

# Metagenome-assembled genome of the alkaliphilic *Cyanobacterium* sp. PNNL-SSL1

Hossein Khadivar,<sup>1,2</sup> Huyen Bui,<sup>1,2</sup> Michael Huesemann,<sup>3</sup> Song Gao,<sup>3</sup> Robin Gerlach<sup>1,2</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 2.

**ABSTRACT** Microalgae and cyanobacteria are promising sources of fuels, chemicals, and bioproducts, but CO<sub>2</sub> supply increases production cost. We present the metagenome-assembled genome of the alkaliphilic *Cyanobacterium* sp. strain PNNL-SSL1 obtained from Soap Lake (Washington, USA). PNNL-SSL1 shows strong potential for biomass production using only atmospheric CO<sub>2</sub>, reducing cultivation expenses.

**KEYWORDS** alkaliphilic cyanobacterium, bioproducts, microalgae, biofuels, high pH

Biomass from microalgae including cyanobacteria is a potential alternative to petroleum (1–3). A limitation for growing microalgae in large scales is supplying sufficient CO<sub>2</sub> in concentrated form, which can result in high production and delivery cost (2, 3). Alkaliphilic microalgae have been investigated because high amounts of inorganic carbon can be provided through bicarbonate–carbonate buffered systems (2–5). Here, we present a linear metagenome-assembled genome (MAG) of an alkaliphilic cyanobacterium, *Cyanobacterium* sp. strain PNNL-SSL1, obtained from a naturally alkaline ecosystem: Soap Lake, WA, USA (47°23'36" N, 119°29'05" W), which has pH ~10 and ~100 mmol/L inorganic carbon concentrations (6). PNNL-SSL1 has been shown to be promising for biomass production, relying solely on atmospherically derived CO<sub>2</sub> (3).

The PNNL-SSL1 MAG was obtained by sequencing the metagenomes of two xenic cultures dominated by PNNL-SSL1 named SSL1-A and SSL1-B. These xenic cultures were maintained in NCMA Spirulina medium (7) at ~22°C under ~450 μmol/m<sup>2</sup>/s with 12 hour light:12 hour dark light cycle and diluted weekly to sustain active growth. A cell pellet from each culture (~150 μL) was resuspended in 150 μL of 2× DNA/RNA Shield buffer (Zymo Research, USA) and stored at –80°C. DNA was extracted with the Zymo-BIOMICS DNA Kit (Zymo Research), quantified using a NanoDrop spectrophotometer (Thermo Scientific), and sent to the Joint Genome Institute (JGI). At JGI, libraries were prepared with insert sizes of 200–300 bps using KAPA Biosystems HyperPrep library kit (Roche) and sequenced on an Illumina NovaSeq S4 platform to obtain 150-bp paired-end reads (331,537,726 reads for SSL1-A; 528,729,554 reads for SSL1-B). Sequencing reads downloaded from the JGI genome portal were filtered and assembled into contigs as a co-assembly using MEGAHIT (v1.2.9) (8). Contigs were binned using the KBase apps CONCOCT (v1.1) and MetaBAT2 (v1.7) and optimized with DAS-Tool (v1.1.2) (9–11). The Anvi'o platform (v7.1) was used for manual bin curation, as well as estimation of completeness and redundancy based on copy numbers of single-copy core genes in each bin (12). Binned contigs were assembled into MAGs and annotated with PROKKA (v1.14.5) (13), and genus-level relative abundances were calculated based on coverage in Anvi'o. Taxonomy was determined using the KBase app GTDB-Tk (v2.3.2) (14). Default parameters of software packages were used.

The co-assembled metagenome contained 19,177 linear contigs ≥1 kb, which were grouped into 52 bins using MetaBAT2 based on the nucleotide content and coverage. The DAS-tool optimization resulted in 29 bins. The only cyanobacterial MAG

**Editor** André O. Hudson, Rochester Institute of Technology, Rochester, New York, USA

Address correspondence to Robin Gerlach, robin\_g@montana.edu.

