Data in Brief

Meta-biological data on prokaryotic diversity of Kunashir Island geothermal spring

Alexei S. Rozanov, Anton V. Korzhuk, Valeria N. Shlyakhtun, Aleksandra A. Shipova, Sergey E. Peltek

Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (ICG SB RAS), Lavrentjeva Ave. 10, Novosibirsk, 630090, Russia
Kurchatov Genomics Center, Institute of Cytology and Genetics, SB RAS, 630090 Novosibirsk, Russia

Data in Brief journal homepage: www.elsevier.com/locate/dib

Article history:
Received 14 July 2020
Revised 20 August 2020
Accepted 25 August 2020
Available online 29 August 2020

Keywords:
Metagenomics
16S rRNA
Acidic hot spring
Kunashir Island

This is data on the microbial diversity of a geothermal spring located on the banks of the acidic creek of Kunashir Island. Data was obtained using 16S rRNA amplicon directed metagenomic sequencing on Illumina MiSeq. The raw sequence data used for analysis is available in NCBI under the Sequence Read Archive (SRA) with the BioProject No. PRJNA637298, PRJNA637447 and SRA accession number SRP265942, SRP266050. The data sequences of the 16S rRNA gene are presented at the accession numbers MT604934-MT604967, MT604911-MT604921 in NCBI GenBank database.

© 2020 Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license. (http://creativecommons.org/licenses/by-nc-nd/4.0/)
1. Data Description

A description of the sampling points is given in Table 1. Photographs of the sampling sites are shown in Fig. 1.

The raw sequencing data contain 19,213 paired-end sequences with a length of 301 bp totalling 11.6 M base pairs for a sample No. 10. Whereas for sample No. 8, 40.380 paired-end sequences with a length of 301 bp totalling 24.3 M base pairs were obtained.

| Point No. | pH  | T, °C | Description |
|-----------|-----|-------|-------------|
| 8         | 2.0 | 38    | The nameless creek originating from a hot spring with a temperature of 73°C and pH 2.0 and flowing into the Kislyi stream. A green bloom and sediment underneath were selected. Coordinates: 43.998029N, 145.767768E |
| 10        | 3.0 | 18    | The sour Kislyi stream. Orange-red bottom sediments selected. Coordinates: 43.997836N, 145.767671E |

Fig. 1. Sampling locations: A - point No. 8, backwater of a hot spring; B - view of the Kislyi stream at the sampling site.
A total of 52 OTUs including 10631 sequences were obtained for the two communities. For community No. 8 from the geothermal stream, 44 OTU and 11 for the microbial community from the bottom sediments of Kislyi Stream. The distribution of the types and classes of received OTUs is shown in Fig. 2.

In community No. 8, about 38% are occupied by archaea, the rest are bacteria, while in community No. 10 only bacteria are present. The following types (or classes in the case of proteobacteria) occupy more than 1% in community No. 8: Thermi, Euryarchaeota, Actinobacteria, Crenarchaeota, Parvarchaeota, Alphaproteobacteria, Firmicutes, Betaproteobacteria, Nitrospirae. In community No. 8, the following types (or classes in the case of proteobacteria) occupy more than 1%: Betaproteobacteria, Bacteroidetes, Alphaproteobacteria, Gammaproteobacteria, Actinobacteria. The studied communities turned out to be completely dissimilar, which is probably associated with various conditions, primarily with the temperature of the media.

2. Experimental Design, Materials and Methods

2.1. Sample collection and DNA extraction

Bottom sediments were taken from the Kislyi stream and the nameless creek flowing into it, originating from a hot spring with a temperature of 73°C and pH 2.0. The samples were taken in sterile 50 ml Falcon tubes and was stored in alcohol at -70°C. Total DNA was isolated from the samples (0.3 g) using the Genomic DNA from soil NucleoSpin® Soil kit (Macherey-Nagel) according to the manufacturer’s protocol.

2.2. Library preparation and next-generation DNA sequencing

To obtain the target fragment comprising the V3-V4 region 16s rRNA, degenerate primers U343F (5’-CCTACGGGGRSGAGCAG-3’ ) and U806R (5’-GGACTACNVGGGTWTCTAAT-3’) were used, which allow the amplification of the 16s rRNA gene of a wide range of microorganisms. Previously, these primers showed wide coverage in terms of amplification of various types of microorganisms. [1]. The regime of first amplification: 96°C - 2'; 25°(96°C-8”; 54°C-20”; 68°C-30”). Purified PCR product was used for a second PCR reaction which attached Illumina sequencing adapters and dual-index barcodes to the amplicon target. The regime of second amplification: 96°C - 2'; 5°(96°C-8”; 54°C-20”; 68°C-30”); 20°(96°C-8”; 60°C-20”; 68°C-30”). Fusion polymerase Q5 (NewEnglandBiolabs) and low primer annealing temperature were used to obtain gene libraries with the smallest number shift. Sequencing was performed at the Genomics Laboratory.
of the IMKB SB RAS using an Illumina MiSeq instrument using the TG MiSeq Reagent Kit v3 (600 cycle) kit. Sequenced reads were automatically divided by barcodes.

