



Metagenomic Sequencing of Microbial Communities from Brackish Water of Pangong Lake of the Northwest Indian Himalayas

Rashmi Rathour,^a Juhi Gupta,^a Madan Kumar,^a Moonmoon Hiloidhari,^a Anil Kumar Mehrotra,^b Indu Shekhar Thakur^a

School of Environmental Sciences, Jawaharlal Nehru University, New Delhi, India^a; Chemical and Petroleum Engineering, Schulich School of Engineering, University of Calgary, Calgary, Alberta, Canada^b

ABSTRACT Pangong is a brackish water lake having environmental conditions that are hostile to supporting life. This is the first report unveiling the microbial diversity of sediment from Pangong Lake, Ladakh, India, using a high-throughput metagenomic approach. Metagenomic data analysis revealed a community structure of microbes in which functional genetic diversity facilitates their survival.

Sediments are rich sources of microbial diversity and represent a special realm in aquatic environments (1). To overcome the limitations of the culture approach in studying these organisms, culture-independent approaches like metagenomics are applied to characterize microbial communities, discover novel genes, and analyze metabolic pathways directly from the environment (2, 3). There is very limited information available on the microbial diversity present at high-altitude cold habitats of the Himalayas (4). The present study investigates, through a metagenomic approach, the functional genetic diversity of microbes present in Pangong Lake, a large brackish water lake situated at a height of 4,250 m above mean sea level in the Himalayas. The microbes present there are halotolerant and cold adapted, and identifying the diversity of the novel cold-active enzymes and secondary metabolites assisting in the survival of these microbes may have great biotechnological potential.

The sediment samples were obtained from Pangong Lake (33°43'04.59"N: 78°53'48.48"E), Ladakh, J&K (Jammu and Kashmir) India, in September 2016 and stored at 4°C until further analysis. The DNA was extracted using the Exgene soil DNA kit (GeneAll Biotechnology Co., Ltd.), and sequencing was performed on the Illumina platform. The paired-end sequencing libraries (2 × 150 bp) were prepared using the Illumina TruSeq Nano DNA library prep kit and were sequenced on the Illumina NextSeq500 platform. The raw data were processed to obtain high-quality clean reads (quality value >20) using Trimmomatic version 0.35 (5). The filtered high-quality reads of the sample were assembled into scaffolds using CLC Genomics Workbench, and genes were predicted using Prodigal version 2.6.3 with default parameters (6). Taxonomic analysis of the predicted genes was carried out using Kaiju (7), a program for sensitive taxonomic classification of high-throughput metagenomics sequencing data. Cognizer (8), which is a comprehensive stand-alone framework that simultaneously provides COG (9), KEGG (10), Pfam (11), GO (12), and SEED (13) subsystem annotations to individual sequences constituting metagenomics data sets, was used for performing the functional analysis of the genes.

The mean of the library fragment size distribution was 486 bp, and ~3 Gb of high-quality data were obtained, with 10,386,213 reads assembled into scaffolds. After assembly, the total size of the scaffolds was 248,068 bp, with an N_{50} value of 635 bp,

Received 21 August 2017 Accepted 23 August 2017 Published 5 October 2017

Citation Rathour R, Gupta J, Kumar M, Hiloidhari M, Mehrotra AK, Thakur IS. 2017. Metagenomic sequencing of microbial communities from brackish water of Pangong Lake of the northwest Indian Himalayas. *Genome Announc* 5:e01029-17. <https://doi.org/10.1128/genomeA.01029-17>.

Copyright © 2017 Rathour 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 Indu Shekhar Thakur, isthakur@mail.jnu.ac.in.

and 337,527 genes, with an average gene length of 401 bp, were predicted. The predicted genes having a length of <300 bp were discarded from taxonomical analysis and functional classification. Taxonomical classification was as follows: bacteria (83.86%), archaea (0.24%), eukaryotes (0.42%), viruses (0.41%), and unclassified (15.02%). The major phyla represented were *Proteobacteria* (54.36%), *Bacteroidetes* (24.01%), *Firmicutes* (1.14%), *Actinobacteria* (0.85%), *Balneolaeota* (0.79%), *Cyanobacteria* (0.59%), *Verrucomicrobia* (0.47%), *Euryarchaeota* (0.21%), *Planctomycetes* (0.19%), and *Ascomycota* (0.10%). At the genus level, *Methylophaga* (10.19%) was found to be the most abundant. Functional analysis of the sequence classified most of the data as being related to carbohydrate metabolism, energy metabolism, lipid metabolism, and nucleotide metabolism.

