Data in brief 28 (2020) 104791

Contents lists available at ScienceDirect

Data in brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Metagenomic data on bacterial diversity profiling of Arabian sea sediment by amplicon sequencing

Harisree P. Nair, Sarita G. Bhat*

Department of Biotechnology, Cochin University of Science and Technology, Kochi, 682022, Kerala, India

ARTICLE INFO

Article history:
Received 9 July 2019
Received in revised form 3 November 2019
Accepted 4 November 2019
Available online 11 November 2019

Keywords:
Community genome
Arabian sea
16S rRNA
Marine

ABSTRACT

This data is about the microbial community genome analysis of Arabian sea sediment by Illumina sequencing by targeting the hypervariable region V3 of 16S rRNA gene. The data analysis revealed the existence of numerous unknown sequences, indicating a large unexploited bacterial diversity in the area. The raw sequence data used for analysis is available in NCBI under the Sequence Read Archive (SRA) with the BioProject No. PRJNA397165 and SRA accession number SRP125840.

1. Data

The largest habitable space for living organisms, particularly microorganisms is the marine realm, covering 70% of the planet surface. These microbial communities are key players in marine ecosystem maintenance [1]. Study of marine microbial biodiversity is of great significance, for it enables understanding biogeochemical cycles prevailing in the area. To harness these enormous genetic diversities in toto, metagenomic procedures can be applied. However, advances in next-generation sequencing methods have accelerated the large-scale exploration of taxonomic diversity of bacterial communities

* Corresponding author.
E-mail addresses: saritagbhat@gmail.com, sgbhat@cusat.ac.in (S.G. Bhat).

https://doi.org/10.1016/j.dib.2019.104791

Specifications Table

Subject area	*Biology*
More specific subject area	*Marine Metagenomics*
Type of data	*Figures*
How data was acquired	*Illumina MiSeq platform*
Data format	*Raw and analyzed*
Experimental factors	*Arabian Sea sediment of 96 m depth were collected, representing the most productive epipelagic zone*
Experimental features	*Metagenomic DNA extraction and amplicon sequencing of V3 region of 16S rRNA gene*
Data source location	*Arabian Sea (9° 59′ 10.9968″ N; 75° 39′ 26.4564″ E)*
Data accessibility	*The sequencing data is available in NCBI under the Sequence Read Archive (SRA) with the BioProject No. PRJNA397165 and SRA accession number SRP125840. The direct URL to data is* https://www.ncbi.nlm.nih.gov/sra/?term=SRP125840

Value of the Data

- The data provides insights into the hidden microbial diversity of Arabian sea sediments which can utilized as a treasure trove of novel biomolecules
- The sequencing data is publicly available for comparative studies of microbial diversity in global oceans.
- The scientific community is informed through the study about the existence of several unidentified sequences indicating the high incidence of novel yet-to-be cultured bacteria in the Arabian sea epipelagic sediments

from diverse environments [2–4]. The complete focus is on presenting the taxonomic profile of bacterial communities of Arabian sea sediment.

6309 Operational Taxonomic Units (OTUs) were identified from the sequencing data and segregated into diverse taxonomic level of bacterial domains, which were classified into 43 bacterial phyla including 18 formally described bacterial phyla (Fig. 1) and 25 candidate phyla (Fig. 2).

