Microbes Environ. 36(3), 2021
https://www.jstage.jst.go.jp/browse/jsme2 doi:10.1264/jsme2.ME21020

Short Communication

Complex History of Aerobic Respiration and Phototrophy in the *Chloroflexota* Class *Anaerolineae* Revealed by High-Quality Draft Genome of *Ca.* Roseilinea mizusawaensis AA3_104

Lewis M. Ward[1,2*], Fátima Li-Hau[2], Takeshi Kakegawa[3], and Shawn E. McGlynn[2*]

[1]*Department of Earth and Planetary Sciences, Harvard University, USA;* [2]*Earth-Life Science Institute, Tokyo Institute of Technology, Japan; and* [3]*Department of Geosciences, Tohoku University, Japan*

(Received March 3, 2021—Accepted June 4, 2021—Published online August 31, 2021)

Roseilinea is a novel lineage of *Chloroflexota* known only from incomplete metagenome-assembled genomes (MAGs) and preliminary enrichments. *Roseilinea* is notable for appearing capable of anoxygenic photoheterotrophy despite being only distantly related to well-known phototrophs in the *Chloroflexia* class such as *Chloroflexus* and *Roseiflexus*. Here, we present a high-quality MAG of a member of *Roseilinea*, improving our understanding of the metabolic capacity and phylogeny of this genus, and resolving the multiple instances of horizontal gene transfer that have led to its metabolic potential. These data allow us to propose a candidate family for these organisms, Roseilineaceae, within the *Anaerolineae* class.

Key words: photosynthesis, thermophile, sulfide, phylogenetics, metagenomics, *Chloroflexi*, green nonsulfur bacteria

Anoxygenic phototrophs in the *Chloroflexota* (formerly *Chloroflexi* or Green Nonsulfur Bacteria) phylum include the well known genera *Chloroflexus* and *Roseiflexus* isolated and characterized from sulfidic hot springs (*e.g.* Thiel *et al.*, 2018), but more recent metagenomic analyses have revealed several additional lineages of phototrophic *Chloroflexota* from environments including iron-rich hot springs (*e.g.* Ward *et al.*, 2019a; Ward *et al.*, 2019b), carbonate tidal flats (Ward *et al.*, 2020), and soda lakes (*e.g.* Grouzdev *et al.*, 2018). While some of these novel phototrophic *Chloroflexota* belong to the *Chloroflexia* class together with *Chloroflexus* and *Roseiflexus* (*e.g.* Grouzdev *et al.*, 2018) (Ward, *et al.* 2019 Genomic evidence for phototrophic oxidation of small alkanes in a member of the Chloroflexi phylum. *bioRxiv* https://doi.org/10.1101/531582), several others belong to the poorly characterized *Anaerolineae* class (*e.g.* Klatt *et al.*, 2011; Ward *et al.*, 2019b). Most phototrophic *Anaerolineae* belong to several discrete lineages within the *Aggregatilineales* order (Ward *et al.*, 2020); however, the first lineage of phototrophic *Chloroflexota* outside of the *Chloroflexia* class ever discovered—*Candidatus* Roseilinea gracile, first observed via metagenomic sequencing of a hot spring microbial mat from Yellowstone National Park (Klatt *et al.*, 2011; Tank *et al.*, 2017)—has never been confidently assigned a taxonomic rank below the class level. The relatively low quality of available genomes related to *Ca.* Roseilinea gracile, together with the absence of a pure culture isolate of this strain, has precluded thorough characterization and confident taxonomic assignment of this novel group of phototrophic *Chloroflexota*. Here, we present a high-quality draft genome of a bacterium within the *Ca.* Roseilinea genus, dubbed here *Ca.* Roseilinea mizusawaensis AA3_104. This genome, recovered from metagenomic sequencing of Mizusawa Onsen in Japan, provides additional high-quality genomic data for characterizing the metabolic potential, evolutionary history, and taxonomic relationships of the *Roseilinea* genus. We propose here the assignment of *Roseilinea* to a novel family Roseilineaceae within the *Chloroflexota* order *Thermoflexales*, with *Ca.* Roseilinea mizusawaensis AA3_104 as type genome.

Samples were collected from Mizusawa Hot Spring in Akita Prefecture, Japan. The sampling location was at 39°45′23.4″N 140°46′42.3″E where the sulfidic hot spring flowed down a steep gulley. The hot spring water was measured as pH 5.6 at the time of sampling. Samples consisted of microbial biomass collected as mottled green and white streamers in areas of flowing water at 50°C and a several mm thick, leathery microbial mat with white, green, red, and orange layers in 43°C water (Fig. 1). Two samples from each site were selected for sequencing.

pH was measured with an Extech DO700 8-in-1 Portable Dissolved Oxygen Meter (FLIR Commercial Systems). Temperature was measured with a Lasergrip 774 IR Thermometer (Etekcity) and a ThermoPop thermometer (ThermoWorks). The presence of sulfide was determined by smell.

Lysis of cells and preservation of DNA was accomplished in the field with Zymo Terralyzer BashingBead Matrix and Xpedition Lysis Buffer. Samples of microbial mat were disrupted immediately after sampling by attaching sample

* Corresponding authors.
Lewis M. Ward: E-mail: lewis_ward@fas.harvard.eduo;
Tel: +1–479–879–3618; Fax: +1–617–384–7396.
Shawn E. McGlynn: E-mail: mcglynn@elsi.jp;
Tel: +85–3–5734–3414; Fax: +81–3–5734–3416.

