

# Microbial mats in the Turks and Caicos Islands reveal diversity and evolution of phototrophy in the Chloroflexota order *Aggregatilineales*

Lewis M. Ward<sup>1</sup>, Usha F. Lingappa<sup>2</sup>, John P. Grotzinger<sup>2</sup>, Woodward W. Fischer<sup>2</sup>

*1. Department of Earth & Planetary Sciences, Harvard University, Cambridge MA 02138.*

*2. Division of Geological & Planetary Sciences, California Institute of Technology, Pasadena, CA 91125.*

**Abstract:** Genome-resolved metagenomic sequencing approaches have led to a substantial increase in the recognized diversity of microorganisms; this included the discovery of novel metabolic pathways in previously recognized clades, and has enabled a more accurate determination of the extant distribution of key metabolisms and how they evolved over Earth history. Here, we present metagenome-assembled genomes of members of the Chloroflexota (formerly Chloroflexi or Green Nonsulfur Bacteria) order *Aggregatilineales* (formerly SBR1031 or *Thermofonsia*) discovered from sequencing of thick and expansive microbial mats present in an intertidal lagoon on Little Ambergris Cay in the Turks and Caicos Islands. These taxa included multiple new lineages of Type 2 reaction center-containing phototrophs that were not closely related to previously described phototrophic Chloroflexota—revealing a rich and intricate history of horizontal gene transfer and the evolution of phototrophy and other core metabolic pathways within this widespread phylum.

## **Main text:**

Most of the known diversity of phototrophic Chloroflexota (formerly Chloroflexi or Green Nonsulfur Bacteria) was derived from isolation- and sequencing-based efforts primarily on hot spring microbial mats (e.g. Thiel et al. 2018, Ward et al. 2018a). However microbial communities with diverse members of phototrophic Chloroflexota are commonly found in other environments, including coastal marine environments and carbonate platforms (e.g. Trembath-Reichert et al. 2016). While cultivation-based efforts have yet to isolate phototrophic Chloroflexota from outside of the Chloroflexia class in pure culture (Thiel et al. 2018), genome-resolved metagenomic sequencing has made substantial progress in uncovering novel diversity of phototrophic lineages that have otherwise remained inaccessible and unknown (e.g. Bryant et al. 2008, Klatt et al. 2011, Ward et al. 2018a, Ward et al. 2019d). In order to better characterize the diversity and evolutionary histories of phototrophic Chloroflexota, we recovered five metagenome-assembled genomes (MAGs) from genome-resolved metagenomic sequencing of microbial mats from the Turks and Caicos Islands which include novel phototrophic lineages of Chloroflexota not closely related to previously described phototrophs.

The Turks and Caicos Islands occur at the southern end of the Bahamian archipelago. The Caicos platform (Fig 1) is a contiguous shallow (<5 m), grainy carbonate platform situated in the trade winds (mean wind velocity of 8 m/s from the east), and has a dry climate with net evaporation in excess of precipitation (Dravis and Wanless 2008). We studied microbial mats found throughout Little Ambergris Cay—a small ~6 km-long uninhabited island near the southern margin of the platform. Little Ambergris Cay contains a bedrock rim, formed by amalgamated cemented beach ridges and fossil eolian dunes, enclosing a tidal lagoon with a dozen large cut channels that communicate between the lagoon with well-mixed platformal waters. The typical diurnal tidal range is ~0.3 m. Polygonal microbial mats occur within the lagoon amongst sparse black mangroves. These mats are inundated daily by high tides. The mats are dissected into individual decimeter-sized heads that take on an upward domed shape by

mature polygons with 60° angles created by episodic desiccation. On average twice a decade, tropical storms transport ooid sediment forming across the platform into the lagoon and onto the mats (Trower et al. 2018)—creating a mode of lamination in addition to that created by microbial succession. Dominant mat building taxa are a thickly-sheathed, heterocystous cyanobacterium of the genus *Scytonema* and the thin-walled cyanobacterium, *Halomicronema* (Trembath-Reichert et al. 2016). The mats host a rich sulfur cycling community (Gomes et al., in press). Deeper layers of the mat contain abundant purple/red- and green- colored microbes visible in exposed cross sections off the mats (Figure 1); these have been confirmed by 16S rRNA gene amplicon sequencing to include diverse and abundant phototrophs in the Proteobacteria and Chloroflexota phyla (Trembath-Reichert et al. 2016).

Methods for metagenomic sequencing and genome binning followed those published previously (Ward et al. 2019a, Ward et al. 2019b) and described briefly here. Samples of microbial mat were collected using an ethanol-sterilized spatula (~0.25 cm<sup>3</sup> of material per sample). Immediately after sampling, cells were lysed and DNA preserved with a Zymo Terralyzer BashingBead Matrix and Xpedition Lysis Buffer. Lysis was achieved by attaching tubes to the blade of a cordless reciprocating saw (Black & Decker, Towson, MD) and operating for 1 minute. Following return to the lab, bulk environmental DNA was extracted and purified with a Zymo Soil/Fecal DNA extraction kit. Purified DNA was submitted to SeqMatic LLC (Fremont, CA) for library preparation and sequencing via Illumina NextSeq.

