Data Article

Metagenomic dataset on lichen *Dirinaria* sp. from the Great Rann of Kutch and tropical moist deciduous Dang forest of Gujarat

Apurvasinh Puvar\textsuperscript{a,b}, Chandrashekar Mootapally\textsuperscript{a}, Chaitanya Joshi\textsuperscript{a}, Madhvi Joshi\textsuperscript{a,*}

\textsuperscript{a} Gujarat Biotechnology Research Centre, Gandhinagar 382011, Gujarat, India  
\textsuperscript{b} Hemchandracharya North Gujarat University, Patan, Gujarat, India

**A R T I C L E   I N F O**

**Article history:**
Received 5 February 2020  
Revised 22 March 2020  
Accepted 23 March 2020  
Available online 14 April 2020

**Keywords:**  
Lichen  
Metagenome  
NGS  
Lichen genomics

**A B S T R A C T**

This paper describes the additional data to our research article “Bacterial line of defense in Dirinaria lichen from two different ecosystems: First genomic insights of its mycobiont *Diriniria* sp. GBRC AP01” by Puvar et al. \cite{1}. In this manuscript we are presenting the data obtained during the annotation of the genome enriched from metagenomic data from the lichen samples.

© 2020 The Author(s). 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                      | Applied microbiology and biotechnology |
|------------------------------|----------------------------------------|
| Specific subject area        | Metagenomics                           |
| Type of data                 | Table                                  |
| How data were acquired       | Graph                                  |
|                             | SRA files                               |
|                             | Shotgun sequencing of metagenomes using Ion Proton with 200bp library chemistry |

* Corresponding author.  
E-mail address: madhvimicrobio@gmail.com (M. Joshi).

https://doi.org/10.1016/j.dib.2020.105504  
2352-3409/© 2020 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY license.  
(http://creativecommons.org/licenses/by/4.0/)
Data format Raw Analyzed

Parameters for data collection

*Dirinaria* sp. collected from two varied geographical coordinates (the Great Rann of Kutch: arid, white salt desert; the Dang: tropical moist deciduous forest) of Gujarat. The samples were collected during post-monsoon and further processed with the similar conditions to avoid handling bias in metagenome samples.

Description of data collection

Collected whole lichen thallus along with substrate (rock) from the field and herbariums were prepared for identification at NBRI, Lucknow, Uttar Pradesh, India. Upon collection, the sample for metagenomic DNA isolation were stored in preservative for the purpose.

Data source location

Institution: Gujarat Biotechnology Research center (GBRC)
City/Town/Region: Gandhinagar-382011, Gujarat,
Country: India

Latitude and longitude (and GPS coordinates) for collected samples/data: 23.936806, 69.814500; 20.737083, 73.492278

Data accessibility

Repository name: Metagenomic raw reads submitted to NCBI, the BioProject number is PRJNA526834
Metagenomic data identification number: SRR8731860-61
Direct URL to metagenomic data:
[https://www.ncbi.nlm.nih.gov/bioproject/526834](https://www.ncbi.nlm.nih.gov/bioproject/526834)
MG-RAST: publically available with MG-RAST ID: 87886
[https://www.mg-rast.org/linkin.cgi?project=mgp87886](https://www.mg-rast.org/linkin.cgi?project=mgp87886)
Skimmed mycobiont *Dirinaria* sp. GBRC AP01 submitted to NCBI, the accession number is SZQE00000000
Mendeley Data: link [http://dx.doi.org/10.17632/dfm22pn63b.2#file-8d344350-1cd3–4188–97a5-555d187de100](http://dx.doi.org/10.17632/dfm22pn63b.2#file-8d344350-1cd3–4188–97a5-555d187de100)

Related research article

Puvar, A. C., Nathani, N. M., Shaikh, I., Bhatt, A. D., Bhargava, P., Joshi, C. G., & Joshi, M. N. (2020). Bacterial line of defense in *Dirinaria* lichen from two different ecosystems: First genomic insights of its mycobiont *Dirinaria* sp. GBRC AP01. Microbiological Research, 233, 126407. DOI:[https://doi.org/10.1016/j.micres.2019.126407](https://doi.org/10.1016/j.micres.2019.126407)

