A novel microbialite-associated phototrophic Chloroflexi lineage exhibiting a quasi-clonal pattern along Depth

Aurélien Saghaï, Yvan Zivanovic, David Moreira, Rosaluz Tavera, Purificación López-García

A Novel Microbialite-Associated Phototrophic Chloroflexi Lineage Exhibiting a Quasi-Clonal Pattern along Depth

Aurélien Saghaï[1,2,*], Yvan Zivanovic[3], David Moreira[1], Rosaluz Tavera[4], and Purificación López-García[1]

[1]Ecologie Systématique Evolution, CNRS, AgroParisTech, Université Paris-Saclay, Orsay, France

[2]Department of Forest Mycology and Plant Pathology, Swedish University of Agricultural Sciences, Uppsala, Sweden

[3]Institut de Biologie Intégrative de la Cellule, CNRS, Université Paris-Saclay, Orsay, France

[4]Departamento de Ecología y Recursos Naturales, Universidad Nacional Autónoma de México, Mexico City, Mexico

*Corresponding author: E-mail: aurelien.saghai@slu.se.

Accepted: 10 June 2020

Abstract

Chloroflexales (Chloroflexi) are typical members of the anoxygenic photosynthesizing component of microbial mats and have mostly been characterized from communities associated to hot springs. Here, we report the assembly of five metagenome-assembled genomes (MAGs) of a novel lineage of Chloroflexales found in mesophilic lithifying microbial mats (microbialites) in Lake Alchichica (Mexico). Genomic and phylogenetic analyses revealed that the bins shared 92% of their genes, and these genes were nearly identical despite being assembled from samples collected along a depth gradient (1–15 m depth). We tentatively name this lineage *Candidatus Lithoflexus mexicanus*. Metabolic predictions based on the MAGs suggest that these chlorosome-lacking mixotrophs share features in central carbon metabolism, electron transport, and adaptations to life under oxic and anoxic conditions, with members of two related lineages, Chloroflexineae and Roseiflexineae. Contrasting with the other diverse microbialite community members, which display much lower genomic conservation along the depth gradient, *Ca. L. mexicanus* MAGs exhibit remarkable similarity. This might reflect a particular flexibility to acclimate to varying light conditions with depth or the capacity to occupy a very specific spatial ecological niche in microbialites from different depths. Alternatively, *Ca. L. mexicanus* may also have the ability to modulate its gene expression as a function of the local environmental conditions during diel cycles in microbialites along the depth gradient.

Key words: Chloroflexi, microbialite, mixotrophy, population genomics.

Significance

Microorganisms capable of photosynthesis in the absence of oxygen are relatively poorly characterized despite their relevance to understand the ecology and evolution of present and past phototrophic microbial ecosystems. We generated five metagenome-assembled genomes corresponding to a novel member of Chloroflexi, *Candidatus Lithoflexus mexicanus*, from microbialites thriving at different depths and locations in the alkaline Lake Alchichica, Mexico. Surprisingly, *Ca. L. mexicanus* displayed a quasi-clonal structure despite differences in light quality and availability with depth. This suggests that *Ca. L. mexicanus* is adapted to varying light conditions, perhaps through diel changes in gene expression, and that Chloroflexi have a different adaptive response to light availability compared with cyanobacteria and other, nonclonal photosynthetic members of the same microbialite communities.

Introduction

Investigations of the structure of bacterial populations have shown that genetically similar organisms often display distinct ecological features (Luo et al. 2011), an observation that has been linked to the existence of an extensive pan-genome (Tettelin et al. 2005; Kashtan et al. 2014). Obtaining such structural genomic information for a broad range of lineages and in multiple ecosystems is essential to identify fine-tuning

evolutionary drivers of microbial diversification. However, knowledge about the genomic heterogeneity for specific lineages in wild populations is restricted to a few abundant lineages in marine (Kashtan et al. 2014) and freshwater (Garcia et al. 2018) environments.

Microbial mats offer an interesting natural target for population genomics studies. Due to their collective metabolic activity, these phylogenetically and functionally diverse benthic microbial communities (Des Marais 1990) generate and maintain steep vertical biogeochemical gradients (van Gemerden 1993; Dupraz and Visscher 2005) that act as strong environmental selectors. Moreover, contemporary microbial mats are often used as analogs of past microbial ecosystems (Ward et al. 1998; Stal 2001). Indeed, the observed functional shifts of conserved core metabolic pathways in mat metagenomes across redox gradients might reflect early metabolic transitions during the oxidation of the atmosphere (Gutiérrez-Preciado et al. 2018). Depending on the specific hydrochemical conditions, these light- and redox-stratified communities can be lithifying, favoring (usually) carbonate precipitation and leading to the formation of microbialites. Although microbialites are relatively uncommon today, they were abundant and globally distributed during the Precambrian and until the Cambrian explosion around 540 Ma (Riding 2000; Tice and Lowe 2004; Allwood et al. 2006).

In contemporary mats, photosynthetic microorganisms play a key role by producing the organic matter that is consumed by (photo-)heterotrophic members of the community (Dupraz and Visscher 2005; Klatt et al. 2013; Kim et al. 2015; Mobberley et al. 2015; Gutiérrez-Preciado et al. 2018). Some evidences of ecological differentiation are available for Cyanobacteria, which have received the most attention for both historical (e.g., distinct morphological features used as evidence of biogenicity in the fossil record, hypothesis that oxygenic photosynthesis was the main driver of microbialite formation) and practical (e.g., abundant community members, relatively simple requirements for cultivation) reasons. Several culture-based studies thus suggest that different ecotypes of *Synechococcus* spp. occur in vertical subsections of individual mats as thin as 80 μm, exhibiting different responses to light (Allewalt et al. 2006; Bhaya et al. 2007; Kilian et al. 2007; Becraft et al. 2015; Nowack et al. 2015; Olsen et al. 2015). Yet, far less is known about the microorganisms capable of anoxygenic photosynthesis, even though they collectively represent a substantial fraction of microbial mat communities (Klatt et al. 2011; Saghaï et al. 2015).