Raw sequence reads from four samples were co-assembled with MegaHit v. 1.02 (Li et al. 2015) and genome bins constructed based on nucleotide composition and differential coverage using MetaBAT (Kang et al. 2015), MaxBin (Wu et al. 2014), and CONCOCT (Alneberg et al. 2013) prior to dereplication and refinement with DAS Tool (Sieber et al. 2018) to produce the final bin set. Annotation was performed using RAST (Aziz et al. 2008). Genome completeness and redundancy/contamination was estimated with CheckM (Parks et al. 2015), and likelihood of presence or absence of metabolic pathways was estimated with MetaPOAP (Ward et al. 2018b). Visualization of the presence of annotated metabolic pathways was done via KEGG-decoder (Graham et al., 2018) following annotation of proteins sequences by GhostKOALA (Kanehisa et al., 2016). Taxonomic assignments were verified with GTDB-Tk (Parks et al. 2018, Chaumeil et al. 2019).

Protein sequences used in analyses (see below) were identified locally with the *tblastn* function of BLAST+(Camacho et al. 2009), aligned with MUSCLE (Edgar 2004), and manually curated in Jalview (Waterhouse 2009). Positive BLAST hits were considered to be full length (e.g. >90% the shortest reference sequence from an isolate genome) with *e*-values better than 1e<sup>-20</sup>. Phylogenetic trees were calculated using RAxML (Stamatakis 2014) on the Cipres science gateway (Miller et al. 2010). Transfer bootstrap support values were calculated by BOOSTER (Lemoine et al. 2018), and trees were visualized with the Interactive Tree of Life viewer (Letunic and Bork 2016). Concatenated ribosomal protein alignments were built following methods from Hug et al. (2016). Evolutionary histories of vertical versus horizontal inheritance of metabolic genes were inferred by comparison of the topologies of organismal and metabolic protein phylogenies (Doolittle 1986, Ward et al. 2018a, Ward et al. 2019c, Ward et al. 2019d).

We recovered five genomes that were taxonomically classified by GTDB-Tk within a lineage of the Chloroflexota phylum currently annotated as order SBR1031 (Table 1, Figure 2). Other genomes previously assigned to this clade included those proposed as the class *Candidatus* Thermofonsia (Ward et al. 2018a) along with an isolate for which the order Aggregatilineales was recently proposed (Nakahara et al. 2019). Following the isolation and characterization of

*Aggregatilinea lenta* (Nakahara et al. 2019), and the clustering of these organisms into an order-level clade by GTDB-Tk, we propose the reassignment of the organisms previously described as “*Candidatus Thermofonsia*” into the order Aggregatilineales. This order is primarily made up of nonphototrophic organisms but also contains some members with full suites of phototrophy genes, which appear to have been derived via horizontal gene transfer from members of the Chloroflexia class of Chloroflexota (Ward et al., 2018a).

The environmental range exhibited by known members of the SBR1031/Aggregatilineales was previously somewhat limited. Previously recovered genomes of members of this group were sourced from hot spring environments (Ward et al., 2018a, Ward et al., 2019), with the single known isolate isolated from seafloor sediments (Nakahara et al. 2019). 16S rRNA sequences of SBR1031 have been recovered from more diverse environments including hot springs (Ward et al., 2017, Ward 2017), contaminated soils (Milton et al. 2010), and wastewater (Björnsson et al. 2002). Recovery of MAGs belonging to this order from carbonate tidal flats therefore expands the available genomic diversity and known range of Aggregatilineales to environments that also have an extensive geological record.

The SBR1031 MAGs reported here encoded similar sets of functional genes to previously reported members of this order (Figure 3). Like previously described members of SBR1031/Aggregatilineales (e.g. Ward et al., 2018a), these organisms encode aerobic respiration via an A-family heme copper O<sub>2</sub> reductase, and contain both a *bc* complex and an alternative complex III (Ward et al. 2018a); based on these electron transport chain complexes, it is likely that these organisms are at least facultatively aerobic. All genomes described here (except for TC\_22, the least complete genome) also encode a *bd* oxidase (O<sub>2</sub> reductase) capable of functioning for O<sub>2</sub> detoxification or respiration at low O<sub>2</sub> concentrations (Borisov et al. 2011)—a trait observed in both aerobic and anaerobic members of the Anaerolineae class (e.g. Hemp et al. 2015, Pace et al. 2015, Ward et al. 2015a, Ward et al. 2018a).

