Data Article

Metagenomics data of microbial communities in bacterial mats and bottom sediments in water bodies within the Kurai Mercury Province (Gorny Altai, Russia)

Alexei S. Rozanov, Irina N. Myagkaya, Anton V. Korzhuk, Nikita I. Ershov, Ivan S. Kirichenko, Maria A. Gustaytis, Bagai-ool Yu. Saryg-ool, Victor I. Malov, Aleksandra A. Shipova, Elena V. Lazareva, Sergey E. Peltek

a Laboratory of molecular biotechnology, Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (ICG SB RAS), 10 Lavrentjeva Ave., Novosibirsk 630090, Russian Federation
b Kurchatov Genomics Center, Institute of Cytology and Genetics, SB RAS, 10 Lavrentjeva Ave., Novosibirsk 630090, Russian Federation
c V.S. Sobolev Institute of Geology and Mineralogy, Siberian Branch of the Russian Academy of Sciences (IGM SB RAS), Novosibirsk 630090, Russian Federation

A R T I C L E   I N F O

Article history:
Received 26 February 2021
Revised 14 April 2021
Accepted 16 April 2021
Available online 30 April 2021

Keywords:
Metagenomics
16s rRNA
Water metagenome
Kurai mercury province

A B S T R A C T

Below is data on the microbial diversity in bottom sediments and microbial mats in water bodies within the Kurai Mercury Province (Ulagan District, Aktash village, Gorny Altai). A database on the geochemical features of water bodies in the study area is presented. 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 BioProject No. PRJNA670076 and SRA accession numbers SRX9316205, SRX9316207, SRX9316208, SRX9316209.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

* Corresponding author at: Laboratory of molecular biotechnology, Institute of Cytology and Genetics, Siberian Branch of the Russian Academy of Sciences (ICG SB RAS), 10 Lavrentjeva Ave., Novosibirsk 630090, Russian Federation.
E-mail address: korshuk_anton@bionet.nsc.ru (A.V. Korzhuk).

https://doi.org/10.1016/j.dib.2021.107099
2352-3409/© 2021 The Author(s). 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/)
**Specifications Table**

| Subject                  | Environmental Genomics and Metagenomics |
|--------------------------|-----------------------------------------|
| Specific subject area    | Metagenomic study on the microbial community in bottom sediments and microbial mats in water bodies within the Kurai Mercury Province |
| Type of data             | Figures, Table and 16S rRNA amplicon sequencing data |
| How data were acquired   | Illumina MiSeq platform, QIME2 v.2020.2, MEGA v.10.1.8 |
| Data format              | Raw and analyzed |
| Parameters for data collection | Metagenomic RNA isolated from bottom sediments and bacterial mats were prepared by amplifying the V3-V4 region of the 16S rRNA gene paired-end sequenced on an Illumina MiSeq platform. |
| Description of data collection | Metagenomic DNA extraction, amplicon sequencing of V3-V4 region of 16S rRNA gene, paired reads processing using QIME2 v.2020.2 platform and building phylogenetic trees with MEGA v.10.1.8 |
| Data source location     | Ulagan District, Aktash village, Corny Altai, Russia: 50.412417 N, 87.586722E (Artesian well); 50.408750 N, 87.598861E (Lake-1); 50.289333 N, 87.667056E (Lake Geyser) |
| Data accessibility       | With the article: The sequences, taxonomy and abundance of the OTU. In a public repository: Repository name: NCBI SRA. The sequencing data is available in NCBI under the Sequence Read Archive (SRA) with the BioProject No. PRJNA670076 and SRA accession numbers SRX9316205, SRX9316207, SRX9316208, SRX9316209. The direct URLs to data are https://www.ncbi.nlm.nih.gov/sra/SRX9316205[accn], https://www.ncbi.nlm.nih.gov/sra/SRX9316207[accn], https://www.ncbi.nlm.nih.gov/sra/SRX9316208[accn], https://www.ncbi.nlm.nih.gov/sra/SRX9316209[accn] |