In a previous study, we generated metagenomes of five microbialite-associated communities from two sites and, at one of them, along a depth gradient (four depths from 1 to 15 m) in Lake Alchichica, Mexico (Saghaï et al. 2015, 2016). The corresponding communities were extremely diverse but, surprisingly, we were able to assemble a noteworthy number of long contigs (>10 kb; 5–8%) affiliating to Chloroflexi, a phylum harboring lineages typically found in the anoxygenic photosynthesizing component of microbial mats (Bryant et al. 2012; Ward et al. 2018). Despite the fact that Chloroflexi were relatively little abundant as reflected by 16S rRNA and conserved single-copy genes (1–3%; Saghaï et al. 2015, 2016), most of these contigs were very long (>200 kb) and represented 40 of the 50 longest contigs across the five metagenomes. The assembly of large contigs from relatively few sequences suggested that these Chloroflexi exhibit little genomic variation. Here, we present comparative analyses of this novel lineage with previous genomic, physiological, and biochemical studies of other Chloroflexi. We also took advantage of the opportunity to assemble genomes directly from metagenomes to gain insight into the natural genetic diversity of Chloroflexi populations in Alchichica microbialites.

Materials and Methods

Sample Collection, DNA Purification, and Sequencing

Microbialite fragments were collected in Winter (January 2012) at two different sites of Lake Alchichica: the Western shore at ca. 0.5–1 m depth (AL-W: 19°25′0.13″ N, 97°24′41.07″ W) and the Northern shore at 1, 5, 10, and 15 m depth (AL-N: 19°25′12.49″ N, 97°24′12.35″ W). Procedures for sample collection, DNA extraction, and shotgun sequencing are described in detail in Saghaï et al. (2015).

Assembly and Binning

Primary assembly was carried out from full individual Illumina AL-W and AL-N read data sets (NCBI Bioproject PRJNA315555) with MEGAHIT (v. 0.3.3; Li et al. 2015) without further filtering and with default parameters except for the k-mer range, which was set to k-min = 27 and k-max = 123 in 10 k-mer size increments. Candidate Chloroflexi contigs were identified as described in Saghaï et al. (2016) and corresponded to contigs longer than 10 kbp, where the majority of genes was affiliated to Chloroflexi and where Chloroflexi genes represented at least 20% of all genes (the actual range was 20–69% of chloroflexal genes). Phylogenetic affinity was determined based on high similarity to sequences in the RefSeq database (minimum best hit identity >50% over at least 80% of the gene length). In these contigs, Chloroflexi genes were found to be far more abundant than genes affiliating to any other taxon (supplementary fig. S1, Supplementary Material online), indicating that they likely were fragments of Chloroflexi genomes. Then, each of the initial read data sets were mapped to their corresponding contig pools with MIRABAIT (MIRA v. 4.9.5, k-mer size = 57; Chevreux et al. 1999) and separately subjected to another round of assembly with SPAdes (v. 3.6.1, default settings; Nurk et al. 2013). The resulting contigs longer than 50 kb were clustered using the ESOM clustering procedure (ESOM tools, v. 1.1; Ultsch and Mörchen 2005) along with six reference genomes chosen for their close phylogenetic

relation to our data set and the availability of a complete genomic sequence: *Anaerolinea thermophila* UNI-1 (NC_014960), *Chloroflexus aggregans* DSM 9485 (NC_011831), *Chloroflexus aurantiacus* J-10-fl (NC_010175), *Roseiflexus castenholzii* DSM 13941 (NC_009767), *Thermomicrobium roseum* DSM 5159 (NC_011959), and *Dehalococcoides mccartyi* 195 (NC_002936). The clustering was performed on nucleotide tetramer frequency distributions computed with the *tetramer_freqs_esom.pl* script (Dick et al. 2009) using a window size of 5,000 nt. ESOM training parameters were set to k-batch training method, 140 × 250 mesh size, radius start 50, 20 training epochs. All other parameters were set to their defaults. The visualization of the ESOM maps was done with a *U*-Matrix background and inverse gray-scale gradient coloring (supplementary fig. S2, Supplementary Material online). The completeness and contamination of the bins were assessed using CheckM with Chloroflexi as lineage marker (Imelfort et al. 2015). Average nucleotide identity (ANI) values were calculated using the OAU program (v 1.2; Yoon et al. 2017). The five assembled bins are available at NCBI Bioproject PRJNA579290.

Reciprocal Recruitment

In order to minimize assembly artifacts, a reciprocal recruitment of sequence reads from each data set to another was performed with the nucmer command implemented in MUMmer (v. 3.21, default settings; Kurtz et al. 2004). Original sequence read data sets were mapped to contigs in each isolated genome bin and used as read pools to cross-match other genome clusters. To ascertain over/under representation of read recruitment, we used the following procedure: for each contig, identity percent values (%idy) of nucmer matched reads were summed over 1 kb windows ($\sum$%idy), allowing the determination of *per contig* median and SD (stddev) values of the $\sum$%idy distribution (supplementary fig. S3, upper panel, Supplementary Material online). Thus, any 1 kb range on a contig with a $\sum$%idy value outside the (median $\pm$ (2× stddev)) range was marked as having a coverage anomaly (supplementary fig. S3, lower panel, Supplementary Material online), thereafter allowing identification of genes spanning tagged regions.

Genome Annotation

Gene prediction was performed on the newly assembled contigs using PRODIGAL (v. 2.6.3, "metagenomic" mode; Hyatt et al. 2010). For a more accurate functional annotation, we used the amino acid sequence of predicted genes. The annotation was done using the BlastP command implemented in DIAMOND (v. 0.7.9, with a maximum e-value of 10^{-5}; Buchfink et al. 2015) with the 2014 edition of the COGs database (Galperin et al. 2015), supplemented with a set of genes from available Chloroflexi genomes (supplementary table 1, Supplementary Material online). In the subsequent analyses, we retained only the best hit to represent each annotated gene.

Phylogenetic Analyses

To reconstruct a reliable phylogenetic tree of the draft Chloroflexi genomes, we selected a set of 30 conserved genes present in single-copy in prokaryotic genomes that we detected in the genome fragments and in publicly available Chloroflexi genomes (supplementary table 2, Supplementary Material online). We also built a phylogenetic tree with the 16S rRNA sequences extracted from each of the draft genomes, together with a selection of the closest Blast hits (Altschul et al. 1990). Amino acid and nucleotide sequences corresponding to the conserved genes were concatenated and then aligned using MAFFT with the L-INS-i algorithm and default parameters (v. 7.310; Katoh and Standley 2013) (supplementary files 1 and 2, Supplementary Material online). 16S rRNA sequences were aligned using the SINA aligner (v. 1.6.0; Pruesse et al. 2012; supplementary file 3, Supplementary Material online). All phylogenetic trees were built with FASTTREE (v. 2.1.3; Price et al. 2010) using either the JTT + CAT approximation for substitution rate heterogeneity or the GRT + CAT model of nucleotide evolution. Finally, the trees were visualized using FIGTREE (v 1.4.4).