Of the five SBR1031/Aggregatilineales genomes reported here, three encode partial or full components necessary for phototrophic energy transduction via a Type 2 reaction center. TC\_71 and TC\_152 encode complete sets of marker genes for phototrophy, including those encoding PufL and PufM subunits of the reaction center and bacteriochlorophyll synthesis (e.g. BchX, BchY, and BchZ). MetaPOAP False Positive estimates for phototrophy in these organisms were low (<0.04), suggesting that it is very unlikely that these genes were recovered as a result of contamination in the MAGs. The TC\_22 genome encodes the BchXYZ complex but did not recover genes for PufL or PufM; MetaPOAP False Positive and False Negative estimates were similarly low (~0.04) for this genome, and based on this analysis it remains unclear whether or not this organism contains a complete set of genes for phototrophy. However, the gene cluster encoding BchX, BchY, and BchZ in TC\_22 is located on the end of a contig, and the region of the chromosome syntenous to that encoding PufL and PufM in other phototrophic Chloroflexota (e.g. 6.5kb upstream of *bchx*, *bchy*, and *bchz* in TC\_152) is missing in the TC\_22 genome. Based on this, it is possible that this organism hosts a complete phototrophy pathway but that some genes simply were not recovered in the MAG. Like previously described phototrophs in SBR1031, these organisms do not encode a BchLNB complex or the capacity for carbon fixation via either the 3-hydroxypropionate bi-cycle or the Calvin cycle (Ward et al. 2018a). It is worth noting, however, that TC\_22 does encode a Form IV rubisco-like protein on a small contig; however enzymes in this family are not capable of catalyzing CO<sub>2</sub> fixation and instead are used for a variety of other functions (Tabita et al., 2008) and this genome does not encode phosphoribulose kinase.

Comparisons of organismal (based on concatenated ribosomal proteins and other vertically inherited markers) and phototrophy protein (e.g. PufL, PufM) phylogenies indicated substantial incongruences between organismal and phototrophy tree topologies. These relationships are indicative of a history of horizontal gene transfer (Figure 2b) (e.g. Raymond et al. 2002). In particular, the reaction centers found in members of SBR1031 branch with those of *Roseiflexus* rather than as a clade separate from those of Chloroflexia (e.g. *Chloroflexus*+*Roseiflexus*), suggesting that horizontal transfer of phototrophy proteins occurred from the *Roseiflexus* branch to members of SBR1031—a pattern previously recognized in other members of the SBR1031/Aggegatilineales (Ward et al. 2018a). Interestingly, like *Roseiflexus* and some other phototrophic members of SBR1031/Aggegatilineales (Ward et al. 2018a), TC\_71 and TC\_152 encode fused *pufL/pufM* genes encoding the two subunits of the Type 2 reaction center heterodimer. Together with reaction center protein phylogenetic relationships, this observation indicated that the reaction centers of these organisms are more closely related to those of the *Roseiflexus* lineage of Chloroflexia than to the more closely related *Ca. Roseilinea gracile*, which encodes unfused *pufL* and *pufM* genes. A corollary of these observations is that phototrophy in SBR1031 must postdate the acquisition and diversification of phototrophy in the Chloroflexia, events that have been estimated to have occurred in the last ~1 billion years (Shih et al. 2017).

While horizontal gene transfer can be confidently determined to have been responsible for the presence of phototrophy in SBR1031, it remains unclear to what extent intra-order horizontal gene transfer played a role in the extant distribution of phototrophy within the clade. Phototrophic lineages of SBR1031 show a polyphyletic distribution, separated by many nonphototrophic lineages (Figure 2a). This distribution could be reasonably explained by the presence of phototrophy in the last common ancestor of SBR1031 followed by extensive loss in most lineages; however differences in the topology of phototrophy proteins and organismal phylogenies within SBR1031 may indicate later acquisition followed by multiple instances of horizontal gene transfer between members of the clade. Discovery and study of additional phototrophic members of SBR1031 will be valuable to confidently resolve phylogenetic relationships of SBR1031 phototrophy proteins to assess organismal relationships in this clade.

The expanded environmental distribution and genetic diversity of Chloroflexota phototrophs described here further reinforces that genome-resolved metagenomic sequencing can provide an effective avenue for discovering novel microbial diversity, particularly of hard-to-culture phototrophs (e.g. Bryant et al., 2007, Ward et al., 2017, Thiel et al., 2018, Holland-Moritz et al., 2019, Ward et al. 2019d). These data also reinforce hypotheses that horizontal gene transfer has been a major mechanism behind the extant distribution of anoxygenic phototrophy (e.g. Fischer et al., 2016, Ward et al., 2018, Ward et al., 2019d). Together with the mounting evidence that most microbial lineages that have ever lived are now extinct (Louca et al., 2018), the continuing lack of discovery of donor lineages for horizontal gene transfer of phototrophy leads toward a consistent hypothesis: most phototrophic lineages that have ever existed have gone extinct, but relatively frequent horizontal gene transfer has allowed phototrophy pathways to persist in new lineages (e.g. Ward and Shih 2019).

**Figure 1:** Geological context of the microbial mats from which the genomes in this study were recovered. A) location of the Turks and Caicos Islands. B) location of Little Ambergris Cay. C) microbial mats and mangroves in tidal flat. D) cross section of microbial mat showing pigmented layers containing phototrophic bacteria.