**Value of the Data**

- The increased natural geochemical background of mercury in water bodies within the Kurai Mercury Province is of interest for studying the features of the distribution of Hg in ecosystem components;
- Hydrochemical data of water bodies can be used to compare the environments (ecological and geochemical) in other mercury provinces of the world;
- The data represent the taxonomic profile of microbial communities in the zone of geochemical barriers of water bodies within the Kurai Mercury Province;
- The data will be interest the phylogography of prokaryotes associated with artesian water and/or searching for attractive organisms/genes for biotechnological use;
- The data may be of interest for the reconstruction and modeling of microbial communities in geochemical barriers that are currently being formed and for understanding the evolution of the ecosystem at the early stages of development;

1. **Data Description**

The paper presents data on microbial communities formed on the outflow of three water sources in the area within the Kurai Mercury Province [1] (Figs. 1, 2). This area is part of neotectonic activity the Kurai Fault Zone [2]. This study area is characterized by an increased geochemical background of mercury (Table 1), including due to emanations of mercury through deep faults characteristic of this area.

The first two water sources (M-1, M-2; Figs. 1, 2) are located in the northern part of a landslide dam of the Cheybek-Kohl Lake (Chibitka River valley). The lake is located among Quaternary glacial, alluvial and mudflow deposits [3]. In the western part of the Cheybek-Kohl Lake, fields of Hg-bearing travertine are found, indicating the activity of faults in this area [4,5]. In the northeastern part of the lake, 900 m from it, there is an artesian well (M-1, Figs. 1, 2a), where
Table 1
Description of the sampling points.

| Point No. | M-1 20.09.2018 | M-2–1 14.06.2019 | M-2–2 20.09.2018 | M-3 20.09.2018 |
|-----------|----------------|------------------|------------------|----------------|
| Data sampling | Microbial mat. Outlet of an artesian well in the northern part of Cheybek-Kohl Lake. | Surface layer of bottom sediments with a predominance of microbial mat. Coastal area of Lake-1. | Bottom sediment. Central part of Lake-1. | Microbial mat. Outlet of a deep underground source, in the waters of which bottom sediments in the form of blue clay and microbial communities of anoxygenic phototrophic microorganisms were formed |
| pH | 7.85 | 8.3 | 8.11 | 8.07 |
| Eh (mV) | 234 | 412 | 294 | 311 |
| T of waters (°C) | 6 | 11 | 6 | 6 |
| TSS (mg/L) | 1.6 | 8.4 | 0.4 | 1.6 |
| TDS (g/L) | 0.39 | 0.22 | 0.27 | 0.29 |
| TOC (mg/L) | 1.6 | 3.5 | 2.4 | 2.7 |
| HCO₃⁻ | 238 | 130 | 200 | 183 |
| SO₄²⁻ | 76 | 3.5 | 10 | 52 |
| Cl | 2.8 | 0.15 | 0.19 | 1.4 |
| F | 0.55 | 0.05 | 0.026 | 0.18 |
| Na_total | 29.28 | 2.2 | 2.1 | 5.1 |
| Mg_total | 9.24 | 28.32 | 10.85 | 11.21 |
| Al_total | 0.04 | 0.02 | 0.01 | 0.01 |
| K_total | 1.63 | 1.6 | 0.86 | 0.91 |
| Ca_total | 29.3 | 56.51 | 38.4 | 44.58 |
| Fe_total | 0.4 | 0.14 | 0.05 | 0.05 |
| As_total (μg/L) | 19.80 | 1.48 | 0.78 | 0.31 |
| Cd_total | 0.01 | 0.04 | 0.004 | 0.01 |
| Pb_total | 0.10 | 0.18 | 0.10 | 0.10 |
| Hg in bottom sediment (g/t) | n/d | 0.23 | 0.28–0.51 | - |
| Hg in bacterial mat | 0.45 | n/d | 0.066 | - |
| BioSample | SAMN16483319 | SAMN16483317 | SAMN16483318 | SAMN16483315 |

Waters characteristics

| pH | 6.5–8.5 |
| Eh (mV) | n/d |
| T of waters (°C) | n/d |
| TSS (mg/L) | 0.25–0.75 |
| TDS (g/L) | 1 |
| TOC (mg/L) | 1 |
| HCO₃⁻ | 30 |
| SO₄²⁻ | 100 |
| Cl | 300 |
| F | 0.75 |
| Na_total | 120 |
| Mg_total | - |
| Al_total | 40 |
| K_total | - |
| Ca_total | 50 |
| Fe_total | 0.04 |
| As_total | 0.9 |
| Cd_total | 0.04 |
| Pb_total | 180 |
| Hg in bottom sediment (g/t) | 0.1 |
| Hg in bacterial mat | 50 |
| BioSample | 0.1 |

