Data Article

Metagenome dataset of lateritic soil microbiota from Sadaipur, Birbhum, West Bengal, India

Meenakshi Mukhopadhyay\textsuperscript{a,*}, Arup Kumar Mitra\textsuperscript{b,*}, Sudeshna Shyam Choudhury\textsuperscript{b}, Sayak Ganguli\textsuperscript{c}

\textsuperscript{a} Department of Botany, Vivekananda College, 269, Diamond Harbour Road, Kolkata, West Bengal 700063, India
\textsuperscript{b} Post Graduate Department of Microbiology, St. Xavier’s College (Autonomous), 30 Mother Teresa Sarani (Park Street), Kolkata, West Bengal 700016, India
\textsuperscript{c} Post Graduate Department of Biotechnology, St. Xavier’s College (Autonomous), 30 Mother Teresa Sarani (Park Street), Kolkata, West Bengal 700016, India

\textbf{A R T I C L E   I N F O}

Article history:
Received 10 February 2021
Revised 28 March 2021
Accepted 1 April 2021
Available online 6 April 2021

Keywords:
Lateritic soil
Metagenomic profile
Illumina
Proteobacteria
Sustainable agriculture

\textbf{A B S T R A C T}

The data represents the bacterial community profile obtained through metagenomic sequencing of soil sample, collected from the ‘Rarh’ region of West Bengal, which is characterized by the lateritic badlands dating back to the late Pliocene. Taxonomic binning and operational taxonomic unit (OTU) prediction of the Illumina sequencing data indicated the abundance Proteobacteria (61%) followed closely by Bacteroidetes (35%). The top two most abundant genera identified, were Sphingobacterium and Acinetobacter respectively. Chemical properties of soil, such as pH, organic carbon content, available nitrogen, phosphorus, and potassium were also analyzed for enabling future researchers to correlate the abundance of microbial taxa with the prevalent conditions. These findings can be effectively used to formulate strategic microbiome engineering through bioaugmentation for a sustainable agricultural system.

© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

\* Corresponding authors.
E-mail addresses: meenakshi.banerjee97@gmail.com (M. Mukhopadhyay), drakmitra01@sxccal.edu (A.K. Mitra).

https://doi.org/10.1016/j.dib.2021.107041
2352-3409/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
Specifications Table

| Subject                      | Microbiology                      |
|------------------------------|-----------------------------------|
| Specific subject area        | Applied Microbiology and Biotechnology |
| Type of data                 | Raw Sequence Reads                |
| How data were acquired       | Shotgun metagenome sequencing and analysis |
| Data format                  | FASTQ file                        |
| Parameters for data collection | Environmental sample; lateritic soil with low nutrient content, susceptible to soil erosion, abandoned for two years after long-term cultivation |
| Description of data collection | Whole bacterial DNA extraction from soil using the PowerSoil® DNA isolation kit. The Illumina platform was used for Shotgun metagenomic sequencing. |
| Data source location         | Latitude and longitude for collected samples/data: Sadaipur, Dubrajpur Block, Birbhum, West Bengal, India, located at 23.3260 °N, 87.4617 °E. |
| Data accessibility           | Repository name: NCBI SRA Data identification number: PRJNA689214 Accession No. SAMN17271210, Raw Sequence Reads_SAM1 Soil metagenome 410658. BioProject ID: PRJNA689214 Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA689214 |

Value of the Data

- The presented dataset is the first report on the status of microbiome in lateritic soil of ‘Rarh region’ of West Bengal. The soil is low in mineral nutrients, organic carbon content and has very low water holding capacity making agricultural practices difficult. Agricultural and environmental biologists will be benefitted from the present data.
- These data will help researchers to mitigate the problem of poor fertility by exploration of different microbial consortia and organic additives incorporated in traditional agricultural methods.
- From this data set, it may be possible to find out potentially beneficial soil bacteria having novel genes coding for enzymes with nutrient enhancing ability. A proper experiment can be designed based on these findings for utilization of these microbes to improve soil productivity for a sustainable agriculture.

1. Data Description

Loose, friable, nutrient-depleted lateritic soil [1] was collected, and analyzed for the chemical properties like pH, organic carbon content, available nitrogen, phosphorus and potassium. (Table 1). The same soil was used for sequencing using Illumina Miseq platform, and a total of 1,60,609 reads were obtained, out of which 197 reads did not pass quality filtering step. Finally, 1,60,412 reads were subsequently used for analysis (Fig. 1a). Only 2% of the reads represented Archaeal members and rest 98% were from Bacterial phylum. Proteobacterial abundance was predominant (Fig. 1b). At the genus level, Sphingobacterium (35%), Acinetobacter (31%) and Pseudomonas (7%) were the top three members in the total distribution of genera in the sample (Fig. 1c). Among the Proteobacterial members, Gammaproteobacteria (59%) was the most abun-

| Soil Parameter                  | Result           | Unit |
|---------------------------------|------------------|------|
| pH at 25 °C                      | 5.99             | -    |
| Total organic carbon            | 0.57± 0.011      | %    |
| Available Nitrogen (N)          | 81.91 ± 0.015    | mg/kg|
| Available Phosphorus (P)        | 6.90 ± 0.01      | mg/kg|
| Available Potassium (K)         | 126.57± 0.021    | mg/kg|
Fig. 1. Data Profile obtained through sequencing. (a) Read Counts as obtained after initial quality check where 197 reads were discarded; (b) Phylum abundances indicating Proteobacterial abundance; (c) Genus level abundances which exhibits the predominance of Sphingobacterium and Acinetobacter (d) Phylogenetic representation using Krona chart to depict the distribution of Proteobacterial members; (e) Phylogenetic representation using Krona chart to depict the distribution of the Bacillus clade at the genus level.
dant followed by alpha, and beta, with predominance of Sphingobacteriaceae, Moraxalleaceae, Enterobacteriaceae, and Pseudomonadiales (Fig. 1d). In the Bacillus clade, *Bacillus cereus* (17%) and *Bacillus thuringensis* (6%) were identified to be the most abundant members (Fig. 1e).

