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

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

*Corresponding authors: jackson.tsuji@uwaterloo.ca; marcus.tank@dsmz.de; jneufeld@uwaterloo.ca

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

Summary

Chlorophyll-based phototrophy is performed using quinone and/or Fe-S type reaction centers[1,2]. Unlike oxygenic phototrophs, where both reaction center classes are used in tandem as Photosystem II and Photosystem I, anoxygenic phototrophs use only one class of reaction center, termed Type II (RCII) or Type I (RCI) reaction centers, separately for phototrophy[3]. Here we report the cultivation and characterization of a filamentous anoxygenic phototroph within the *Chloroflexota* (formerly *Chloroflexi*[4]) phylum, provisionally named '*Candidatus* Chlorohelix allophototropha', that performs phototrophy using a distinct fourth clade of RCI protein, despite placing sister taxonomically to RCII-utilizing *Chloroflexota* members. '*Ca.* Chx. allophototropha' contains chlorosomes, uses bacteriochlorophyll *c*, and encodes the FMO protein like other RCI-utilizing phototrophs in the *Chlorobiales* and *Chloracidobacterales* orders[5]. '*Ca.* Chx. allophototropha' also encodes the potential for carbon fixation using the Calvin-Benson-Bassham (CBB) cycle, unlike all known RCI-utilizing phototrophs[6]. The discovery of '*Ca.* Chx. allophototropha', as the first representative of a novel *Chloroflexota* order (i.e., '*Ca.* Chloroheliales'), sheds light on longstanding questions about the evolution of photosynthesis, including the origin of chlorosomes among RCII-utilizing *Chloroflexota* members[5,7]. The *Chloroflexota* is now the only phylum outside the *Cyanobacteria* containing genomic potential for both quinone and Fe-S type reaction centers and can thus serve as an additional system for exploring fundamental questions about the evolution of photosynthesis.

Main text

The evolution of photosynthesis remains poorly understood[8,9]. Among chlorophyll-based phototrophic organisms, two distinct classes of photosynthetic reaction center, either quinone-type or Fe-S type, are used for light energy conversion by members of at least eight microbial phyla, yet the evolutionary history of these reaction centers is unclear[1,3,10,11]. Members of a single microbial phylum, the *Cyanobacteria* (along with chloroplasts in eukaryotes), use both quinone and Fe-S type reaction centers in tandem as Photosystem II (PSII) and Photosystem I (PSI), respectively, to generate molecular oxygen through the oxidation of water[2]. Diverse anoxygenic phototrophs making up the remaining known phyla perform phototrophy using a single reaction center class, either a Type II (RCII) or Type I (RCI) reaction center related to PSII or PSI, respectively[3]. Combined phylogenetic, biochemical, and physiological evidence suggests that many photosynthesis-associated genes, including reaction center genes, are ancient in origin and may have been transferred between distantly related microorganisms multiple times over evolutionary history[5,12]. Such lateral gene transfer is presumed to be a key factor allowing for the existence of both oxygenic and anoxygenic photosynthesis[13]. However, modern case studies of such distant lateral gene transfer events are limited, hindering understanding of the fundamental mechanisms allowing for chlorophototrophy to evolve[14].

Among phototrophic microorganisms, anoxygenic phototrophs within the *Chloroflexota* phylum (formerly the *Chloroflexi* phylum[4,15]; also called filamentous anoxygenic phototrophs or "green non-sulfur bacteria") have been particularly cryptic in their evolutionary origin. Although these bacteria use RCII for photosynthesis, several phototrophic *Chloroflexota* members also contain chlorosomes, a light-harvesting apparatus that is otherwise found only in RCI-utilizing bacteria[16]. How chlorosomes were adopted by both RCI-containing phototrophs and RCII-containing phototrophic *Chloroflexota* members remains highly speculative, with some suggesting interactions between RCI and RCII-utilizing bacteria in the evolutionary past[5,7]. However, concrete evidence for such interactions has never been demonstrated in nature. Here we report the cultivation of a highly novel phototroph within the

Chloroflexota phylum that uses RCI and chlorosomes for phototrophy. We describe the physiological and genomic properties of the novel bacterium and its implications for the evolution of photosynthesis.

Water was collected from the illuminated and anoxic water column of an iron-rich Boreal Shield lake to enrich for resident bacterial phototrophs. After approximately two months of incubation under light, enrichment cultures began showing ferrous iron oxidation activity (Supplementary Note 1), and 16S ribosomal RNA (16S rRNA) gene amplicon sequencing from some cultures indicated the presence of a novel bacterium distantly related to known *Chloroflexota* members (Supplementary Data 1). Further enrichment in deep agar dilution series eliminated a purple phototrophic bacterium related to the metabolically versatile *Rhodopseudomonas palustris*[17] and a green phototrophic bacterium belonging to the *Chlorobium* genus. Metagenome sequencing and assembly recovered the nearly closed genome of the *Chloroflexota* member, which represented the only detectable phototroph and comprised 43.4% of raw metagenome sequences (Extended Data Fig. 1), along with nearly closed genomes of the two major heterotrophic populations remaining in the culture (Supplementary Note 2). Given the detection of genes for RCI in the *Chloroflexota* member's genome, we provisionally classify this bacterium as, '*Candidatus* Chlorohelix allophototropha' strain L227-S17 (Chlo.ro.he'lix. Gr. adj. *chloros*, green; Gr. fem. n. *helix*, spiral; N.L. fem. n. *Chlorohelix*, a green spiral. al.lo.pho.to.tro'pha. Gr. adj. *allos*, other, different; Gr. n. *phos*, -*otos*, light; N.L. suff. -*tropha*, feeding; N.L. fem. adj. *allophototropha*, phototrophic in a different way).

Physiologically, '*Ca.* Chx. allophototropha' resembled other cultured phototrophs belonging to the *Chloroflexota* phylum (Fig. 1). Cells grew well in soft agar (Fig. 1a) and developed as long, spiralling filaments (Fig. 1b), similar to those observed for *Chloronema giganteum*[18], that were composed of cells with dimension of ~2-3 µm x 0.6 µm (Fig. 1c-d). In addition to the presence of light-harvesting chlorosomes (Fig. 1e-f), '*Ca.* Chx. allphototropha' cells were pigmented, with the dominant pigment being bacteriochlorophyll *c* based on the 433 and 663 nm peaks observed in absorption spectrum data from acetone:methanol pigment extracts (Fig. 1g; see Supplementary Note 3)[19].

Despite physiological similarities to cultured *Chloroflexota* members, genome sequencing data revealed that '*Ca.* Chx. allophototropha' encoded a remote homolog of a *pscA*-like RCI gene (Fig. 2) rather than the expected *pufL*/*pufM* RCII genes. A *pscA*-like gene was also detected in a second genome bin, named '*Ca.* Chloroheliales bin L227-5C', which we recovered from a second enrichment culture that was subsequently lost (see Supplementary Notes 1-2). Our detection of RCI genes in '*Ca.* Chx. allophototropha' makes the *Chloroflexota* phylum unique among all known phototroph-containing microbial phyla, which otherwise contain representatives that use strictly RCI or RCII for phototrophy (Fig. 2a). Based on a maximum-likelihood amino acid sequence phylogeny (Fig. 2b; Extended Data Fig. 2), the PscA-like predicted protein of '*Ca.* Chx. allophototropha' represents a clearly distinct fourth class of RCI protein that places between clades for PscA-like proteins of *Chloracidobacterales* members and PshA proteins of *Heliobacterales* members. Homology modelling indicated that the '*Ca.* Chx. allophototropha' PscA-like protein contains the expected six N-terminal transmembrane helices for coordinating antenna pigments and five C-terminal transmembrane helices involved in core electron transfer[20] (Fig. 2c). The *pscA*-like gene occurred on a long scaffold (3.29 Mb in length) and was flanked by other genes whose predicted primary sequences had closest BLASTP hits to other *Chloroflexota* members (Supplementary Data 2). The genome did not contain the *pscBCD* photosystem accessory genes found among *Chlorobia* members[5], nor a cyanobacterial *psaB*-like paralog, implying that '*Ca.* Chx. allophototropha' uses a homodimeric RCI complex[2] like other known anoxygenic phototrophs.

Based on the Genome Taxonomy Database (GTDB) classification system[4,21] used throughout this work, '*Ca.* Chx. allophototropha' represents the first cultivated member of a novel order within the *Chloroflexota* phylum (Fig. 3), which we provisionally name the '*Ca.* Chloroheliales' (see Supplementary Note 4). This order placed immediately basal to the *Chloroflexales* order, which contains the canonical RCII-utilizing phototroph clade in the phylum that includes the *Chloroflexaceae*, *Roseiflexaceae*, and *Oscillochloridaceae* families[15]. The *Chloroflexales* order also contains a basal non-

phototrophic family, the *Herpetosiphonaceae*, which was formerly placed in its own order prior to
110 reassignment in the GTDB[4]. The close phylogenetic placement of RCI-utilizing '*Ca.* Chloroheliales' members and RCII-utilizing *Chloroflexales* members implies the potential for past genetic interaction between these phototroph groups.

The genome of '*Ca.* Chx. allophototropha' encoded numerous phototrophy-associated genes with remote homology to phototrophy genes of known organisms (Fig. 3; Supplementary Data 3). We
115 detected a highly novel homolog of *fmoA* encoding the Fenna-Matthews-Olson (FMO) protein involved in energy transfer from chlorosomes to RCI[22] (Extended Data Fig. 3). This finding supports that '*Ca.* Chx. allophototropha' relies on RCI and chlorosomes for phototrophic energy conversion and makes the *Chloroflexota* the third known phylum to use the FMO protein in phototrophy[23]. Using both bidirectional BLASTP and profile hidden Markov model-based searches, we also detected a potential
120 homolog of the key chlorosome structural gene *csmA*[5,16] with ~33% identity at the predicted amino acid level to the CsmA primary sequence of *Chloroflexus aurantiacus* J-10-fl (Fig. 3). This potential homolog had a similar predicted tertiary structure to the CsmA protein of *Chlorobium tepidum*[24] based on homology modelling and shared several highly conserved residues with other known CsmA sequences, although the potential homolog lacked the H25 histidine hypothesized to be required for
125 bacteriochlorophyll *a* binding[25] (Extended Data Fig. 4). No other predicted open reading frames in the genome had close homology to known CsmA sequences, implying that this sequence represents a highly novel CsmA variant. We also detected potential homologs of *csmM* and *csmY*[5] in the genome (Fig. 3).

