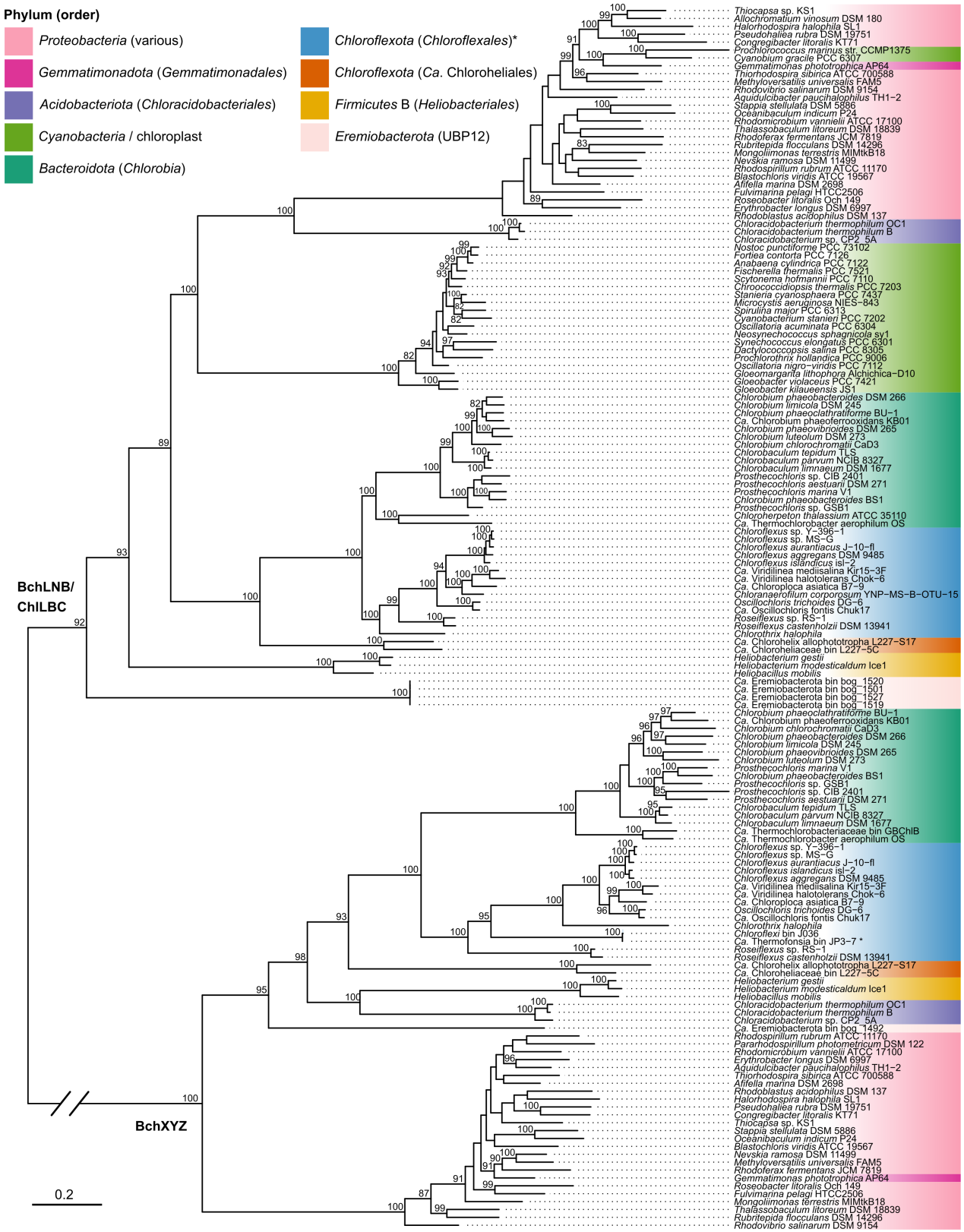
Extended Data Fig. 7 | Highly expressed genes among "*Ca.* Chloroheliales"-associated genome bins within Lake 221 and 304 metatranscriptome data. The top 50 protein-coding genes with highest normalized expression values are shown for genome bins ELA319 and 729. Error bars represent the standard deviation of the $\log_2$ expression ratio based on triplicate metatranscriptomes (n=3). Genes potentially involved in phototrophy, carbon fixation, and buoyancy are highlighted. Normalized expression values for all genes, along with their predicted amino acid sequences and annotations, are included in Supplementary Data 5.