**Figure 2:** **A)** Concatenated ribosomal protein phylogeny of the Chloroflexota, focusing on order SBR1031 (Aggregatilineales). Organisms encoding phototrophy with a Type 2 reaction center highlighted in green (including TC\_22, which did not recover reaction center genes but may be a phototroph, as discussed in the text). Genomes first described here noted with pink circles. Strains are labelled with MAG IDs (this study) or NCBI WGS database IDs (others), followed by taxonomy as derived from GTDB-Tk. Clades not the focus of this study have been collapsed and labeled with GTDB-Tk taxonomy. **B)** Tanglegram showing phylogenetic (in)congruence between concatenated ribosomal proteins (left) reflecting organismal relationships with PufM (right) as a marker of the horizontal gene transfer of phototrophy proteins. Dotted lines show topological congruence within some lineages of Chloroflexia (in black) and incongruence between *Roseiflexus*, *Roseilinea*, and phototrophic members of SBR1031 (in red). This is indicative of horizontal gene transfer of phototrophy proteins from the *Roseiflexus* lineage to *Roseilinea* and SBR1031.

**Figure 3:** Heatmap of metabolic functions of Aggregatilineales genomes produced by KEGG-decoder. Genome IDs as used in this study (TC##) or as WGS identifiers (all others). Note that the apparent presence of Calvin cycle genes in TC\_22 and other members of Aggregatilineales is due to the presence of a Form IV rubisco-like protein that does not catalyze CO<sub>2</sub> fixation, as described in the text.

**Declarations:**

**Ethics approval and consent to participate:**

Not applicable

**Consent for publication:**

Not applicable

**Availability of data and materials:**

The datasets generated during and analysed during the current study are available in the NCBI WGS repository under project ID PRJNA602167.

**Competing interests:**

The authors declare that they have no competing interests

**Funding:**

This work was made possible with support from the Agouron Institute, NSF IOS project # 1833247, and the Caltech Center for Environment-Microbe Interactions. LMW acknowledges support from an Agouron Institute Postdoctoral Fellowship and the Simons Foundation Collaboration on Marine Microbial Ecology. UFL was supported by an NSF Graduate Research Fellowship.

**Authors contributions:**

UFL, JPG, and WWF performed field work and collected samples. UFL and LMW processed and analyzed data. LMW wrote the manuscript with assistance from UFL, JPG, and WWF. All authors read and approved the final manuscript.