The authors declare no conflict of interest.

See the funding table on p. 2.

**Received** 1 December 2025

**Accepted** 2 March 2026

**Published** 19 March 2026

Copyright © 2026 Khadivar et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

**TABLE 1** Genome assembly statistics for the metagenome-assembled genome of *Cyanobacterium* sp. strain PNNL-SSL1

| PNNL-SSL1 key statistics | Values      |
|--------------------------|-------------|
| Genome size (bp)         | 3,073,714   |
| Number of contigs        | 69          |
| GC content (%)           | 38.38       |
| N50 (bp)                 | 82,559      |
| Number of CDS            | 2,779       |
| Number of tRNAs          | 38          |
| Number of tmRNAs         | 1           |
| Number of reads          | 328,724,316 |
| Read length (bp)         | 150         |
| Completeness (%)         | 97.18       |

obtained in the co-assembled metagenome was estimated to be 97.18% complete with 1.41% redundancy. Features of the resulting genome and its annotation are listed in Table 1. The MAG was identified to belong to the genus *Cyanobacterium*, which is consistent with the primary classification of SSL1 by PNNL (3). The closest placement taxonomy, *Cyanobacterium* sp002813895, was determined by GTDB-Tk based on the Genome Taxonomy Database (14). Realignment of sequencing reads to the co-assembled metagenome (Bowtie v2.3.2 in KBase) showed that 328,724,316 reads were aligned to the cyanobacterial MAG, equivalent of ~16,042× coverage (15).

## ACKNOWLEDGMENTS

*Cyanobacterium* SSL1 was isolated at PNNL as part of the Incubator Project funded by the U.S. Department of Energy (DOE), Office of Energy Efficiency and Renewable Energy, Bioenergy Technology Office, under award no. EE0007004 (3).

This work was supported by the US DOE through DE-EE0009676. Partial financial support was also provided by the National Science Foundation (NSF) (grant nos. 2125083 and 2125748). Metagenome sequencing for this work was performed under project no. 505760 by the U.S. Department of Energy Joint Genome Institute (<https://ror.org/04xm1d337>), a DOE Office of Science User Facility, supported by the Office of Science of the U.S. Department of Energy operated under contract no. DE-AC02-05CH11231.

The findings presented here are those of the authors and do not necessarily reflect the views of the DOE or the NSF.

## AUTHOR AFFILIATIONS

<sup>1</sup>Center for Biofilm Engineering, Montana State University-Bozeman, Bozeman, Montana, USA

<sup>2</sup>Department of Chemical and Biological Engineering, Montana State University-Bozeman, Bozeman, Montana, USA

<sup>3</sup>Marine and Coastal Research Laboratory, Pacific Northwest National Laboratory, Sequim, Washington, USA

## AUTHOR ORCIDs

Hossein Khadivar  <http://orcid.org/0000-0001-6673-8540>

Robin Gerlach  <http://orcid.org/0000-0002-7669-3072>

## FUNDING

| Funder                    | Grant(s)                   | Author(s)                      |
|---------------------------|----------------------------|--------------------------------|
| U.S. Department of Energy | DE-EE0009676, DE-EE0007004 | Huyen Bui<br>Michael Huesemann |

| Funder                      | Grant(s)         | Author(s)                                      |
|-----------------------------|------------------|------------------------------------------------|
|                             |                  | Song Gao<br>Robin Gerlach                      |
| National Science Foundation | 2125083, 2125748 | Huyen Bui<br>Hossein Khadivar<br>Robin Gerlach |
| Joint Genome Institute      | 505760           | Huyen Bui<br>Robin Gerlach                     |

## AUTHOR CONTRIBUTIONS

Hossein Khadivar, Data curation, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review and editing | Huyen Bui, Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Writing – review and editing | Michael Huesemann, Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Writing – review and editing | Song Gao, Writing – review and editing | Robin Gerlach, Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Writing – review and editing