Similarly to *Oscillochloris trichoides*, the '*Ca.* Chx. allophototropha' genome encoded genes for
130 the Calvin-Benson-Bassham (CBB) cycle (Fig. 3), including a deep-branching Class IC/ID "red type" *cbbL* gene[26] representing the large subunit of RuBisCO (Extended Data Fig. 5). Carbon fixation via the CBB cycle has never before been reported for an RCI-utilizing phototroph, and this finding further highlights the novelty of '*Ca.* Chx. allophototropha' compared to known phototrophs. The novel

genome did not encode the potential for carbon fixation via the 3-hydroxypropionate (3HP) bicycle, which is thought to represent a more recent evolutionary innovation within the *Chloroflexales* order[27]. The '*Ca.* Chx. allophototropha' genome also encoded the biosynthesis pathway for bacteriochlorophylls *a* and *c*[5] and had the genomic potential for nitrogen fixation based on the presence of a *nifHI*$_1$*I*$_2$*BDK* gene cluster[28] (Fig. 3). Although the closest BLASTP hits against RefSeq for predicted primary sequences of *nifH*, *nifB*, *nifD*, and *nifK* were to *Chloroflexota*-associated sequences, the *nifI*$_1$ and *nifI*$_2$ genes involved in regulation of nitrogen metabolism[29] were not detectable among other *Chloroflexota* members. Other members instead encoded a *nifHBDK* gene cluster, implying that the '*Ca.* Chx. allophototropha' *nif* cluster could represent a more ancestral *nif* cluster variant compared to those of its relatives. The genome bin of the second '*Ca.* Chloroheliales' member enriched in this study (Fig. 3; bin L227-5C) encoded a similar set of phototrophy-related gene homologs as '*Ca.* Chx. allophototropha', which further supports the robustness of our detection of these novel phototrophy-related genes.

Previously, the genetic content of RCII-utilizing phototrophic *Chloroflexota* members has been enigmatic because several *Chloroflexota*-associated phototrophy genes appear to be most closely related to genes of RCI-utilizing phototrophs. Our results lead us to hypothesize that '*Ca.* Chx. allophototropha' represents a missing transition form in the evolution of phototrophy among the *Chloroflexota* phylum that resolves this enigma by providing a direct link between RCI- and RCII-utilizing phototrophic taxa. Along with encoding chlorosomes, RCII-utilizing *Chloroflexota* members have been observed to encode bacteriochlorophyll synthesis genes most closely related to those of RCI-utilizing *Chlorobia* members[5]. Using maximum likelihood phylogenies of the (bacterio)chlorophyll synthesis proteins BchIDH/ChlIDH (Extended Data Fig. 6) and BchLNB/ChlLNB/BchXYZ (Extended Data Fig. 7), we observed that '*Ca.* Chloroheliales' sequences placed immediately adjacent to this sister grouping of RCII-utilizing *Choroflexota* and RCI-utilizing *Chlorobia* members. A phylogeny of the chlorosome structural protein CsmA also indicated closer evolutionary relatedness between '*Ca.*

Chloroheliales', RCII-utilizing *Chloroflexota*, and RCI-utilizing *Chlorobia* members compared to RCI-
160 utilizing *Acidobacteriota* members such as *Chloracidobacterium thermophilum*[23] (Extended Data Fig. 8). The phylogenetic clustering of RCI- and RCII-utilizing *Chloroflexota* members with RCI-utilizing *Chlorobia* members suggests a shared genetic history of phototrophy between these groups despite using different core reaction center components. Thus, it is possible that an ancestor of '*Ca.* Chx. allophototropha' received genes for RCI, chlorosomes, and bacteriochlorophyll synthesis, along with
165 genes for carbon or nitrogen fixation, via lateral gene transfer and then provided the photosynthesis accessory genes needed by RCII-containing *Chloroflexales* members via either direct vertical inheritance or a subsequent lateral gene transfer event.

The cultivation of '*Ca.* Chx. allophototropha' also opens substantial possibilities for understanding the evolution of oxygenic photosynthesis. The *Chloroflexota* now represents the only
170 phylum outside of the *Cyanobacteria* where both quinone and Fe-S type photosynthetic reaction centers can be found (Fig. 2a), although in different organisms. Genes for RCII and RCI might be readily incorporated into a single member of the *Chloroflexota* phylum, for example given the apparent lateral transfer of RCII genes to basal *Chloroflexota* clades (Fig. 3) described previously[30]. One model for the origin of oxygenic photosynthesis suggests that a single organism received quinone and Fe-S
175 type reaction centers via lateral gene transfer and then developed the capacity to use these two reaction center classes in tandem as PSII and PSI[1,13]. The physiological effect of receiving both reaction center classes might now be testable in a laboratory setting. Ancestral state reconstructions could also yield insights into the genomic properties of the common ancestor of RCI-utilizing '*Ca.* Chloroheliales' and RCII-utilizing *Chloroflexales* members and whether this common ancestor was itself phototrophic.
180 Generally, discovering '*Ca.* Chx. allophototropha' demonstrates the possibility for massive photosynthesis gene movements over evolutionary time, which is required by many evolutionary models but has been sparsely demonstrated in nature[12,14].

With the cultivation of '*Ca.* Chx. allophototropha', the *Chloroflexota* now represents the first known phylum of phototrophs having members that use either RCI or RCII for phototrophic energy conversion. '*Ca.* Chx. allophototropha' is unique among known phototrophs, because it encodes a distinct fourth class of RCI protein, forms the third known phototroph clade to utilize the FMO protein for energy transfer, and represents the only known RCI-utilizing phototroph to encode the CBB cycle for carbon fixation. Future research examining the biochemical properties of the reaction center and chlorosomes of '*Ca.* Chx. allophototropha', the potentially photoferrotrophic metabolism of this strain (Supplementary Note 1), and the diversity and biogeography of '*Ca.* Chloroheliales' members, will provide fundamental insights into how photosynthesis operates and evolves. Cultivation of '*Ca.* Chx. allophototropha' allows the *Chloroflexota* to serve as a model system for exploring photosystem gene movements over evolutionary time.

References

1. Hohmann-Marriott, M. F. & Blankenship, R. E. Evolution of photosynthesis. *Annu Rev Plant Biol* **62**, 515–548 (2011).
2. Fischer, W. W., Hemp, J. & Johnson, J. E. Evolution of oxygenic photosynthesis. *Annu Rev Earth Pl Sc* **44**, 647–683 (2016).
3. Thiel, V., Tank, M. & Bryant, D. A. Diversity of chlorophototrophic bacteria revealed in the omics era. *Annu Rev Plant Biol* **69**, 21–49 (2018).
4. 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).
5. 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.
6. Tang, K.-H., Tang, Y. J. & Blankenship, R. E. Carbon metabolic pathways in phototrophic bacteria and their broader evolutionary implications. *Front Microbiol* **2**, (2011).
7. Hohmann-Marriott, M. F. & Blankenship, R. E. Hypothesis on chlorosome biogenesis in green photosynthetic bacteria. *FEBS Lett* **581**, 800–803 (2007).
8. 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.
9. Cardona, T. Thinking twice about the evolution of photosynthesis. *Open Biol* **9**, 180246 (2019).
10. Ward, L. M., Cardona, T. & Holland-Moritz, H. Evolutionary implications of anoxygenic phototrophy in the bacterial phylum *Candidatus* Eremiobacterota (WPS-2). *Front. Microbiol.* **10**, (2019).

11. 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).

12. 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).

13. Soo, R. M., Hemp, J. & Hugenholtz, P. Evolution of photosynthesis and aerobic respiration in the cyanobacteria. *Free Radical Bio Med* **140**, 200–205 (2019).

14. 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).

15. 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.

16. 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).

17. Jiao, Y., Kappler, A., Croal, L. R. & Newman, D. K. Isolation and characterization of a genetically tractable photoautotrophic Fe(II)-oxidizing bacterium, *Rhodopseudomonas palustris* strain TIE-1. *Appl Environ Microbiol* **71**, 4487–4496 (2005).

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

19. 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.

20. Gisriel, C. *et al.* Structure of a symmetric photosynthetic reaction center–photosystem. *Science* **357**, 1021–1025 (2017).

21. 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).

22. 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.

23. Bryant, D. A. *et al. Candidatus* Chloracidobacterium thermophilum: an aerobic phototrophic Acidobacterium. *Science* **317**, 523–526 (2007).

24. Pedersen, M. Ø., Underhaug, J., Dittmer, J., Miller, M. & Nielsen, N. C. The three-dimensional structure of CsmA: A small antenna protein from the green sulfur bacterium *Chlorobium tepidum*. *FEBS Lett* **582**, 2869–2874 (2008).

25. 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).

26. 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).

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 Nat Acad Sci USA* **114**, 10749–10754 (2017).

28. Raymond, J., Siefert, J. L., Staples, C. R. & Blankenship, R. E. The natural history of nitrogen fixation. *Mol Biol Evol* **21**, 541–554 (2004).

29. Enkh-Amgalan, J., Kawasaki, H. & Seki, T. Molecular evolution of the *nif* gene cluster carrying *nifI*$_1$ and *nifI*$_2$ genes in the Gram-positive phototrophic bacterium *Heliobacterium chlorum*. *Int J Syst Evol Microbiol* **56**, 65–74 (2006).

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, a seasonally anoxic and ferruginous Boreal Shield lake at the International Institute for Sustainable Development Experimental Lakes Area (IISD-ELA; near Kenora, Canada). Lake 227 develops up to ~100 μM concentrations of dissolved ferrous iron in its anoxic water column[31], and anoxia is more pronounced than expected naturally due to the long-term experimental eutrophication of the lake[32]. The lake's physico-chemistry has been described in detail previously.[31] We collected water from the illuminated portion of the upper anoxic zone of Lake 227 in September 2017, at 3.875 and 5 m depth, and transported this water anoxically and chilled in 120 mL glass serum bottles, sealed with black rubber stoppers (Geo-Microbial Technology Company; Ochelata, Oklahoma, USA), to the laboratory.

Water was supplemented with 2% v/v of a freshwater medium containing 8 mM ferrous chloride[33] and was distributed anoxically into 120 mL glass serum bottles, also sealed with black rubber stoppers (Geo-Microbial Technology Company), that had a headspace of dinitrogen gas at 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[34] 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). Cultures were monitored regularly for ferrous iron concentration using the ferrozine assay[35] and were amended with additional freshwater medium or ferrous chloride (e.g., in 0.1-1 mM increments) when ferrous iron levels dropped, presumably due to iron oxidation.

The '*Ca.* Chx. allophototropha' enrichment culture was gradually acclimatized to higher media concentrations 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; LED Grow Lights Depot; Portland, Oregon, USA). A deep agar dilution series[36] was used to eliminate contaminants from the culture. The same freshwater medium was used for this dilution series, amended with an acetate feeding solution[37]. 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. Over multiple rounds of optimization, the growth medium was adjusted to contain 2 mM ferrous chloride and 1.2 mM sodium acetate and to be a 1:1 dilution of the original freshwater medium, kept at a pH of ~7.5. In addition, cultures were generally incubated in soft agar (0.2-0.3%; w/v)[38] 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[19]. Additionally, a higher medium pH of ~7.5-8.5 was used for a time to select against green phototrophic bacteria belonging to the *Chlorobia* class, because these bacteria are known to grow poorly in moderately basic conditions[39]. 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).

