



## Data in Brief

## Deciphering the microbiota of Tuwa hot spring, India using shotgun metagenomic sequencing approach

Amitsinh Mangrola<sup>a,\*</sup>, Pravin Dudhagara<sup>b</sup>, Prakash Koringa<sup>c</sup>, C.G. Joshi<sup>c</sup>, Mansi Parmar<sup>d</sup>, Rajesh Patel<sup>d</sup><sup>a</sup> Department of Biochemistry, Shri Alpesh N. Patel Post Graduate Institute, Anand, India<sup>b</sup> Department of Biotechnology, Veer Narmad South Gujarat University, Surat, India<sup>c</sup> Department of Animal Biotechnology, Anand Agricultural University, Anand, Gujarat, India<sup>d</sup> Department of Life Sciences, Hemchandracharya North Gujarat University, Patan, India

## ARTICLE INFO

## Article history:

Received 9 April 2015

Accepted 12 April 2015

Available online 20 April 2015

## Keywords:

Hot spring

Metagenome

Shotgun sequencing

Community analysis

Functional analysis

## ABSTRACT

Here, we report metagenome from the Tuwa hot spring, India using shotgun sequencing approach. Metagenome consisted of 541,379 sequences with 98.7 Mbps size with 46% G + C content. Metagenomic sequence reads were deposited into the EMBL database under accession number ERP009321. Community analysis presented 99.1% sequences belong to bacteria, 0.3% of eukaryotic origin, 0.2% virus derived and 0.05% from archaea. Unclassified and unidentified sequences were 0.4% and 0.07% respectively. A total of 22 bacterial phyla include 90 families and 201 species were observed in the hot spring metagenome. Firmicutes (97.0%), Proteobacteria (1.3%) and Actinobacteria (0.4%) were reported as dominant bacterial phyla. In functional analysis using Cluster of Orthologous Group (COG), 21.5% drops in the poorly characterized group. Using subsystem based annotation, 4.0% genes were assigned for stress responses and 3% genes were fit into the metabolism of aromatic compounds. The hot spring metagenome is very rich with novel sequences affiliated to unclassified and unidentified lineages, suggesting the potential source for novel microbial species and their products.

© 2015 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

## Specifications

|                           |   |
|---------------------------|---|
| Organism/cell line/tissue | Metagenome of Tuwa hot spring water   |
| Sex                       | Not applicable  |
| Sequencer or array type   | Ion-Torrent PGM platform  |
| Data format               | Raw data: FASTQ file  |
| Experimental factors      | Environmental sample  |
| Experimental features     | Shotgun metagenome sequencing of hot spring community DNA using Ion-Torrent PGM platform followed by community and functional metagenome analysis using MG-RAST server. |
| Consent                   | Not applicable  |
| Sample source location    | Water sample, Tuwa hot spring, Gujarat State, India   |

## Direct link to deposited data

<http://www.ebi.ac.uk/ena/data/view/ERP009321>

Culturability of microorganisms in extreme habitats is usually very poor, which creates the need to explore the culture-independent tech-

niques to evaluate the taxonomic and functional status of the microbes. Currently, either 16S rRNAs amplicon based metagenome sequencing or shotgun metagenome sequencing is largely used for the assessment of the microbial diversity [1,2]. Such culture-independent techniques have been successfully used for comprehensive analysis of microbial populations in hot spring [3]. Hot springs comprise several microbial habitats including thermal fluids, microbial mats and sediments. Metagenomic study of hot springs produced huge information, providing holistic images of the microbial communities present in hot springs [4]. The temperature exerts a strong control on microbial diversity in hot spring [5]. Therefore, hot spring microbial communities largely encompass the thermophilic, hyperthermophilic and polyextremophilic microbes.

India is endowed with 340 thermal hot water springs, presenting an opportunity to study their microbial ecology to trap the novel species and genes. Hot spring of Indian subcontinent offers an attractive and challenging platform for researchers due to the existence of unknown and untapped microbial communities. Many hidden microbial species and genes have been revealed in culture-independent microbial diversity assessment of Tulsi Shyam and Lasundra hot spring, India [6,7]. Tuwa hot spring (22°47'58"N and 73°27'37"E) is one of the unexplored hot water springs situated in Panchmahal District, Gujarat State, India. The temperature of Tuwa hot spring is usually recorded between 54 to

\* Corresponding author at: Department of Biochemistry, Shri Alpesh N. Patel Post Graduate Institute, Anand 388001, India. Tel.: +91 9879626009.