Results and Discussion

Reconstruction and Phylogenetic Affiliation of Chloroflexi Genomes

We applied compositional binning to sequence reads mapping to contigs affiliating to Chloroflexi from five microbialite metagenomic data sets in Lake Alchichica (AL-W, AL-N-1, AL-N-5, AL-N-10, and AL-N-15) and several reference genomes. The ESOM visualization maps (supplementary fig. S2, Supplementary Material online) of all data sets showed two major features. First, each reference genome was contained in a single, well-defined cluster. Second, the candidate Chloroflexi contigs displayed a similar pattern across all five data sets, with three to five clusters (named A–E; supplementary fig. S4, Supplementary Material online). However, comparison of the cumulative size of the cluster contigs, estimated completeness and estimated contamination collectively suggested that only cluster A contigs could be considered as candidate bins (supplementary tables 1 and 3, Supplementary Material online). Interestingly, an analysis of sequence divergence revealed a very high level of similarity across the five data sets (AL-W and AL-N-1, AL-N-5, AL-N-10, and AL-N-15) in this cluster. Both amino acid and nucleotide sequences showed that 92% of all genes were shared between data sets and nearly identical (i.e., average and median percentages of gene alignments maximum bitscores approaching 100% for amino acids and nucleotide

A

	AL-N-1	AL-N-5	AL-N-10	AL-N-15	AL-W
AL-N-1	100.00	99.26	98.57	95.72	96.24
AL-N-5	95.84	100.00	95.29	92.08	93.44
AL-N-10	97.74	97.88	100.00	95.43	95.73
AL-N-15	98.74	98.38	99.23	100.00	97.98
AL-W	94.25	94.80	94.54	93.04	100.00

B

	AL-N-1	AL-N-5	AL-N-10	AL-N-15	AL-W
AL-N-1	100.00	98.87	98.08	95.28	95.80
AL-N-5	95.41	100.00	95.06	91.75	93.07
AL-N-10	97.27	97.68	100.00	95.12	95.34
AL-N-15	98.29	98.08	98.95	100.00	97.57
AL-W	93.81	94.45	94.15	92.66	100.00

Fig. 1.—Pairwise percentage of genes with 100% identity over theHSP, as defined by Blast, in (*A*) nucleotide and (*B*) amino acid sequences between the five bins. The genome bins correspond to those forming the cluster A in ESOM analyses.

sequences; fig. 1 and supplementary fig. S5*A–F*, Supplementary Material online). Furthermore, computed ANI values between all draft genome pairs (>99%; supplementary fig. S5*G*, Supplementary Material online) were higher than commonly used speciation cutoff values (95–96%; Goris et al. 2007; Richter and Rosselló-Móra 2009). Even though the contigs could not be merged to form a single reliable scaffold, a synteny analysis conducted on the individual contigs at nucleotide level showed that these were highly conserved along most of their length among the different data sets (fig. 2 and supplementary figs. S5*H*, *I*, and S6, Supplementary Material online). Altogether, these observations support the idea that cluster A contigs actually constituted a set of five highly related chloroflexal genomes that derived from a single lineage with variable gene content. Phylogenetic reconstructions based on a set of 30 conserved single-copy genes supported this hypothesis (at both amino acid and nucleotide levels), revealing that the five metagenomic sequences clustered together within the Chloroflexales and formed a distantly related sister group to Roseiflexineae and *Kouleothrix aurantiaca* (fig. 3*A* and *B*). We tentatively propose to consider this novel Chloroflexi quasiclonal lineage as a new species, *Candidatus Lithoflexus mexicanus*.

Comparison of Low Coverage Regions

Despite the high similarity of the five *Ca. L. mexicanus* genome drafts, they differed in size (7.9–8.4 Mb, table 1). To understand the origin of these differences, we analyzed contig regions with poor read coverage in the other data sets (see Reciprocal Recruitment in Materials and Methods). These regions, and the genes they contain (missing or divergent, abbreviated M/D hereafter) might represent genomic islands present in one strain and absent in others (accounting for 8–25% of the genes), but they might also simply result from greater sequence divergence in those regions. Individual and cumulative gene counts (supplementary fig. S7,

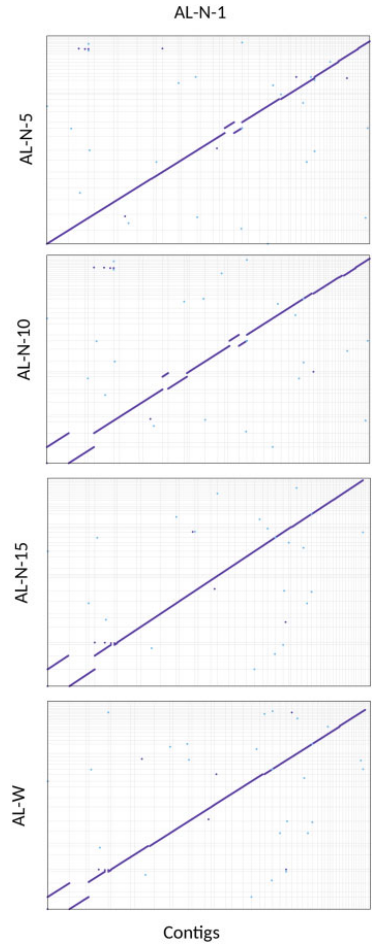

Fig. 2.—Cluster A contigs synteny plots between AL-N-1 (vertical lines) and the four other bins (horizontal lines). Contigs pairs scatter plots were calculated using nucmer/mummerplot alignment and plotting features, and the layout was optimized to simplify the colinear display of contigs. Note that the apparent ordering of contigs is not related to their true physical linkage. The full set of synteny plots is available in supplementary fig. S6, Supplementary Material online.