Solid matter

| Hg in bottom sediment (g/t) | 0.28–0.51 |
| Hg in bacterial mat | - |

Note: 1 – Decree No. 552 of 13 December 2016 on Quality of Water for Fishery, Including Maximum Permissible Level of Pollutants in Lakes and Rivers Used for Fishery; 2 – World Health Organization (2017); n/d – no data; * - Not of health concern at levels found in drinking-water.
in October 2018, microbial mats were collected that formed on the outflow of the well. Another lake (Lake-1) is located in the northwest direction, 250 m from the Cheybek-Kohl Lake (M-2; Figs. 1, 2b). The bottom sediment was sampled here in October 2018, and microbial mats were also collected in June 2019. The third water body is known as Lake Geyser (M-3, Figs. 1, 2c). It is located in the southern part of the Kurai fault zone (5 km from Aktash village) among Quaternary alluvial deposits in the Chuya River valley. The origin of this lake has not been established. It is formed by underground sources; the water temperature in the lake is not higher than 10 °C.
in summer. In the central part of the lake, at the bottom, air emissions are periodically observed, forming a bizarre pattern at the bottom. In October 2018, samples of microbial mats from this lake were taken. A description of the sampling points is given in Table 1.

At the outflow of artesian springs, conditions are formed that are unique for surface waters, representing geochemical barriers to the transition from low to high redox potential (Eh) values. The most interesting natural sources that form local niches for the development of microbial communities that differ significantly from the environment are Lake Geyser and Lake-1.
Fig. 3. Phylogenetic tree of the obtained OTUs and their related sequences of SSU rRNA from the NCBI database.
Both lakes have a similar appearance; microbial communities are observed in them, forming bottom sediments in the form of blue clay with purple inclusions formed by the anoxygenic phototrophic bacteria Chloroflexi.

Raw sequencing data of the sample M-1 contain 13,975 paired-end reads, the sample M-2–1 - 9375, the sample M-2–2–3559 and the sample M-3–7162. The read length is 301 bp.

In total, after processing and cleaning, 42 OTUs including 1370 sequences were obtained. The sample M-1 contains 6 OTU (including 151 sequences), the sample M-2–1–24 OTU (including 854 sequences), the sample M-2–2–7 OTU (including 112 sequences) and the sample M-3–7 OTU (including 253 sequences). Fig. 3 shows a phylogenetic tree of the obtained OTUs and their related 16S rRNA sequences from the NCBI database. The OTU sequences are presented in Supplementary Table 1.

The ratio of the sequences of different taxa is shown in Fig. 4. Euryarchaeota (Methanobacterium and Methanoseta) and Crenarchaeota (order pGrfC26 in the MCG candidate class) in the studied microbial community formed in Lake Geyser were identified. The sequence Crenarchaeota MCG (pGrfC26) was also found in the microbial community of bottom sediments of an artesian well. Bacteria were mainly represented by Bacteroidetes, Actinobacteria, Alphaproteobacteria, Betaproteobacteria, Verrucomicrobia, and Acidobacteria. Phototrophic bacteria were represented by Chloroflexi (Anaerolinaceae and the candidate class Ellin6529). The phylums Chlorobi, Firmicutes, Gemmatimonadetes, and Deltaproteobacteria were less represented. More detailed information is presented in Supplementary Table 1.

2. Experimental Design, Materials and Methods

At each of the water sources, a complex of geochemical water sampling was implemented, according to our work [6]. Water samples were taken 30 m from the shore.

The pH and Eh values of water samples were measured in situ using a portable Infraspak-Analit Anion 7051 water analyzer (Russia), with an electrode calibrated against standard buffer
solutions of pH = 1.68–4.01–6.86–9.18. The absolute errors did not exceed ±0.02 for pH and ±2 for Eh. The temperature was measured with a mercury thermometer.