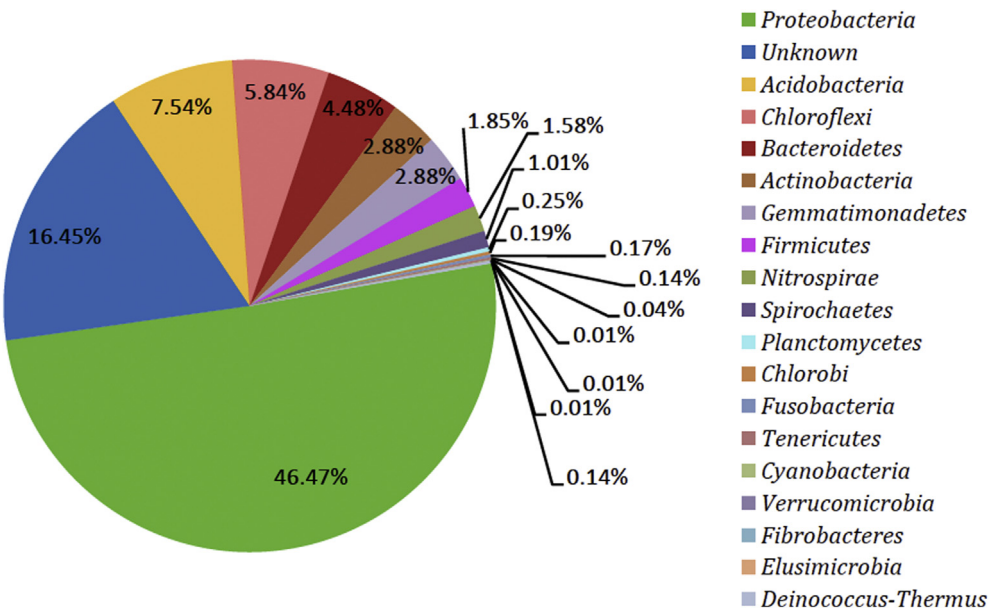

Fig. 1. Taxonomic distribution of OTUs at phylum level from Arabian sea sediment.

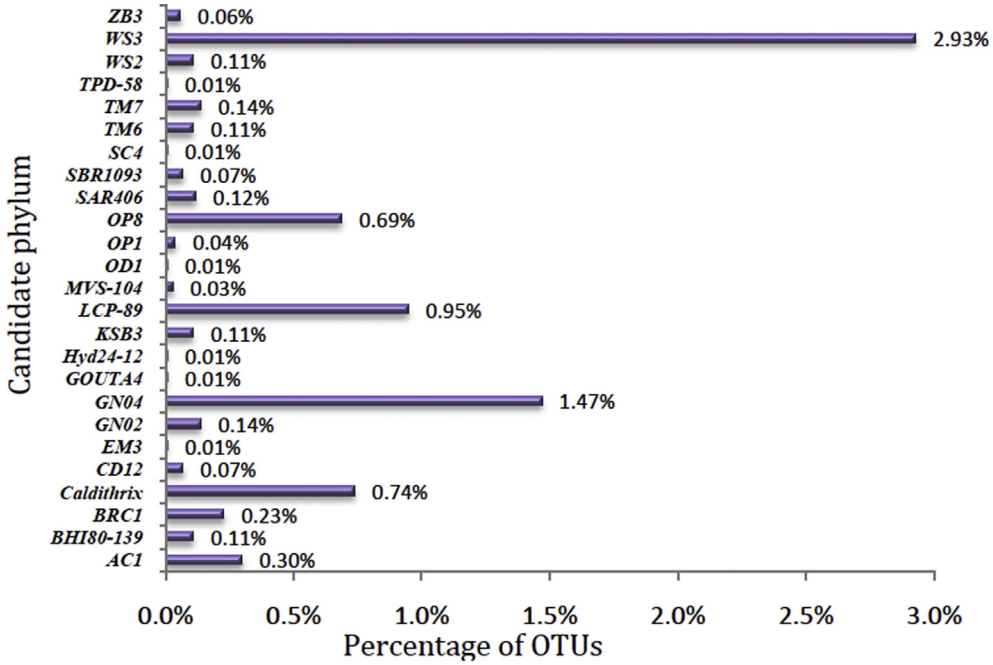

Fig. 2. Taxonomic classification of OTUs at candidate phylum level from Arabian sea sediment.