Value of the data

- The lichen is complex ecosystem having numerous species with majorly fungi and algae or cyanobacteria as symbionts. The Whole shotgun sequence data of lichen from two different eco geographic region revels their diversity, ecosystem and functional attributes in the ecosystem.
- The information supports to the claims by Puvar et al. [1] for KEGG pathway, COG and SEED classifications of the lichen metagenomes.
- Metagenomic data of *Dirinaria* lichen reveals the association of bacteria and mycobiont symbiosis along with associated diversity in response to their natural habitat.
- Functional attributes observed in the lichen metagenomic data will give insight for further elucidation of microbiome role in local biogeochemical cycling.
- The skimmed mycobiont data from lichen metagenome enhances the fungal reference genome database for *Dirinaria* genus.

1. Data description

Metagenomic assemblies from raw data generated in [1] are described in Table 1. which depicts the assembly statistics of the lichen metagenomes. Draft genome was reconstructed using combined approach and basic assembly statistics are described in Table 2, which summarizes the skimmed mycobiont *Dirinaria* sp. GBRC AP01. Both metagenomic data were analyzed for taxonomic distribution and their top 20 taxa represented as bar chart in Fig. 1. Similarly, both
metagenome assemblies were mapped against KEGG, COG and SEED databases and their relative percent abundance is described in Figs. 2–4 respectively.

2. Experimental design, materials and methods

2.1. Sampling

*Dirinaria* sp. from two different geographic regions from Gujarat were collected along with their substratum. (a) DG18 was collected from the Dangs district located in southern part of Gujarat covered with semi to dense moist forest. (b) KK01 was collected from Kutch district located western part of the state situated in white salt desert and thorny forest. Both samples were collected during post monsoon seasons along with their substratum i.e. rock.

2.2. DNA extraction

Metagenomic DNA was isolated using combined protocol optimized using Qiagen Powersoil kit and Qiagen Plant Mini kit. DNA from each sample (lichen whole thallus) was extracted multiple time to achieve sufficient quantity to perform whole shotgun sequencing i.e. 1 μg DNA in 100 μl volume.

2.3. Metagenomic sequencing

About 1 μg of total metagenomic DNA from both samples was used to prepare libraries of 200 bp fragment length using Ion Xpress™ Plus Fragment Library kit and were sequenced using Ion Proton system using Ion PI v2 chip kit [1].

2.4. Data analysis

Metagenome assembly was performed using Megahit [3] with the default parameters except minimum contig length as 200 base pairs and the assembly statistics are described in the Table 1. Both assemblies of *Dirinaria* sp, were submitted to MG-RAST server for taxonomic and functional annotations. The taxonomic (Fig. 1) and functional (Fig. 2–4) annotations of both metagenomes were derived using MG-RAST [4–7] with default parameters. The mycobiont

| Table 1 | Metegenomic assemblies of data. |
|------------------|-------------------------------|
| Sampling site    | Lab ID | Total number of reads (Million) | Total contig number | Total length (Kbp) | Largest contig (Kbp) | N50 (Kbp) |
| Dang             | DG18   | 5.3 | 21,532 | 32,753.60 | 21.6 | 2.9 |
| Kutch            | KK01   | 6.3 | 45,952 | 39,719.95 | 18.0 | 1.3 |

| Table 2 | Assembly statistics of reconstructed draft genome of *Dirinaria* sp. |
|---------|-----------------------------------------------------------------------|
| Size (Mbp) | Scaffolds (Nos) | N50 (Kbp) | Longest Contig (Kbp) | Genome Completeness |
| 31.67 | 8624 | 4.51 | 27.91 | 70.3% |
Fig. 1. describes the comparative taxonomic profile (top 20 at phylum level) of lichen metagenomes.
Fig. 2. shows the percent difference of top 20 KO (at level2) from both the metagenomes.
Fig. 3. demonstrates variation of COG (top20) at level2 among metagenomes.
Fig. 4. shows top20 abundance difference at level1 of SEED hits in metagenomes.
Diriniria sp. GBRC AP01 was skimmed using modified sequence dependent binning approach as described by Albertsen et al., 2013 [8] and genome completeness (Table 2) was assessed using BUSCOs [9].