Citation: Ward, L. M., Li-Hau, F., Kakegawa, T., and McGlynn, S. E. (2021) Complex History of Aerobic Respiration and Phototrophy in the *Chloroflexota* Class *Anaerolineae* Revealed by High-Quality Draft Genome of *Ca.* Roseilinea mizusawaensis AA3_104. *Microbes Environ* **36**: ME21020.
https://doi.org/10.1264/jsme2.ME21020

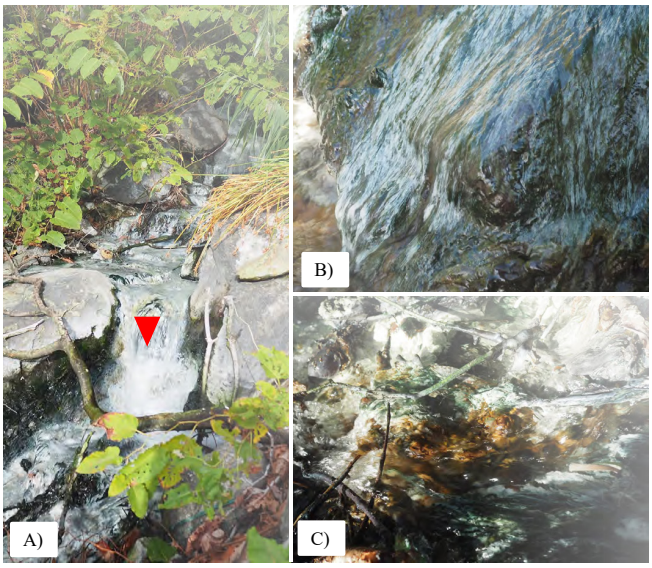

Fig. 1. A) Overview of Mizusawa Hot Spring. B) White streamers in area of high flow. C) Thick laminated mat with multiple colored layers.

tubes to a cordless reciprocating saw (Makita JR101DZ) and operating for 1 min. Bulk environmental DNA was extracted and purified after return to the lab with a Zymo Soil/Fecal DNA extraction kit. Quantification of DNA was performed with a Qubit 3.0 fluorimeter (Life Technologies) according to manufacturer's instructions. Purified DNA was submitted to the Integrated Microbiome Resource for library preparation and sequencing following established protocols (Comeau *et al.*, 2017) with 2×150 bp Illumina NextSeq. Raw sequence reads were quality controlled with BBTools (Bushnell, 2014) and coassembled with MegaHit v. 1.02 (Li *et al.*, 2016). Genome bins were constructed based on differential coverage using MetaBAT (Kang *et al.*, 2015). Completeness and contamination/redundancy were determined with CheckM v1.1.2 (Parks *et al.*, 2015). The genome was uploaded to RAST v2.0 for annotation and characterization (Aziz *et al.*, 2008). Presence or absence of metabolic pathways of interest was predicted using MetaPOAP v1.0 (Ward *et al.*, 2018c). Taxonomic assignment was determined with GTDB-Tk v1.2 (Parks *et al.*, 2018; Chaumeil *et al.*, 2020; Parks *et al.*, 2020). Genomes were compared with AAI (Rodriguez-R and Konstantinidis, 2014) to verify species and genus-level divisions. Organismal phylogenies were built using concatenated ribosomal proteins following methods derived from Hug *et al.* (2016) using the software pipeline described below. Protein sequences were extracted from genomes using the *tblastn* function of BLAST+ (Camacho *et al.*, 2009), aligned using MUSCLE (Edgar, 2004), and manually trimmed and curated using Jalview (Waterhouse *et al.*, 2009). Trees were calculated using RAxML v8.2.12 (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). All software was run with default parameters.

Metagenomic sequencing of 4 samples from Mizusawa Onsen produced 119,132,162 reads totaling 5,755,889,978 bp. Coassembly of these data produced 881,712 contigs totaling 842,365,641 bp with and N50 of 1,467 bp. AA3_104 was recovered as a metagenome-assembled genomes from binning of this dataset.

The AA3_104 MAG was recovered as 3,855,811 nt in 59 contigs with an N50 of 137,185 nt. The genome is 61.6% GC and encodes 49 RNAs (*e.g.* rRNAs and tRNAs) and 3,396 coding sequences. The genome was estimated to be 97.71% complete, 1.1% redundant/contamination, and 0% strain heterogeneity by CheckM. The MAG encodes 49 RNAs and complete 16S, 23S, and 5S rRNA gene sequences. This MAG meets current standards for a high-quality draft genome (Bowers *et al.*, 2017). Raw metagenomic reads mapped to the AA3_104 genome were recovered with ~1.7-fold coverage in the two samples from streamers collected from 50°C water and at up to 13.2-fold coverage in the samples from microbial mats from 43°C.

