Data Article

Metagenomic 16S rDNA amplicon data of microbial diversity and its predicted metabolic functions in the Southern Ocean (Antarctic)

Balamurugan Sadaiappan a, Sivakumar Kannan b, Sivasankar Palaniappan c, Radhakrishnan Manikkam d, Balagurunathan Ramasamy c, Anilkumar N. e, Mahendran Subramanian f,*

a Plankton Ecology Laboratory, CSIR National Institute of Oceanography, Goa, India
b Centre of Advanced Study in Marine Biology, Annamalai University, Tamil Nadu, India
c Department of Microbiology, Periyar University, Salem, India
d Centre for Drug Discovery and Development, Sathyabama Institute of Science and Technology, Chennai, India
e National Centre for Antarctic and Ocean Research, Goa, India
f Department of Bioengineering and Department of Computing, Imperial College London, United Kingdom

A R T I C L E   I N F O

Article history:
Received 14 August 2019
Received in revised form 30 October 2019
Accepted 18 November 2019
Available online 23 November 2019

Keywords:
Microbial diversity
Southern ocean
Metabolic functions
Proteobacteria

A B S T R A C T

Antarctica holds about 70% of all the freshwater on the planet in the form of ice. The seawater, it chills, affect the currents and temperature everywhere. Global warming risks the melting of the icecaps as it has already increased the ocean temperature by 1 °C to the West Antarctic peninsula since 1955. A better understanding of the microbial community in this extreme environment of utmost importance is of interest to the scientific community. Herein, we document our metagenomics analysis of the microbial diversity and abundance in the Southern Ocean [Lat 55° 33’ 396 S; Lon 55° 31’ 448 E] using Next Generation Sequencing (NGS), QIIME 1.9.1, Silvangs and a naïve Bayesian classifier. Such metagenomics data hold the potential to aid predictive analysis, which is critical to our understanding of the dynamics of the microbial communities and their role in the Southern Ocean at present and in the future.

© 2019 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 author.
E-mail address: m.subramanian@imperial.ac.uk (M. Subramanian).

https://doi.org/10.1016/j.dib.2019.104876
2352-3409/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
1. Data

The illumina Miseq sequencer produce 724,150 paired end reads, after quality filter and contigs assembly with average read length of 250bp we obtained 362,075 reads. The Qiime191 analysis showed that the reads represented two microbial Kingdoms (Bacteria and Archaea). Overall, the reads represented 38 phyla, 86 classes, and 412 genera. The bacterial diversity was found to be higher, with 96.8% of the total reads, followed by Archaea with only 1.7%. Among the diversity the bacterial diversity varied widely and included 33 phyla viz. Acidobacteria, Actinobacteria, Bacteroidetes, Chloroflexi, Cyanobacteria, Deinococcus-Thermus, Firmicutes, Nitrospirae, Planctomycetes, Proteobacteria, Verrucomycetes and 1.5% of the total reads were unclassified (Fig. 1).

Within the bacterial fractions, the Phylum Proteobacteria was dominant, with 87% of the total reads consisted of 3 classes’ viz. Gammaproteobacteria (58%), Alphaproteobacteria (29%), and Deltaproteobacteria (0.4%) (Fig. 2). Followed by phylum Bacteroidetes (4.2%), Firmicutes (1.7%). Archaeal diversity, consisted of 5 Phyla such as Crenarchaeota, Diapherotrites, Euryarchaeota, Hadesarchaeota and Thaumarchaeota.

2. Experimental design, material and methods

2.1. Sample collection

The surface seawater samples were collected during the Southern Ocean (Antarctica) expedition SOE 8 (Nov 2016) at the coordinates of 55° 33’ 396 S; 55° 31’ 448 E, using Niskin water sampler on board and the temperature, salinity and oxygen were obtained using on board Seabird CTD (Table 1). The water samples were stored in –80 °C and transported to the CAS in Marine Biology laboratory for further studies.
Fig. 1. Occurrence and abundance of different Phyla in the Southern Ocean (Antarctica).
2.2. DNA isolation

Water samples were filtered through 47mm dia 0.2 μm polycarbonate filter paper, and the eDNA was isolated using DNeasy PowerWater Kit (Qiagen). Briefly, the filter papers were homogenised with 5ml of lysis buffer and incubated for 90 min, and the mixture was centrifuged at 4000×g for 5 mins at 4 °C. The supernatant was collected in a separate fresh tube, and the DNA was precipitated by adding ice-cooled 0.7 vol isopropanol. The isolated DNA was pooled together, and the quality was checked in agarose and NanoDrop 2000/2000c (Thermos Scientifics). The extracted DNA was sent to the Xcelris Labs Ltd, Ahmedabad, India, for amplicon sequencing.