Metagenomic analysis revealed a diverse domain of microbial communities thriving in harsh conditions, creating a base for further microbial exploration to improve the efficacy of bioprospecting metagenomics of soil and sediment, which may lead to the discovery of novel enzymes and bioactivities.

Accession number(s). The nucleotide sequences reported here have been submitted to the NCBI Sequence Read Archive (SRA) under accession number [SRX2861366](https://www.ncbi.nlm.nih.gov/sra/SRX2861366).

ACKNOWLEDGMENTS

R.R. is grateful to the UGC:NFSC, New Delhi, for a Junior Research Fellowship. I.S.T. extends thanks to the Shastri Indo-Canadian Institute (SICI), New Delhi, India, for financial support as a visiting scientist during the period in which the manuscript was prepared.

We have not received grants from other funding agencies. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

REFERENCES

- Wang Y, Sheng HF, He Y, Wu JY, Jiang YX, Tam NF, Zhou HW. 2012. Comparison of the levels of bacterial diversity in freshwater, intertidal wetland, and marine sediments by using millions of Illumina tags. *Appl Environ Microbiol* 78:8264–8271. <https://doi.org/10.1128/AEM.01821-12>.
- Vollmers J, Wiegand S, Kaster AK. 2017. Comparing and evaluating metagenome assembly tools from a microbiologist's perspective—not only size matters. *PLoS One* 12:e0169662. <https://doi.org/10.1371/journal.pone.0169662>.
- Wobus A, Bleul C, Maassen S, Scheerer C, Schuppler M, Jacobs E, Röske I. 2003. Microbial diversity and functional characterization of sediments from reservoirs of different trophic state. *FEMS Microbiol Ecol* 46: 331–347. [https://doi.org/10.1016/S0168-6496\(03\)00249-6](https://doi.org/10.1016/S0168-6496(03)00249-6).
- Gangwar P, Alam SI, Bansod S, Singh L. 2009. Bacterial diversity of soil samples from the western Himalayas, India. *Can J Microbiol* 55:564–577. <https://doi.org/10.1139/w09-011>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- 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>.
- Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7:11257. <https://doi.org/10.1038/ncomms11257>.
- Bose T, Haque MM, Reddy C, Mande SS. 2015. COGNIZER: a framework for functional annotation of metagenomic datasets. *PLoS One* 10: e0142102. <https://doi.org/10.1371/journal.pone.0142102>.
- Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res* 28:33–36. <https://doi.org/10.1093/nar/28.1.33>.
- Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res* 44:D457–D462. <https://doi.org/10.1093/nar/gkv1070>.
- Finn RD, Bateman A, Clements J, Coghill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42: D222–D230. <https://doi.org/10.1093/nar/gkt1223>.
- Reference Genome Group of the Gene Ontology Consortium. 2009. The Gene Ontology's Reference Genome Project: a unified framework for functional annotation across species. *PLoS Comput Biol* 5:e1000431. <https://doi.org/10.1371/journal.pcbi.1000431>.
- Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M, de Crécy-Lagard V, Diaz N, Disz T, Edwards R, Fonstein M, Frank ED, Gerdes S, Glass EM, Goesmann A, Hanson A, Iwata-Reuyl D, Jensen R, Jamshidi N, Krause L, Kubal M, Larsen N, Linke B, McHardy AC, Meyer F, Neuweger H, Olsen G, Olson R, Osterman A, Portnoy V, Pusch GD, Rodionov DA, Rückert C, Steiner J, Stevens R, Thiele I, Vassieva O, Ye Y, Zagnitko O, Vonstein V. 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res* 33:5691–5702. <https://doi.org/10.1093/nar/gki866>.