## Acknowledgements:

Not applicable

## References:

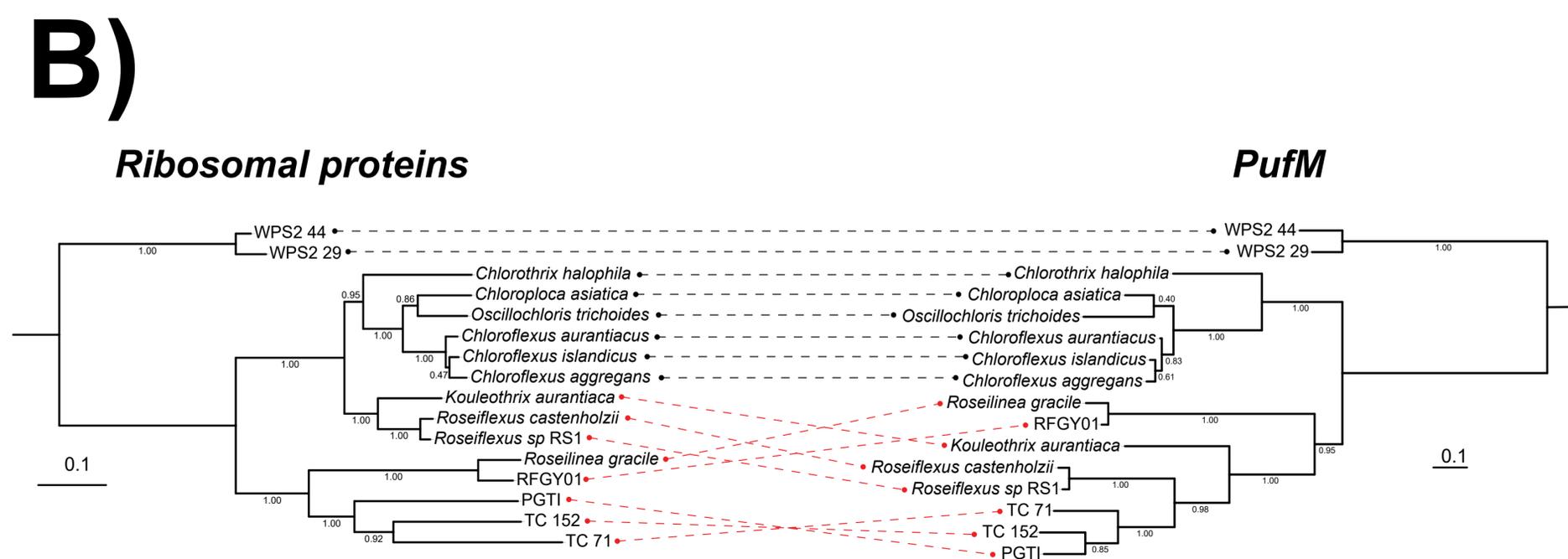
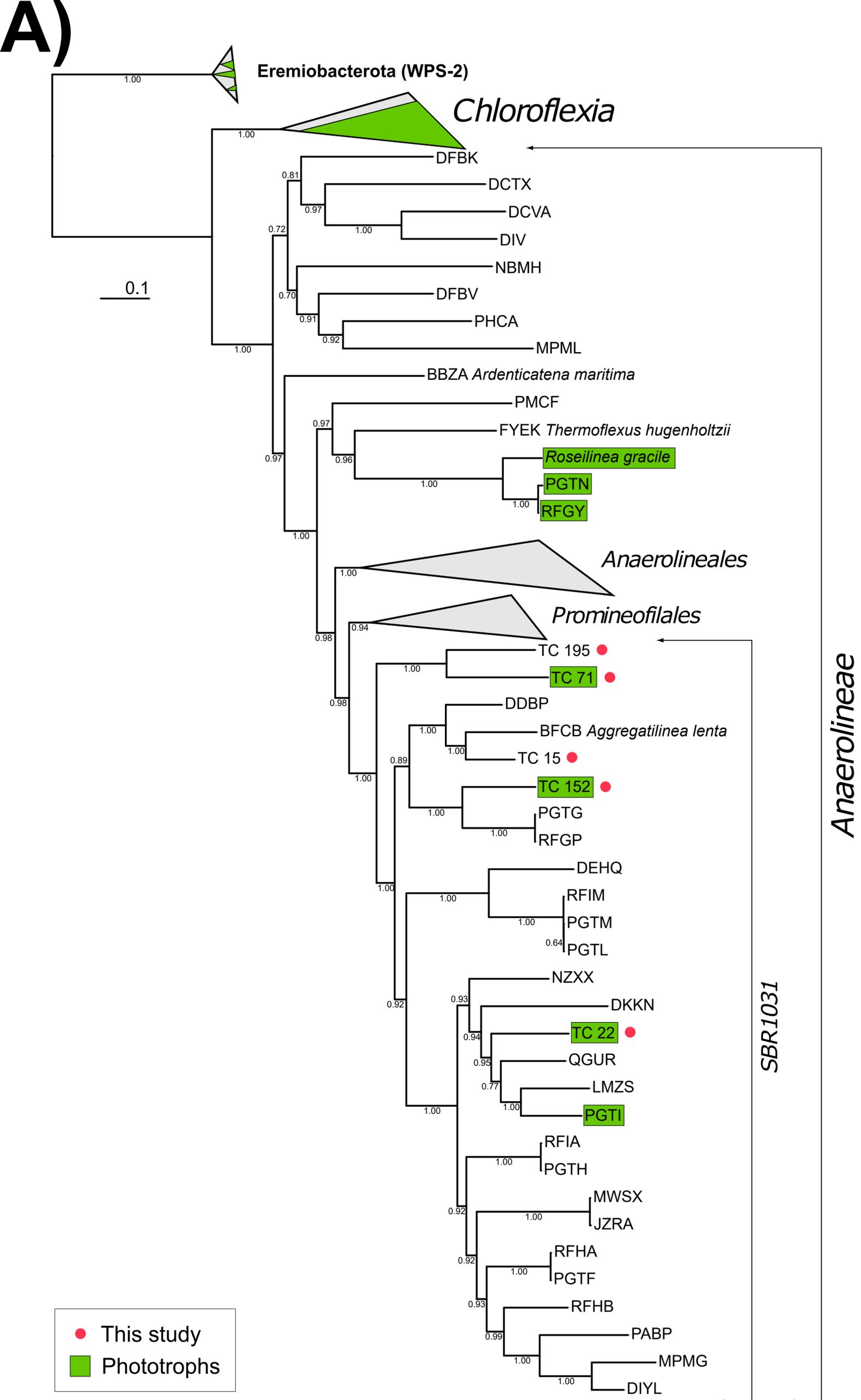
1. Alneberg, J., Bjarnason, B.S., de Bruijn, I., Schirmer, M., Quick, J., Ijaz, U.Z., Loman, N.J., Andersson, A.F. and Quince, C., 2013. CONCOCT: clustering contigs on coverage and composition. *arXiv preprint arXiv:1312.4038*.
2. Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M. and Meyer, F., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC genomics*, 9(1), p.75.
3. Björnsson, L., Hugenholtz, P., Tyson, G.W. and Blackall, L.L., 2002. Filamentous Chloroflexi (green non-sulfur bacteria) are abundant in wastewater treatment processes with biological nutrient removal. *Microbiology*, 148(8), pp.2309-2318.
4. Borisov, V. B., Gennis, R. B., Hemp, J., and Verkhovsky, M. I. (2011). The cytochrome bd-respiratory oxygen reductases. *Biochim. Biophys. Acta* 1807, 1398–1413. doi: 10.1016/j.bbabi.2011.06.016
5. Bryant, D. A. *et al.* Candidatus Chloracidobacterium thermophilum: An aerobic phototrophic acidobacterium. *Science* 317, 523-526, doi:10.1126/science.1143236 (2007).
6. Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden (2009). "BLAST+: architecture and applications." *Bmc Bioinformatics* 10: 421. DOI: 10.1186/1471-2105-10-421.
7. Chaumeil, P.A., Mussig, A.J., Hugenholtz, P. and Parks, D.H., 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*.
8. Doolittle, R. F. (1986). Of urfs and orfs: A primer on how to analyze derived amino acid sequences. Mill Valley, California, University Science Books.
9. Dravis, J. J., and Wanless, H. R., Caicos platform models of Quaternary carbonate deposition controlled by stronger easterly Trade Winds—application to petroleum exploration, *in* Proceedings Developing models and analogs for isolated carbonate platform—Holocene and Pleistocene carbonates of Caicos Platform, British West Indies. SEPM Core Workshop2008, Volume 22, SEPM.
10. Gomes, M., Reidman, L.A., O'Reilly, S.S., Lingappa, U., Metcalfe, K., Fike, D., Grotzinger, J.P., Fischer, W.W., and Knoll, A.H., *submitted*, Microbial Mats on Little Ambergris Cay, Turks and Caicos Islands: Taphonomy and the Selective Preservation of Biosignatures. *Geobiology*.
11. Graham ED, Heidelberg JF, Tully BJ. (2018) Potential for primary productivity in a globally-distributed bacterial phototroph. *ISME J* 350, 1–6
12. Hemp, J., Ward, L. M., Pace, L. A., and Fischer, W. W. (2015). Draft genome sequence of *Levilinea saccharolytica* KIBI-1, a member of the Chloroflexi class Anaerolineae. *Genome Announc.* 3:e01357-15. doi: 10.1128/genomeA.01357-15
13. Holland-Moritz, H., Stuart, J., Lewis, L.R., Miller, S., Mack, M.C., McDaniel, S.F. and Fierer, N., 2018. Novel bacterial lineages associated with boreal moss species. *Environmental microbiology*, 20(7), pp.2625-2638.
14. Kanehisa, M., Sato, Y., and Morishima, K. (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726-731.