E-mail address: [avmang@gmail.com](mailto:avmang@gmail.com) (A. Mangrola).

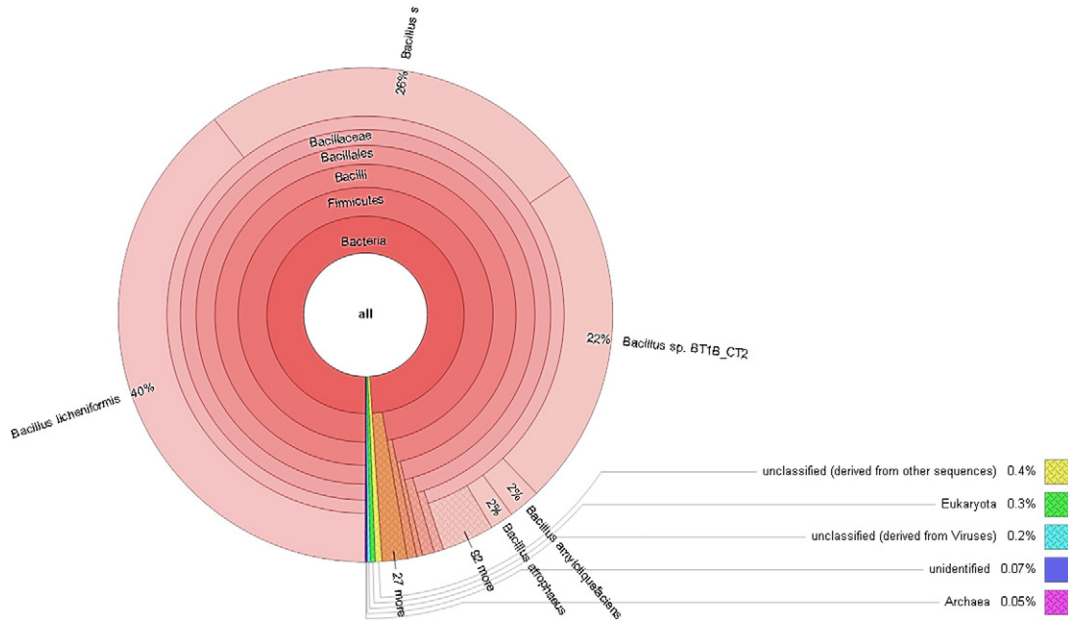


Fig. 1. Community structure of Tuwa hot spring metagenome.

65 °C and water is alkaline with high salt and minerals content. Visitor and local community often take the bath due to its therapeutic belief. The bottom of the reservoir has hard rocks and photosynthetic microbes thrive in the water column. To the best of our knowledge, the microbial diversity and biogeological study of Tuwa hot water springs has not yet been investigated. The present research is the first investigation to describe the microbial diversity by metagenomics approach.

The water sample was collected in thermos bottle and transported to the laboratory for the physicochemical analysis. The microbial community DNA from the hot spring water was isolated and purified by silica spin columns using HiPurA™ water DNA isolation kit (MB547; Hi-Media, Mumbai, India). Isolated metagenomic DNA was subjected to the sequencing after the testing of quality and purity using agarose gel electrophoresis and spectrophotometer. Shotgun metagenomic sequencing was performed with an Ion Torrent PGM

platform using Ion Express Template 300 chemistry on a 318 chip as per the manufacturer's instructions. Analysis and annotation of output data were carried out by MG-RAST, an online server using default parameter [8].

During the sampling of water from Tuwa hot springs, pH was reported ranging from 8.2 to 9.0 and conductivity was 19 µS/cm. The elemental composition of hot spring water (mg/L) showed the Sodium 19.93, Potassium 14.5, Calcium 0.5, Chloride 15.8 and Sulphate 1.83, Carbonate 2.4 and Bicarbonate 0.4. The output data encompassed a total of 541,379 sequences comprised of 98,773,736 bps with 46% G + C content. The microbial community analysis revealed predominance of domain bacteria (99.1%), followed by unclassified sequences (0.4%) eukaryota (0.3%), viruses (0.2%), unidentified sequences (0.07%), and archaea (0.05%).