Fig. 2. Relative abundances of bacterial (96.8%) archaeal (1.7%) classes and unique reads (1.5%) in water sample of Southern Ocean.

Table 1
Physico-chemical parameters.

| Parameter (units) | Values      |
|-------------------|-------------|
| O2 (μM)           | 347.9563    |
| pH                | 7.85        |
| NO3 (μM)          | 26.54       |
| NO2 (μM)          | 0.44        |
| PO4 (μM)          | 0.41        |
| SiO4 (μM)         | 22.34       |
2.3. DNA sequence

The amplicon libraries were prepared using the Nextera XT Index Kit (Illumina inc.) as per the 16S Metagenomics Sequencing Library preparation protocol (Part # 15044223 Rev. B). The V3–V4 hyper-variable region of 16S rDNA gene of bacteria and archaea primer was designed in Xcelris NGS Bioinformatics Lab [1]. These primers were synthesised within the Xcelris PrimeX facility. The amplicons with the Illumina adaptors were amplified, using i5 and i7 primers that add multiplexing index sequences as well as standard adapters required for cluster generation (P5 and P7) as per the standard Illumina protocol. The amplicon libraries were purified by 1X AmpureXP beads and checked on Agilent High Sensitivity (HS) chip on Bioanalyzer 2100 and quantified on fluorimeter by Qubit dsDNA HS Assay kit (Life Technologies). The library was loaded onto the Illumina Platform at an appropriate concentration (10–20pM) for cluster generation and sequencing. Paired-End sequencing allows the template fragments to be sequenced in both the forward and reverse directions.

2.4. Sequence analysis

The sequences were analysed in QIIME 1.9.1 [2] and only the sequence with minimum 150bp length was used for further analysis. The bacterial and archaeal 16S rDNA sequences were clustered at 97% using both the open and closed-reference OTU picking a strategy and the OTUs were classified using Greengene 13_8 16S reference database [3]. The taxonomy assignment was made to each Operational Taxonomic Unit (OTU) using the RDP classifier [4] and Silvangs [5]. The sequence coverage was evaluated by rarefaction analysis (Fig. 3), and the species richness (H = 5.7) and diversity index values (Chao1 = 1439) were calculated. The microbial metabolic pathways were estimated based on the 16S rDNA gene data using Parallel-Meta3 software [6] were a core-housekeeping function such as amino acid metabolism and lipid metabolism was dominant followed by the biosynthesis of other secondary metabolites, xenobiotics, degradation and metabolism and nucleotide metabolism (Fig. 4). NSTI value was determined in the present study, and it was 0.10085 indicating that the predicted functions of the

Fig. 3. Rarefaction curve showing the inadequate sampling.
microbial taxa are very close to the microbial reference genome database, and it shows the accuracy of the prediction.

Acknowledgement

Authors received financial assistance to carry out this work from the Department of Biotechnology, Government of India, New Delhi. Currently, MS is funded by the EPSRC and Imperial College London (EP/N509486/1: 1,979,819). Authors thank their funders for the successful completion of this work.

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.G. Caporaso, C.L. Lauber, W.A. Walters, D. Berg-Lyons, J. Huntley, N. Fierer, S.M. Owens, J. Betley, L. Fraser, M. Bauer, N. Gormley, J.A. Gilbert, G. Smith, R. Knight, Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms, ISME J. 6 (8) (2012) 1621–1624.

[2] J.G. Caporaso, J. Kuczynski, J. Stombaugh, K. Bittinger, F.D. Bushman, E.K. Costello, 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.

[3] T.Z. DeSantis, P. Hugenholtz, N. Larsen, M. Rojas, E.L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, G.L. Andersen, Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB, Appl. Environ. Microbiol. 72 (2006) 5069–5072, https://doi.org/10.1128/AEM.03006-05.

[4] E. Pruesse, C. Quast, K. Knittel, B.M. Fuchs, W. Ludwig, J. Peplies, F.O. Glöckner, SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB, Nucleic Acids Res. 35 (21) (2007) 7188–7196, https://doi.org/10.1093/nar/gkm864.

[5] Q. Wang, G.M. Garrity, J.M. Tiedje, J.R. Cole, Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy, Appl. Environ. Microbiol. 73 (16) (2007) 5261–5267, https://doi.org/10.1128/AEM.00062-07.

[6] G. Jing, Z. Sun, H. Wang, Y. Gong, S. Huan, K. Ning, J. Xu, X. Su, Parallel-META 3: comprehensive taxonomical and functional analysis platform for efficient comparison of microbial communities, Sci. Rep. 7 (2017) 40371, https://doi.org/10.1038/srep40371.