Supplementary Material online) showed that there was between 96 and 600 M/D genes when comparing pairwise reads/contigs data sets (562–1,844 when comparing the number of M/D genes in any data set individually compared with other reads data sets). In brief, these numbers were consistent with the total nucleotide length, contig, and gene number of each data set, AL-W being the least and AL-N-15 the most profuse data sets, respectively. However, these differences did not seem to be randomly distributed among COG categories (supplementary fig. S8, Supplementary Material online). Leaving aside the "RNA processing and modification" category that represents only five genes, the most salient features were as follows. First, the two largest bins, AL-W and AL-N-5 (table 1), showed that M/D genes were

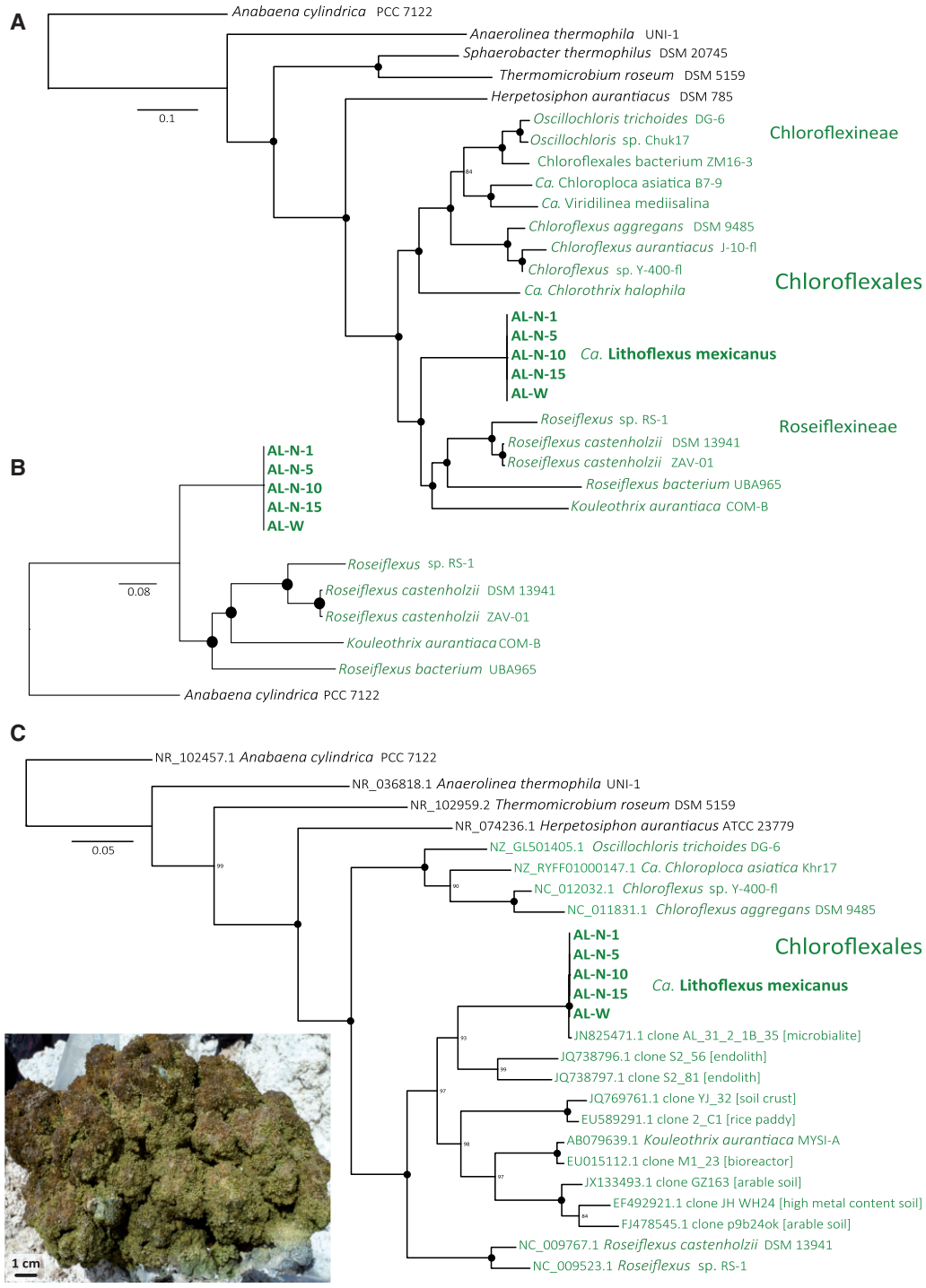

Fig. 3.—Approximate maximum likelihood phylogenetic tree of concatenated amino acid (*A*) and nucleotide (*B*) sequences corresponding to 30 conserved single-copy genes present in the five draft genomes and in closely related genomes. (*C*) Approximate maximum likelihood phylogenetic tree of 16S rRNA sequences present in the five draft genomes and their closest environmental sequences in GenBank. Black dots indicate 100% support. A picture of the AL-N-1 microbialite is shown (scale bar corresponds to 1 cm).

Table 1
Genome Statistics of *Ca. Lithoflexus mexicanus*

	AL-W	AL-N-1	AL-N-5	AL-N-10	AL-N-15
No. of paired end reads[a]	306,787,264	130,153,392	110,282,579	97,849,068	131,562,915
No. of reads in cluster A	5,919,172	1,835,724	6,039,184	5,387,044	2,244,450
Prop. cluster A reads (%)	0.96	0.71	2.74	2.75	0.85
No. of contigs	41	49	37	42	40
Cumulative contig length (bp)	8,288,841	8,154,893	8,421,798	8,242,010	7,914,973
Mean contig size (bp)	202,166	166,426	227,616	196,238	197,874
N50 (bp)	370,194	279,010	375,049	367,192	405,938
% GC	49.14	49.01	49.06	48.92	48.84
No. of predicted genes	7,423	7,266	7,522	7,331	7,040
No. COG hits	5,355	5,210	5,395	5,233	5,055
Perc. COG hits	72.14	71.70	71.72	71.38	71.80
No. of single-copy gene families[b]	39	39	39	38	38
Total occurrences of single-copy gene families	68[c]	44	43	42	43

[a]Raw number of total pairs of reads produced (it implies the same number of forward and reverse sequences).
[b]See supplementary table 2, Supplementary Material online.
[c]Due to the presence of a partially duplicated ribosomal proteins operon.

distributed across all categories, with only a few exceptions. Second, AL-W appears to have an extra set of genes in the "Translation, ribosomal structure and biogenesis" and, to a lesser extent, the "Amino acid transport and metabolism" categories, which is due to the presence of a duplicated ribosomal protein operon in this data set. Third, the pattern of "Mobilome, prophages and transposons" category of M/D genes was uneven across data sets, despite the fact that overall gene counts in this category remained comparable ($n = 30$–40, mainly including rayT, IS5, and TnpA transposases; supplementary table 4, Supplementary Material online).