Physiological characterization

Filaments of '*Ca.* Chx. allophototropha' cells were identified based on pigment autofluorescence using epifluorescence microscopy. An Eclipse 600 light microscope equipped for fluorescence microscopy (Nikon; Shinagawa, Tokyo, Japan) was used with a xenon light source. Autofluorescence in the wavelength range expected for bacteriochlorphyll *c* was detected using an excitation filter of 445(±45) nm, dichroic mirror of 520 nm, and emission filter of 715 (LP) nm (all from Semrock;

Rochester, New York, 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).

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 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 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 (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 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.

355 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 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 360 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

The microbial community composition of early enrichment cultures was assessed using 16S 365 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 amplified from extracted DNA using the 370 universal prokaryotic PCR primers 515F-Y[40] and 926R[41], using primers with attached index sequences and sequencing adaptors as described previously[42–44]. Amplification was performed in singlicate. Library pooling, cleanup, and sequencing on a MiSeq (Illumina; San Diego, California, USA) was performed as described previously[42]. Sequence data analysis was performed using QIIME2[45] version 2019.10 via the AXIOME[46] pipeline, commit 1ec1ea6 (https://github.com/neufeld/axiome3), with

default analysis parameters. Briefly, paired-end reads were trimmed, merged, and denoised using DADA2[47] to generate an amplicon sequence variant (ASV) table. Taxonomic classification of ASVs was performed using QIIME2's naive Bayes classifier[48] trained against the Silva SSU database[49,50], release 132. The classifier training file was prepared using QIIME2 version 2019.7.

Metagenome sequencing

The functional gene content of early liquid enrichment cultures was assessed via shotgun metagenome sequencing. Genomic DNA was extracted as above, and library preparation and sequencing was performed at The Centre for Applied Genomics (TCAG; The Hospital for Sick Children, Toronto, Canada). The Nextera DNA Flex Library Prep Kit (Illumina) kit was used for metagenome library preparation, and libraries were sequenced using a fraction of a 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). Read cloud DNA sequencing library preparation was performed using the TELL-Seq WGS Library Prep Kit[51] (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.

Genome assembly and binning

Read cloud metagenome sequencing data for the '*Ca.* Chx. allophototropha' enrichment culture were demultiplexed and quality control was performed on index reads using the Tell-Read pipeline, version 0.9.7 (Universal Sequencing Technology), with default settings. Demultiplexed reads were then assembled using Tell-Link, version 1.0.0 (Universal Sequencing Technology), using global and local kmer lengths of 65 and 35, respectively. Metagenome data for this sample, together with metagenome data for an early enrichment of the same culture, were then partitioned into genome bins and analyzed using the ATLAS pipeline, version 2.2.0[52] (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) version 0.3.3, which relied on GTDB release 89[4,21]. 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[53], and selected portions of the genome were checked for functional genes using the BlastKOALA webserver version 2.2[54].

Short read metagenome sequencing data generated from a second early enrichment culture was also analyzed. This second enrichment was lost early in the cultivation process but contained another member of the '*Ca.* Chloroheliales' (Supplementary Note 1). The single enrichment culture metagenome was processed using ATLAS version 2.2.0[52,55] for read quality control, assembly and genome binning. A single genome bin was classified using the GTDB-Tk[21] to the *Chloroflexota* phylum. This genome bin was manually curated (Supplementary Methods) and was then annotated using Prokka[53] as above. All ATLAS settings for both runs are available in the code repository associated with this work.

To assess the community composition of the enrichment cultures, the relative abundances of each recovered genome bin from ATLAS were assessed by calculating the number of reads mapped to each bin divided by the total reads mapped to assembled scaffolds. To cross-validate these results, short

protein fragments were predicted directly from unassembled read data using FragGeneScanPlusPlus[56] 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[57] version 0.92 with default settings. The HMM for *rpoB* (2009 release) was downloaded from FunGene[58], and taxonomic classification was performed using the RefSeq database, release 80 (March 2017)[59]. Output of MetAnnotate was analyzed using the metannoviz R library, version 1.0.2 (https://github.com/metannotate/metannoviz) with an e-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[60] version 3.1b2 was used to search predicted proteins from the bins using HMMs downloaded from Pfam[61]. In particular, genes encoding the core Type I reaction center (*pscA*/*pshA*/*psaAB*; PF00223), a Type I reaction center-associated protein (*pscD*; PF10657), chlorosomes structural units (*csmAC*; PF02043, PF11098), and a chlorosome energy transfer protein (*fmoA*; PF02327), were queried. Detected homologs in the genome bins were assessed for their alignment to known RCI-associated genes encoded by phototrophic microorganisms belonging to other phyla. The tertiary structures of the detected homologs were also predicted using I-TASSER[62], and phylogenetic placement was assessed (see below). Custom HMMs were built for selected homologs using available input sequences. Primary sequences predicted from the genomes for each gene of interest were aligned using Clustal Omega[63] version 1.2.3, and HMMs were generated using hmmbuild[60] version 3.1b2. Custom HMMs and homology models generated by I-TASSER are available in the code repository associated with this work.

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[4], 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 *Chloroflexaceae* family[15], the '*Ca.* Thermofonsia' order[30], and the '*Ca.* Roseiliniales' order[64]). Selected genomes of known phototrophs that were deposited in other genome databases were also downloaded (see details in the code repository associated with this work). In addition, any genome bins listed in the GTDB that belonged to the '54-19' order (which represented the name of the '*Ca.* Chloroheliales' order in this database) were downloaded. Non-phototrophic lineages were subsequently pruned to one representative per genus, with the exception of non-phototrophic lineages closely related to phototrophic clades. In total, this left 58 genomes, including genomes of 28 known phototrophs. A concatenated core protein phylogeny was constructed for this genome collection using GToTree[65] version 1.4.11. The 'Bacteria.hmm' collection of 74 single-copy marker genes was used, with a minimum of 30% of the genes required in each genome in order for the genome to be kept in the final phylogeny. IQ-TREE[66] version 1.6.9 was used to build the phylogeny from the masked and concatenated multiple sequence alignment. ModelFinder[67] was used to determine the optimal evolutionary model for phylogeny building. Branch supports were calculated using 1000 ultrafast bootstraps[68].

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[69] and Bryant and colleagues[5]. Reference sequences for these genes were selected from genomes of well-studied representatives of the phylum, *Chloroflexus aurantiacus*[70], *Oscillochloris trichoides*[71], and *Roseiflexus castenholzii*[72]. Bidirectional BLASTP[73] was performed on these reference sequences against the entire *Chloroflexota* genome collection (above) to detect potential orthologs. The BackBLAST pipeline[74], version 2.0.0-alpha3

470 (doi:10.5281/zenodo.3697265), was used for bidirectional BLASTP, and cutoffs for the *e* value, percent identity, and query coverage of hits were empirically optimized to 10^{-3}, 20%, and 50%, respectively. Genes involved in iron oxidation or reduction were also searched using FeGenie[75], commit 30174bb, with default settings.

475 *Phylogenetic assessment of photosynthesis-associated genes*

Selected photosynthesis-associated genes were tested against genes from other phyla to examine the phylogenetic relationship of the novel '*Ca.* Chloroheliales' sequences to those of other phototrophic organisms. Genes for RCI/PSI (*pscA*/*pshA*/*psaAB*)[76], RCI-chlorosome attachment (*fmoA*)[22], chlorosome structure (*csmA*)[25], (bacterio)chlorophyll synthesis (*bchIDH*/*chlIDH* and *bchLNB*/*chlLNB*/*bchXYZ*)[5],
480 and RuBisCo (*cbbL*)[77], were examined. Genomes potentially containing these genes of interest were determined using a combination of automated detection via AnnoTree[78] and descriptions in the literature. Annotree was also used to summarize the GTDB reference tree for downstream data visualization. Representative genomes from this initial genome set were selected from the GTDB, based on genome quality and taxonomic novelty, before being downloaded from NCBI. Potential
485 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 3) via BackBLAST[74] version 2.0.0-alpha3. In addition, for some sequence sets (i.e., Type I reaction centers; Bch proteins; CbbL), additional reference sequences were manually added based on literature references[5,76,77]. Sequence sets were then pruned to remove closely related
490 sequences. Identified primary sequences were aligned using Clustal Omega[63] 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[79] 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 1). Maximum likelihood protein phylogenies were built from masked sequence

alignments using IQ-TREE[66] version 1.6.9. Evolutionary rate model selection was performed using ModelFinder[67], and 1000 rapid bootstraps were used to calculate branch support values[68]. Rate models used are summarized in Extended Data Table 1.

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.

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).

References

31. Schiff, S. L. *et al.* Millions of Boreal Shield lakes can be used to probe Archaean Ocean biogeochemistry. *Sci Rep* **7**, 46708 (2017).

32. 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).

33. 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).

34. 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).

35. Stookey, L. L. Ferrozine - a new spectrophotometric reagent for iron. *Anal Chem* **42**, 779–781 (1970).

36. 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).

37. 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.

38. Gaisin, V. A. *et al.* '*Candidatus* Viridilinea mediisalina', a novel phototrophic Chloroflexi bacterium from a Siberian soda lake. *FEMS Microbiol Lett* **366**, (2019).

39. 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.

40. 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.

41. Quince, C., Lanzen, A., Davenport, R. J. & Turnbaugh, P. J. Removing noise from pyrosequenced amplicons. *BMC Bioinform* **12**, 38 (2011).

42. 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).

43. 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).

44. Cavaco, M. A. *et al.* Freshwater microbial community diversity in a rapidly changing High Arctic watershed. *FEMS Microbiol Ecol* **95**, fiz161 (2019).

45. Bolyen, E. *et al.* Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* **37**, 852–857 (2019).

46. Lynch, M. D., Masella, A. P., Hall, M. W., Bartram, A. K. & Neufeld, J. D. AXIOME: automated exploration of microbial diversity. *GigaScience* **2**, 3–3 (2013).

47. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).

48. 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).

49. 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).

50. 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).

51. 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* gr.260380.119 (2020) doi:10.1101/gr.260380.119.

52. 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).

53. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).

54. 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).

55. 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).

56. Singh, R. G. *et al.* Unipept 4.0: functional analysis of metaproteome data. *J Proteome Res* **18**, 606–615 (2018).

57. 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).

58. Fish, J. A. *et al.* FunGene: the functional gene pipeline and repository. *Front Microbiol* **4**, 291 (2013).

59. 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).

60. Eddy, S. R. Accelerated profile HMM searches. *PLOS Comput Biol* **7**, e1002195 (2011).

61. Finn, R. D. *et al.* The Pfam protein families database: towards a more sustainable future. *Nucl. Acids Res.* **44**, D279–D285 (2016).

62. Yang, J. *et al.* The I-TASSER Suite: protein structure and function prediction. *Nat Methods* **12**, 7–8 (2015).

63. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* **7**, 539 (2011).

64. 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.

65. Lee, M. D. GToTree: a user-friendly workflow for phylogenomics. *Bioinformatics* **35**, 4162–4164 (2019).

66. 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).

67. 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).

68. 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).

69. Tang, K.-H. *et al.* Complete genome sequence of the filamentous anoxygenic phototrophic bacterium *Chloroflexus aurantiacus*. *BMC Genomics* **12**, 334 (2011).

