

Draft genome of *Chloroflexus* sp. MS-CIW-1, of the *Chloroflexus* sp. MS-G group from Mushroom Spring, Yellowstone National Park

Amanda N. Shelton,¹ Feiqiao B. Yu,^{1,2} Freddy Bunbury,¹ Jia Yan,² Carlos Rivas,¹ Arthur Grossman,¹ Devaki Bhaya¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Chloroflexus* sp. MS-CIW-1 was isolated from a phototrophic mat in Mushroom Spring, an alkaline hot spring in Yellowstone National Park, WY, USA. We report the draft genome of 4.8 Mb consisting of 6 contigs with 3755 protein-coding genes and a GC content of 54.45%.

KEYWORDS *Chloroflexus*, phototrophic mat, chloroflexota, hot spring, anoxygenic phototroph, phototrophs, genomes, Mushroom Spring

*C*hloroflexus sp. MS-CIW-1 was isolated from a mat core sampled at 60°C in the runoff channels of Mushroom Spring, an alkaline hot spring in Yellowstone National Park, WY, USA (44.538714, -110.798022) on 2006/09/12. A liquid enrichment culture in DH10 medium (1) with 500 mg/L yeast extract and tryptone was serially diluted and then plated onto PE medium (2) at 50°C under continuous white light (50 $\mu\text{mol m}^{-2} \text{ s}^{-1}$). The edges of orange motile colonies were picked and passaged twice more in liquid and on plates. Axenicity was confirmed by light microscopy and by genome sequencing. This isolate is routinely used in our lab to study the physiology of filamentous anoxygenic phototrophs and their interactions with other members of the mat community. Cells are maintained in liquid PE medium without shaking. Our original analysis of the partial 16S sequence using the BLAST webserver (3) suggested the closest two 16S sequences were from *Chloroflexus* sp. Y-396-1 (GCF_000516515), from the nearby Octopus Spring, and *Chloroflexus* sp. MS-G (GCF_000735195), also from Mushroom Spring (4).

We isolated DNA by a phenol-chloroform-based method following bead beating (5). Illumina libraries were created with the Nextera XT Kit and were sequenced using a 2 \times 150 bp configuration on Illumina's iSeq, yielding 5,094,597 reads. For long-read sequencing, DNA was not sheared prior to library preparation with Oxford Nanopore's SQK-LSK109 PCR-Free Ligation Kit, and sequencing was performed with the Oxford Nanopore MinION flow cell (FLO-MIN106) to obtain 391,464 reads with a median length of 7,000 bp. Basecalling was performed with Guppy. Read trimming and filtering were performed by BBTools (v38.73) (6), quality assessed with FastQC (v0.11.9) (7), followed by hybrid assembly using Unicycler (v0.4.8) (8), and quality control by QualiMap (v.2.2.2) (9) and QUAST (v5.0.2) (10) using the workflow as described by Jin et al. (11). The high-quality draft genome consists of 6 contigs, with an N50 of 3,270,211 bp, with a total length of 4,838,670 bp. Mean coverage was 209 \times for the short reads and 621 \times for long reads. The GC content was 54.45%.

We compared this genome to the available *Chloroflexus* genomes in RefSeq and found the closest relative by ANI [fastANIv0.1.3 on KBase (12, 13)] was strain MS-G with an identity of 98.4%. An alternative approach using GTDB-Tk [v1.7.0 on KBase (14, 15)] placed MS-CIW-1 in "s__Chloroflexus sp000735195," the MS-G species group. All five genomes in this species group on GTDB R207 (16, 17) are MAGs or isolates from

Editor Julia A. Maresca, SUNY College of Environmental Science and Forestry, USA

Address correspondence to Devaki Bhaya, dbhaya@carnegiescience.edu.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 1 August 2023

Accepted 15 January 2024

Published 1 February 2024

Copyright © 2024 Shelton et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Yellowstone National Park. Two of the three 16S loci in MS-CIW-1 are identical, the third is 99.73% identical. These 16S are 99.52%–99.66% identical to MS-G ([KR230107](#)) and 98.06%–98.27% identical to Y-396-1 ([AJ308498](#)) [GeneiousPrime v2022.1.1 and MUSCLE v3.8.425 (18)]. MS-CIW-1 is the most contiguous genome available for the *Chloroflexus* sp. MS-G group.

We annotated MS-CIW-1 using the NCBI prokaryotic genome annotation pipeline (PGAP v6.5) (19), predicting 3,755 protein-coding genes (3,778 including pseudogenes), 3 16S rRNA loci, 50 tRNA, and 4 CRISPR arrays. The CRISPR systems were classified as type IIIA and type I by CRISPRCasfinder (accessed 06 July 2023) (20, 21). Pangenome analysis [KBase Compute Pangenome v0.0.7 (13)] with MS-G found 3,399 homolog families are shared between MS-G and MS-CIW-1. Genome comparison using Mauve aligner (22) on GeneiousPrime reveals different placements of some transposons and nearby genes between the two genomes, suggesting active transposition since these genomes split.