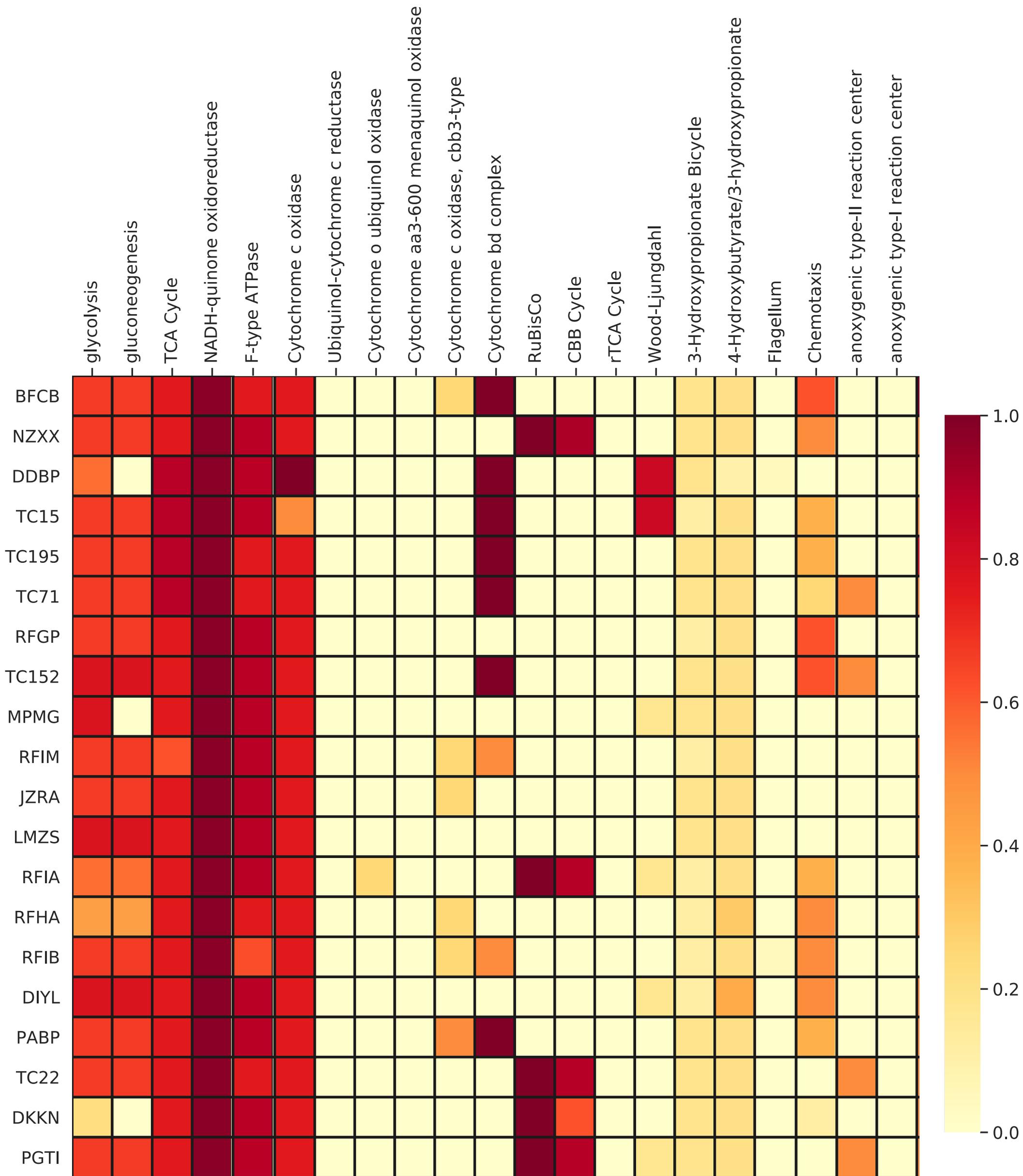
15. Kang, D.D., Froula, J., Egan, R. and Wang, Z., 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, 3, p.e165.
16. Klatt, C. G. *et al.* Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. *ISME J.* 5, 1262-1278, doi:10.1038/ismej.2011.73 (2011).
17. Lemoine, F., J. B. Domelevo Entfellner, E. Wilkinson, D. Correia, M. Davila Felipe, T. De Oliveira and O. Gascuel (2018). "Renewing Felsenstein's phylogenetic bootstrap in the era of big data." *Nature* 556(7702): 452-456. DOI: 10.1038/s41586-018-0043-0.
18. Letunic, I. and P. Bork (2016). "Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees." *Nucleic Acids Res* 44(W1): W242-W245. DOI: 10.1093/nar/gkw290.
19. Li, D., Liu, C.M., Luo, R., Sadakane, K. and Lam, T.W., 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10), pp.1674-1676.
20. Louca, S., Shih, P.M., Pennell, M.W., Fischer, W.W., Parfrey, L.W. and Doebeli, M., 2018. Bacterial diversification through geological time. *Nature ecology & evolution*, 2(9), p.1458.
21. Militon, C., Boucher, D., Vachelard, C., Perchet, G., Barra, V., Troquet, J., Peyretilade, E. and Peyret, P., 2010. Bacterial community changes during bioremediation of aliphatic hydrocarbon-contaminated soil. *FEMS Microbiology Ecology*, 74(3), pp.669-681.
22. Miller, M. A., W. Pfeiffer and T. Schwartz (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop (GCE).
23. Nakahara, N., Nobu, M.K., Takaki, Y., Miyazaki, M., Tasumi, E., Sakai, S., Ogawara, M., Yoshida, N., Tamaki, H., Yamanaka, Y. and Katayama, A., 2019. *Aggregatilinea lenta* gen. nov., sp. nov., a slow-growing, facultatively anaerobic bacterium isolated from subseafloor sediment, and proposal of the new order *Aggregatilineales* ord. nov. within the class *Anaerolineae* of the phylum *Chloroflexi*. *International journal of systematic and evolutionary microbiology*, 69(4), pp.1185-1194.
24. Pace, L. A., Hemp, J., Ward, L. M., and Fischer, W. W. (2015). Draft genome of *Thermanaerothermophilus daxensis* GNS-1, a thermophilic facultative anaerobe from the *Chloroflexi* class *Anaerolineae*. *Genome Announc.* 3, e01354–e01315. doi: 10.1128/genomeA.01354-15
25. Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P. and Tyson, G.W., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research*, 25(7), pp.1043-1055.
26. Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.A. and Hugenholtz, P., 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature biotechnology*.
27. Raymond, J., Zhaxybayeva, O., Gogarten, J.P., Gerdes, S.Y. and Blankenship, R.E., 2002. Whole-genome analysis of photosynthetic prokaryotes. *Science*, 298(5598), pp.1616-1620.
28. Shih, P.M., Ward, L.M. and Fischer, W.W., 2017. Evolution of the 3-hydroxypropionate bicycle and recent transfer of anoxygenic photosynthesis into the *Chloroflexi*. *Proceedings of the National Academy of Sciences*, 114(40), pp.10749-10754.
29. Sieber, C.M., Probst, A.J., Sharrar, A., Thomas, B.C., Hess, M., Tringe, S.G. and Banfield, J.F., 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.*, p.1.