Phylogenetic analysis using concatenated ribosomal protein phylogenies (Fig. 2) clustered AA3_104 with *Ca.* Roseilinea gracile and J036, enigmatic phototrophic *Chloroflexota* only distantly related to other phototrophs in this phylum. *Ca.* Roseilinea gracile and its relatives have previously been proposed to belong to the *Anaerolineae* (Klatt *et al.*, 2011), *Ca.* Thermofonsia (Ward *et al.*, 2018a), and a novel order-level lineage (Ward *et al.*, 2019b). Taxonomic assignment of AA3_104, *Ca.* Roseilinea gracile, and J036 by GTDB-Tk determined that these organisms belongs to the *Thermoflexales* order within the *Anaerolineae* class of the *Chloroflexota* phylum, but could not place these genomes in a defined clade at the family rank or below. Together with their apparent phylogenetic divergence from other defined *Chloroflexota* orders suggests that these organisms represent a novel family within the *Thermoflexales* order of the *Anaerolineae* class of *Chloroflexota*. These genomes are 60–63% similar by AAI and 75–88% similar by ANI, consistent with these genomes representing three distinct species within a single genus (referred hereafter collectively as *Roseilinea*).

Analysis of the metabolic potential of AA3_104 suggests similar capacities to other members of *Roseilinea*; importantly, however, the relatively high completeness and quality of the AA3_104 MAG relative to its relatives allows more confident determination of what proteins and pathways are or are not actually encoded by these organisms.

Like previously described *Roseilinea* strains, AA3_104 appears to be capable of photoheterotrophy via a Type 2 Reaction Center encoded by unfused *pufL* and *pufM* genes. AA3_104 appears to encode synthesis of bacteriochlorophyll *a* through an atypical pathway previously proposed for phototrophic *Anaerolineae* in which *bchM*, *bcheE*, and the *bchLNB* complex are absent (*e.g.* Ward *et al.*, 2018a). The high completeness of AA3_104 (>97%) greatly increases confidence that the apparent absence of these genes is not a false negative caused by failure to recover sequences in the MAG (MetaPOAP false negative probability $<10^{-9}$). Known alternatives to missing bacteriochlorophyll synthesis genes (*e.g. acsF* for *bchE*, genes encoding the light-dependent POR enzyme for genes encoding the BchLNB complex, Chew and Bryant, 2007) were also absent. Despite the apparent absence of several components of the classical bacteriochlorophyll synthesis pathway from *Roseilinea* genomes,

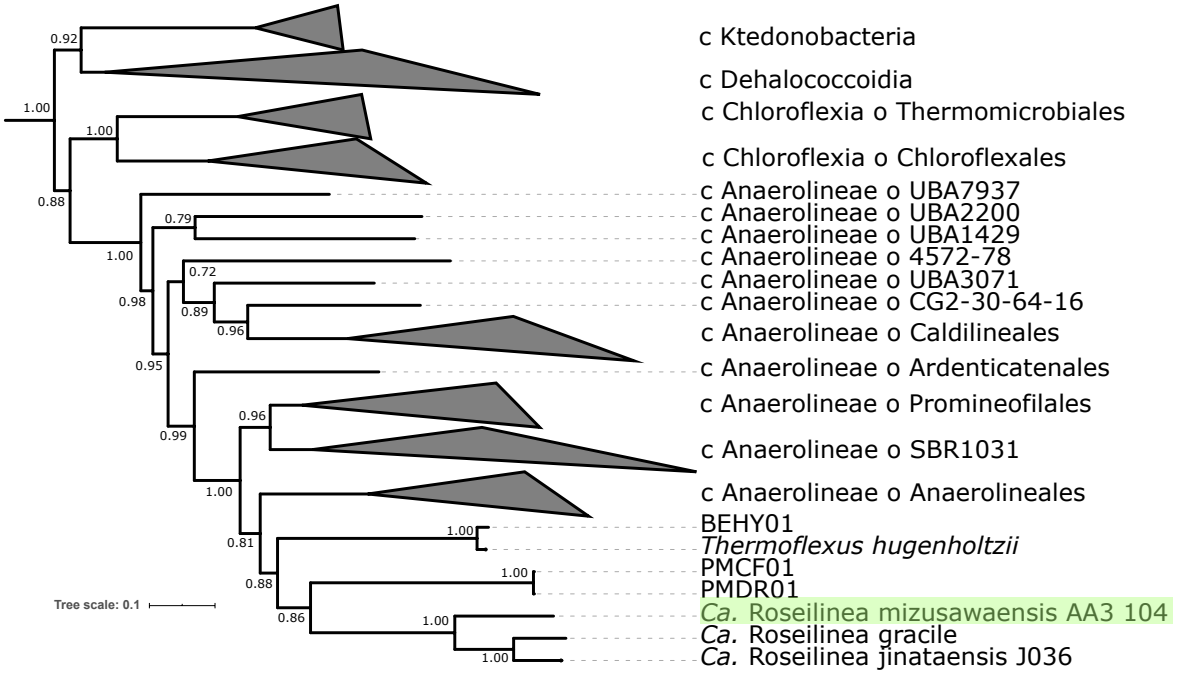

Fig. 2. Phylogeny the *Chloroflexota*, built with concatenated ribosomal proteins following Hug *et al.*, 2016 and rooted with the closely related phyla *Eremiobacterota* (Ward *et al.*, 2019a) and *Armatimonadetes* (Ward *et al.*, 2017). The *Thermoflexales* order of the *Anaerolineae* class is shown at the strain level, while other clades are collapsed at the order (*Chloroflexia* and *Anaerolineae* classes) or class level.

observations of cells putatively identified as *Ca.* Roseilinea gracile do show autofluorescence consistent with bacteriochlorophyll a (Tank *et al.*, 2017), suggesting that this pigment is in fact synthesized in these organisms. Genes such as *bchK*, *bchU*, and *bchQ* are absent from AA3_104, consistent with this organism not producing bacteriochlorophylls c, d, or e. Taken all together, these data are consistent with previous proposals that phototrophic *Anaerolineae* (both *Roseilinea* and members of the order *Aggregatilineales*) utilize a novel bacteriochlorophyll a synthesis pathway that makes use of uncharacterized or multifunctional enzymes to perform steps typically performed by BchE, BchM and the BchLNB complex (*e.g.* a multifunctional BchXYZ complex for the conversion of both protochlorophyllide a to chlorophyllide a and the conversion of chlorophyllide a to 3-vinyl-bacteriochlorophyllide a) (Ward *et al.*, 2018a; Ward and Shih, 2021).