ACKNOWLEDGMENTS

We thank Mary Bateson and all members of the Ward lab for their assistance in sampling. We acknowledge support by BBSRC-NSF/BIO #1921429 awarded to Devaki Bhaya and Arthur Grossman, NSF/EF #2125965 awarded to Devaki Bhaya, and the Carnegie Institution for Science.

Sampling was performed under NPS Park Permits: YELL-5494 (to David Ward, multiyear), YELL-5660 (to Devaki Bhaya, 2007–2008), YELL 5694 (to Devaki Bhaya 2007–2009).

AUTHOR AFFILIATIONS

¹Department of Plant Biology, Division of Biosphere Sciences and Engineering, Carnegie Institution for Science, Stanford, California, USA

²Chan Zuckerberg Biohub, Stanford, California, USA

AUTHOR ORCIDs

Amanda N. Shelton  <http://orcid.org/0000-0001-9296-8331>

Devaki Bhaya  <http://orcid.org/0000-0001-7965-4258>

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation (NSF)	1921429	Arthur Grossman
		Devaki Bhaya
National Science Foundation (NSF)	2125965	Devaki Bhaya
Carnegie Institution for Science (CIS)		Arthur Grossman
		Devaki Bhaya

DATA AVAILABILITY

This Whole-Genome Shotgun project has been deposited in DDBJ/ENA/GenBank under the accession no. [JAUBWR0000000000](#). The version described in this paper is version [JAUBWR0100000000](#). Reads have been deposited in SRA under accessions [SRR25023911](#) and [SRR25024071](#). All project data are available under the BioProject accession number [PRJNA985735](#).

REFERENCES

1. Castenholz RW. 1981. Isolation and cultivation of thermophilic cyanobacteria. In *The Prokaryotes* from Japanese hotspring using an improved isolation procedure. *J Gen Appl Microbiol* 41:119–130. <https://doi.org/10.2323/jgam.41.119>
2. Hanada S, Hiraishi A, Shimada K, Matsuura K. 1995. Isolation of *Chloroflexus aurantiacus* and related thermophilic phototrophic bacteria

3. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>
4. Thiel V, Hamilton TL, Tomsho LP, Burhans R, Gay SE, Schuster SC, Ward DM, Bryant DA. 2014. Draft genome sequence of a sulfide-oxidizing, autotrophic filamentous anoxygenic phototrophic bacterium, *Chloroflexus* sp. strain MS-G (Chloroflexi). *Genome Announc* 2:e00872-14. <https://doi.org/10.1128/genomeA.00872-14>
5. Steunou AS, Bhaya D, Bateson MM, Melendrez MC, Ward DM, Brecht E, Peters JW, Kühl M, Grossman AR. 2006. *In situ* analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. *Proc Natl Acad Sci USA* 103:2398–2403. <https://doi.org/10.1073/pnas.0507513103>
6. Bushnell B. 2022. BBMap. SourceForge. Available from: <https://sourceforge.net/projects/bbmap>. Retrieved 16 Feb 2023.
7. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Available from: <https://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Retrieved 2022.
8. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
9. Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics* 32:292–294. <https://doi.org/10.1093/bioinformatics/btv566>
10. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>
11. Jin X, Yu FB, Yan J, Weakley AM, Dubinkina V, Meng X, Pollard KS. 2023. Culturing of a complex gut microbial community in mucin-hydrogel carriers reveals strain- and gene-associated spatial organization. *Nat Commun* 14:3510. <https://doi.org/10.1038/s41467-023-39121-0>
12. Jain C, Rodriguez-R LM, Phillip AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>
13. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al. 2018. KBase: the United States department of energy systems biology knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>
14. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH, Borgwardt K. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38:5315–5316. <https://doi.org/10.1093/bioinformatics/btac672>
15. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH, Hancock J. 2019. GTDB-TK: a toolkit to classify genomes with the genome taxonomy database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz284>
16. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 50:D785–D794. <https://doi.org/10.1093/nar/gkab776>
17. Parks DH, Chuvochina M, Chaumeil P-A, Rinke C, Mussig AJ, Hugenholtz P. 2020. A complete domain-to-species taxonomy for bacteria and archaea. *Nat Biotechnol* 38:1079–1086. <https://doi.org/10.1038/s41587-020-0501-8>
18. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797. <https://doi.org/10.1093/nar/gkh340>
19. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>
20. Grissa I, Vergnaud G, Pourcel C, Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyripides NC, Hugenholtz P, Janssen P, Goldovsky L, Kunin V, Darzentas N, a. OC, Yu CS, Lin CJ, Hwang JK. 2007. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic Acids Res* 35:W52–7. <https://doi.org/10.1093/nar/gkm360>
21. Couvin D, Bernheim A, Toffano-Nioche C, Touchon M, Michalik J, Néron B, Rocha EPC, Vergnaud G, Gautheret D, Pourcel C. 2018. CRISPRCasFinder, an update of CRISPRFinder, includes a portable version, enhanced performance and integrates search for CAS proteins. *Nucleic Acids Res* 46:W246–W251. <https://doi.org/10.1093/nar/gky425>
22. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res* 14:1394–1403. <https://doi.org/10.1101/gr.2289704>