2. Experimental Design, Materials and Methods

2.1. Soil collection and chemical analysis

Soil sample was collected using the protocol as recommended by TNAU-2013 [2]. Chemical analysis was performed to estimate the pH, organic carbon content, available nitrogen, potassium, and phosphorus using the standard methods as followed in [3].

2.2. Next generation sequencing and metagenomic analysis

Metagenomic DNA extraction from the sieved soils was carried out using the PowerSoil™ DNA isolation kit (MOBIO), following which the sample was sequenced using the manufacturer's protocol. DNA quality was analyzed by Nanodrop and then evaluated on agarose gel. It was quantified using QUBIT. The library preparation was carried out using Illumina standardized V3-V4 regions of the 16S rRNA gene library protocol. The enriched library was further quantified and validated using qPCR and Agilent Bioanalyzer (DNA 1000 chip). The library generated containing V3-V4 amplicons [Primer Details: 16S Amplicon PCR Forward Primer = 5’TCCTCGGAGACTICTGATTTATAAGAGACAGCCTACGGGNGGCWGCAG; 16S Amplicon PCR Reverse Primer = 5’GTCTCTGCTTCAGATGTTATAAGAGACAGCCTACGGGNGGCWGCAG] and adaptor sequences: Forward overhang: 5’ TCGTCGACGCTATGATCAAGACAGCAGAGACTACHVGGGTATCTATCTATCC and reverse overhang: Reverse overhang: 5’ GTCTCGTACTCAGATGTTATAAGACAGCAGAGACTACHVGGGTATCTATCTATCC (locus specific sequence) Reverse overhang: 5’ GTCTCTGCTTCAGATGTTATAAGACAGCAGAGACTACHVGGGTATCTATCTATCC was sequenced on Illumina MiSeq using reagent kit V3 for generating 2 × 300 bp read length. The sequenced raw data files were passed through Quality check using the FASTQC pipeline, and sequences which passed the quality screening were then used for assembly using the SILVSngs (1.3) platform. It involved homopolymer removal [4] along with discarding of artifacts and contaminations [5]. Qiime was used to cluster the operational taxonomic units (OTUs), and KRONA charts were generated to analyze the microbial abundances. The pipeline followed was in accordance with the methods described in [6] and [7].

Funding

This work was financially supported by the Department of Higher Education, Science Technology and Biotechnology, Govt. of West Bengal, India [Memo No. 528 (Sanc.)/ST/P/S&T/2G-31/2018 Dated 26.03.2019]

Ethics Statement

Not applicable.

CRediT Author Statement

*Meenakshi Mukhopadhyay*: Conceptualization, Investigation, Data curation, Funding acquisition; Writing - Original Draft, Writing - Review & Editing; Formal Analysis; *Arup Kumar Mi*
M. Mukhopadhyay, A.K. Mitra and S.S. Choudhury et al. / Data in Brief 36 (2021) 107041

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The authors are grateful to the Principal, Vivekananda College, and St Xavier’s College, Kolkata for providing the necessary facilities to carry out the research work.

References

[1] District Census Handbook Birbhum, Census of India 2011, Series20, Part XII A 2011, Page 20: Soil and Cropping Pattern; Directorate of Census Operation, West Bengal. https://www.censusindia.gov.in/2011census/dchb/DCHB_A/19/1908_PART_A_DCHB_BIRBHUM.pdf (Retrieved on 2017).
[2] Tamil Nadu Agricultural University, India, TANU Agritech Portal, Resource Management: Soil: Soil Sampling Procedure. https://agritech.tnau.ac.in/agriculture/agri_soil_sampling.html/TNAU–2013.
[3] S. Mandal, M. Karmakar, F.H. Rahman, Study on soil nutrient status according to global positioning system in different blocks of Birbhum district of West Bengal, Curr. J. Appl. Sci. Tech. 39 (21) (2020) 64–70, doi:10.9734/cjast/2020/v39i2130824.
[4] R.P.A. Pereira, J. Peplies, I. Brettar, et al, Development of a genus-specific next generation sequencing approach for sensitive and quantitative determination of the legionella microbiome in freshwater systems, BMC Microbiol. 17 (79) (2017), doi:10.1186/s12866-017-0987-5.
[5] C. Lepinay, A. Mihajlovski, S. Touron, D. Seyer, F. Bousta, et al., Bacterial diversity associated with saline efflorescences damaging the walls of a French decorated prehistoric cave registered as a World cultural heritage site, J. IBiod 130 (2018) 55–64 Elsevier, doi:10.1016/j.ibiod.2018.03.016/hal-01951117.
[6] S. Ganguli, S. Rahaman, A.R. Bera, V. Vishal, S. Malik, K. Roopalakshmi, P.K. Singh, Rhizospheric metagenome of the terrestrial mangrove fern Acrostichum from Indian Sunderbans, Genom. Data 14 (2017) 53–55, doi:10.1016/j.j.gdata.2017.09.001.
[7] W. Chakraborty, S. Ganguli, P. Nandy (Dutta), Rhizospheric metagenome dataset of the terrestrial mangrove nypa fruticans wurmb. from Indian Sunderbans, J. Environ. Sociobiol. 17 (1) (2020) 79–82.