

Draft Genome Sequence of the Thermotolerant Cyanobacterium *Desertifilum* sp. IPPAS B-1220

Kirill S. Mironov,^a Maria A. Sinetova,^a Kenzhegul Bolatkhan,^b Bolatkhan K. Zayadan,^b Vera V. Ustinova,^c Elena V. Kupriyanova,^a Alexandra N. Skrypnik,^a Natalya E. Gogoleva,^d Yuriy V. Gogolev,^d Dmitry A. Los^a

Institute of Plant Physiology, Russian Academy of Science, Moscow, Russia^a; Department of Biotechnology, Faculty of Biology and Biotechnology, Al-Farabi Kazakh National University, Almaty, Kazakhstan^b; Microbiology Department, Central Tuberculosis Research Institute RAMS, Moscow, Russia^c; Kazan Institute of Biochemistry and Biophysics, Kazan Science Centre, Russian Academy of Sciences, Kazan, Russia^d

Here, we report the draft genome of the filamentous cyanobacterium *Desertifilum* sp. strain IPPAS B-1220, isolated from Lake Shar-Nuur, Mongolia. The genome of 6.1 Mb codes for 5,113 genes. Genome mining revealed 10 clusters for the synthesis of bioactive compounds (nonribosomal peptides, polyketides, bacteriocins, and lantipeptides) with potential biotechnological or medical importance.

Received 25 September 2016 Accepted 30 September 2016 Published 17 November 2016

Citation Mironov KS, Sinetova MA, Bolatkhan K, Zayadan BK, Ustinova VV, Kupriyanova EV, Skrypnik AN, Gogoleva NE, Gogolev YV, Los DA. 2016. Draft genome sequence of the thermotolerant cyanobacterium *Desertifilum* sp. IPPAS B-1220. *Genome Announc* 4(6):e01304-16. doi:10.1128/genomeA.01304-16.

Copyright © 2016 Mironov 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 Kirill S. Mironov, kirill.mironoff@gmail.com.

Cyanobacteria produce a broad range of secondary metabolites with diverse chemical structures. The majority of these metabolites are the products of nonribosomal peptide synthetase or polyketide synthase pathways (1). We sequenced the genome of the cyanobacterium *Desertifilum* sp. strain IPPAS B-1220 (family, *Oscillatoriaceae*), which was newly isolated from the freshwater Lake Shar-Nuur, Mongolia.

Genomic DNA was isolated from cells grown to stationary phase by incubation with saturated iodide solution followed by lysozyme treatment and 2% SDS lysis at 70°C (2). DNA was purified by phenol-chloroform extraction. Isolated DNA was fragmented by adaptive focused acoustics technology using a Covaris S220 ultrasonicator (Covaris, Woburn, MA, USA). Parameters for DNA shearing were adapted to obtain 500-bp fragments. A DNA library was prepared with a NEBNext Ultra DNA library prep kit (NEB, Ipswich, MA, USA) for Illumina sequencing, which was performed on the MiSeq system with MiSeq reagent kit version 3 in a 600-cycle paired-end format. The sequence quality was analyzed by FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Adapters and low-quality nucleotides were trimmed with Trimmomatic (3). The genome was assembled using AbySS (4) and SPAdes (5) software. The quality of the draft genome assembly was analyzed by QUAST (6). The draft genome coverage was no less than 30×, with N_{50} statistics $>10^5$. The approximate genome size is 6.1 Mb, with an estimated average G+C content of 48.7%.

The genome was annotated using the automated NCBI Prokaryotic Genome Annotation Pipeline (PGAP). It contains 5,113 genes: 4,964 genes coding for proteins, 86 pseudogenes, six rRNA-coding sequences (four for 5S, one for 16S, and one for 23S), 53 tRNAs, and four noncoding RNAs. Seven CRISPR arrays were found in the genome. The genome sequence was analyzed with the antiSMASH tool (7) for the presence of gene clusters coding for secondary metabolite biochemical pathways. Eight contigs con-

tained gene clusters for secondary metabolite biosynthesis (non-ribosomal peptides, polyketides, bacteriocins, lantipeptides, etc.). Thus, genome mining of *Desertifilum* sp. IPPAS B-1220 revealed a set of putative bioactive compounds that may have further biotechnological or pharmaceutical applications.

Accession number(s). The genome sequence of *Desertifilum* sp. strain IPPAS B-1220 was deposited at NCBI with the following attributes: SAMN05788062 for BioSample, PRJNA343432 for BioProject, and SRR4255595 for the SRA database. The genome PGAP file accession number is [MJGC00000000](https://ncbi.nlm.nih.gov/assembly/MJGC00000000/).

ACKNOWLEDGMENTS

We are grateful to the Common-Use Center “Biotechnology” of the All-Russian Scientific Research Institute of Agricultural Biotechnology (Moscow, Russia) for help with bioinformatic analysis. This work was supported by a grant from the Russian Science Foundation (no. 14-24-00020) to D.A.L. and by a grant from the Ministry of Education and Science of Republic of Kazakhstan to B.K.Z.

FUNDING INFORMATION

This work, including the efforts of Dmitry Los, was funded by Russian Science Foundation (RSF) (14-24-00020). This work, including the efforts of Bolatkhan K. Zayadan, was funded by Ministry of Education and Science of the Republic of Kazakhstan (Ministry of Education and Science, Republic of Kazakhstan) (1582/GF4).

REFERENCES

- Dittmann E, Gugger M, Sivonen K, Fewer DP. 2015. Natural product biosynthetic diversity and comparative genomics of the cyanobacteria. *Trends Microbiol* 23:642–652. <http://dx.doi.org/10.1016/j.tim.2015.07.008>.
- Campbell WS, Laudenbach DE. 1995. Characterization of four superoxide dismutase genes from a filamentous cyanobacterium. *J Bacteriol* 177:964–972.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <http://dx.doi.org/10.1093/bioinformatics/btu170>.

4. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I. 2009. ABySS: a parallel assembler for short read sequence data. *Genome Res* 19: 1117–1123. <http://dx.doi.org/10.1101/gr.089532.108>.
5. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <http://dx.doi.org/10.1089/cmb.2012.0021>.
6. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUILT: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <http://dx.doi.org/10.1093/bioinformatics/btt086>.
7. Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA, Weber T, Takano E, Breitling R. 2011. antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences. *Nucleic Acids Res* 39:W339–W346. <http://dx.doi.org/10.1093/nar/gkr466>.