Photochemical Apparatus

Candidatus Lithoflexus mexicanus lacks homologs of known chlorosome envelope proteins (*csmA*, *csmM*, and *csmN*) and genes encoding bacteriochrolophyll c (*bchK* and *bchU*), which are typical of Chloroflexineae (Orf and Blankenship 2013). In turn, similar to all known Roseiflexineae species (Hanada and Pierson 2006), *Ca. L. mexicanus* possesses a light-harvesting core antenna complex. This protein–pigment complex is associated with two spectral types (B800–880 in *R. castenholzii*; Collins et al. 2009, 2010) of bacteriochlorophyll a (BChl a; supplementary table 5, Supplementary Material online). In addition, *Ca. L. mexicanus* contains genes encoding the three subunits of type II reaction centers (RC), L (*pufL*), M (*pufM*), and the tetraheme Cyt c_{554} (*pufC*; Bryant et al. 2012). Both *pufL* and *pufM* genes are fused as observed in *Roseiflexus* spp. (Yamada et al. 2005). Finally, *Ca. L. mexicanus* can synthetize carotenoids (*crtBIOPU* and *cruA*), which are involved in photoprotection and light harvesting (Maresca et al. 2008). Whereas some Chloroflexales have CrtY- (Chloroflexineae and *K. aurantiaca*) or CrtL-type (*Roseiflexus* spp.) lycopene cyclases (catalyzing the final step in carotenoid biosynthesis), *Ca. L. mexicanus* encodes a homolog of the CruA-type lycopene cyclase found in the nonphototrophic *Herpetosiphon aurantiacus* (Kiss et al. 2011; Bryant et al. 2012).

Proteins Adapted to Aerobic versus Anaerobic Conditions

Consistently with the diel variations in oxygen within the microbialites, and similar to other mat-inhabiting Chloroflexales, *Ca. L. mexicanus* harbors a set of enzymes specialized to work in the presence or absence of oxygen (Tang et al. 2011; Bryant et al. 2012). We detected genes encoding both AcsF and BchE proteins, which catalyze the isocyclic ring formation of chlorophyll under aerobic and anaerobic conditions, respectively (Ouchane et al. 2004; Tang et al. 2009). *Candidatus Lithoflexus mexicanus* also encodes the three subunit of the Mg-chelatase (*bchHID*), responsible for the insertion of Mg in porphyrin during the first step of BChl synthesis. It has been proposed that distinct *bchH* gene products may participate to the biosynthesis of different BChl in some organisms (Eisen et al. 2002). However, BChl-encoding genes different from BChl a appear to be absent in *Ca. L. mexicanus*. Therefore, the presence of two and three nonidentical homologs of *bchH* and *bchI*, respectively, might rather reflect differential expression under aerobic and anaerobic conditions. Likewise, *Ca. L. mexicanus* encodes two forms of the cobalt chelatase, a key component in the synthesis of cobalamin, acting under both aerobic (*cobNST*) and anaerobic (*cbiK*) conditions. Last, we also identified genes encoding protein pairs functioning in either aerobic or anaerobic conditions involved in pyruvate metabolism and the TCA cycle (supplementary table 5, Supplementary Material online).

Electron Transport

Candidatus Lithoflexus mexicanus encodes core electron transport components, including a succinate dehydrogenase/fumarate reductase (*sdhBAC*) and a cytochrome c oxidase (supplementary table 5, Supplementary Material online). Similar to many anoxygenic photosynthetic bacteria (Proteobacteria and other Chloroflexales), *Ca. L. mexicanus* harbors two sets of loosely arranged genes (*nuoA* to *nuoN*) encoding enzymes for the NADH: quinone oxidoreductase, which catalyzes the transport of electrons in the oxidative phosphorylation pathway. In addition, it possesses an electron transport complex named alternative complex III (ACIII), which uses menaquinone as liposoluble electron and proton carrier (Hale et al. 1983; Gao et al. 2009, 2010; Majumder et al. 2013). In *Ca. L. mexicanus*, like in *Roseiflexus* spp., the ACIII is encoded by a single six-gene operon (*actABCDEF*) that lacks the gene encoding the component G in *Chloroflexus* spp. (Hanada et al. 2002; Tang et al. 2011). It has been proposed that the electrons from this complex reduce auracyanin, a blue-copper protein serving as mobile electron carrier from ACIII to the RC, where it could act as an electron donor for the cytochrome c (Trost et al. 1988; McManus et al. 1992; Tsukatani et al. 2009). We also detected genomic fragments containing Ni–Fe hydrogenases suggesting that *Ca. L. mexicanus* can oxidize hydrogen and use it as an electron source (van der Meer et al. 2010). Carbon monoxide could represent another electron source in aerobic and semiaerobic conditions, as the draft genomes encoded a putative molybdopterin-containing carbon monoxide dehydrogenase (*coxLM*).

Mixotrophy in *Ca. L. mexicanus*

Comparative genomic analyses of the five draft *Ca. L. mexicanus* genomes and those of other Chloroflexales readily allowed the detection of genes involved in the first cycle of the 3HP pathway (supplementary table 5, Supplementary Material online). However, a deeper mining in the five full metagenomic data sets revealed the presence of sequences >50% identical at amino acid level with known Chloroflexales proteins involved in the second cycle of this pathway, suggesting that *Ca. L. mexicanus* likely has the genetic capacity to use the complete 3HP bicycle. In contrast, genes encoding key proteins involved in the Calvin-Benson cycle (e.g., ribulose 1,5-biphosphate carboxylase) were lacking.