70. 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).

71. Keppen, O. I., Baulina, O. I. & Kondratieva, E. N. *Oscillochloris trichoides* neotype strain DG-6. *Photosynth Res* **41**, 29–33 (1994).

72. 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).

73. 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).

74. 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).

75. 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).

76. Cardona, T. Early Archean origin of heterodimeric Photosystem I. *Heliyon* **4**, e00548 (2018).

77. 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).

78. Mendler, K. *et al.* AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res* **47**, 4442–4448 (2019).

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

Acknowledgments

We 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, and R. Henderson for assistance with lake sampling; 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; and N. Tran and K. Shimada for discussion and advice about biochemistry. 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.

Author contributions

S.L.S., J.D.N., and J.J.V. conceived the study. J.M.T., M.T., and N.A.S. 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 genomic DNA, and performed metagenome library preparation and sequencing. J.M.T. analyzed 16S rRNA gene and metagenome sequencing data, conducted phylogenetic analyses, and visualized the data. J.M.T., S.N., M.T., S.H., and J.D.N. interpreted the data and its impacts on understanding the evolution of photosynthesis. J.D.N., M.T., and S.H. supervised research. J.M.T. wrote the paper with J.D.N., N.A.S., S.N., and comments from all other authors. All authors have read, revised and approved the manuscript submission.

Ethics declarations

Competing interests

The authors declare no competing interests.

Supplementary Information | PDF file containing Supplementary Notes 1-4, Supplementary Methods, and Supplementary References.

Supplementary Data 1 | Amplicon sequencing variant table of early phototroph enrichment cultures from 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.

Supplementary Data 2 | 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 3 | Bidirectional BLASTP results for photosynthesis-related genes among the *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. 3.

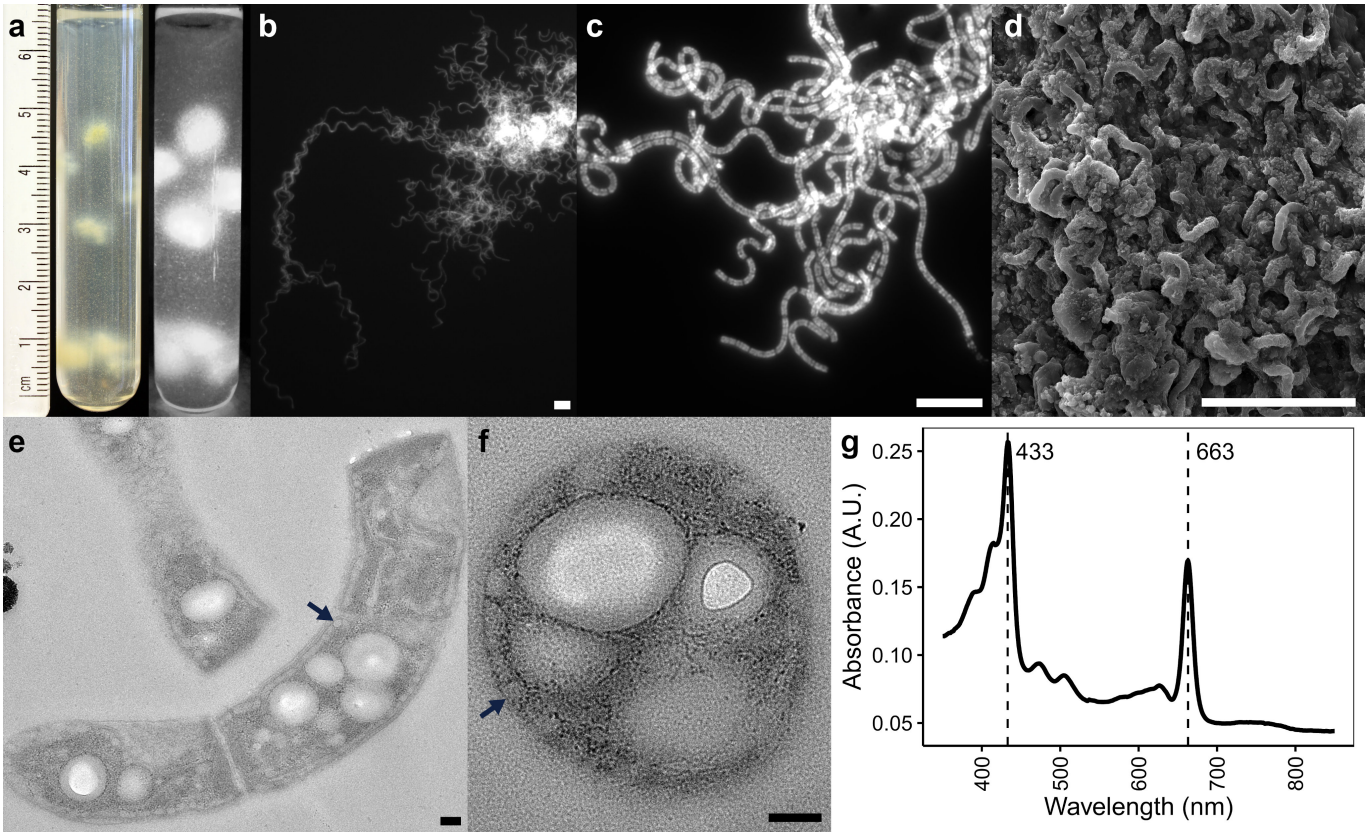


Fig. 1 | Physiology of the '*Ca.* Chx. allophototropha culture'. a, Example growth of the '*Ca.* Chx. allophototropha' enrichment in soft agar. The left panel is a photograph of a test tube with a ruler shown for scale, and the right panel shows the same tube when viewed under a pigment fluorescence detection system optimized for bacteriochlorophyll *c* (Supplementary Methods; Extended Data Fig. 9) **b-c**, Epifluorescence microscopy images of '*Ca.* Chx. allophototropha' cells. Cell autofluorescence is shown, with light excitation and detection wavelengths optimized for bacteriochlorophyll *c*. **d**, Scanning electron microscopy image of an aggregate of '*Ca.* Chx. allophototropha' cells. **e-f**, Transmission electron microscopy images showing longitudinal and cross sections, respectively, of '*Ca.* Chx. allophototropha' cells. Example chlorosomes are marked with arrows. **g**, Absorption spectrum (350-850 nm) of an acetone:methanol extract of pigments from the '*Ca.* Chx. allophototropha' culture. Major absorption peaks are marked by dashed lines. Scale bars on the top row of images (**b-d**) represent 10 μm, and scale bars on the bottom row of images (**e-f**) represent 0.1 μm.

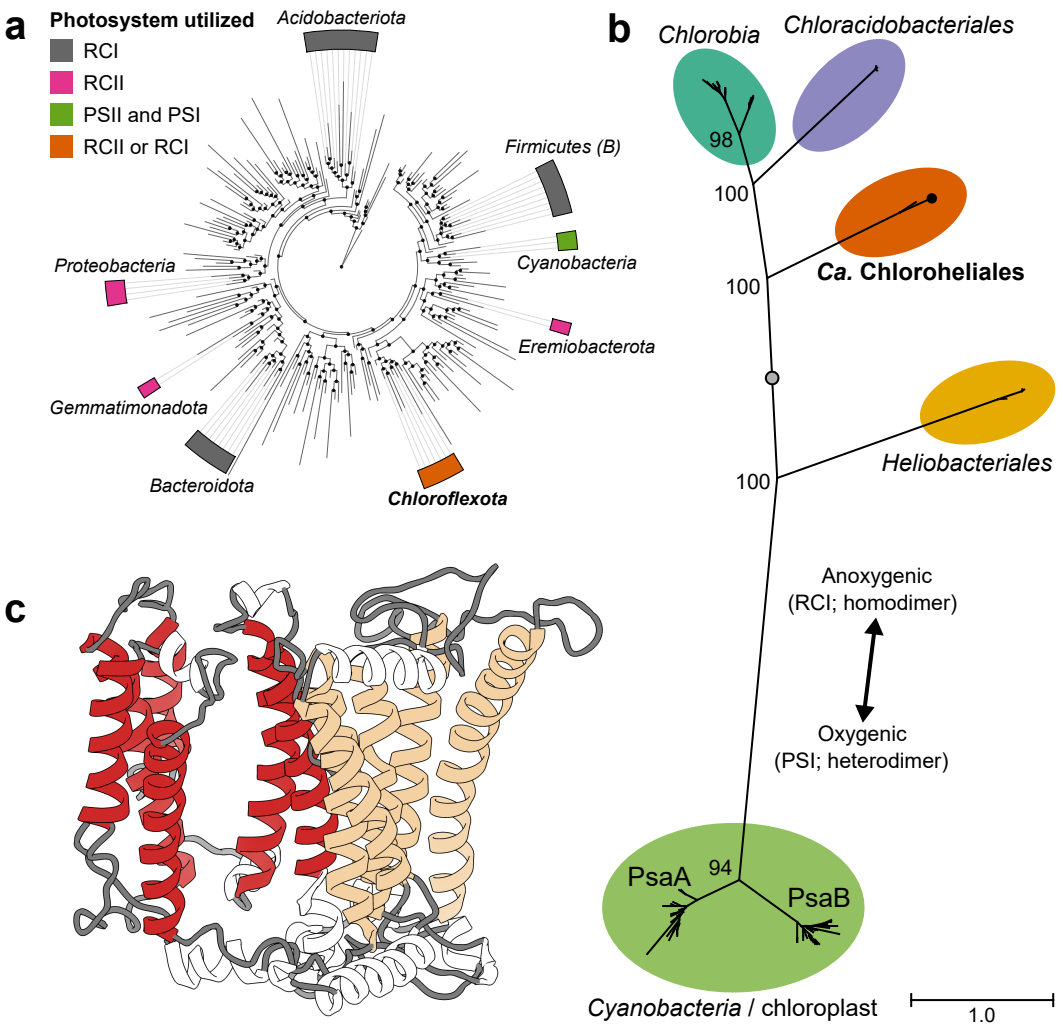


Fig. 2 | Functional novelty of the '*Ca.* Chlorohelix allophototropha' photosystem. a, Overview of the eight bacterial phyla containing known chlorophototrophs. A class-level summary of the Genome Tree Database (GTDB) bacterial reference tree (release 89) is shown, with chlorophototrophic phyla highlighted based on the photosynthetic reaction center utilized. **b**, Maximum likelihood phylogeny of Fe-S type reaction center primary sequences. Ultrafast bootstrap values for major branch points are shown out of 100 based on 1000 replicates, and the scale bar shows the expected proportion of amino acid change across the 548-residue masked sequence alignment (Extended Data Table 1). Chlorophototrophic lineages are summarized by order name. The placement of the PscA-like sequence of '*Ca*. Chx. allophototropha' is indicated by a black dot. All clades above they grey dot utilize chlorosomes. **c**, Predicted tertiary structure of the novel '*Ca*. Chlorohelix allophototropha' PscA-like primary sequence based on homology modelling. The six N-terminal and five C-terminal transmembrane helices expected for RCI are coloured in red and tan, respectively.