Like previously described phototrophic *Anaerolineae* (including *Roseilinea* and phototrophic members of *Aggregatilineales*), but unlike phototrophic *Chloroflexia*, AA3_104 lacks key marker genes for carbon fixation pathways including the Calvin cycle (*e.g.* rubisco) and the 3-hydroxypropionate bi-cycle (*e.g.* the trifunctional malyl-CoA/beta-methylmalyl-CoA/citramalyl-CoA lyase). AA3_104 is therefore most likely incapable of autotrophy and instead makes a living as an anoxygenic photoheterotroph.

AA3_104 appears to be capable of at least facultative aerobic respiration via both an A- and a B-family heme copper O_2 reductase (HCO), two *bc* Complex IIIs, and a *bd* oxidase. This complement of respiration genes is consistent with that previously described for members of *Roseilinea* (Ward *et al.*, 2018a; Ward *et al.*, 2019b).

The presence of a B-family HCO in phototrophic *Roseilinea* is a trait shared with all other known phototrophic *Chloroflexota* as well as phototrophic members of the closely related phylum *Eremiobacterota*, although closely related nonphototrophic strains lack a B-family HCO (*e.g.* Ward *et al.*, 2018a; 2019a; 2020). The functional link between the B-family HCO and phototrophy in these organisms is not well understood, but may relate to the adaptation of B-family HCOs to relatively low oxygen concentrations (Han *et al.*, 2011) and the oxygen sensitivity of proteins involved in anoxygenic phototrophy (*e.g.* Hamilton, 2019) linking anoxygenic photoheterotrophy in these organisms to low-oxygen environments. Interestingly, phylogenetic relationships of B-family HCO proteins are incongruent with organismal relationships and with relationships among phototrophy proteins (Fig. 3 and 4). This suggests that B-family HCOs and phototrophy proteins have independent histories of horizontal gene transfer in the *Chloroflexota* despite their apparent functional link. B-family HCO genes are not collocated with phototrophy genes and do not display conserved synteny between phototrophic *Chloroflexota* lineages, consistent with these genes having an independent history of horizontal gene transfer.

Unlike other phototrophic *Chloroflexota*, AA3_104 and other *Roseilinea* do not encode an Alternative Complex III. While the absence of genes encoding this protein from previous MAGs could have been ascribed to their low completeness, the high quality of AA3_104 makes it quite unlikely that this enzyme would not have been recovered in the MAG if it were found in the source genome (MetaPOAP false negative estimate <2.5% for AA3_104 alone, $<10^{-4}$ for all three available *Roseilinea* genomes considered together). It therefore appears that members of *Roseilinea*, in contrast to other phototrophic *Chloroflexota*, use a *bc* Complex III in their phototrophic electron transport chain instead of an

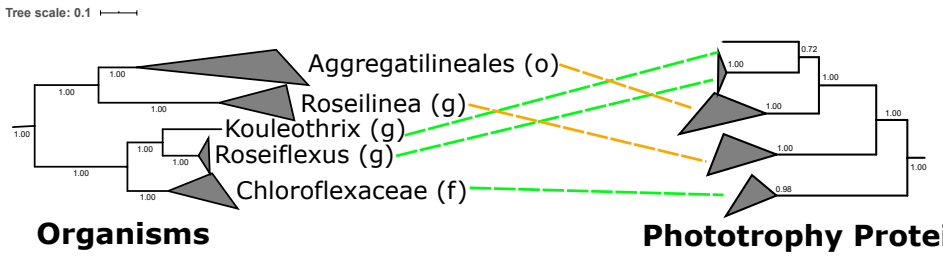

Fig. 3. Tanglegram demonstrating phylogenetic incongruity between organismal phylogeny (left, built with concatenated ribosomal proteins) and phototrophy protein phylogeny (right, built with concatenated PufL, PufM, BchX, BchY, and BchZ sequences) of phototrophic *Chloroflexota*, collapsed at the genus (*Roseilinea*, *Roseiflexus*, and *Kouleothrix*), family (*Chloroflexaceae*) or order (*Aggregatilineales*) levels. While phototrophy proteins appear to have been vertically inherited within the *Chloroflexia* class after the divergence of the basal nonphototrophic *Herpetosiphonaceae* family (Ward *et al.*, 2015; Ward *et al.*, in press), the incongruent positioning of phototrophic *Aggregatilineales* and *Roseilinea* suggests that these lineages acquired phototrophy via horizontal gene transfer from members of the *Chloroflexia* (Ward *et al.*, 2018). Secondary HGT events may have been occurred within *Aggregatilineales* (Ward *et al.*, 2020)

Alternative Complex III. The evolutionary and biochemical logic for this difference is not yet apparent, but it does appear to confirm that an Alternative Complex III is not essential for phototrophy in the *Chloroflexota*.