30. Stamatakis, A. (2014). "RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies." *Bioinformatics* 30(9): 1312-1313. DOI: 10.1093/bioinformatics/btu033.
31. Tabita, F.R., Satagopan, S., Hanson, T.E., Kreel, N.E. and Scott, S.S., 2008. Distinct form I, II, III, and IV Rubisco proteins from the three kingdoms of life provide clues about Rubisco evolution and structure/function relationships. *Journal of experimental botany*, 59(7), pp.1515-1524.
32. Thiel, V., Tank, M. and Bryant, D.A., 2018. Diversity of chlorophototrophic bacteria revealed in the omics era. *Annual review of plant biology*, 69, pp.21-49.
33. Trembath-Reichert, E., Ward, L.M., Slotznick, S.P., Bachtel, S.L., Kerans, C., Grotzinger, J.P. and Fischer, W.W., 2016. Gene sequencing-based analysis of microbial-mat morphotypes, Caicos Platform, British West Indies. *Journal of Sedimentary Research*, 86(6), pp.629-636.
34. Trower, E.J., O'Reilly, S.S., Gomes, M., Cantine, M., Stein, N., Grotzinger, H., Strauss, J.V., Lamb, M.P., Grotzinger, J.P., Knoll, A.H., and Fischer, W.W., 2018, Active ooid growth driven by sediment transport in a high energy shoal, Little Ambergris Cay, Turks and Caicos, British Overseas Territories: *Journal of Sedimentary Research*, v. 88, p. 1132-1151.
35. Ward, L. M., Hemp, J., Pace, L. A., and Fischer, W. W. (2015). Draft genome sequence of *Leptolinea tardivitalis* YMTK-2, a mesophilic anaerobe from the Chloroflexi class Anaerolineae. *Genome Announc.* 3:e01356-15. doi: 10.1128/genomeA.01356-15
36. Ward, L.M., 2017. *Microbial evolution and the rise of oxygen: the roles of contingency and context in shaping the biosphere through time* (Doctoral dissertation, California Institute of Technology).
37. Ward, L.M., Idei, A., Terajima, S., Kakegawa, T., Fischer, W.W. and McGlynn, S.E., 2017. Microbial diversity and iron oxidation at Okuoku-hachikurou Onsen, a Japanese hot spring analog of Precambrian iron formations. *Geobiology*, 15(6), pp.817-835.
38. Ward, L.M., McGlynn, S.E. and Fischer, W.W., 2017. Draft genome sequence of *Chloracidobacterium* sp. CP2\_5A, a phototrophic member of the phylum Acidobacteria recovered from a Japanese hot spring. *Genome Announc.*, 5(40), pp.e00821-17.
39. Ward, L.M., Hemp, J., Shih, P.M., McGlynn, S.E. and Fischer, W.W., 2018a. Evolution of phototrophy in the Chloroflexi phylum driven by horizontal gene transfer. *Front. Microbiol.*, 9, p.260.
40. Ward, L.M., Shih, P.M. and Fischer, W.W., 2018b. MetaPOAP: presence or absence of metabolic pathways in metagenome-assembled genomes. *Bioinformatics*, 34(24), pp.4284-4286.
41. Ward, LM, and PM Shih. The evolution of carbon fixation pathways in response to changes in oxygen concentration over geological time. *Free Radical Biology and Medicine*, 140 pp.188-199.
42. Ward, L., Shih, P.M., Hemp, J., Kakegawa, T., Fischer, W.W. and McGlynn, S.E., 2019a. Phototrophic Methane Oxidation in a Member of the Chloroflexi Phylum. *bioRxiv*, p.531582.
43. Ward, LM, A Idei, M Nakagawa, Y Ueno, WW Fischer, SE McGlynn. 2019b. Geochemical and metagenomic characterization of Jinata Onsen, a Proterozoic-analog hot spring, reveals novel microbial diversity including iron-tolerant phototrophs and thermophilic lithotrophs. *Microbes and Environments*, 34(3): 278-292
44. Ward, L.M., Johnston, D. and Shih, P.M., 2019c. Phanerozoic Radiation of Ammonia Oxidizing Bacteria. *bioRxiv*, p.655399.