Unfiltered water samples were taken to determine the cationic and anionic composition, TOC, mercury and potentially toxic elements (PTE). Water for anionic composition and TOC were poured into polypropylene vials and did not acidify. All water samples for Hg were poured into borosilicate glass Pyrex vials, for other PTEs – into polypropylene vials. The samples for both major ion and Hg analyses were acidified with double-distilled HNO₃ (1 ml per 200 ml of water). To determine the total suspended solids (TSS), the solutions were filtered through membrane filters of 0.45 µm.

Bottom sediment samples and mats were taken into sterile 50 ml Falcon tubes, fixed with an equal volume of 96% ethanol, and stored at −70 °C. The sampling was carried out in the places of the source outlet and the outflow of the artesian well. The sampling depth is about 40 cm.

Carbon was measured by IR spectroscopy on a Shimadzu Total Organic Carbon Analyzer, TOC-V_CSH (Japan). The detection limit was 4 ppb and the analytical accuracy was better than 1.5%.

The anion composition of water samples was determined by capillary electrophoresis (CE) on a Lumex Kapel 103P instrument (Russia). The detection limits were from 0.1 to 10 mg/L for different ions, the relative error did not exceed 15%.

The contents of cations in the water samples of 2018 were determined by the flame atomic absorption spectroscopy (FAAS). The FAAS assay was performed on a Thermo Electron Solar M6 spectrometer with a Zeeman and deuterium background correction systems (USA). Relative analytical errors were ±35% for element concentrations at 5·10⁻⁶ wt.% and ±10% for higher concentrations up to 10 wt.%, at P = 0.95 confidence probability. The cations in the waters sampled in 2019 were determined by inductively-coupled plasma mass spectrometry (ICP-MS) on an Agilent 7500 spectrometer. Relative analytical errors were 5%, at P = 0.95 confidence probability.

The Hg contents in all samples were measured by the cold-vapor method, using SnCl₂ as a reducing agent, with subsequent FASS assay, on a RA- 915 M analyzer with a Lumex RP-92 system (Russia). The detection limit was 0.02 µg/L (waters) and 0.01 g/t (solid samples), and the relative error was 20%.

To isolate total DNA, 0.3 g of the solid fraction of the samples were taken. The liquid fraction was removed after centrifugation. Total DNA was isolated using NucleoSpin® Soil kit (Macherey-Nagel). The procedure was carried out in accordance with the manufacturer’s protocol. DNA was extracted separately for all samples.

We used the degenerate primers U806R (5′-GGACTACNVGGGTWTCTAAT-3′) and U343F (5′-CCTACCGGRSGCGAGCAG-3′) to obtain the target 16S rRNA fragment (region V3-V4). The primers used have previously proven their ability to amplify a wide range of microorganisms from various phyla [7–9]. To obtain a library of gene fragments with the smallest number shift, we used low-temperature annealing of primers and Fusion Polymerase Q5 from New England Biolabs. The first amplification regime: 96 °C-2′; 25°C(96 °C-8′; 54 °C-20′; 68 °C-30′). The product of the first amplification was purified and used for the second PCR reaction. The second amplification regime: 96 °C-2′; 5°C(96 °C-8′; 54 °C-20′; 68 °C-30′); 20°C(96 °C-8′; 60 °C-20′; 68 °C-30′). Illumina sequencing adapters and dual-index barcodes were attached to the target amplicon during the second PCR reaction. 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.

Bioinformatic processing of paired reads of the 16S rRNA gene was performed on the QIIME2 v.2020.2 platform [10]. Using the DADA2 tool, noise was removed, paired reads were integrated, and OTUs were built. Taxonomic classification of the obtained OTUs was carried out using the scikit-learn classifier, which was trained on fragments of 16S rRNA from the Greengenes v.13.8 database, limited by the used primers. Phylogenetic tree was constructed in the MEGA v. 10.1.8 [11] using the maximum likelihood method. The Kimura 2-parameter model with gamma distribution and invariant sites was chosen, the rest parameters were default. Tree tested by Bootstrap method (1000 replications). The replacement model was selected using the MEGA’s built-in best model search tool.
Ethics Statement

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

CRediT Author Statement

Alexei S. Rozanov: Visualization, Methodology; Irina N. Myagkaya: Investigation, Writing original draft, Writing review & editing, Project administration; Anton V. Korzhuk: Writing original draft, Visualization; Nikita I. Ershov: Data curation; Ivan S. Kirichenko: Visualization; Maria A. Gustaytis: Investigation; Bagai-ool Yu. Saryg-oool: Investigation; Victor I. Malov: Investigation; Aleksandra A. Shipova: Investigation; Elena V. Lazareva: Conceptualization; Sergey E. Peltek: Supervision.