We detected a complete set of genes for the oxidative TCA cycle as well as genes involved in the assimilation of low-molecular weight organic compounds typically released by mat-associated cyanobacteria during, for instance, photorespiration (e.g., glycolate; Bateson & Ward 1988) and fermentation (e.g., lactate; Anderson et al. 1987; Nold and Ward 1996; Kim et al. 2015; supplementary table 5, Supplementary Material online). *Candidatus Lithoflexus mexicanus* might thus share the ability to simultaneously incorporate organic and inorganic carbon with *Roseiflexus* and *Chloroflexus* spp., in which both 3HP and TCA cycles have been shown to share similar gene and protein expression patterns over a diel period (Klatt et al. 2013; Kim et al. 2015). We also found genes encoding both phosphoenolpyruvate carboxykinase (*pckA*) and phosphoenolpyruvate carboxylase (*ppc*), suggesting that *Ca. L. mexicanus* can assimilate inorganic carbon and also refuel the TCA cycle through CO_2-anaplerotic pathways, as shown in other phototrophic bacteria (Evans et al. 1966; Tang and Blankenship 2010; Tang et al. 2011).

Nitrogen and Sulfur Metabolism

Candidatus Lithoflexus mexicanus harbors genes encoding an ammonia transporter (*amtB*) and a set of typical enzymes involved in ammonia production (supplementary table 5, Supplementary Material online) but lacks genes related to both nitrogen fixation and nitrate reduction. In that regard, it is more similar to *Chloroflexus* spp., which only use ammonia and some amino acids as nitrogen sources (Hanada and Pierson 2006; Tang et al. 2011), whereas *Roseiflexus* spp. encode a nitrogenase (van der Meer et al. 2010).

Candidatus Lithoflexus mexicanus is unlikely to grow autotrophically on sulfide given the absence of homologs to genes involved in dissimilatory sulfur metabolism (e.g., *dsr*, *sox*). However, our findings suggest that *Ca. L. mexicanus* has the potential to perform assimilatory sulfate reduction in a similar way to other *Roseiflexus* and *Chloroflexus* spp. (Bryant et al. 2012). Indeed, its genome encodes a fusion of a Sat-type ATP-sulfurylase and a CysC-type APS kinase that can transform sulfate directly into 3′-phosphoadenosine-5′-phosphosulfate (PAPS; Bryant et al. 2012). PAPS may be used to produce a variety of cellular sulfate compounds or be reduced to sulfite by the PAPS reductase (*cysH*; Tang et al. 2011). This sulfite can then be reduced into sulfide by a sulfite reductase and incorporated into cysteine using cysteine synthases (*cysKM*).

Ecological implications of the low genetic variation in *Ca. L. mexicanus*

Our phylogenetic and functional prediction analyses show that the chloroflexal bins reconstructed from five metagenomes coming from two sites and four different depths correspond to a single, low abundant (0.7–2.8% of the reads; table 1) and highly similar population (table 2). The 16S rRNA gene fragments extracted from *Ca. L. mexicanus* bins were closely related to a sequence recovered in a shallow microbialite sampled in a previous campaign in Lake Alchichica, AL_31_2_1B_35 (Couradeau et al. 2011) and formed a clade with two environmental sequences from endolithic communities (fig. 3C; Roush et al. 2017). This suggests that this clade is specialized in endolithic lifestyle.

Table 2
Main Functional Characteristics of *Ca. Lithoflexus mexicanus*

		Ca. L. mexicanus
Photosynthetic apparatus	Bchl a	+
	Bchl c	–
	Chlorosomes	–
	B808–866 core antenna complex	+
	Type II reaction center	+
Electron transport systems	Alternative complex III	+
	Auracyanin	+
	Oxidative phosphorylation pathway	+
Central carbon metabolism	3HP bicycle	+
	Calvin–Benson–Benson cycle	–
	CO_2 anaplerotic pathways	+
Nitrogen metabolism	N_2 fixation	–
	Ammonia and amino acid assimilation	+
Sulfur metabolism	Autotrophic growth on sulfide	–
	Assimilatory sulfate reduction	+

Because *Ca. L. mexicanus* is phototrophic and largely depends on light to grow, several alternative hypotheses can be evoked to explain its apparent low genetic diversity. First, *Ca. L. mexicanus* could be able to grow at various light intensities, as observed in chlorosome-containing Chloroflexales (Taisova et al. 2014). However, antenna equipped with BChl a absorb light only in a narrow range of the infrared (Taisova et al. 2014) and all five bins display 100% identity at the nucleotide level for various components of the photosynthetic apparatus (e.g., *crtB*, *cruA*, alpha and beta subunits of the core antenna; data not shown), suggesting a similar potential for light harvesting. More likely hypotheses might thus be that *Ca. L. mexicanus* occupies slightly different depths within microbialites (deeper in upper microbialites, more superficial in deeper microbialites) or regulates its metabolic activity according to the fluctuations in resource availability (e.g., light, low-molecular weight organic compounds) and chemical gradients (e.g., oxygen) occurring over the diel cycle. Interestingly, metatranscriptomic and metabolomic studies conducted over entire microbial mats have suggested that both Chloroflexales and Cyanobacteria exhibit maximal metabolic rates for specific processes at different times during the diel cycle (Klatt et al. 2013; Kim et al. 2015). One can thus hypothesize that the timing of the regulation patterns of *Ca. L. mexicanus* could differ depending on the microbialite depth in the water column. Alternatively, differentiated diel migration patterns could also explain the presence of a highly similar population along the depth gradient. Further metatranscriptomic studies combining both spatial (i.e., different microbialite depth/layers) and temporal (i.e., different time points in the diel cycle) scales could give valuable indications on the spatial distribution and metabolic activity of *Ca. L. mexicanus* within Alchichica microbialites.

Conclusions

The present study expands our knowledge on the diversity and ecology of an important group of anoxygenic phototrophs in contemporary microbial mats. *Candidatus Lithoflexus mexicanus* forms an ecologically coherent clade with organisms specialized from endolithic communities, which is distinct from the well-characterized *Chloroflexus* and *Roseiflexus* spp. typically associated to nonlithifying mats. More similar to Roseiflexineae in its light-harvesting apparatus and electron transport pathways, it carried all the genes necessary to a mixotrophic lifestyle under both aerobic and anaerobic conditions. Finally, *Ca. L. mexicanus* displayed unique population genomics features compared with other community members, as reflected by the high degree of genomic conservation along the depth gradient, suggesting adaptations to a very specific ecological niche within Alchichica microbialites.

Acknowledgments

This work was supported by the European Research Council Grants *ProtistWorld* (PI P.L.G., Grant No. 322669) under the European Union's Seventh Framework Program, the RTP Génomique environnementale of the CNRS (project MetaStrom, PI D.M.), and the ANR project Microbialites (PI P.L.G.).