Funding Information

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

We are grateful to GBRC, GSBTM and DST for institutional/infrastructural support for conducting this work. We are thankful to Dr. Sanjeeva Nayaka, Principal Scientist, CSIR-National Botanical Research Institute (Govt. of India) for providing guidance in taxonomic identification.

Conflict of Interest

The authors declare that they have no known competing financial interests or personal relationships which have, or could be perceived to have, influenced the work reported in this article.

References

[1] A.C. Puvar, N.M. Nathani, I. Shaikh, A.D. Bhatt, P. Bhargava, C.G. Joshi, M.N. Joshi, Bacterial line of defense in Dirinaria lichen from two different ecosystems: first genomic insights of its mycobiont Dirinaria sp. GBRC AP01, Microbiol. Res. 233 (2020) 126407, doi: 10.1016/j.micres.2019.126407.

[2] A. Puvar, C.S. Mootapally, M. Joshi, Supporting information to data in brief article “Metagenomic dataset on lichen Dirinaria sp. from the Great Rann of Kutch and tropical moist deciduous Dang forest of Gujarat,” 2019. 10.17632/dfm22pm63b.2.

[3] D. Li, R. Luo, C.-M. Liu, C.-M. Leung, H.-F. Ting, K. Sadakane, H. Yamashita, T.-W. Lam, MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices, Methods 102 (2016) 3–11, doi: 10.1016/j.ymeth.2016.02.020.

[4] E.M. Glass, J. Wilkening, A. Wilke, D. Antonopoulous, F. Meyer, Using the metagenomics rast server (MG-RAST) for analyzing shotgun metagenomes, Cold Spring Harb. Protoc. 2010 (2010) pdb.prot5368-pdb.prot5368., doi: 10.1101/pdb.prot5368.

[5] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, M. Hirakawa, KEGG for representation and analysis of molecular networks involving diseases and drugs, Nucleic Acids Res. 38 (2010) D355–D360, doi: 10.1093/nar/gkp896.

[6] R. Overbeek, T. Begley, R.M. Butler, J.V. Choudhuri, H.-Y. Chuang, M. Cohoon, V. de Crécy-Lagard, N. Diaz, T. Disz, R. Edwards, M. Fonstein, E.D. Frank, S. Gerdes, E.M. Glass, A. Goessmann, A. Hansen, D. Iwata-Reuyl, R. Jensen, N. Jamshidi, I. Krause, M. Kubal, N. Larsen, B. Linke, A.C. McHardy, F. Meyer, H. Neuveguide, G. Olsen, R. Olson, A. Osterman, V. Portnoy, G.D. Pusch, D.A. Rodionov, C. Rücker, J. Steiner, R. Stevens, I. Thiele, O. Vassieva, Y. Ye, O. Zagnitko, V. Vonstein, The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes, Nucleic Acids Res. 33 (2005) 5691–5702, doi: 10.1093/nar/gki866.

[7] R.L. Tatusov, M.Y. Galperin, D.A. Natale, E.V. Koonin, The COG database: a tool for genome-scale analysis of protein functions and evolution, Nucleic Acids Res. 28 (2000) 33–36, doi: 10.1093/nar/28.1.33.

[8] M. Albertsen, P. Hugenholtz, A. Skarszewski, K.L. Nielsen, G.W. Tyson, P.H. Nielsen, Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes, Nat. Biotechnol. 31 (2013) 533–538, doi: 10.1038/nbt.2579.

[9] F.A. Simão, R.M. Waterhouse, P. Ioannidis, E.V. Kriventseva, E.M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs, Bioinformatics 31 (2015) 3210–3212, doi: 10.1093/bioinformatics/btv351.