Genomics Data 5 (2015) 13–14

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

Illumina-based analysis of bacterial community in Khuangcherapuk cave of Mizoram, Northeast India

Surajit De Mandal, Amrita Kumari Panda, Esther Lalnunmawii, Satpal Singh Bisht, Nachimuthu Senthil Kumar *

Department of Biotechnology, Mizoram University, Aizawl 796004, Mizoram, India

ARTICLE INFO

Article history:
Received 9 April 2015
Accepted 28 April 2015
Available online 8 May 2015

Keywords:
Khuangcherapuk cave
Illumina
Metagenome

ABSTRACT

Bacterial community of the Khuangcherapuk cave sediment was assessed by Illumina amplicon sequencing. The metagenome comprised of 533,120 raw reads with an average base quality (Phred score) 36.75 and G + C content is 57.61%. A total of 18 bacterial phyla were detected with following abundant genus — *Mycobacterium* (21.72%), *Rhodococcus* (7.09%), *Alteromonas* (1.42%), *Holomonas* (0.7%) and *Salinisphaera* (0.20%). Majority portion of the sequences (68%) is unclassified at the genus level indicating the possibilities for the presence of novel species in this cave. This study reports the cave bacterial diversity from the biodiversity hotspot region of Eastern Himalayas. Metagenome sequence data are available at NCBI under the Bioproject database with accession no. SRP056890.

Specifications	
Organism/cell line/tissue	Illumina-based analysis of bacterial community in Khuangcherapuk cave of Mizoram, Northeast India
Sex	Not applicable
Sequencer or array type	Illumina
Data format	Analyzed
Experimental factors	Environmental sample
Experimental features	V3 hypervariable region of 16S rDNA was sequenced using paired end Illumina Mi-Seq technology and the sequence was analyzed using QIIME data analysis package.
Consent	Not applicable
Sample source location	Sediment sample, Khuangcherapuk cave, Mizoram, Northeast India

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/?term=SRP056890.

2. Experimental design, materials and methods

Most of the diversity study on cave microbiology is based on the cultivation approach which can determine less than 1% of the microbes [1] With the advancement of culture independent technique using next generation sequencing or clone library construction, it is now possible to analyze the entire population in the community as well as their functional potentiality in extreme environments [2–4] Therefore, the present research was intended to analyze the bacterial community using Illumina based metagenomic approach in Khuangcherapuk cave, which is devoid of any light source and thrives under energetically unfavorable and nutrient-poor conditions.

Samples were collected during February 2014 from the Khuangcherapuk Cave (23°41′30″ N, 92°37′5″E), Ailawng village, Mizoram, Northeast India. The cave is 162 m long with a vertical range of 10 m depth and is considered as the biggest cave in Mizoram. Ten individual composite sediment samples were collected from different places of the cave floor and DNA was extracted using the Fast DNA spin kit (MP Biomedical, Solon, OH, USA). The extracted DNA was purified twice using 0.5% low melting point agarose gel and mixed to prepare a composite sample.

The V3 hypervariable region of the 16S rRNA gene was amplified using F 341/R518 primer combination (5′-CCTACGGGAGGCAGCAG-3′; 5′-ATTACCGCGGCTGCTGG-3′). Amplicon metagenomic sequencing was performed using the Illumina Mi-Seq platform and the analysis and annotation of output data were carried out by QIIME data analysis package [5]. Raw sequences were filtered based on base quality score, average base content per read and GC distribution in the reads. Reads that did not cluster with other sequences i.e. singletons (abundances <2) were removed. Chimeras were also removed using UCHIME program [6] The pre-processed consensus V3 sequences were finally grouped into operational taxonomic units (OTUs) using the clustering program UCLUST at a similarity threshold of 0.97 [7] All the pre-processed reads were used to identify the OTUs using QIIME program and the representative sequences were aligned against the Greengenes core set reference database using PyNAST program [8]. Representative

* Corresponding author. Tel.: +91 9436352574 (mobile).
E-mail address: nskmzu@gmail.com (N.S. Kumar).

http://dx.doi.org/10.1016/j.gdata.2015.04.023

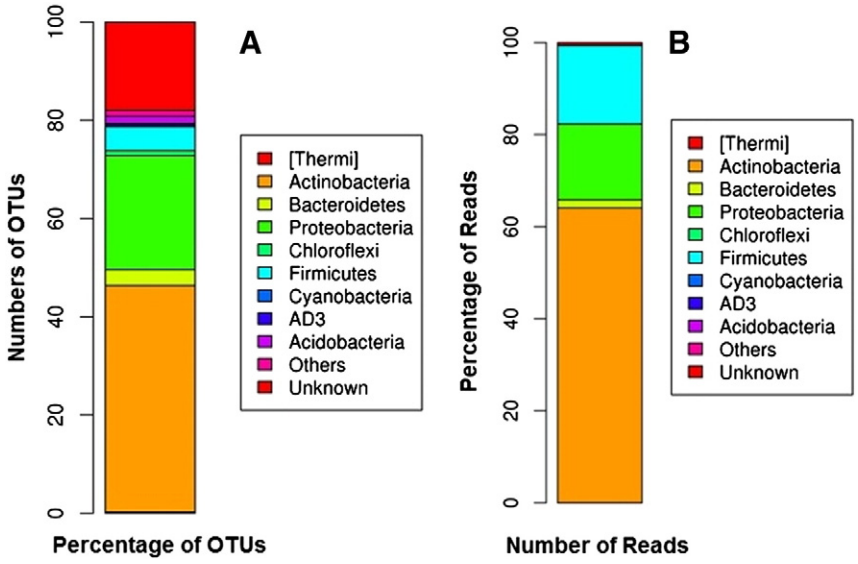

Fig. 1. Taxonomy classification of reads at phylum level (A), OTUs at phylum level (B) for the sample. Only top 10 enriched class categories are shown in the figure.