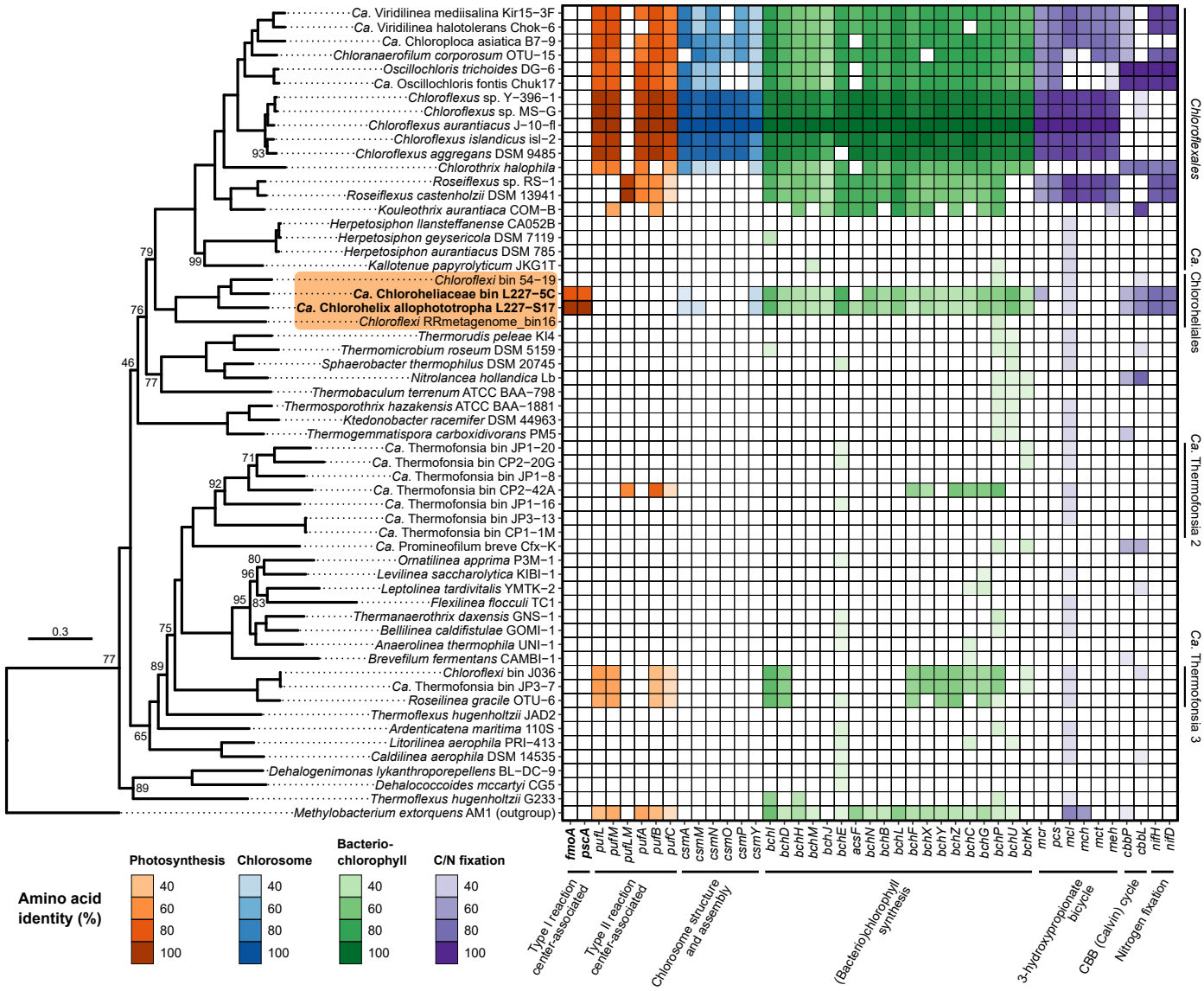
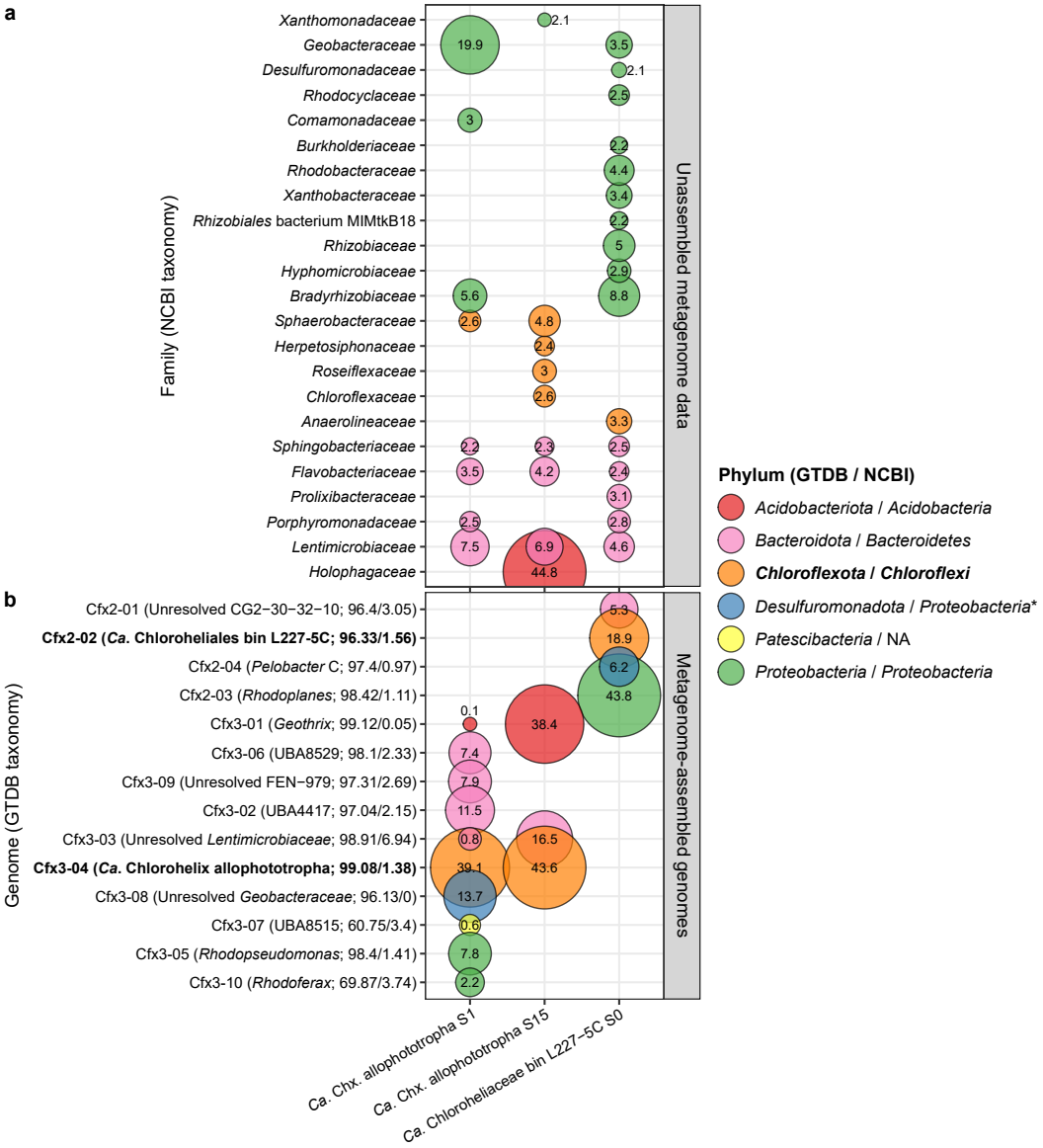
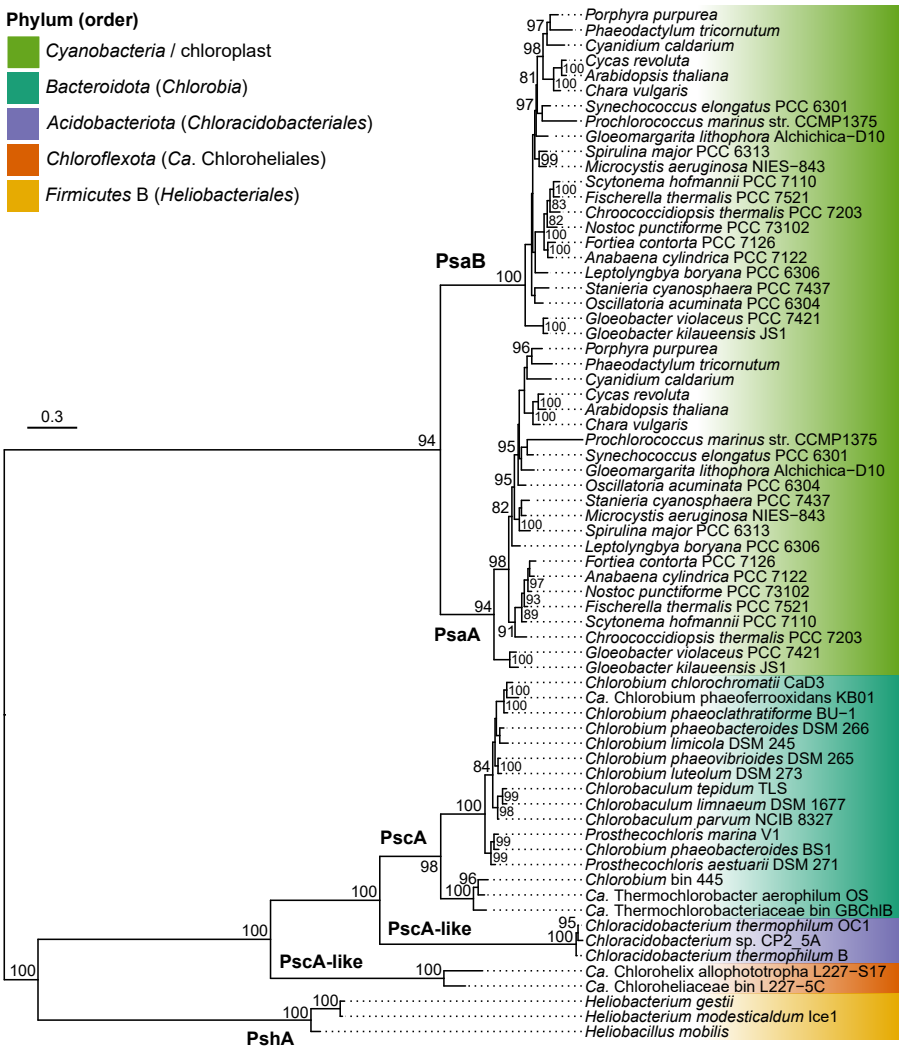