Phylum *Proteobacteria* with 2932 OTUs was most abundant (at 46.47% of the total diversity). 16.45% (1038 OTUs) represented novel yet to be cultured organisms in Arabian Sea sediments awaiting discovery. 476 OTUs belonged to *Acidobacteria*, 369 OTUs to *Chloroflexi*, 283 OTUs to *Bacteroidetes*, while 182 OTUs from *Actinobacteria* and *Gemmatimonadetes* were also identified. *Firmicutes*, *Nitrospirae, Spirochaetes, Planctomycetes, Chlorobi, Fusobacteria, Tenericutes, Cyanobacteria, Verrucomicrobia, Fibrobacteres, Deinococcus-Thermus* and *Elusimicrobia* contributed less than 2% of the total identified OTUs.

2. Experimental design, materials, and methods

Marine sediments were collected from eastern Arabian Sea (9°59′10.9968″ N; 75° 39′ 26.4564″ E) onboard the research vessel FORV Sagar Sampada (Cruise No: 305) during August 2012 using grab at a depth of 96 m. Community DNA was isolated by modifying the classical method by utilizing liquid nitrogen for grinding sediment sample [5,6]. The V3 hypervariable region of 16S rRNA gene was amplified using 341F 5′-CCTACGGGAGGCAGCAG-3′ and 518R 5′-ATTACCGCGGCTGCTGG-3′ primer pairs with appropriate dilution of metagenomic DNA as template. Purified PCR product was used for a second PCR reaction which attached Illumina sequencing adapters and dual-index barcodes to the amplicon target. Sequencing reactions (151 bp × 2 paired end reads) were performed using the MiSeq platform (Illumina, Inc., CA, USA) following manufacturer's instructions. Raw sequencing data obtained were quantity filtered and processed using QIIME [7].

Acknowledgments

Authors acknowledge the financial support from University Grants Commission, Govt. of India through Major research project vide F.No. 41/527/2012 (SR).

Conflict of 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.

References

[1] J.A. Fuhrman, Microbial community structure and its functional implications, Nature 459 (2009) 193–199, https://doi.org/10.1038/nature08058.
[2] Z. Qiu, N. Li, X. Lu, Z. Zheng, M. Zhang, X. Qiao, Characterization of microbial community structure and metabolic potential using Illumina MiSeq platform during the black garlic processing, Food Res. Int. 106 (2018) 428–438, https://doi.org/10.1016/j.foodres.2017.12.081.
[3] C. Chen, X. Peng, S. Huang, Y. Wang, S. Liao, Y. Wei, Data on microbial community composition of sludge from high altitude wastewater treatment plants determined by 16S rRNA gene sequencing, Data Brief 23 (2019) 103739, https://doi.org/10.1016/j.dib.2019.103739.
[4] P.G. Marshall, S.M. Karst, P.H. Nielsen, B.B. Jørgensen, Metagenomes from deep Baltic Sea sediments reveal how past and present environmental conditions determine microbial community composition, Mar. Genomics 37 (2018) 58–68, https://doi.org/10.1016/j.margen.2017.08.004.
[5] T. Volossiouk, E.J. Robb, R.N. Nazar, Direct DNA extraction for PCR-mediated assays of soil organisms, Appl. Environ. Microbiol. 61 (11) (1995) 3972–3976.
[6] H.P. Nair, H. Vincent, S.G. Bhat, Evaluation of five in situ lysis protocols for PCR amenable metagenomic DNA from mangrove soils, Biotechnol. Rep. 4 (2014) 134–138, https://doi.org/10.1016/j.btre.2014.09.008.
[7] 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, G.A. Huttley, S.T. Kelley, D. Knights, J.E. Koenig, R.E. Ley, C.A. Lozupone, D. McDonald, B.D. Muegge, M. Pirrung, J. Reeder, J.R. Sevinsky, P.J. Turnbaugh, W.A. Walters, J. Widmann, T. Yatsunenko, J. Zaneveld, R. Knight, QIIME allows analysis of high-throughput community sequencing data, Nat. Methods 7 (5) (2010) 335–336, https://doi.org/10.1038/nmeth.f.303.