At phyla level, a total of 22 diverse microbial phyla together with unclassified category were depicted in the metagenome. Most abundant

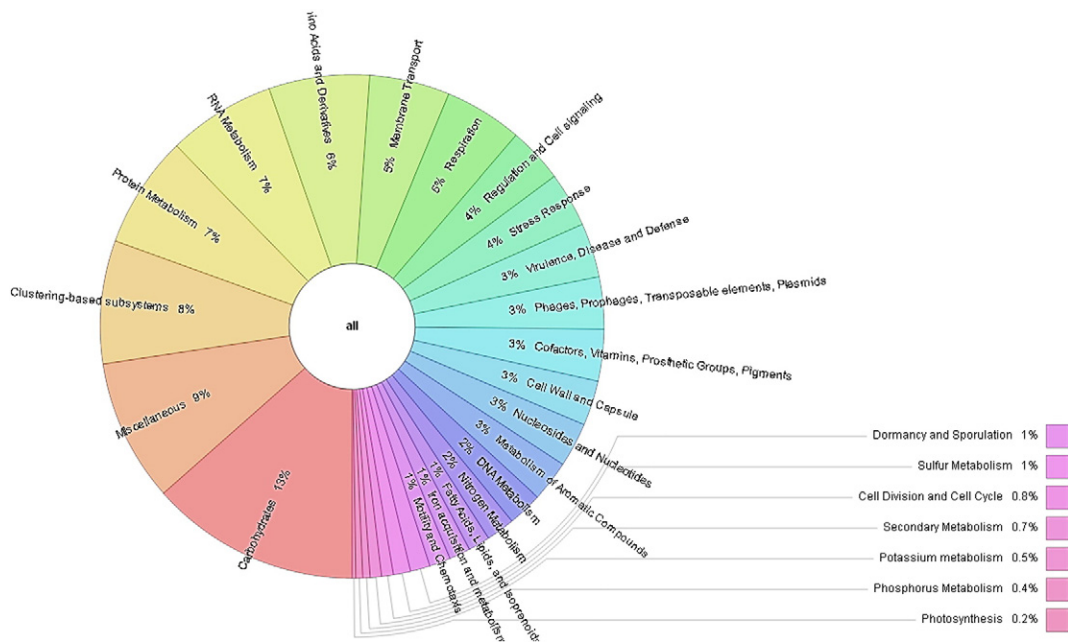


Fig. 2. Functional structure of Tuwa hot spring metagenome using subsystem annotation.

sequences affiliated to bacterial phyla were Firmicutes (97.0%) followed by Proteobacteria (1.3%) and Actinobacteria (0.4%). Whereas Ascomycota (0.08%) and Chordata (0.08%) were abundant eukaryotic phyla and prominent archaeal phylum was Euryarchaeota (0.05%). Metagenomes comprised a total of 90 families with the abundant existence of Bacillaceae (94.6%) followed by Paenibacillaceae (0.4%) and Clostridiaceae (0.4%). At the species level, the most common bacterial species reported were *Bacillus licheniformis* (40%) and *Bacillus subtilis* (26%) (Fig. 1). Thermophilic alkaline serine protease producing *B. licheniformis* was reported from the same territory at Unnai hot spring, India [9].

Functional analysis was also performed using MG-RAST server. Out of the 350,069 sequences that exceeded quality control, 263,997 sequences (75.4%) produced a total of 170,681 predicted protein coding regions. Of these 170,681 predicted protein features, 127,064 (74.4%) of features have been assigned an annotation using at least one of the protein databases (M5NR) and rest of 25.6% of features have no significant similarities to the protein database.

In the functional annotation, a total of 218,926 features were assigned to functional categories with COG approach and 47,049 sequences (21.5%) falls in the poorly characterized group. Out of them, 50% were characterized as an unknown function. Likewise, the annotation with Non-supervised Orthologous Groups (NOG) approach, 16,754 predicted functions assigned and interestingly 13,619 sequences (81.3%) falls in the poorly characterized group and out of them 93% were an unknown function. The result indicated the possibilities of gaining novel genes from the metagenome. In subsystem based annotation approach, 4% genes belonged to stress responses for the oxidative stress, osmotic stress, heat shock, acid shock, periplasmic stress, detoxification, desiccation stress and cold shock. Likewise, 3% genes were assigned for the metabolism of aromatic compounds (Fig. 2). Furthermore, many key biotechnological process and product encoding genes were detected in the hot spring metagenome. The result indicates the shared activities of thermophilic microbes useful to access the metagenome as a resource of novel molecular sequence space [10].