sequence for each OTU was classified using RDP classifier and Greengenes OTU database.

The output file comprised 161 MB data with a total of 533,120 raw reads having 57.61% GC content. A total of 18 bacterial phyla were detected in our analysis. The most dominant prokaryotic phylum was Actinobacteria (64.07%), a broad class of high G + C, Gram-positive bacteria commonly found in caves and soils [9]. In this phylum, 34.26% reads were classified under the genus *Mycobacterium*. Other dominant phyla were Firmicutes (17.06%), Proteobacteria (16.43%), Bacteroidetes (1.75%) and Chloroflexi (0.02%) (Fig. 1B). At the family level, Mycobacteriaceae (21.95%) was dominant followed by Bacillaceae (17.04%), Sphingomonadaceae (9.74%), Alteromonadaceae (1.53%), Salinisphaeraceae (0.44%), Xanthomonadaceae (0.39%), Flavobacteriaceae (0.18%) and Moraxellaceae (0.005%). The leading genera were *Mycobacterium* (21.72%), *Rhodococcus* (7.09%), *Alteromonas* (1.42%), *Holomonas* (0.7%) and *Salinisphaera* (0.20%) (Supplementary Figs. 1 and 2). Among the identified species *Rhodococcus fascians* was present in high numbers which is reported to participate in Calcite Biomineralization process [10] Our data provides the first scientific report on diverse group of bacteria, using Illumina sequencing method, from the unexplored Khuangcherapuk cave located in a lesser known Northeastern Indian region. The most dominated phylum in this study was actinomycetes which are known to produce valuable secondary metabolites useful for biotechnological applications. This study also detected a huge number of unclassified bacteria which might be representative of novel species.

3. Nucleotide sequence accession number

Metagenome sequence data are available at NCBI accession no. SRP056890.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2015.04.023.

4. Competing interests

The authors declare that there are no competing interests.

Acknowledgements

This research was funded by a grant from the Bioinformatics Infrastructure Facility sponsored by Department of Biotechnology, Govt. of India, New Delhi.

References

[1] R.I. Amann, B.J. Binder, R.J. Olson, S.W. Chisholm, R. Devereux, D.A. Stahl, Combination of 16S rRNA targeted oligonucleotide probes with flow-cytometry for analyzing mixed microbial populations. Appl. Environ. Microbiol. 56 (1990) 1919–1925.
[2] A.V. Mangrola, P. Dudhagara, P. Koringa, C.G. Joshi, R.K. Patel, Shotgun metagenomic sequencing based microbial diversity assessment of Lasundra hot spring, India. Genomics Data 4 (2015) 73–75.
[3] W. Xie, F. Wang, L. Guo, Z. Chen, S.M. Sievert, J. Meng, G. Huang, Y. Li, Q. Yan, S. Wu, X. Wang, S. Chen, G. He, X. Xiao, A. Xu, Comparative metagenomics of microbial communities inhabiting deep-sea hydrothermal vent chimneys with contrasting chemistries. ISME J. 5 (3) (2011) 414–426.
[4] A. Ghelani, R. Patel, A. Mangrola, P. Dudhagara, Cultivation-independent comprehensive survey of bacterial diversity in Tulsi Shyam hot springs, India. Genomics Data 4 (2015) 54–56.
[5] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, N. Fierer, A.G. Pena, J.K. Goodrich, J.I. Gordon,, et al., QIIME allows analysis of high-throughput community sequencing data. Nat. Methods 7 (2010) 335–336.
[6] Edgar, et al., UCHIME improves sensitivity and speed of chimera detection. Bioinformatics 27 (16) (2011) 2194–2200.
[7] R.C. Edgar, et al., Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26 (2010) 2460–2461.
[8] T.Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E.L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, G.L. Anderson, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl. Environ. Microbiol. 72 (2016) 5069–5072.
[9] M. Goodfellow, S.T. Williams, Ecology of actinomycetes. Annu. Rev. Microbiol. 37 (1983) 189–216.
[10] A. Rusznyák, D.M. Akob, S. Nietzsche, K. Eusterhues, K. Totsche, T. Neu, T. Frosch, P. Popp, R. Keiner, J. Geletneky, L. Katzschmannn, E.D. Schule, K. Küsel, Calcite biomineralization by bacterial isolates from the recently discovered pristine karstic herrenberg cave. Environ. Microbiol. 78 (2012) 1157–1167.