2.3. Taxonomic analysis

Paired reads obtained during sequencing of 16S rRNA libraries were processed using QIIME2 v.2020.2 platform [2]. Noise removal, integration of paired reads, and construction of OTUs were performed using the DADA2 algorithm. For taxonomic classification of OTUs, the scikit-learn classifier was used, trained on fragments of 16S rRNA of the Greengenes v13_8 database, limited by the primers used.

Ethics Statement

In the work were not involve the use of human subjects, animals, cell lines and endangered species of wild fauna and flora.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

The reported study was funded by ICG SB RAS Budget Project no. 0324-2019-0040-C-01 "Genetic basis of biotechnology and bioinformatics" and Kurchatov Genomics Center of IC&G (075-15-2019-1662). This work was done at the Center for collective use "A collection of biotechnological microorganisms as a source of novel promising objects for biotechnology and bioengineering" of Federal Research Center "Institute of Cytology and Genetics of the Siberian Branch of the RAS".

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.dib.2020.106244.

References

[1] A.S. Rozanov, A.V. Bryanskaya, T.K. Malup, I.A. Meshcheryakova, E.V. Lazareva, O.P. Taran, T.V. Ivanisenko, V.A. Ivanisenko, S.M. Zhmodik, N.A. Kolchanov, S.E. Peltek, Molecular analysis of the benthos microbial community in Zavarin thermal spring (Uzon Caldera, Kamchatka, Russia), BMC Genom. 15 (2014) S12, doi:10.1186/1471-2164-15-S12-S12.

[2] E. Bolyn, J.R. Rideout, M.R. Dillon, N.A. Bokulich, C.C. Abnet, G.A. Al-Ghalith, H. Alexander, E.J. Alm, M. Arumugam, F. Assicar, Y. Bai, J.E. Bisanz, K. Bittinger, A. Brejnrod, C.J. Bristaw, C.T. Brown, B.J. Callahan, A.M. Caraballo-Rodriguez, J. Chase, E.K. Cope, R. Da Silva, C. Diener, P.C. Dorrestein, G.M. Douglas, D.M. Durall, C. Duvallal, C.F. Edwardson, M. Ernst, M. Estaki, J. Fouquier, J.M. Gauglitz, S.M. Gibbons, D.L. Gibson, A. Gonzalez, K. Gorlick, J. Guo, B. Hillmann, S. Holmes, H. Holste, C. Hutterhoffer, G.A. Huttley, S. Janssen, A.K. Jarmsuch, L. Jiang, B.D. Kaehler, K. Bin Kang, C.R. Keefe, P. Keim, S.T. Kelley, D. Knights, J. Koester, T. Kosciolak, J. Kreps, M.G.I. Langille, J. Lee, R. Ley, Y.X. Liu, E. Loftfield, C. Lozupone, M. Maher, C. Marotz, B.D. Martin, D. McDonald, L.J. McIver, A.V. Melnik, J.L. Metcalf, S.C. Morgan, J.T. Morton, A.T. Naimey, J.A. Navas-Molina, L.F. Nothias, S.B. Orphanian, T. Pearson, S.L. Peoples, D. Petras, M.L. Preuss, E. Pruesse, L.B. Rasmussen, A. Rivers, M.S. Robeson, P. Rosenthal, N. Segata, M. Shaffer, A. Shiffer, R. Sinha, S.J. Song, J.R. Spear, A.D. Swafford, L.R. Thompson, P.J. Torres, P. Trinh, A. Tripathi, P.J. Turnbaugh, S. Ul-Hasan, J.J.J. van der
Hooft, F., Vargas, Y. Vázquez-Baeza, E. Vogtmann, M. von Hippel, W. Walters, Y. Wan, M. Wang, J. Warren, K.C. Weber, C.H.D. Williamson, A.D. Willis, Z.Z. Xu, J.R. Zaneveld, Y. Zhang, Q. Zhu, R. Knight, J.G. Caporaso, Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2, Nat. Biotechnol. 37 (2019) 852–857, doi:10.1038/s41587-019-0209-9.