The study presented a considerable microbial diversity and offers insights into microbial community composition, distribution and abundance. A number of DNA sequences remained taxonomically and functionally unanswered representing potentially novel microorganisms and genes in the hot spring.

## Nucleotide sequence accession number

Metagenome sequence data are available on EMBL Metagenomics under the accession no. ERP009321.

## Acknowledgements

Authors are thankful to Ome Research Facility (ORF), Anand Agricultural University, Anand, India for providing the facility of the next generation sequencing.

## References

- [1] P. Dudhagara, A. Ghelani, R. Patel, R. Chaudhari, S. Bhatt, Bacterial tag encoded FLX titanium amplicon pyrosequencing (bTEFAP) based assessment of prokaryotic diversity in metagenome of Lonar soda lake, India. *Genomics Data 4* (2015) 8–11, <http://dx.doi.org/10.1016/j.gdata.2015.01.010>.
- [2] R. Patel, V. Mevada, D. Prajapati, P. Dudhagara, P. Koringa, C.G. Joshi, Metagenomic sequence of saline desert microbiota from wild ass sanctuary, Little Rann of Kutch, Gujarat, India. *Genomics Data 2* (2015) 137–139, <http://dx.doi.org/10.1016/j.gdata.2015.01.003>.
- [3] C.S. Chan, K.-G. Chan, Y.-L. Tay, Y.-H. Chua, K.M. Goh, Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front. Microbiol.* 6 (2015) 177, <http://dx.doi.org/10.3389/fmicb.2015.00177>.
- [4] O. Lopez-Lopez, M.E. Cerdan, M.I. Gonzalez-Siso, Hot spring metagenomics. *Life* 3 (2013) 308–320, <http://dx.doi.org/10.3390/life3020308>.
- [5] C.E. Sharp, A.L. Brady, G.H. Sharp, S.E. Grasby, M.B. Stott, P.F. Dunfield, Humboldt'sspa: microbial diversity is controlled by temperature in geothermal environments. *ISME J.* 8 (2014) 1166–1174, <http://dx.doi.org/10.1038/ismej.2013.237>.
- [6] A. Ghelani, R. Patel, A. Mangrola, P. Dudhagara, Cultivation-independent comprehensive survey of bacterial diversity in Tulsii Shyam Hot Springs, India. *Genomics Data 4* (2015) 54–56, <http://dx.doi.org/10.1016/j.gdata.2015.03.003>.
- [7] A. Mangrola, P. Dudhagara, P. Koringa, C.G. Joshi, R. Patel, Shotgun metagenomic sequencing based microbial diversity assessment of Lasundra hot spring, India. *Genomics Data 4* (2015) 73–75, <http://dx.doi.org/10.1016/j.gdata.2015.03.005>.
- [8] F. Meyer, D. Paarmann, M. D'Souza, R. Olson, E.M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, R.A. Edwards, The metagenomics RAST server-public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinforma.* 9 (2008) 386, <http://dx.doi.org/10.1186/1471-2105-9-386>.
- [9] P. Dudhagara, S. Bhavasar, A. Ghelani, S. Bhatt, Isolation, characterization and investing the industrial applications of thermostable and solvent tolerant serine protease from Hot Spring isolated thermophilic *Bacillus licheniformis* U1. *Int. J. Appl. Sci. Biotechnol.* 2 (2014) 75–82, <http://dx.doi.org/10.3126/ijasbt.v2i1.9519>.
- [10] P. Lorenz, K. Liebeton, F. Niehaus, J. Eck, Screening for novel enzymes for biocatalytic processes: accessing the metagenome as a resource of novel functional sequence space. *Curr. Opin. Biotechnol.* 13 (2002) 572–577, [http://dx.doi.org/10.1016/S0958-1669\(02\)00345-2](http://dx.doi.org/10.1016/S0958-1669(02)00345-2).