



Draft Genome Sequence of the Extremely Desiccation-Tolerant Cyanobacterium *Gloeocapsopsis* sp. Strain AAB1

 Fernando Puente-Sánchez,^a Carlos González-Silva,^b Victor Parro,^c Javier Tamames,^a Armando Azua-Bustos^{c,d}

^aSystems Biology Program, Centro Nacional de Biotecnología, Consejo Superior de Investigaciones Científicas (CSIC), Madrid, Spain

^bCentro de Investigación del Medio Ambiente (CENIMA), Universidad Arturo Prat, Iquique, Chile

^cCentro de Astrobiología (CSIC-INTA), Madrid, Spain

^dInstituto de Ciencias Biomédicas, Facultad de Ciencias de la Salud, Universidad Autónoma de Chile, Santiago, Chile

ABSTRACT *Gloeocapsopsis* sp. strain AAB1 is an extremely desiccation-tolerant cyanobacterium isolated from translucent quartz stones from the Atacama Desert (Chile). Here, we report its draft genome sequence, which consists of 137 contigs with an ~5.4-Mb genome size. The annotation revealed 5,641 coding DNA sequences, 38 tRNA genes, and 5 rRNA genes.

Members of the *Cyanobacteria* are among the most ecophysiologicaly resilient microorganisms, and they have been extensively studied in order to understand the limits of life on Earth (1). *Gloeocapsopsis* sp. strain AAB1 is a cyanobacterium isolated from a hypolithic community found under translucent quartz stones dispersed on the Coastal Range of the Atacama Desert, the driest and oldest desert on Earth. Although rain events in the Atacama are rare, cyanobacterium-dominated biofilms are able to colonize the underside of quartz stones, using fog as the main source of water (2). The recent physiological, morphological, and molecular characterization of isolate AAB1 revealed that it is extremely tolerant to desiccation (3) and represents a good model for understanding tolerance to low water availability.

Although *Gloeocapsopsis* cells were regularly grown using liquid BG-11 media, for the purpose of sequencing, DNA was extracted from 5 mg of dried cells. We used dried cells because, although we were confident of having a single cyanobacterial isolate (3), we detected 3 or 4 additional heterotrophic bacteria that we could not eliminate despite attempting different cleaning protocols. Thus, cells were immediately dried, with the expectation that the desiccation process would favor the preferential preservation of cyanobacterial DNA. DNA was extracted using the Qiagen DNeasy PowerSoil kit (Qiagen, Venlo, Netherlands) according to the manufacturer's instructions, except that at the cell lysis step only a single 120-s pulse was performed, using a FastPrep24 5G homogenizer (MP Biomedicals, Santa Ana, CA). The sequencing library was prepared using a NEBNext Ultra II DNA library prep kit for Illumina (New England Biolabs, Ipswich, MA), following the manufacturer's instructions, and sequenced in a 2 × 300 pair-end sequencing run on a MiSeq sequencer (Illumina, San Diego, CA).

Raw reads were quality trimmed with Prinseq Lite v0.20.4 (4). The resulting 5,830,860 pairs were processed with the SqueezeM pipeline (J. Tamames, unpublished data). Briefly, they were assembled using MEGAHIT v1.1.2 (5), and open reading frames (ORFs) and rRNA coding sequences were predicted with Prodigal v2.6.2 (6) and barrnap v0.9-dev (<https://github.com/tseemann/barrnap>), respectively. ORFs were aligned against the Kyoto Encyclopedia of Genes and Genomes (KEGG) (7) and Clusters of Orthologous Groups of proteins (COG) (8) databases using DIAMOND (9), functionally annotated based on their best hits, and taxonomically annotated based on a last-common-

Received 1 March 2018 Accepted 16 March 2018 Published 26 April 2018

Citation Puente-Sánchez F, González-Silva C, Parro V, Tamames J, Azua-Bustos A. 2018. Draft genome sequence of the extremely desiccation-tolerant cyanobacterium *Gloeocapsopsis* sp. strain AAB1. *Genome Announc* 6:e00216-18. <https://doi.org/10.1128/genomeA.00216-18>.

Copyright © 2018 Puente-Sánchez et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Fernando Puente-Sánchez, puente@cnb.csic.es.

ancestor algorithm over all hits within 20% of the bit score of the first hit. Furthermore, whole contigs were assigned a consensus taxonomy based on the annotation of their constituent ORFs. A total of 137 high-coverage contigs, encompassing a total of 5,443,570 bases, were identified as having a clear cyanobacterial origin and selected for further analyses. Among these contigs, the N_{50} value was 73,596 bases, the average coverage was 509 \times , and the GC content was 42.42%. CheckM v1.0.7 (10) reported a genome completeness of 100% and genome contamination of 1.36%, after being run on the cyanobacterial contigs. Finally, gene prediction and annotation of the *Gloeocapsopsis* sp. AAB1 contigs were carried out using the Rapid Annotations using Subsystems Technology (RAST) pipeline (11) as provided by the Pathosystems Resource Integration Center (PATRIC) server (12). A total of 5,641 coding DNA sequences (CDSs), 38 tRNA genes, and 5 rRNA genes were identified.

Accession number(s). The nucleotide sequence and annotation data for the *Gloeocapsopsis* sp. strain AAB1 draft genome have been deposited at DDBJ/ENA/GenBank under the accession number [PTJN00000000](https://doi.org/10.1007/0-306-46855-7_13). The version described in this paper is version PTJN01000000.

ACKNOWLEDGMENTS

F.P.-S. was supported by the Spanish Ministerio de Economía y Competitividad grant CTM2016-80095-C2-1-R. A.A.-B. thanks the European Research Council for starting grant 307496. This work was partially supported by the Spanish Ministerio de Economía y Competitividad grant ESP2015-69540-R.

REFERENCES

- Wynn-Williams DD. 2000. Cyanobacteria in deserts—life at the limit?, pp. 341–366. *In* The ecology of cyanobacteria. Springer, Dordrecht, The Netherlands. https://doi.org/10.1007/0-306-46855-7_13.
- Azua-Bustos A, González-Silva C, Mancilla RA, Salas L, Gómez-Silva B, McKay CP, Viciña R. 2011. Hypolithic cyanobacteria supported mainly by fog in the coastal range of the Atacama Desert. *Microb Ecol* 61:568–581. <https://doi.org/10.1007/s00248-010-9784-5>.
- Azua-Bustos A, Zúñiga J, Arenas-Fajardo C, Orellana M, Salas L, Viciña R. 2014. *Gloeocapsopsis* AAB1, an extremely desiccation-tolerant cyanobacterium isolated from the Atacama Desert. *Extremophiles* 18:61–74. <https://doi.org/10.1007/s00792-013-0592-y>.
- Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27:863–864. <https://doi.org/10.1093/bioinformatics/btr026>.
- 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>.
- Hyatt D, Chen GL, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>.
- Kanehisa M. 2002. The KEGG database. *Novartis Found Symp* 247: 91–103. <https://doi.org/10.1002/0470857897>.
- Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA. 2003. The COG database: an updated version includes eukaryotes. *BMC Bioinformatics* 4:41. <https://doi.org/10.1186/1471-2105-4-41>.
- Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59. <https://doi.org/10.1038/nmeth.3176>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid annotations using subsystems technology. *BMC Genomics* 9:75. <https://doi.org/10.1186/1471-2164-9-75>.
- Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ, Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R, Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A, Will R, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* 42:D581–D591. <https://doi.org/10.1093/nar/gkt1099>.