Metagenome Sequencing Revealed *Rhodococcus* Dominance in Farpuk Cave, Mizoram, India, an Eastern Himalayan Biodiversity Hot Spot Region

Surajit De Mandal, Zothan Sanga, Nachimuthu Senthil Kumar

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

The present study employed 16S rRNA amplicon sequencing to survey the prokaryotic microbiota on Farpuk Cave, revealing a diverse bacterial community with 4,021 operational taxonomical units (OTUs), mainly dominated by the genus *Rhodococcus*. Moreover, 18.17% of the OTUs were unclassified at the phylum level, suggesting the existence of novel bacterial species.

Address correspondence to Nachimuthu Senthil Kumar, nskmzu@gmail.com.

Mizoram, India, falling under the biodiversity hot spot regions of eastern Himalayas, is not well known for its geologically unique cave habitats (1). Farpuk Cave provides unique opportunities to understand the community structure of the large reservoirs of unknown microbial life. In the present study, we performed 16S rRNA amplicon metagenomic sequencing of sediment samples collected from Farpuk Cave located in Champhai, Mizoram, in northeast India (23.10N 92.53E). Sediment samples were collected in sterilized containers from 10 sites of Farpuk Cave, and the genomic DNA was extracted using the FastDNA spin kit for soils (MP Biomedicals, Solon, OH, USA) and finally mixed to prepare a composite sample. The V3 hypervariable region of the 16S rRNA gene was amplified using the F341/R518 primer combination, and amplicon metagenomic sequencing was performed using the Illumina MiSeq platform followed by the analysis and annotation of output data using the QIIME data analysis package (2, 3). The sequencing yield was 381.05 Mb of data consisting of 1,261,787 reads and a G+C content of 59.03%. The average base quality (Phred score) was 36.79, and the individual sequence length was 150 bp. After removing singletons (abundances <2) and chimeras, 875,614 preprocessed consensus V3 reads were grouped into 4,021 operational taxonomical units (OTUs) at a similarity threshold of 0.97 using UCHIME and UCLUST (4, 5). The OTU representative sequence was aligned using the PyNAST tool (6), and the reference sequence of each OTU was classified using the Ribosomal Database Project (RDP) classifier and Greengenes OTU database (7, 8).

Among the 11 phyla detected in the cave metagenome, *Actinobacteria* (81.43%) was the most abundant phylum of bacteria. Other sequences were classified as follows: *Firmicutes* (10.41%), *Proteobacteria* (2.83%), *Acidobacteria* (2.39%), *Gemmatimonadetes* (0.30%), and *Bacteroidetes* (0.14%). These detected phyla were common inhabitants of the cave microbial community and found in other subsurface environments (9, 10). About 79.48% of the identified genera fell under *Rhodococcus*, an aerobic, nonsporulating, nonmotile Gram-positive bacterium that can catabolize a wide range of compounds. They are also known to produce bioactive steroids, acrylamide, and acrylic acid and are involved in fossil fuel biodesulfurization (11, 12). *Rhodococcus fascians*, *Propionibacterium acnes*, *Glaciecola polaris*, *Mycobacterium celatum*, *Virgisporangium ochraceum*, *Actinomadura vinacea*, and *Bacillus foraminis* were the main bacterial species in the cave sediments. However, a large number of reads did not classify at the phylum level, suggesting the existence of novel bacteria in Farpuk Cave. Further studies with whole-metagenome sequencing will resolve the industrially important novel genes and metabolic pathways.

Nucleotide sequence accession number. The sequences obtained in this project have been deposited in the NCBI Short Read Archive under the accession no. SRP057997.

REFERENCES

1. **Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GA, Kent J.** 2000. Biodiversity hotspots for conservation priorities. Nature **403:** 853–858. http://dx.doi.org/10.1038/35002501.
2. **Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE, Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R.** 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods **7:** 335–336. http://dx.doi.org/10.1038/nmeth.f.303.
3. **De Mandal S, Zothansanga, Panda AK, Bisht SS, Kumar NS.** 2015. First report of bacterial community from a bat guano using Illumina next-generation sequencing. Genomics Data **4:** 99–101. http://dx.doi.org/10.1016/j.gdata.2015.04.001.
4. **Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R.** 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27:** 2194–2200. http://dx.doi.org/10.1093/bioinformatics/btr381.
5. **Edgar RC.** 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26:** 2460–2461. http://dx.doi.org/10.1093/bioinformatics/btq461.

6. **Caporaso JG, Bittinger K, Bushman FD, DeSantis TZ, Andersen GL, Knight R**. 2010. PyNAST: a flexible tool for aligning sequences to a template alignment. Bioinformatics **26**:266–267. http://dx.doi.org/10.1093/bioinformatics/btp636.

7. **Wang Q, Garrity GM, Tiedje JM, Cole JR**. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol **73**:5261–5267. http://dx.doi.org/10.1128/AEM.00062-07.

8. **DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL**. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol **72**:5069–5072. http://dx.doi.org/10.1128/AEM.03006-05.

9. **Cuezva S, Fernandez-Cortes A, Porca E, Pašić L, Jurado V, Hernandez-Marine M, Serrano-Ortiz P, Hermosin B, Cañaveras JC, Sanchez-Moral S, Saiz-Jimenez C**. 2012. The biogeochemical role of *Actinobacteria* in Altamira Cave, Spain. FEMS Microbiol Ecol **81**:281–290. http://dx.doi.org/10.1111/j.1574-6941.2012.01391.x.

10. **Pašić L, Kovče B, Sket B, Herzog-Velikonja B**. 2010. Diversity of microbial communities colonizing the walls of a karstic cave in Slovenia. FEMS Microbiol Ecol **71**:50–60. http://dx.doi.org/10.1111/j.1574-6941.2009.00789.x.

11. **Vander-Geize R, Dijkhuizen L**. 2004. Harnessing the catabolic diversity of rhodococci for environmental and biotechnological applications. Curr Opin Microbiol **7**:255–261. http://dx.doi.org/10.1016/j.mib.2004.04.001.

12. **McLeod MP, Warren RL, Hsiao WW, Araki N, Myhre M, Fernandes C, Miyazawa D, Wong W, Lillquist AL, Wang D, Dosanjh M, Hara H, Petrescu A, Morin RD, Yang G, Stott JM, Schein JE, Shin H, Smailus D, Siddiqui AS, Marra MA, Jones SJM, Holt R, Brinkman FSL, Miyauchi K, Fukuda M, Davies JE, Mohn WW, Eltis LD**. 2006. The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. Proc Natl Acad Sci U S A **103**:15582–15587. http://dx.doi.org/10.1073/pnas.0607048103.