Data in Brief

Bacterial diversity of Murlen National Park located in Indo-Burman Biodiversity hotspot region: A metagenomic approach

Surajit De Mandal [a], Zothansanga [a], H.T. Lalremsanga [b], Nachimuthu Senthil Kumar [a,b,*]

[a] Department of Biotechnology, Mizoram University, Aizawl 796004, Mizoram, India
[b] Department of Zoology, Mizoram University, Aizawl 796004, Mizoram, India

ABSTRACT

Paired end Illumina Mi-Seq sequencing of 16S rRNA gene amplicon was carried out to study the bacterial community in the soil of Murlen National Park located in Indo-Burman Biodiversity hotspot region. Metagenome consisted of 302,416 reads with 151.81 Mb data and G + C content of 56.48%. More than 85% sequence was having a Phred score >=Q30 and individual sequence length was 251 bp. Metagenome sequence data are available at NCBI under the Bioproject database with accession no. SRP057136. Community metagenomics revealed a total of 1802 species belonging to 29 different phyla dominated by Acidobacteria (39.45%), Proteobacteria (26.95%) and Planctomycetes (7.81%). Our data detected a wide group of bacterial community which will be useful in further isolating and characterizing the economic importance of bacteria from this region.

Specifications	
Organism/cell line/tissue	Illumina-based analysis of bacterial community in Murlen National Park of Mizoram, Northeast India
Sex	Not applicable
Sequencer or array type	Illumina
Data format	Analyzed
Experimental factors	Environmental sample
Experimental features	16S rRNA Illumina sequencing
Consent	Not applicable
Sample source location	Soil sample, Murlen National Park, Mizoram, Northeast India

1. Direct link to deposited data

http://www.ncbi.nlm.nih.gov/sra/?term=SRP057136.

2. Experimental design, materials and methods

Microbes are the biggest reservoir of biodiversity in natural forests and are involved in nutrient transformations thereby maintaining the forest ecosystem [1]. There are many natural forest types present in the North Eastern Region of India and many are still far from human understanding and any kind of anthropogenic interference. Murlen National Park is an undisturbed natural forest falling within the Indo-Burman Biodiversity hotspot zone. This region contains diverse ecological niches with leading hotspots in terms of endemics and provides unique niches for the evolution of novel microorganisms although this ecologically important region has been poorly explored in terms of soil microflora [2]. The ecology and functional role of most of the microbes are unknown due to their cultural constrains.

A recent advance in molecular microbial ecology has provided means to identify microbial life in many ecosystems without the need to culture the microorganisms through a metagenomic approach [3–4]. Next-generation sequencing technology has been successfully applied to study bacterial diversity in different environments such as bat guano, hot spring, lake, cave sediments, etc. [5–8].

Soil was collected from Murlen forest (21° 49′ 53.581″ N, 88° 36′ 44.860″ E) in a sealed sterile container and the metagenomic DNA was extracted separately using MP-Biomedical Soil kit and finally mixed to prepare a composite sample. V4 region of bacterial 16S rRNA gene was amplified by PCR according to De Mandal et al.[5]. Sequencing was performed on Illumina MiSeq platform and the analysis was carried out by QIIME data analysis package [5,9].

Metagenome consisted of 151.81 Mb data consisting of 302,416 reads with a G + C content of 56.48%. More than 85% of the sequence had a Phred score >=Q30 and individual sequence length was 251 bp (Supplementary Table 1). A total of 29 phyla were detected in our analysis and the Dominant phyla were Acidobacteria (39.45%)

* Corresponding author at: Department of Biotechnology, Mizoram University, Aizawl 796004, Mizoram, India. Tel.: +91 9436352574 (Mobile).
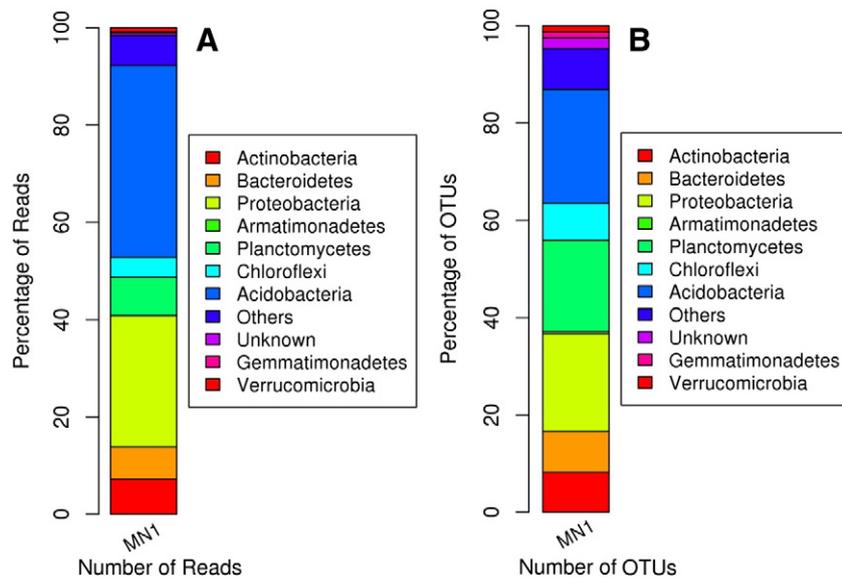E-mail address: nskmzu@gmail.com (N.S. Kumar).