In most lineages of phototrophic bacteria, it appears that the capacity for aerobic respiration was acquired before the acquisition of phototrophy (Fischer *et al.*, 2016). This trend has been confirmed for phototrophs in the *Chloroflexota* orders *Chloroflexales* and *Aggregatilineales*, in which A-family HCOs and other components of electron transport chains were acquired to enable aerobic respiration before the acquisition of phototrophy and B-family HCOs (Shih *et al.*, 2017; Ward *et al.*, 2020). However, at present it is impossible to confirm whether this trend extends to *Roseilinea*. While genes for aerobic respiration are widespread in even apparently obligately anaerobic members of the *Anaerolineae* class of *Chloroflexota* (*e.g.* Hemp *et al.*, 2015; Pace *et al.*, 2015; Ward *et al.*, 2018b), it appears that these genes were acquired independently in many lineages within this class subsequent to their divergence (*e.g.* Ward *et al.*, 2020). Members of Roseilineaceae described so far consist of a single apparent genus at the end of a relatively long branch whose closest relatives appear to be a family of *Thermoflexales* provisionally identified by GTDB as Fen-1058 (Fig. 2). Fen-1058 does encode aerobic respiration; however, respiratory proteins in the families of *Thermoflexaceae*—Roseilineaceae, Fen-1058, and *Thermoflexales*—are not closely related (Fig. 4). This suggests that each of the known families of *Thermoflexales* acquired respiration independently after they diverged. Both aerobic respiration and phototrophy therefore appear to have been acquired along the long branch leading to crown group *Roseilinea*; in the absence of additional information it is therefore impossible to determine which of these traits was acquired first along this branch. Resolving this uncertainty will require recovering additional Roseilineaceae diversity that breaks up the long branch leading to *Roseilinea*. However, it is clear from phylogenetic relationships of proteins that components of the electron transport chain in Roseilineaceae were acquired through independent HGT events from different sources, suggesting that the capacity for respiration and phototrophy was not acquired in a single large HGT event.

We propose the assignment of AA3_104, described here, and J036, described in Ward *et al.*, 2019, to the *Roseilinea* genus proposed by Tank *et al.* (2017). We propose the specific epithets *Ca.* Roseilinea mizusawaensis for AA3_104 in recognition of Mizusawa Onsen as the source of this organism, and *Ca.* Roseilinea jinataensis for J036 in recognition of its discovery in Jinata Onsen. Given the apparent divergence of *Roseilinea* from other members of the phylum *Chloroflexota* as determined by concatenated ribosomal protein phylogenies as well as analysis via GTDB-Tk, we propose the assignment of these organisms to a novel family in the *Thermoflexales* order of the *Anaerolineae* class of the *Chloroflexota* phylum, Roseilineaceae, fam. nov.. As it is currently the highest quality MAG available for this clade, we propose *Ca.* Roseilinea mizusawaensis AA3_104 as the type genome for Roseilineaceae until such time as an isolate and/or a complete genome is available, following recent recommendations for *candidatus* taxa (Chuvochina *et al.*, 2019).

The family Roseilineaceae is so far known only as putatively at least facultatively aerobic photoheterotrophs from hot springs. Members of Roseilineaceae are currently known from geographically and geochemically diverse hot springs in the United States and Japan, including iron-rich and moderately acidic intertidal hot springs in Tokyo Prefecture (Ward *et al.*, 2019b), moderately acidic and sulfidic hot springs in Akita Prefecture (this study), and alkaline siliceous hot springs in Yellowstone National Park (Klatt *et al.*, 2011). However, 16S rRNA gene amplicon data suggests that this clade may have a wider environmental distribution. Highly similar 16S sequences (>97%) to that of *Ca.* Roseilinea mizusawaensis AA3_104 have been reported from wastewater treatment systems (Karst *et al.*, 2018) while somewhat similar (>94%) sequences have been reported from environments including contaminated aquifers (Thavamani *et al.*, 2012) and soil (Xiao *et al.*, 2009). Given the long phylogenetic branch between crown group *Roseilinea* and the divergence of Roseilineaceae from *Thermoflexus* and other members of *Thermoflexales*, together with current understanding of the diversification rates of bacterial lineages through time (Louca *et al.*, 2018), it should be expected that much more diversity of Roseilineaceae has existed through geologic time. Ignoring possibilities of apparently rare extreme extinction events and population bottlenecks, it seems highly likely that additional

