1   Draft genome sequence of the thermotolerant cyanobacterium *Desertifilum* sp. IPPAS B-

2   1220

3

4   Kirill S. Mironov[a]#, Maria A. Sinetova[a], Kenzhegul Bolatkhan[b], Bolatkhan K. Zayadan[b],

5   Vera V. Ustinova[c], Elena V. Kupriyanova[a], Alexandra N. Skrypnik[a], Natalya E.

6   Gogoleva[d], Yuriy V. Gogolev[d], Dmitry A. Los[a]

7

8   Institute of Plant Physiology, Russian Academy of Science, Moscow, Russia[a];

9   Department of Biotechnology, Faculty of Biology and Biotechnology, Al-Farabi Kazakh

10   National University, Almaty, Kazakhstan[b]; Microbiology Department, Central

11   Tuberculosis Research Institute RAMS, Moscow, Russia[c]; Kazan Institute of

12   Biochemistry and Biophysics, Kazan Science Centre, Russian Academy of Sciences,

13   Kazan, Russia[d]

14

15   Running Head: Draft genome of the cyanobacterium *Desertifilum* sp.

16

17   #Correspondence to Kirill S. Mironov: kirill.mironoff@gmail.com

18

19

20

21

22

23   **Abstract**

24   **We report the draft genome of the filamentous cyanobacterium *Desertifilum* sp.**

25   **strain IPPAS B-1220 isolated from Lake Shar-Nuur, Mongolia. The genome of 6.1**

26   **Mb codes for 5113 genes. Genome mining revealed 10 clusters for the synthesis of**

27   **bioactive compounds (non-ribosomal peptides, polyketides, bacteriocins,**

28   **lantipeptides) with potential biotechnological or medical importance**.

29

30   Cyanobacteria produce a broad range of secondary metabolites with diverse chemical

31   structures. The majority of these metabolites are the products of non-ribosomal peptide

32   synthetase (NRPS) or polyketide synthase (PKS) pathways (1). We sequenced the

33   genome of the cyanobacterium *Desertifilum* sp. strain IPPAS B-1220 (Oscillatoriaceae),

34   which was newly isolated from fresh-water Lake Shar-Nuur, Mongolia.

35   Genomic DNA was isolated from cells grown at stationary phase by incubation

36   with saturated iodide solution followed by lysozyme treatment and 2% SDS lysis at 70ºC

37   (2). DNA was purified by phenol/chloroform extraction. Isolated DNA was fragmented

38   by adaptive focused acoustics technology using Covaris S220 ultrasonicator (Covaris,

39   Woburn, MA, USA). Parameters for DNA shearing were adapted to obtain 500 bp

40   fragments. DNA library was prepared with NEBNext® Ultra™ DNA Library Prep Kit

41   for Illumina (NEB, Ipswich, MA, USA). Sequencing was performed on MiSeq system

42   with MiSeq reagent kit v3 in 600 cycle paired-end format. Sequence quality were

43   analysed by FastQC (3). Adapters and low-quality nucleotides were trimmed with

44   Trimmomatic (4). Genome was assembled using AbySS (5) and SPAdes (6) software.

45   The draft genome assembly quality was analyzed by QUAST (7). Draft genome coverage

46   was no less than 30, N50 statistics $> 10^5$. Approximate genome size is 6.1 Mb, the

47   average G+C content was estimated as 48.7%.

48   Genome was annotated using automated NCBI Prokaryotic Genome Annotation

49   Pipeline. It contains 5113 genes: 4964 genes coding for proteins, 86 pseudo genes, 6

50   rRNA-coding sequences (4 for 5S, 1 for 16S, and 1 for 23S), 53 tRNAs, 4 non-coding

51   RNAs. Seven CRISPR arrays have been found in the genome. The genome sequence was

52   analyzed with antiSMASH tool (8) for the presence of gene clusters coding for secondary

53   metabolite biochemical pathways. Eight contigs contained gene clusters for secondary

54   metabolites biosynthesis (non-ribosomal peptides, polyketides, bacteriocins,

55   lantipeptides, *etc.*). Thus, genome mining of *Desertifilum* sp. IPPAS B-1220 revealed a

56   set of putative bioactive compounds that may have further biotechnological or

57   pharmaceutical applications.

58   **Accession number(s).** Genome sequence of *Desertifilum* sp. strain IPPAS B-

59   1220 was deposited at NCBI with the following attributes: SAMN05788062 for

60   BioSample, PRJNA343432 for BioProject, and SRR4255595 for SRA database. Genome

61   PGAP file accession number is MJGC00000000.

62

72

73   **References**

74   1. **Dittmann E, Gugger M, Sivonen K, Fewer DP.** 2015. Natural product biosynthetic

75   diversity and comparative genomics of the cyanobacteria. Trends Microbiol **23**:642-52.

76   2. **Campbell WS, Laudenbach DE.** 1995. Characterization of four superoxide dismutase

77   genes from a filamentous cyanobacterium. J Bacteriol **177:**964-72.

78   3. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

79   4. **Bolger AM, Lohse M, Usadel B.** 2014. Trimmomatic: A flexible trimmer for Illumina

80   sequence data. Bioinformatics **30**:2114-20.

81   5. **Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I.** 2009. ABySS: A

82   parallel assembler for short read sequence data. Genome Res **19:**1117-1123.

83   6. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin**

84   **VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N,**

85   **Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: A new genome assembly

86   algorithm and its applications to single-cell sequencing. J Comput Biol **19**:455-477.

87   7. **Gurevich A., Saveliev V., Vyahhi N., Tesler G.** 2013. QUAST: quality assessment

88   tool for genome assemblies. Bioinformatics **29:**1072-1075.

89   8. **Medema MH, Blin K, Cimermancic P, de Jager V, Zakrzewski P, Fischbach MA,**

90   **Weber T, Breitling R, Takano E.** 2011. antiSMASH: Rapid identification, annotation

91   and analysis of secondary metabolite biosynthesis gene clusters. Nucleic Acids Res

92   **39:**W339-W346.