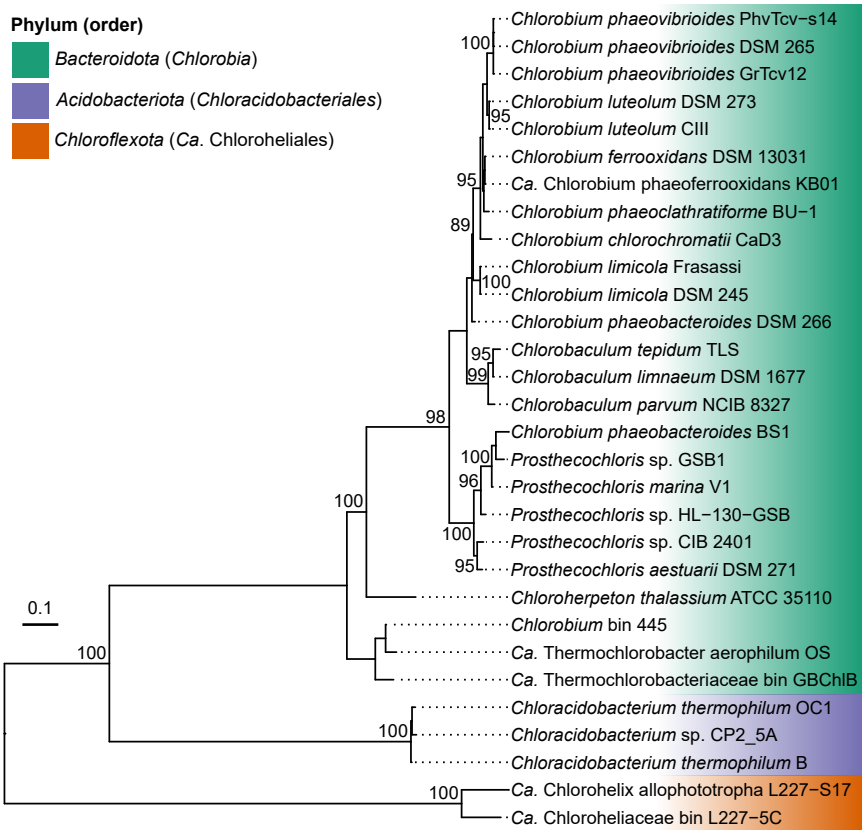
Fig. 3 | Genomic potential for phototrophy among members of the *Chloroflexota* phylum. A maximum likelihood phylogeny of key members of the phylum is shown on the left based on a set of 74 concatenated core bacterial proteins. Ultrafast bootstrap values are shown out of 100 based on 1000 replicates, and the scale bar shows the expected proportion of amino acid change across the 11 613-residue masked sequence alignment (Extended Data Table 1). The '*Ca.* Chloroheliales' clade is highlighted in orange. On the right, a heatmap shows the presence/absence of genes involved in photosynthesis or related processes based on bidirectional BLASTP. The intensity of the fill colour of each tile corresponds to the percent identity of the BLASTP hit compared to the query sequence. 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. Raw results of bidirectional BLASTP are summarized in Supplementary Data 3. Orders containing potential phototrophs are labelled on the right of the heatmap.



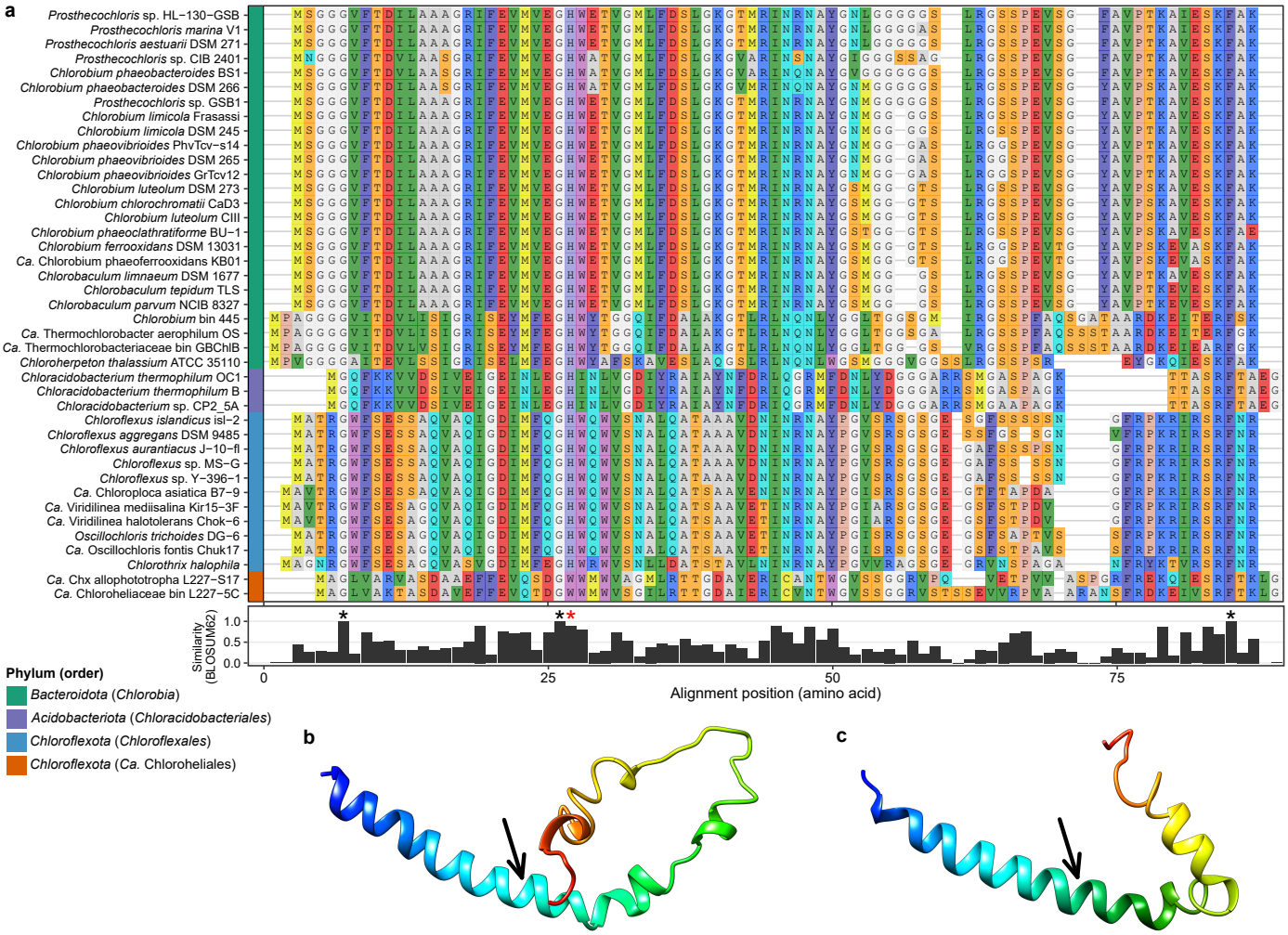
Extended Data Fig. 1 | Composition of the '*Ca.* Chloroheliales' enrichment cultures described in this study. a, Bubble plot showing the relative abundances of taxa within '*Ca.* Chloroheliales' enrichment culture metagenomes based on classification of partial RpoB fragments recovered from unassembled 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.* Chloroheliales bin L227-5C'. **b**, Bubble plot showing the relative abundances of uncurated genome bins recovered from the same metagenomes. Relative abundances are expressed as the percentage of reads mapped to the genome bin compared to the total number of assembled reads. Bubbles are coloured by phylum according to GTDB taxonomy and are shown for all entries greater than 0.05% relative abundance. Listed beside each genome bin name are the GTDB genus that the bin was classified to and the percent completeness and contamination of the bin according to CheckM. Statistics shown for the '*Ca.* Chx. allophototropha' bin differ slightly compared to Supplementary Note 2, which describes the statistics for the bin after manual bin curation.



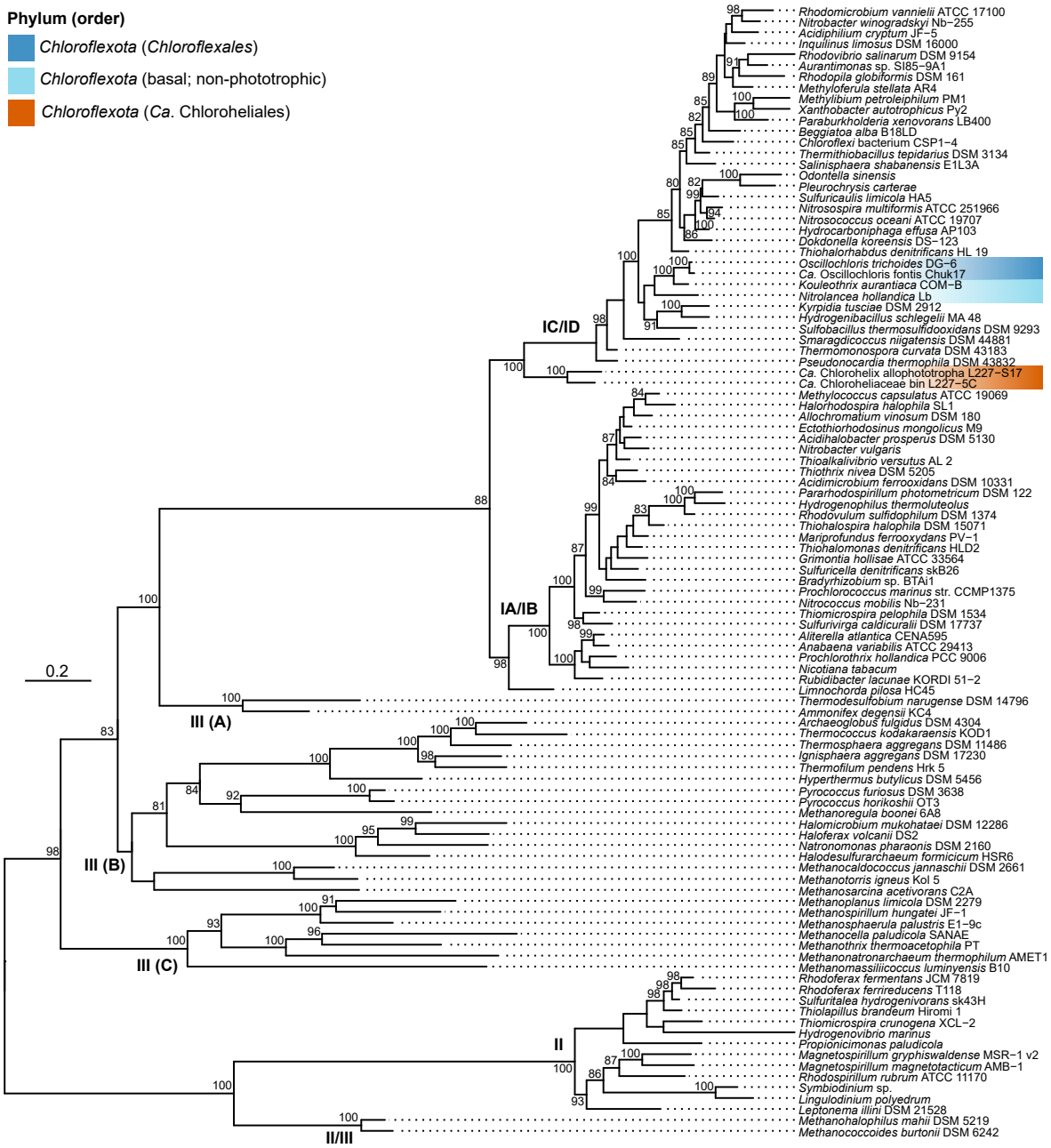
Extended Data Fig. 2 | Maximum likelihood phylogeny of oxygenic and anoxygenic Type I reaction center predicted protein sequences. A simplified depiction of the same phylogeny is shown in Fig. 2b. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale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