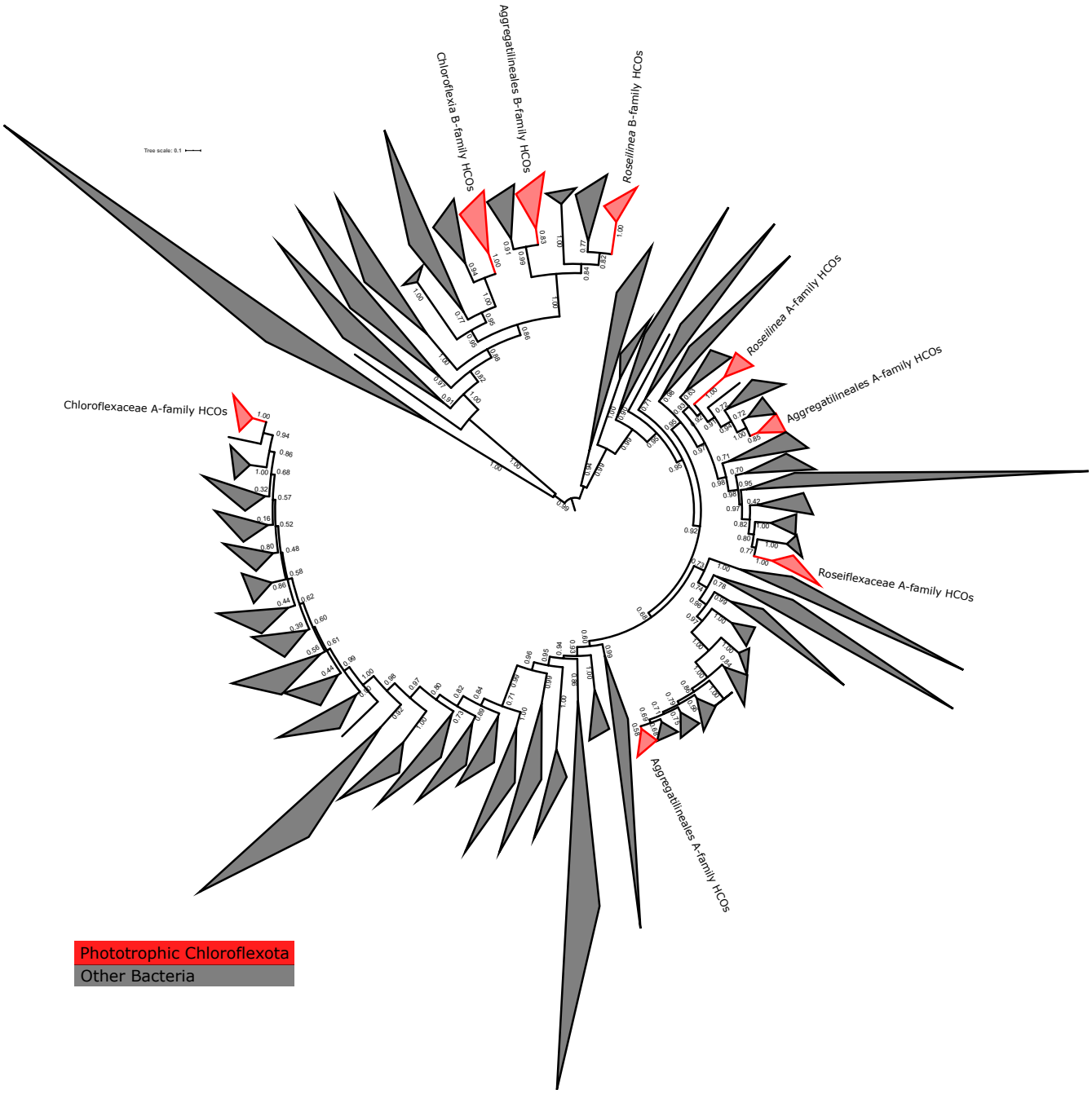

Fig. 4. Phylogeny of Heme-Copper Oxidoreductase (HCO) protein sequences with clades consisting of sequences from phototrophic *Chloroflexota* highlighted in red and labeled. While all known phototrophic *Chloroflexota* encode both an A- and a B-family HCO for O_2 reduction, these appear to have been acquired independently, with HCO phylogenies incongruent with both organismal relationships as well as with relationships between phototrophy proteins. Clades containing HCOs from other members of *Thermoflexales* are also labeled to highlight their phylogenetic distance from those of *Roseilinea*.

Roseilineaceae lineages exist in the environment today. Additional genome-resolved metagenomic sequencing of diverse environments is likely to yield additional diversity of this group, potentially breaking up the relatively long phylogenetic branch leading to *Roseilinea* and allowing for better resolution of the evolutionary transitions that have led to the significant divergence of *Roseilinea* from its closest known relatives.

Ca. Roseilinea mizusawaensis AA3_104 improves the available genomic diversity of anoxygenic phototrophic *Chloroflexota* outside of the *Chloroflexia* class, and provides the best available metagenome-assembled genome within the newly proposed Roseilineaceae family (Table 1). Improving sampling across the diversity of *Chloroflexota*—particularly within novel phototrophic lineages—provides valuable insight into the evolutionary trends leading to the extant diversity of phototrophs within this phylum and across the entire tree of life. In particular, the apparently extensive role of horizontal gene transfer in shaping the distribution of phototrophy and related pathways across the

Table 1. Genome statistics and GTDB-Tk taxonomic assignments of *Roseilinea* and other *Thermoflexales* strains.