Extended Data Fig. 8 | Maximum likelihood phylogeny of CsmA predicted protein sequences. A simplified depiction of the same phylogeny is shown in Fig. 4a. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 74-residue masked sequence alignment (Extended Data Table 2).



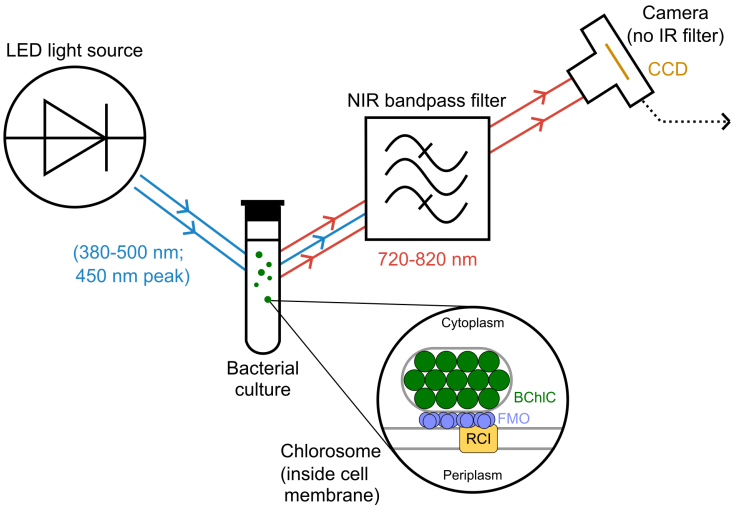
Extended Data Fig. 9 | Maximum likelihood phylogeny of BchIDH/ChlIDH protein sequences. A simplified depiction of the same phylogeny is shown in Fig. 4b. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 1702-residue masked sequence alignment (Extended Data Table 2). No sequences associated with the *Eremiobacterota* phylum are shown, because the complete BchIDH gene set could not be detected in any of the genome bins available from that phylum using bidirectional BLASTP.



935

Extended Data Fig. 10 | Maximum likelihood phylogeny of predicted protein sequences of the paralogs BchLNB/ChlLNB and BchXYZ. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 819-residue masked sequence alignment (Extended Data Table 2). One sequence among the *Chloroflexales* cluster for BchXYZ, associated with "*Ca.* Thermofonsia bin JP3-7", belongs to a genome bin that places in the basal clades of the *Chloroflexota* phylum (see Fig. 2) and is thus marked with an asterisk.

940



Extended Data Fig. 11 | Schematic of a pigment fluorescence detection system optimized for bacteriochlorophyll *c*. Light is depicted as straight lines with arrows, and pigment-containing microbial biomass inside the tube is depicted as green dots. The fluorescently active chlorosome complex of an example Type I reaction center-containing cell is shown as an inset. An example photograph using the system is shown in Extended Data Fig. 3a. Abbreviations: LED, light-emitting diode; IR, infrared; NIR near infrared; CCD, charge-coupled device; BChlC, bacteriochlorophyll *c*; FMO, Fenna-Matthews-Olson protein; RCI, Type I photosynthetic reaction center.

Extended Data Table 1 | Genes potentially involved in phototrophy or carbon/nitrogen fixation among genomes of "*Ca.* Chloroheliales" members recovered in this study. The GenBank accession numbers for the predicted primary sequence of each gene are shown. Results correspond to those shown in Fig. 2 except that homologs associated with the incomplete 3-hydroxypropionate bicycle are omitted for clarity. Raw results of bidirectional BLASTP are available in the code repository associated with this work, and queries for bidirectional BLASTP are summarized in Supplementary Data 2.

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

Extended Data Table 2 | Properties of sequence alignments and phylogenetic trees presented in this study. The unmasked alignment length of the core protein set was not available due to the constraints of the software pipeline used to generate the final masked alignment.

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