



# Metagenomic Analysis of Microbial Diversity in Landfill Lysimeter Soil of Ghazipur Landfill Site, New Delhi, India

Juhi Gupta, Rashmi Rathour, Madan Kumar, Indu Shekhar Thakur

School of Environmental Sciences, Jawaharlal Nehru University, New Delhi, India

**ABSTRACT** We report the soil microbial diversity and functional aspects related to degradation of recalcitrant compounds, determined using a metagenomic approach, in a landfill lysimeter prepared with soil from Ghazipur landfill site, New Delhi, India. Metagenomic analysis revealed the presence and functional diversity of complex microbial communities responsible for waste degradation.

Soil provides an intricate and challenging environment for microbiologists, with a large percentage of uncharacterized microbes that can be elucidated only through a metagenomic approach (1, 2). A lysimeter is a simulation of existing landfill conditions for waste degradation and leachate treatment that helps to resolve the problems associated with unengineered waste dumping sites (3). The aim of this study was to investigate through a metagenomic approach the functional genetic diversity of the diverse microbial communities responsible for degradation in a landfill lysimeter. Samples were collected in August 2016 from a landfill site (28°37'25.11"N, 77°19'36.1"E) located at the Ghazipur area in New Delhi, India. A composite sample was prepared from sampling positions and transferred to the physical setup of a lysimeter to study the leachate characteristics in an open dump.

DNA was extracted from a landfill lysimeter soil sample using an Exgene soil DNA kit (GeneAll Biotechnology Co., Ltd.). Paired-end sequencing libraries (2 × 150 bp) were prepared using an Illumina TruSeq Nano DNA library pep kit followed by sequencing using the Illumina NextSeq 500 platform. The raw reads were processed through Trimmomatic v 0.35 (4) to obtain high-quality clean reads devoid of adapter sequences and low-quality reads. The high-quality reads were assembled through CLC Genomics Workbench. Prodigal 2.6.3 (5) was used with default parameters to predict genes from the assembled scaffolds. Kaiju (6) (a program for sensitive taxonomic classification of high-throughput metagenomic sequencing data) was used for taxonomical analysis of the predicted genes, and functional analysis of genes from the sample was carried out using COGNIZER (7) (a comprehensive standalone framework) enabled to simultaneously provide COG (8), KEGG (9), Pfam (10), GO (11), and SEED (12) subsystems annotation to individual sequences constituting metagenomic data sets.

A mean library fragment size of 490 bp and ~3 Gb high-quality data were obtained. A total of 10,254,512 high-quality reads were assembled into scaffolds. A scaffold of 406,004 bp was retrieved, with an  $N_{50}$  value of 454 bp. The software predicted a total of 258,260 genes with an average length of 337 bp. The predicted genes with <300 bp were excluded from taxonomical analysis and functional classification. Taxonomical classification was performed, and the results consisted of 65.80% bacteria, 2.93% *Archaea*, 0.92% eukaryotes, 0.13% viruses, and 30.23% unclassified organisms. The phylum *Proteobacteria*, with a percentage abundance of 28.24%, was found to be the most dominant phylum. The other major phyla and percentage abundances were *Actinobacteria* 9.71%, *Bacteroidetes* 4.66%, *Chloroflexi* 3.87%, *Firmicutes* 2.47%, *Gemmatimonadetes* 2.09%, *Planctomycetes* 1.61%, *Acidobacteria* 1.52%, *Ignavibacteriae* 1.51%,

Received 4 September 2017 Accepted 6 September 2017 Published 19 October 2017

**Citation** Gupta J, Rathour R, Kumar M, Thakur IS. 2017. Metagenomic analysis of microbial diversity in landfill lysimeter soil of Ghazipur landfill site, New Delhi, India. *Genome Announc* 5:e01104-17. <https://doi.org/10.1128/genomeA.01104-17>.

**Copyright** © 2017 Gupta 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](mailto:isthakur@mail.jnu.ac.in).

*Euryarchaeota* 1.31%, *Thaumarchaeota* 1.21%, *Cyanobacteria* 1.13%, “*Candidatus* Woesebacteria” 0.58%, *Deinococcus-Thermus* 0.52%, *Ascomycota* 0.35%, *Nitrospirae* 0.35%, *Verrucomicrobia* 0.34%, and *Balneolaeota* 0.30%. Other phyla were present in much lower percentages. At the genus level, the *Streptomyces* genus, with a percentage abundance of 1.85%, was found to be the most abundant. During the functional analysis of the metagenome, the leading pathway subclasses found included those involving metabolism of carbohydrate, amino acids, energy, nucleotides, and lipids, xenobiotic biodegradation.

The metagenomic analysis revealed the presence of diverse microbial communities and the functional analysis suggested the presence of various genes related to degradation of xenobiotic compounds under prevailing environmental conditions.

**Accession number(s).** The obtained nucleotide sequences were submitted to the NCBI Sequence Read Archive (SRA) under the accession number [SRX2861368](https://www.ncbi.nlm.nih.gov/sra/SRX2861368).

## ACKNOWLEDGMENTS

We thank the University Grants Commission (UGC), New Delhi, India, for providing a Fellowship and UPE-II, University Grants Commission, New Delhi, India, for financial support. We have not received grants from funding agencies. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

- Mocali S, Benedetti A. 2010. Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. *Res Microbiol* 161:497–505. <https://doi.org/10.1016/j.resmic.2010.04.010>.
- Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC. 2011. Integrative analysis of environmental sequences using MEGAN4. *Genome Res* 21:1552–1560. <https://doi.org/10.1101/gr.120618.111>.
- Kumari M, Ghosh P, Thakur IS. 2016. Landfill leachate treatment using bacto-algal coculture: an integrated approach using chemical analyses and toxicological assessment. *Ecotoxicol Environ Saf* 128:44–51. <https://doi.org/10.1016/j.ecoenv.2016.02.009>.
- 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, Coggill 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>.