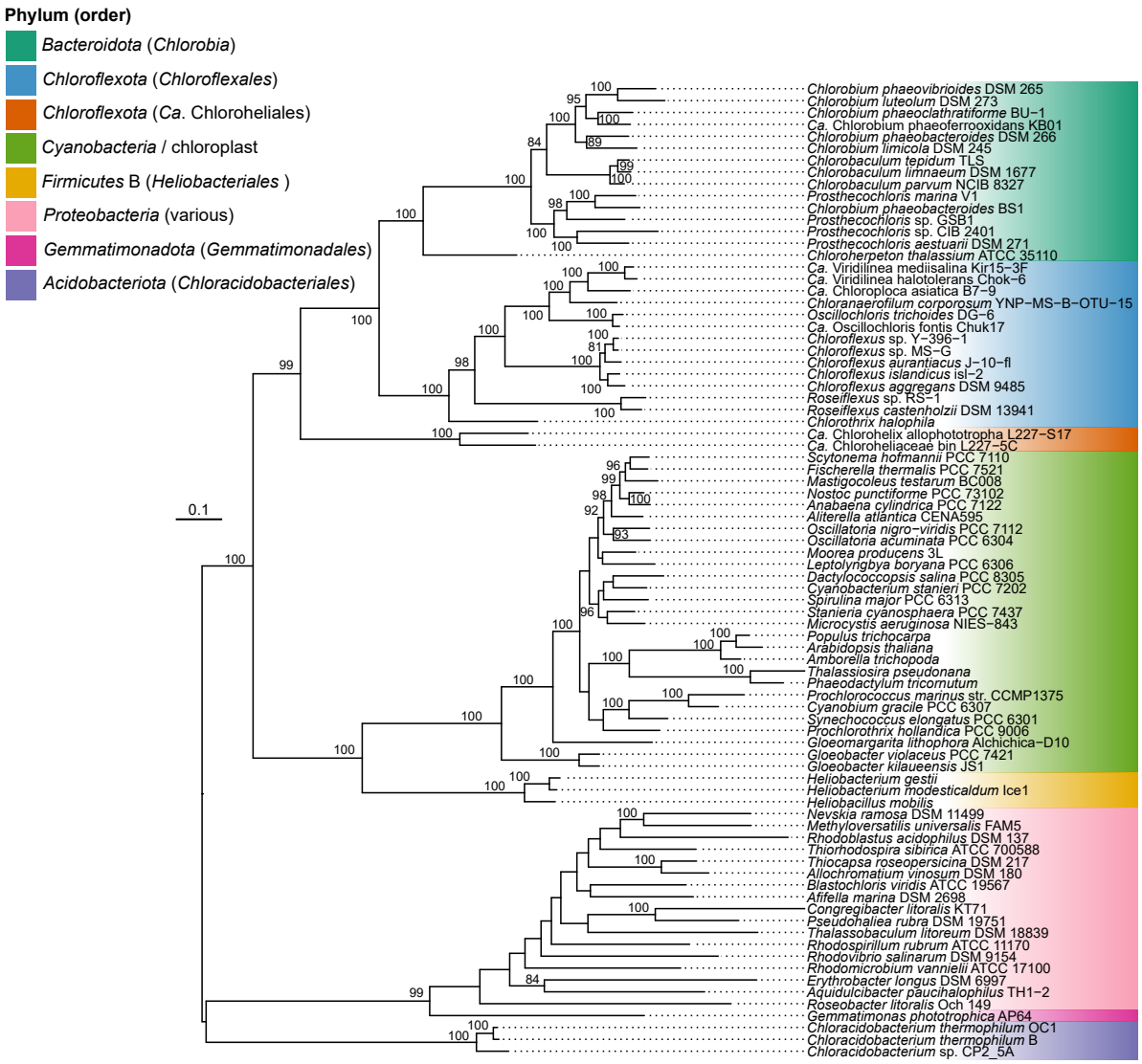
Extended Data Fig. 3 | Maximum likelihood phylogeny of Fenna-Matthews-Olson (FMO) protein (FmoA) sequences. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 356-residue masked sequence alignment (Extended Data Table 1).



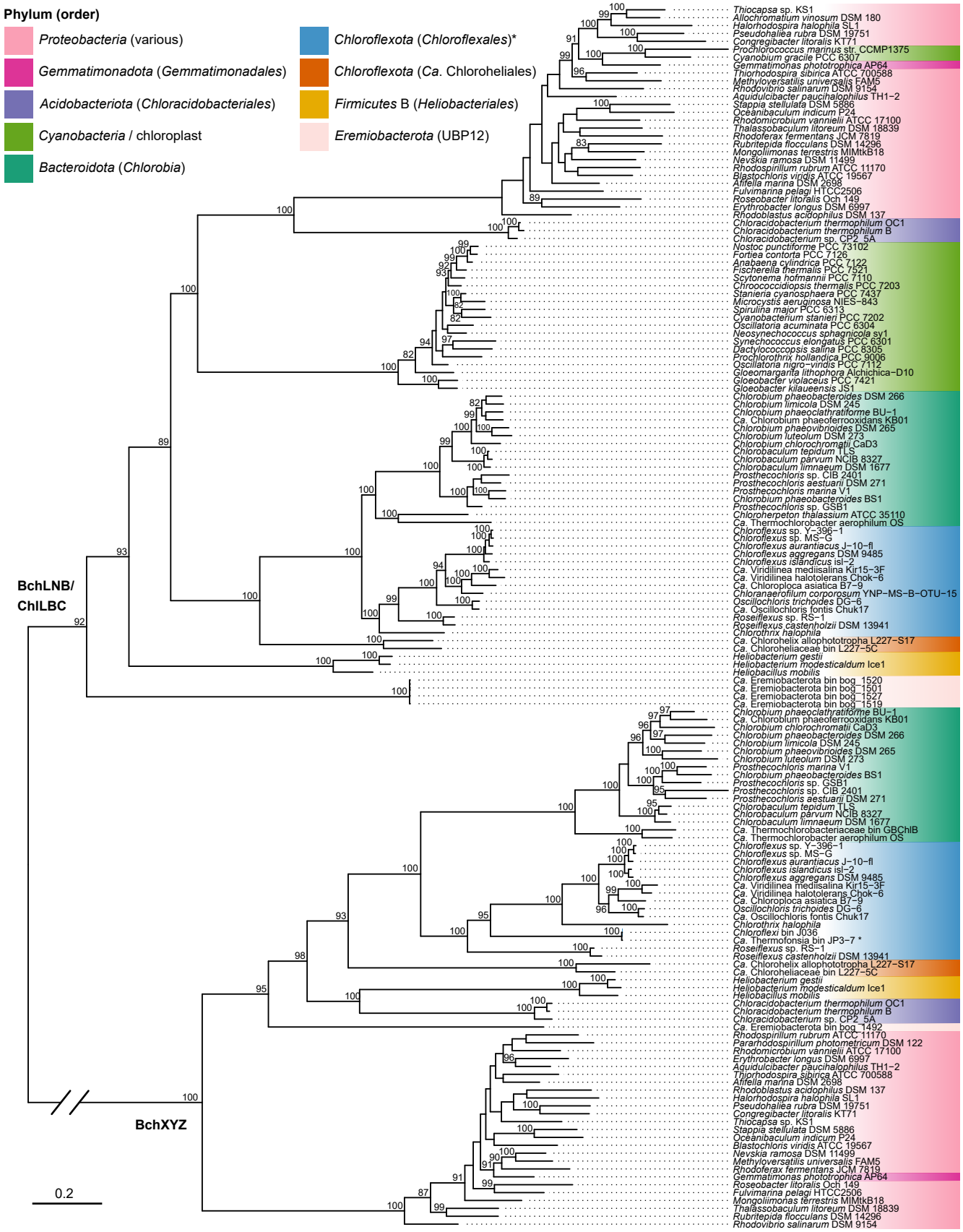
Extended Data Fig. 4 | Structural properties of the predicted chlorosome protein CsmA in '*Ca.* Chx. allophototropha'. a, Multiple sequence alignment of the CsmA primary sequence including representatives from known chlorosome-containing phyla. The entire CsmA primary sequence is shown, and a colour 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. 6. The BLOSUM62 sequence similarity score is shown underneath the alignment. Residues that are conserved across all sequences are marked with a black asterisk, and the H25 histidine residue predicted to be involved in bacteriochlorophyll *a* binding is marked with a red asterisk. **b**, Homology model showing the possible tertiary structure of the '*Ca.* Chx. allophototropha' CsmA protein as predicted by I-TASSER. The tryptophan residue corresponding to the typical H25 residue is indicated by an arrow. **c**, Tertiary structure of the CsmA protein from *Chlorobium tepidum* (PDB accession 2K37); the H25 site is marked with an arrow.



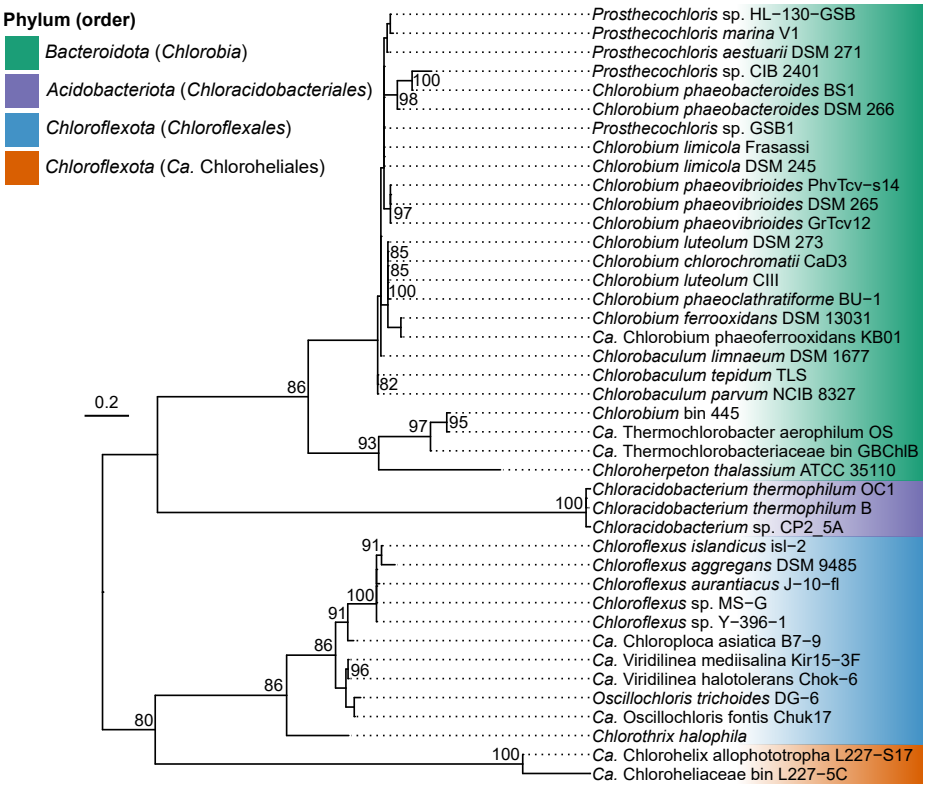
Extended Data Fig. 5 | Maximum likelihood phylogeny of RuBisCO large subunit (CbbL) predicted protein sequences. Group I to III CbbL sequences are shown, and group IV sequences, which do not form proteins involved in carbon fixation, are omitted for conciseness. The phylogeny is midpoint rooted, and 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 1).