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 fieldwork was partly performed on a government assignment to the V.S. Sobolev Institute of Geology and Mineralogy. Geochemical data were obtained at the Analytical Center for Multi-Elemental and Isotope Research (Novosibirsk, Russia). Geochemical data and microbiological sequencing were obtained with the support of the Russian Science Foundation # 18–77–10056. Data processing was carried out with the support of the ICG SB RAS Budget Project no. 0324–2019–0040–C-01 “Genetic basis of biotechnology and bioinformatics”. 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.2021.107099.

References

[1] I.V. Gaskov, Features of endogenous metallogeny of mountainous and ore Altai (Russia), Geol. Geophys. 59 (2018) 1254–1270 (in Russian), doi:10.15372/gig20180809.

[2] E.V. Deev, I.V. Turova, A.P. Borodovskiy, I.D. Zolnikov, L. Oleszczak, Unknown large ancient earthquakes along the Kurai fault zone (Gorny Altai): new results of palaeoseismological and archaeoseismological studies, Int. Geol. Rev. 59 (2017) 293–310, doi:10.1080/00206814.2016.1258675.

[3] S.I. Fedak, Y.A. Turkin, A.I. Gusev, S.P. Shokalsky, G.G. Rusanov, B.A. Borisov, G.M. Belyaev, E.M. Leontyeva, State Geological Map of the Russian federation. Scale 1:1000000. Third generation. Series: Altai-Sayan. Sheet M-45–Gorno-Altaisk. Explanatory note. VSEGEI, St.Petersburg, 2011 (in Russian).

[4] E.V. Deev, E.V. Sokol, Yu.M. Ryapolova, S.N. Koch, G.G. Rusanov, Quaternary travertines of the Kurai fault zone (Gorny Altai), Rep. Acad. Sci. 473 (2017) 54–59 (in Russian), doi:10.7868/s0869565217070155.

[5] S.N. Kokh, E.V. Sokol, E.V. Deev, Yu.M. Ryapolova, Indicator characteristics of continental carbonates by the example of paleotravertines in Gorny Altai, Met. Anc. Mod. Ocean. 1 (2016) 243–247 https://www.elibrary.ru/item.asp?id=25963740. (in Russian).
[6] I.N. Myagkaya, E.V. Lazareva, M.A. Gustaytis, S.M. Zhmodik, Gold and silver in a system of sulfide tailings. part 1: migration in workflow, J. Geochem. Explor. 160 (2016) 16–30, doi:10.1016/j.gexplo.2015.10.004.

[7] 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 Zavarzin thermal spring (Uzon Caldera, Kamchatka, Russia), BMC Genom. 15 (2014) S12, doi:10.1186/1471-2164-15-S12-S12.

[8] A.S. Rozanov, A.V. Korzhuk, V.N. Shlyakhtun, A.A. Shipova, S.E. Peltek, Metagenomic data on prokaryotic diversity of Kunashir island geothermal spring, Data Br. 32 (2020) 106244, doi:10.1016/j.dib.2020.106244.

[9] A.V. Korzhuk, I.N. Myagkaya, A.S. Rozanov, N.I. Ershov, B. ool Y. Saryg-ool, V.I. Malov, M.A. Gustaytis, A.A. Shipova, E.V. Lazareva, S.E. Peltek, Metagenomics data of microbial communities of natural organic matter from the dispersion train of sulfide tailings, Data Br. 35 (2021) 106720, doi:10.1016/j.dib.2021.106720.

[10] E. Bolyen, et al., Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2, Nat. Biotechnol. 37 (2019) 852–857, doi:10.1038/s41587-019-0209-9.

[11] S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: molecular evolutionary genetics analysis across computing platforms, Mol. Biol. Evol. 35 (2018) 1547–1549, doi:10.1093/molbev/msy096.