Literature Cited

Allewalt JP, Bateson MM, Revsbech NP, Slack K, Ward DM. 2006. Effect of temperature and light on growth of and photosynthesis by *Synechococcus* isolates typical of those predominating in the Octopus Spring microbial mat community of Yellowstone National Park. Appl Environ Microbiol. 72(1):544–550.

Allwood AC, Walter MR, Kamber BS, Marshall CP, Burch IW. 2006. Stromatolite reef from the early Archaean era of Australia. Nature 441(7094):714–718.
Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol. 215(3):403–410.
Anderson KL, Tayne TA, Ward DM. 1987. Formation and fate of fermentation products in hot spring cyanobacterial mats. Appl Environ Microbiol. 53(10):2343–2352.
Bateson M, Ward DM. 1988. Photoexcretion and fate of glycolate in a hot spring cyanobacterial mat. Appl Environ Microbiol. 54(7):1738–1743.
Becraft ED, et al. 2015. The molecular dimension of microbial species: 1. Ecological distinctions among, and homogeneity within, putative ecotypes of *Synechococcus* inhabiting the cyanobacterial mat of Mushroom Spring, Yellowstone National Park. Front Microbiol. 6:590.
Bhaya D, et al. 2007. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. ISME J. 1(8):703–713.
Bryant DA, et al. 2012. Comparative and functional genomics of anoxygenic green bacteria from the taxa *Chlorobi*, *Chloroflexi*, and *Acidobacteria*. In: Burnap R, Vermaas W, editors. Functional genomics and evolution of photosynthetic systems. Dordrecht (Netherlands): Springer. p. 47–102.
Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods. 12(1):59–60.
Chevreux B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information. German Conf Bioinformatics. 99: 45–56.
Collins AM, et al. 2010. Light-harvesting antenna system from the phototrophic bacterium *Roseiflexus castenholzii*. Biochemistry 49(35):7524–7531.
Collins AM, Xin Y, Blankenship RE. 2009. Pigment organization in the photosynthetic apparatus of *Roseiflexus castenholzii*. Biochim Biophys Acta. 1787(8):1050–1056.
Couradeau E, et al. 2011. Prokaryotic and eukaryotic community structure in field and cultured microbialites from the alkaline Lake Alchichica (Mexico). PLoS One. 6(12):e28767.
Des Marais DJ. 1990. Microbial mats and the early evolution of life. Trends Ecol Evol. 5(5):140–144.
Dick GJ, et al. 2009. Community-wide analysis of microbial genome sequence signatures. Genome Biol. 10(8):R85.
Dupraz C, Visscher PT. 2005. Microbial lithification in marine stromatolites and hypersaline mats. Trends Microbiol. 13(9):429–438.
Eisen JA, et al. 2002. The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. Proc Natl Acad Sci U S A. 99(14):9509–9514.
Evans MCW, Buchanan BB, Arnon DI. 1966. New cyclic process for carbon assimilation by a photosynthetic bacterium. Science 152(3722):673.
Galperin MY, Makarova KS, Wolf YI, Koonin EV. 2015. Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43(Database issue):D261–D269.
Gao X, Xin Y, Bell P, Wen J, Blankenship RE. 2010. Structural analysis of alternative complex III in the photosynthetic electron transfer chain of *Chloroflexus aurantiacus*. Biochemistry 49(31):6670–6679.
Gao X, Xin Y, Blankenship RE. 2009. Enzymatic activity of the alternative complex III as a menaquinol: auracyanin oxidoreductase in the electron transfer chain of *Chloroflexus aurantiacus*. FEBS Lett. 583(19):3275–3279.
Garcia SL, et al. 2018. Contrasting patterns of genome-level diversity across distinct co-occurring bacterial populations. ISME J. 12(3):742–755.
Goris J, et al. 2007. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 57(1):81–91.
Gutiérrez-Preciado A, et al. 2018. Functional shifts in microbial mats recapitulate early Earth metabolic transitions. Nat Ecol Evol. 2(11):1700–1708.
Hale MB, Blankenship RE, Fuller RC. 1983. Menaquinone is the sole quinone in the facultatively aerobic green photosynthetic bacterium *Chloroflexus aurantiacus*. Biochim Biophys Acta. 723(3):376–382.
Hanada S, Pierson BK. 2006. Proteobacteria: delta, epsilon subclass. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The prokaryotes. Vol. 7. New York (NY): Springer. p. 815–842.
Hanada S, Takaichi S, Matsuura K, Nakamura K. 2002. *Roseiflexus castenholzii* gen. nov., sp. nov., a thermophilic, filamentous, photosynthetic bacterium that lacks chlorosomes. Int J Syst Evol Microbiol. 52(Pt 1):187–193.
Hyatt D, et al. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11(1):119.
Imelfort M, Skennerton CT, Parks DH, Tyson GW, Hugenholtz P. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25(7):1043–1055.
Kashtan N, et al. 2014. Single-cell genomics reveals hundreds of coexisting subpopulations in wild Prochlorococcus. Science 344(6182):416–420.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 30(4):772–780.
Kilian O, et al. 2007. Responses of a thermophilic *Synechococcus* isolate from the microbial mat of Octopus Spring to light. Appl Environ Microbiol. 73(13):4268–4278.
Kim YM, et al. 2015. Diel metabolomics analysis of a hot spring chlorophototrophic microbial mat leads to new hypotheses of community member metabolisms. Front Microbiol. 6:209.
Kiss H, et al. 2011. Complete genome sequence of the filamentous gliding predatory bacterium *Herpetosiphon aurantiacus* type strain (114-95T). Stand Genomic Sci. 5(3):356–370.
Klatt CG, et al. 2013. Temporal metatranscriptomic patterning in phototrophic Chloroflexi inhabiting a microbial mat in a geothermal spring. ISME J. 7(9):1775–1789.
Klatt CG, et al. 2011. Community ecology of hot spring cyanobacterial mats: predominant populations and their functional potential. ISME J. 5(8):1262–1278.
Kurtz S, et al. 2004. Versatile and open software for comparing large genomes. Genome Biol. 5(2):R12.
Li D, Liu C-M, Luo R, Sadakane K, 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):1674–1676.
Luo C, et al. 2011. Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. Proc Natl Acad Sci U S A. 108(17):7200–7205.
Majumder ELW, King JD, Blankenship RE. 2013. Alternative complex III from phototrophic bacteria and its electron acceptor auracyanin. Biochim Biophys Acta. 1827(11–12):1383–1391.
Maresca JA, Graham JE, Bryant DA. 2008. The biochemical basis for structural diversity in the carotenoids of chlorophototrophic bacteria. Photosynth Res. 97(2):121–140.
McManus JD, et al. 1992. Isolation, characterization, and amino acid sequences of auracyanins, blue copper proteins from the green photosynthetic bacterium *Chloroflexus aurantiacus*. J Biol Chem. 267(10):6531–6540.
Mobberley JM, et al. 2015. Inner workings of thrombolites: spatial gradients of metabolic activity as revealed by metatranscriptome profiling. Sci Rep. 5:12601.
Nold SC, Ward DM. 1996. Photosynthate partitioning and fermentation in hot spring microbial mat communities. Appl Environ Microbiol. 62(12):4598–4607.