45. Ward, LM, T Cardona, and H Holland-Moritz. 2019d. Evolutionary Implications of Anoxygenic Phototrophy in the Bacterial Phylum Candidatus Eremiobacterota (WPS-2). *Frontiers in Microbiology*, 10:1658. doi: 10.3389/fmicb.2019.01658.
46. Waterhouse, A. M., J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton (2009). "Jalview Version 2-a multiple sequence alignment editor and analysis workbench." *Bioinformatics* 25(9): 1189-1191. DOI: 10.1093/bioinformatics/btp033.
47. Wu, Y.W., Tang, Y.H., Tringe, S.G., Simmons, B.A. and Singer, S.W., 2014. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome*, 2(1), p.26.







Bin Id	Classification
TC_15	d__Bacteria;p__Chloroflexota;c__Anaerolineae;o__SBR1031;f__UBA2029;g__;s__
TC_152	d__Bacteria;p__Chloroflexota;c__Anaerolineae;o__SBR1031;f__;g__;s__
TC_195	d__Bacteria;p__Chloroflexota;c__Anaerolineae;o__SBR1031;f__;g__;s__
TC_22	d__Bacteria;p__Chloroflexota;c__Anaerolineae;o__SBR1031;f__A4b;g__;s__
TC_71	d__Bacteria;p__Chloroflexota;c__Anaerolineae;o__;f__;g__;s__

Phototrophy markers (of 5)	0	5	0	3	5
N50	10293	17396	11646	6808	8059
RNAs	43	40	44	30	44
Contigs	662	312	431	746	576
GC%	60.1	55.7	61.5	61.7	56.9
Coding sequences	5127	3445	4178	4167	4216
Genome size (mb)	5.45	3.86	4.06	4.43	4.06
Strain heterogeneity	0	0	0	0	33.33
Contamination	2.94	2.75	0	1.83	1.33
Completeness	88.07	92.66	95.41	77.88	87.61

Phototrophy False Negative estimate	2.40E-05	N/A	1.98E-07	0.04896	N/A
Phototrophy False Positive estimate	N/A	0.03717	N/A	0.04242	0.03902