Genome ID	GTDB Classification	# Contigs	Size	N50	Largest contig	GC%	Completeness	Contamination	Strain Heterogeneity	Source
AA3_104	d__Bacteria p__*Chloroflexota* c__*Anaerolineae* o__*Thermoflexales* f__ g__ s__	59	3,855,811	137,185	436,788	0.61647	97.71	1.1	0	Sulfidic hot spring
BEHY01	d__Bacteria p__*Chloroflexota* c__*Anaerolineae* o__*Thermoflexales* f__*Thermoflexaceae* g__*Thermoflexus* s__*Thermoflexus* sp002898735	175	2,931,246	34,479	127,152	0.66051	90.37	1.1	25	Ammonia-oxidizing enrichment culture
FYEK01	d__Bacteria p__*Chloroflexota* c__*Anaerolineae* o__*Thermoflexales* f__*Thermoflexaceae* g__*Thermoflexus* s__*Thermoflexus hugenholtzii*	87	3,216,440	139,933	294,380	0.67347	97.25	0.92	0	Hot spring
J036	d__Bacteria p__*Chloroflexota* c__*Anaerolineae* o__*Thermoflexales* f__ g__ s__	150	3,068,507	30,577	104,612	0.63702	93.12	0.1	0	Iron-rich hot spring
PMCF01	d__Bacteria p__*Chloroflexota* c__*Anaerolineae* o__*Thermoflexales* f__Fen-1058 g__Fen-1058 s__Fen-1058 sp003154115	516	7,694,415	25,596	118,786	0.57589	95.77	4.59	0	Permafrost
PMDR01	d__Bacteria p__*Chloroflexota* c__*Anaerolineae* o__*Thermoflexales* f__Fen-1058 g__Fen-1058 s__Fen-1058 sp003154115	982	6,553,527	8,786	44,993	0.57971	79.93	3.67	0	Permafrost
Roseilinea_gracile	d__Bacteria p__*Chloroflexota* c__*Anaerolineae* o__*Thermoflexales* f__ g__ s__	2329	2,103,554	1,147	4,984	0.63651	76.33	1.83	0	Sulfidic hot spring

tree of life (*e.g.* Raymond *et al.*, 2002; Hohmann-Marriott and Blankenship, 2011; Fischer *et al.*, 2016; Shih *et al.*, 2017; Ward *et al.*, 2018a; 2019a; Ward and Shih, 2019) provides an opportunity to investigate the history and evolution of functional traits even though most bacteria that have possessed these traits through geologic time are now extinct (Louca *et al.*, 2018).

Utilizing the extant diversity and evolutionary relationships of microorganisms to understand the early evolution of metabolic traits requires adequate sampling across extant diversity. The recovery of *Ca.* Roseilinea mizusawaensis AA3_104 and other members of Roseilineaceae provide an excellent example of the taxonomic and metabolic novelty that exists in the environment which can be discovered via genome-resolved metagenomic sequencing and which will provide crucial data for comparative phylogenetic analyses that can help answer longstanding questions about the evolution of phototrophy, respiration, and other traits deep in the tree of life.

Data Availability: The AA3_104 genome has been uploaded to the NCBI WGS database under the submission ID SUB8655613 and will be publicly available immediately following processing.

Acknowledgements

LMW acknowledges support from a Simons Foundation Postdoctoral Fellowship in Marine Microbial Ecology. SEM acknowledges support from the Astrobiology Center Program of the National Institutes of Natural Sciences (grant no. AB311013).

References

Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., *et al.* (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* **9**: 75.

Bowers, R.M., Kyrpides, N.C., Stepanauskas, R., Harmon-Smith, M., Doud, D., Reddy, T.B.K., *et al.* (2017) Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* **35**: 725–731.

Bushnell, B. (2014) BBTools software package. URL http://sourceforge.net/projects/bbmap

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009) BLAST+: architecture and applications. *BMC Bioinf* **10**: 421.

Chaumeil, P.A., Mussig, A.J., Hugenholtz, P., and Parks, D.H. (2020) GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* **36**: 1925–1927.

Chew, A.G.M., and Bryant, D.A. (2007) Chlorophyll biosynthesis in bacteria: the origins of structural and functional diversity. *Annu Rev Microbiol* **61**: 113–129.

Chuvochina, M., Rinke, C., Parks, D.H., Rappé, M.S., Tyson, G.W., Yilmaz, P., *et al.* (2019) The importance of designating type material for uncultured taxa. *Syst Appl Microbiol* **42**: 15–21.

Comeau, A.M., Douglas, G.M., and Langille, M.G.I. (2017) Microbiome Helper: A custom and streamlined workflow for microbiome research. *mSystems* **2**: e00127-16.

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792–1797.

Fischer, W.W., Hemp, J., and Johnson, J.E. (2016) Evolution of oxygenic photosynthesis. *Annu Rev Earth Planet Sci* **44**: 647–683.

Grouzdev, D.S., Rysina, M.S., Bryantseva, I.A., Gorlenko, V.M., and Gaisin, V.A. (2018) Draft genome sequences of '*Candidatus* Chloroploca asiatica' and '*Candidatus* Viridilinea mediisalina', candidate representatives of the *Chloroflexales* order: phylogenetic and taxonomic implications. *Stand Genomic Sci* **13**: 24.

Hamilton, T.L. (2019) The trouble with oxygen: The ecophysiology of extant phototrophs and implications for the evolution of oxygenic photosynthesis. *Free Radical Biol Med* **140**: 233–249.

Han, W., Kwan, K.Y., Shim, S., Lam, M.M., Shin, Y., Xu, X., *et al.* (2011) TBR1 directly represses Fezf2 to control the laminar origin and development of the corticospinal tract. *Proc Natl Acad Sci U S A* **108**: 3041–3046.

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.

Hohmann-Marriott, M.F., and Blankenship, R.E. (2011) Evolution of photosynthesis. *Annu Rev Plant Biol* **62**: 515–548.

Hug, L.A., Baker, B.J., Anantharaman, K., Brown, C.T., Probst, A.J., Castelle, C.J., *et al.* (2016) A new view of the tree of life. *Nat Microbiol* **1**: 16048.

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**: e1165.

Karst, S.M., Dueholm, M.S., McIlroy, S.J., Kirkegaard, R.H., Nielsen, P.H., and Albertsen, M. (2018) Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. *Nat Biotechnol* **36**: 190.