Nowack S, et al. 2015. The molecular dimension of microbial species: 2. *Synechococcus* strains representative of putative ecotypes inhabiting different depths in the Mushroom Spring microbial mat exhibit different adaptive and acclimative responses to light. Front Microbiol. 6:626.
Nurk S, et al. 2013. Assembling single-cell genomes and mini-metagenomes from chimeric MDA products. J Comput Biol. 20(10):714–737.
Olsen MT, et al. 2015. The molecular dimension of microbial species: 3. Comparative genomics of *Synechococcus* strains with different light responses and in situ diel transcription patterns of associated putative ecotypes in the Mushroom Spring microbial mat. Front Microbiol. 6:604.
Orf GS, Blankenship RE. 2013. Chlorosome antenna complexes from green photosynthetic bacteria. Photosynth Res. 116(2–3):315–331.
Ouchane S, Steunou AS, Picaud M, Astier C. 2004. Aerobic and anaerobic Mg-protoporphyrin monomethyl ester cyclases in purple bacteria: a strategy adopted to bypass the repressive oxygen control system. J Biol Chem. 279(8):6385–6394.
Price MN, Dehal PS, Arkin AP. 2010. FastTree 2: approximately maximum-likelihood trees for large alignments. PLoS One. 5(3):e9490.
Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28(14):1823–1829.
Richter M, Rosselló-Móra R. 2009. Shifting the genomic gold standard for the prokaryotic species definition. Pro Natl Acad Sci U S A. 106(45):19126–19131.
Riding R. 2000. Microbial carbonates: the geological record of calcified bacterial–algal mats and biofilms. Sedimentology 47:179–214.
Roush D, Couradeau E, Guida B, Neuer S, Garcia-Pichel F. 2017. A new niche for anoxygenic phototrophs as endoliths. Appl Environ Microbiol. 84:e02055–17.
Saghaï A, et al. 2016. Comparative metagenomics unveils functions and genome features of microbialite-associated communities along a depth gradient. Environ Microbiol. 18(12):4990–5004.
Saghaï A, et al. 2015. Metagenome-based diversity analyses suggest a significant contribution of non-cyanobacterial lineages to carbonate precipitation in modern microbialites. Front Microbiol. 6:797.
Stal LJ. 2001. Coastal microbial mats: the physiology of a small-scale ecosystem. South African J Bot. 67(3):399–410.
Taisova AS, Yakovlev AG, Fetisova ZG. 2014. Size variability of the unit building block of peripheral light-harvesting antennas as a strategy for effective functioning of antennas of variable size that is controlled in vivo by light intensity. Biochemistry (Mosc). 79(3):251–259.
Tang K-H, et al. 2011. Complete genome sequence of the filamentous anoxygenic phototrophic bacterium *Chloroflexus aurantiacus*. BMC Genomics. 12(1):334.
Tang K-H, Blankenship RE. 2010. Both forward and reverse TCA cycles operate in green sulfur bacteria. J Biol Chem. 285(46):35848–35854.
Tang K-H, Tang Y, Blankenship RE. 2011. Carbon metabolic pathways in phototrophic bacteria and their broader evolutionary implications. Front Microbiol. 2:165.
Tang K-H, Wen J, Li X, Blankenship RE. 2009. Role of the AcsF protein in *Chloroflexus aurantiacus*. J Bacteriol. 191(11):3580–3587.
Tettelin H, et al. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial 'pan-genome'. Proc Natl Acad Sci U S A. 102(39):13950–13955.
Tice MM, Lowe DR. 2004. Photosynthetic microbial mats in the 3,416-Myr-old ocean. Nature 431(7008):549–552.
Trost JT, McManus JD, Freeman JC, Ramakrishna BL, Blankenship RE. 1988. Auracyanin, a blue copper protein from the green photosynthetic bacterium *Chloroflexus aurantiacus*. Biochemistry. 27(20):7858–7863.
Tsukatani Y, et al. 2009. Characterization of a blue-copper protein, auracyanin, of the filamentous anoxygenic phototrophic bacterium *Roseiflexus castenholzii*. Arch Biochem Biophys. 490(1):57–62.
Ultsch A, Mörchen F. 2005. ESOM-Maps: tools for clustering, visualization, and classification with Emergent SOM. Tech. Rep. Dept. Math. Comput. Sci. Univ. Marburg, Ger. 1–7.
van Gemerden H. 1993. Microbial mats: a joint venture. Mar. Geol. 113(1-2):3–25.
van der Meer MTJ, et al. 2010. Cultivation and genomic, nutritional, and lipid biomarker characterization of *Roseiflexus* strains closely related to predominant in situ populations inhabiting Yellowstone hot spring microbial mats. J Bacteriol. 192(12):3033–3042.
Ward DM, Ferris MJ, Nold SC, Bateson MM. 1998. A natural view of microbial biodiversity within hot spring cyanobacterial mat communities. Microbiol Mol Biol Rev. 62(4):1353–1370.
Ward LM, Hemp J, Shih PM, McGlynn SE, Fischer WW. 2018. Evolution of phototrophy in the *Chloroflexi phylum* driven by horizontal gene transfer. Front Microbiol. 9:260.
Yamada M, et al. 2005. Structural and spectroscopic properties of a reaction center complex from the chlorosome-lacking filamentous anoxygenic phototrophic bacterium *Roseiflexus castenholzii*. J Bacteriol. 187(5):1702–1709.
Yoon SH, et al. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 110(10):1281–1286.

Associate editor: Esther Angert