## DATA AVAILABILITY

The raw data for SSL1-A ([SRX29282892](https://www.ncbi.nlm.nih.gov/sra/SRX29282892)) and SSL1-B ([SRX29282893](https://www.ncbi.nlm.nih.gov/sra/SRX29282893)) are available in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) database under BioProject accession number [PRJNA1280515](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1280515). The obtained co-assembly was deposited on Dryad; <https://datadryad.org/dataset/doi:10.5061/dryad.31zcrjf28>. *Cyanobacterium* sp. PNNL-SSL1's BioSample is deposited under the accession number [SAMN50708828](https://www.ncbi.nlm.nih.gov/biosample/SAMN50708828). This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession [JBSZCP000000000](https://www.ncbi.nlm.nih.gov/assembly/JBSZCP000000000). The version described in this paper is version [JBSZCP000000000.1](https://www.ncbi.nlm.nih.gov/assembly/JBSZCP000000000.1). The individual metagenome assemblies of SSL1-A (aka SSL1-WT culture) and SSL1-B (aka SSL1-Turbo culture) were originally deposited in the Department of Energy Joint Genome Institute Integrated Microbial (IMG) Genomes database under [IMG 291693](https://img.jgi.doe.gov/IMG_291693) and [IMG 291694](https://img.jgi.doe.gov/IMG_291694), respectively.

## REFERENCES

- Malhi Y, Franklin J, Seddon N, Solan M, Turner MG, Field CB, Knowlton N. 2020. Climate change and ecosystems: threats, opportunities and solutions. *Phil Trans R Soc B* 375:20190104. <https://doi.org/10.1098/rstb.2019.0104>
- Ataeian M, Liu Y, Canon-Rubio KA, Nightingale M, Strous M, Vadlamani A. 2019. Direct capture and conversion of CO<sub>2</sub> from air by growing a cyanobacterial consortium at pH up to 11.2. *Biotechnol Bioeng* 116:1604–1611. <https://doi.org/10.1002/bit.26974>
- Gao S, Pittman K, Edmundson S, Huesemann M, Greer M, Louie W, Chen P, Nobles D, Benemann J, Crowe B. 2023. A newly isolated alkaliphilic cyanobacterium for biomass production with direct air CO<sub>2</sub> capture. *J CO<sub>2</sub> Util* 69:102399. <https://doi.org/10.1016/j.jcou.2023.102399>
- Vadlamani A, Pendyala B, Viamajala S, Varanasi S. 2019. High CO<sub>2</sub> input. *ACS Sustainable Chem Eng* 7:1933–1943. <https://doi.org/10.1021/acssuschemeng.8b04094>
- Vadlamani A, Viamajala S, Pendyala B, Varanasi S. 2017. Cultivation of microalgae at extreme alkaline pH Conditions: a novel approach for biofuel production. *ACS Sustainable Chem Eng* 5:7284–7294. <https://doi.org/10.1021/acssuschemeng.7b01534>
- Sorokin DY, Foti M, Pinkart HC, Muyzer G. 2007. Sulfur-oxidizing bacteria in Soap Lake (Washington State), a meromictic, haloalkaline lake with an unprecedented high sulfide content. *Appl Environ Microbiol* 73:451–455. <https://doi.org/10.1128/AEM.02087-06>
- Shuichi Aiba TO, Schlösser UG. 2025. *Spirulina* Medium (Modified). National Center for Marine Algae, Bigelow National Center for Marine Algae and Microbiota, Algae Media Recipes. Retrieved 20 Nov 2025.
- Li D, Liu CM, Luo R, Sadakane K, Lam TW. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
- Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods* 11:1144–1146. <https://doi.org/10.1038/nmeth.3103>
- Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>
- Siebert CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3:836–843. <https://doi.org/10.1038/s41564-018-0171-1>
- Eren AM, Kiehl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, et al. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. *Nat Microbiol* 6:3–6. <https://doi.org/10.1038/s41564-020-00834-3>
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2021. 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>
15. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>