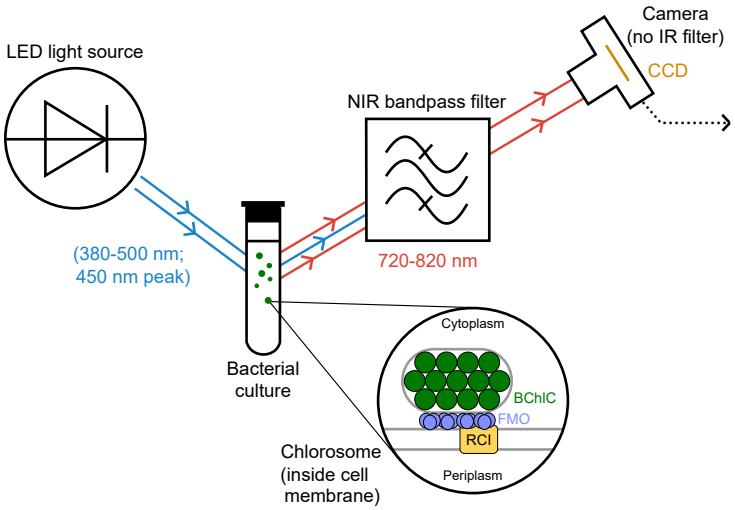
Extended Data Fig. 6 | Maximum likelihood phylogeny of BchIDH/ChlIDH protein sequences. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 1702-residue masked sequence alignment (Extended Data Table 1). No sequences associated with the *Eremiobacterota* phylum are shown, because the complete BchIDH gene set could not be detected in any of the genome bins available from that phylum using bidirectional BLASTP.


Phylum (order)
Proteobacteria (various)
Gemmatimonadota (Gemmatimonadales)
Acidobacteriota (Chloracidobacteriales)
Cyanobacteria / chloroplast
Bacteroidota (Chlorobia)
Chloroflexota (Chloroflexales)*
Chloroflexota (Ca. Chloroheliales)
Firmicutes B (Heliobacteriales)
Eremiobacterota (UBP12)
BchLNB/
ChlLBC
BchXYZ
0.2
Thiocapsa sp. KS1
Allochromatium vinosum DSM 180
Halorhodospira halophila SL1
Pseudohaliea rubra DSM 19751
Congregibacter litoralis KT71
Prochlorococcus marinus str. CCMP1375
Cyanobium gracile PCC 6307
Gemmatimonas phototrophica AP64
Thiorhodospira sibirica ATCC 700588
Methyloversatilis universalis FAM5
Rhodovibrio salinarum DSM 9154
Aquidulcibacter paucihalophilus TH1−2
Stappia stellulata DSM 5886
Oceanibaculum indicum P24
Rhodomicrobium vannielii ATCC 17100
Thalassobaculum litoreum DSM 18839
Rhodoferax fermentans JCM 7819
Rubritepida flocculans DSM 14296
Mongoliimonas terrestris MIMtkB18
Nevskia ramosa DSM 11499
Rhodospirillum rubrum ATCC 11170
Blastochloris viridis ATCC 19567
Afifella marina DSM 2698
Fulvimarina pelagi HTCC2506
Roseobacter litoralis Och 149
Erythrobacter longus DSM 6997
Rhodoblastus acidophilus DSM 137
Chloracidobacterium thermophilum OC1
Chloracidobacterium thermophilum B
Chloracidobacterium sp. CP2_5A
Nostoc punctiforme PCC 73102
Fortiea contorta PCC 7126
Anabaena cylindrica PCC 7122
Fischerella thermalis PCC 7521
Scytonema hofmannii PCC 7110
Chroococcidiopsis thermalis PCC 7203
Stanieria cyanosphaera PCC 7437
Microcystis aeruginosa NIES−843
Spirulina major PCC 6313
Cyanobacterium stanieri PCC 7202
Oscillatoria acuminata PCC 6304
Neosynechococcus sphagnicola sy1
Synechococcus elongatus PCC 6301
Dactylococcopsis salina PCC 8305
Prochlorothrix hollandica PCC 9006
Oscillatoria nigro−viridis PCC 7112
Gloeomargarita lithophora Alchichica−D10
Gloeobacter violaceus PCC 7421
Gloeobacter kilaueensis JS1
Chlorobium phaeobacteroides DSM 266
Chlorobium limicola DSM 245
Chlorobium phaeoclathratiforme BU−1
Ca. Chlorobium phaeoferrooxidans KB01
Chlorobium phaeovibrioides DSM 265
Chlorobium luteolum DSM 273
Chlorobium chlorochromatii CaD3
Chlorobaculum tepidum TLS
Chlorobaculum parvum NCIB 8327
Chlorobaculum limnaeum DSM 1677
Prosthecochloris sp. CIB 2401
Prosthecochloris aestuarii DSM 271
Prosthecochloris marina V1
Chlorobium phaeobacteroides BS1
Prosthecochloris sp. GSB1
Chlorotherpeton thalassium ATCC 35110
Ca. Thermochlorobacter aerophilum OS
Chloroflexus sp. Y−396−1
Chloroflexus sp. MS−G
Chloroflexus aurantiacus J−10−fl
Chloroflexus aggregans DSM 9485
Chloroflexus islandicus isl−2
Ca. Viridilinea mediisalina Kir15−3F
Ca. Viridilinea halotolerans Chok−6
Ca. Chloroploca asiatica B7−9
Chloranaerofilum corporosum YNP−MS−B−OTU−15
Oscillochloris trichoides DG−6
Ca. Oscillochloris fontis Chuk17
Roseiflexus sp. RS−1
Roseiflexus castenholzii DSM 13941
Chlorothrix halophila
Ca. Chlorohelix allophototropha L227−S17
Ca. Chloroheliaceae bin L227−5C
Heliobacterium gestii
Heliobacterium modesticaldum Ice1
Heliobacillus mobilis
Ca. Eremiobacterota bin bog_1520
Ca. Eremiobacterota bin bog_1501
Ca. Eremiobacterota bin bog_1527
Ca. Eremiobacterota bin bog_1519
Chlorobium phaeoclathratiforme BU−1
Ca. Chloroblum phaeoferrooxidans KB01
Chlorobium chlorochromatii CaD3
Chlorobium phaeobacteroides DSM 266
Chlorobium limicola DSM 245
Chlorobium phaeovibrioides DSM 265
Chlorobium luteolum DSM 273
Prosthecochloris marina V1
Chlorobium phaeobacteroides BS1
Prosthecochloris sp. GSB1
Prosthecochloris sp. CIB 2401
Prosthecochloris aestuarii DSM 271
Chlorobaculum tepidum TLS
Chlorobaculum parvum NCIB 8327
Chlorobaculum limnaeum DSM 1677
Ca. Thermochlorobacteriaceae bin GBChlB
Ca. Thermochlorobacter aerophilum OS
Chloroflexus sp. Y−396−1
Chloroflexus sp. MS−G
Chloroflexus aurantiacus J−10−fl
Chloroflexus islandicus isl−2
Chloroflexus aggregans DSM 9485
Ca. Viridilinea mediisalina Kir15−3F
Ca. Viridilinea halotolerans Chok−6
Ca. Chloroploca asiatica B7−9
Oscillochloris trichoides DG−6
Ca. Oscillochloris fontis Chuk17
Chlorothrix halophila
Chloroflexi bin J036
Ca. Thermofonsia bin JP3−7 *
Roseiflexus sp. RS−1
Roseiflexus castenholzii DSM 13941
Ca. Chlorohelix allophototropha L227−S17
Ca. Chloroheliaceae bin L227−5C
Heliobacterium gestii
Heliobacterium modesticaldum Ice1
Heliobacillus mobilis
Chloracidobacterium thermophilum OC1
Chloracidobacterium thermophilum B
Chloracidobacterium sp. CP2_5A
Ca. Eremiobacterota bin bog_1492
Rhodospirillum rubrum ATCC 11170
Pararhodospirillum photometricum DSM 122
Rhodomicrobium vannielii ATCC 17100
Erythrobacter longus DSM 6997
Aquidulcibacter paucihalophilus TH1−2
Thiorhodospira sibirica ATCC 700588
Afifella marina DSM 2698
Rhodoblastus acidophilus DSM 137
Halorhodospira halophila SL1
Pseudohaliea rubra DSM 19751
Congregibacter litoralis KT71
Thiocapsa sp. KS1
Stappia stellulata DSM 5886
Oceanibaculum indicum P24
Blastochloris viridis ATCC 19567
Nevskia ramosa DSM 11499
Methyloversatilis universalis FAM5
Rhodoferax fermentans JCM 7819
Gemmatimonas phototrophica AP64
Roseobacter litoralis Och 149
Fulvimarina pelagi HTCC2506
Mongoliimonas terrestris MIMtkB18
Thalassobaculum litoreum DSM 18839
Rubritepida flocculans DSM 14296
Rhodovibrio salinarum DSM 9154

Extended Data Fig. 7 | 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 1). One sequence among the *Chloroflexales* cluster for BchXYZ, associated with '*Ca.* Thermofonsia bin JP3-7', actually belongs to a genome bin that places in the basal clades of the *Chloroflexota* phylum (see Fig. 3) and is thus marked with an asterisk.



Extended Data Fig. 8 | Maximum likelihood phylogeny of CsmA predicted protein sequences. The phylogeny is midpoint rooted, and ultrafast bootstrap values of at least 80/100 are shown. The scale bar represents the expected proportion of amino acid change across the 74-residue masked sequence alignment (Extended Data Table 1).



Extended Data Fig. 9 | 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 Fig. 1a. 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 | Properties of sequence alignments and phylogenetic trees presented in this study.

Protein set	Alignment length (unmasked)	Alignment length (masked)	Evolutionary rate model used
Core proteins (*Chloroflexota* phylum)	NA	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