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.

followed by Proteobacteria (26.95%), Planctomycetes (7.81%), Actinobacteria (7.18%), Bacteroidetes (6.65%), Chloroflexi (4.11%) and Nitrospirae (3.33%). However, 0.51% of the reads were not classified at the phylum level (Fig. 1). The analysis identified 100 bacterial orders dominated by iii1–15 (23.94%), Rhizobiales (13.97%), Actinomycetales (5.83%), Solibacterales (5.06%), Burkholderiales (4.83%), Saprospirales (4.61%), Nitrospirales (3.33%), Gemmatales (3.40%), Ellin6513 (2.94%), Thermogemmatisporales (2.74%) and Acidobacteriales (2.55%). A total of 138 genus were detected and the most dominating among them were DA101 (7.02%), followed by Gemmata (1.03%), Candidatus Xiphinematobacter (1.02%), and Rhodoplanes (0.92%). However, 84.44% reads were not identified at the genus level which suggested that the uniqueness and unidentified bacterial community structure might lead to reveal novel bacterial populations with some unique properties (Supplementary Fig. 1).

Our study, which is the first to study bacterial communities using high-throughput methods in Murlen national forest soil, a pristine reserve forest, revealed the presence of diverse bacterial population. This will be helpful in cataloging and describing the bacteria diversity of the forests of Mizoram that needs to be conserved and simultaneously could be the habitat for large number of economically important microbes.

Metagenome sequence data are available at NCBI Accession No. SRP057136.

Competing interests

The authors declare that there are no competing interests.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2015.04.025.

References

[1] D.S. Nichols, K. Sanderson, A. Buia, J. Kamp, P. Halloway, J.P. Bowman, M. Smith, C. Mancuso, P.D. Nichols, T.A. McMeekin, The Antarctic: Past, Present and Future. in: G.J. Jabour, M. Haward (Eds.),Antarctic CRC. Res. Report Hobart. 2002, pp. 85–103.

[2] N. Myers, R.A. Mittermeier, C.G. Mittermeier, G.A.B. da Fonseca, J. Kent, Biodiversity hotspots for conservation priorities. Nature 403 (2000) 853–858.

[3] V. Torsvik, L. Ovreas, Microbial diversity and function in soil: from genes to ecosystems. Curr. Opin. Microbiol. 5 (2002) 240–245.

[4] J. Handelsman, Metagenomics: application of genomics to uncultured microorganisms. Microbiol. Mol. Biol. Rev. 6 (2004) 669–685.

[5] S. De Mandal, A.K.P. Zothansanga, S.S. Bisht, N. Senthil Kumar, First report of bacterial community from a bat guano using Illumina next-generation sequencing. Genomics Data 4 (2015) 99–101 http://dx.doi.org/10.1016/j.gdata.2015.04.001.

[6] A.K. Panda, S.S. Bisht, N. Senthil Kumar, S. De Mandal, Investigation on microbial diversity of Jakrem hot spring, Meghalaya, India using cultivation-independent approach. Genomics Data (2015) http://dx.doi.org/10.1016/j.gdata.2015.04.016 Accepted in Genomics Data.

[7] A. Pramanik, P. Basak, S. Banerjee, S. Sengupta, D. Chattopadhyay, M. Bhattacharyya, Pyrosequencing based profiling of the bacterial community in the Chilika lake, the largest lagoon of India. Genomics Data (2015) http://dx.doi.org/10.1016/j.gdata.2015.04.003.

[8] S. De Mandal, E. Lalnunmawii, A.K. Panda, S.S. Bisht, N. Senthil Kumar, Illumina-based analysis of bacterial community in Khuangcherapuk cave of Mizoram, Northeast India. Genomics Data (2015) http://dx.doi.org/10.1016/j.gdata.2015.04.023 (Communicated to Genomics Data).

[9] 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.