Klatt, C.G., Wood, J.M., Rusch, D.B., Bateson, M.M., Hamamura, N., Heidelberg, J.F., *et al.* (2011) Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. *ISME J* **5**: 1262–1278.

Lemoine, F., Entfellner, J.B.D., Wilkinson, E., Correia, D., Felipe, M.D., Oliveira, T., and Gascuel, O. (2018) Renewing Felsenstein's phylogenetic bootstrap in the era of big data. *Nature* **556**: 452.

Letunic, I., and Bork, P. (2016) Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* **44**: W242–W245.

Li, D., Luo, R., Liu, C.M., Leung, C.M., Ting, H.F., Sadakane, K., *et al.* (2016) MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods (Amsterdam, Neth)* **102**: 3–11.

Louca, S., Shih, P.M., Pennell, M.W., Fischer, W.W., Parfrey, L.W., and Doebeli, M. (2018) Bacterial diversification through geological time. *Nat Ecol Evol* **2**: 1458.

Miller, M.A., Pfeiffer, W., and Schwartz, T. (2010) Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *2010 Gateway Computing Environments Workshop (GCE)*. New Orleans, LA. *Institute of Electrical and Electronics Engineers (IEEE)*, pp. 1–8.

Pace, L.A., Hemp, J., Ward, L.M., and Fischer, W.W. (2015) Draft genome of *Thermanaerothrix daxensis* GNS-1, a thermophilic facultative anaerobe from the *Chloroflexi* class *Anaerolineae*. *Genome Announc* **3**: e01354-15.

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 Res* **25**: 1043–1055.

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. *Nat Biotechnol* **36**: 996–1004.

Parks, D.H., Chuvochina, M., Chaumeil, P.A., Rinke, C., Mussig, A.J., and Hugenholtz, P. (2020) A complete domain-to-species taxonomy for Bacteria and Archaea. *Nat Biotechnol* **38**: 1079–1086.

Raymond, J., Zhaxybayeva, O., Gogarten, J.P., Gerdes, S.Y., and Blankenship, R.E. (2002) Whole-genome analysis of photosynthetic prokaryotes. *Science* **298**: 1616–1620.

Rodriguez-R, L.M., and Konstantinidis, K.T. (2014) Bypassing cultivation to identify bacterial species. *Microbe* **9**: 111–118.

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. *Proc Natl Acad Sci U S A* **114**: 10749–10754.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**: 1312–1313.

Tank, M., Thiel, V., Ward, D.M., and Bryant, D.A. (2017) 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*. Cham: Springer, pp. 87–137.

Thavamani, P., Malik, S., Beer, M., Megharaj, M., and Naidu, R. (2012) Microbial activity and diversity in long-term mixed contaminated soils with respect to polyaromatic hydrocarbons and heavy metals. *J Environ Manage* **99**: 10–17.

Thiel, V., Tank, M., and Bryant, D.A. (2018) Diversity of chlorophototrophic bacteria revealed in the omics era. *Annu Rev Plant Biol* **69**: 21–49.

Ward, L.M., Hemp, J., Pace, L.A., and Fischer, W.W. (2015) Draft genome sequence of *Herpetosiphon geysericola* GC-42, a nonphototrophic member of the *Chloroflexi* class *Chloroflexia*. *Genome Announc* **3**: e01352-15.

Ward, L.M., McGlynn, S.E., and Fischer, W.W. (2017) Draft genome sequences of a novel lineage of Armatimonadetes recovered from Japanese hot springs. *Genome Announc* **5**: e00820-17.

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**: 260.

Ward, L.M., McGlynn, S.E., and Fischer, W.W. (2018b) Draft genome sequences of two basal members of the Anaerolineae class of Chloroflexi from a sulfidic hot spring. *Genome Announc* **6**: e00570-18.

Ward, L.M., Shih, P.M., and Fischer, W.W. (2018c) MetaPOAP: presence or absence of metabolic pathways in metagenome-assembled genomes. *Bioinformatics* **34**: 4284–4286.

Ward, L.M., Cardona, T., and Holland-Moritz, H. (2019a) Evolutionary implications of anoxygenic phototrophy in the bacterial phylum Candidatus Eremiobacterota (WPS-2). *Front Microbiol* **10**: 1658.

Ward, L.M., Idei, A., Nakagawa, M., Ueno, Y., Fischer, W.W., and McGlynn, S.E. (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 Environ* **34**: 278–292.

Ward, L.M., and Shih, P.M. (2019) The evolution and productivity of carbon fixation pathways in response to changes in oxygen concentration over geological time. *Free Radical Biol Med* **140**: 188–199.

Ward, L.M., Lingappa, U.F., Grotzinger, J.P., and Fischer, W.W. (2020) Microbial mats in the Turks and Caicos Islands reveal diversity and evolution of phototrophy in the Chloroflexota order Aggregatilineales. *Environ Microbiome* **15**: 1–9.

Ward, L.M., and Shih, P.M. (2021) Granick revisited: synthesizing evolutionary and ecological evidence for the late origin of bacteriochlorophyll via ghost lineages and horizontal gene transfer. *PLoS One* **16**: e0239248.

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., and Barton, G.J. (2009) Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**: 1189–1191.

Xiao, S., Xie, X., and Liu, J. (2009) Microbial communities in acid water environments of two mines, China. *Environ Pollut* **157**: 1045–1050.