Metagenomics-based bacterial community analysis: assessment by amplicon sequencing of Tuva Timba and Dholera hot-water springs, Gujarat, India

Disha N. Patel[1,3], Jalpa R. Thakkar[2], Vincent J. Braganza[3] and Hasmukh A. Modi[1]

[1]Department of Lifesciences, Gujarat University, Ahmedabad 380 009, India
[2]Department of Animal Biotechnology, Anand Agricultural University, Anand 388 110, India
[3]Xavier Research Foundation, Loyola Centre for Research and Development, Ahmedabad 380 009, India

Amplicon sequencing of extreme locations helps reveal the presence and abundance of diverse extremophiles within extreme environments. Firmicutes, Proteobacteria and Actinobacteria were identified as a major phyla using 16S rDNA amplicons from the soil and water samples of Tuva Timba and water samples of Dholera in Gujarat, India. Season-wise data when compared with location-wise data showed less diversity. The presence of thermophiles and sulphate-reducing taxa reveals a strong correlation of taxonomy with temperature and sulphate content of the samples. The clear cluster analysis within three sample groups reveals the importance of geographic as well as environmental factors for diverse bacterial community colonization.

Keywords: Amplicon sequencing, bacterial community analysis, hot springs, metagenomics.

ORGANISMS that can survive under extreme temperature, salinity, acidic or alkaline pH are known as extremophiles. The ability to adapt to extreme conditions is displayed by the oldest inhabitants on Earth, enabling them to colonize it with enormous diversity[1]. Among them, thermophiles are organisms which can survive under very high temperatures such as those of hot-water springs. It is known that the extremophiles possess biomolecules, namely extremozymes which are catalytically active and are useful for industrial as well as biotechnological purposes. However, the molecular mechanism for their survival under extreme conditions is still unknown[2]. These catalytically active compounds called biocatalysts have various applications in industries like food, pharmaceuticals, fuels and cosmetics.

India harbours a total of 340 thermal hot-water springs in different regions of the country, with different ecological zones and representing novel species and genera of extremophiles[3]. The metagenomics and next-generation sequencing technologies have been used widely to study the complexity and versatility of genomic complexes of unculturable extremophilic communities in order to discover novel thermophiles and their role in various applications[4]. Many hidden microbial species and genes contributing to extremophiles from TulsiShyam, Lasudra, Unkeshwar, Manikaram, Jakrem hot spring in India have been studied using culture-independent techniques[3,5–9]. The Tuva Timba hot-water spring in Panchmahals district, Gujarat, India and Dholera located 30 km away in the southwest from Dhandhuka village, Ahmedabad district, Gujarat and 60 km away in the north from Bhavnagar, Gujarat were selected in this study to compare the effect of seasons on extremophiles detected using 16S amplicon sequencing strategies. Seasonal microbial diversity has not been studied earlier using ion torrent platform-based metagenomics approach.

Materials and methods

Sample collection

Water and soil samples were collected from Tuva Timba and Dholera at three different seasons: monsoon, winter and summer. Dholera water samples were collected in February, October and May 2016, whereas Tuva Timba water and soil samples were collected in January, May and October 2016. Due to unavailability of soil samples, only water samples were collected from Dholera. The study includes six samples from Tuva Timba and three samples from Dholera; thus a total of nine samples were collected.

Physico-chemical analysis

Temperature and pH of the collected samples were recorded at the time of sampling using field thermometer

and portable pH meter. Sulphate concentration of the collected samples was measured in accordance with IS:3025 (Part 24)[10]. Principal component analysis (PCA) was also performed using physico-chemical parameters for both the thermal springs using PAST 3.0 software.

Metagenomic DNA extraction and 16s rDNA metagenome sequencing

The metagenomes from water and soil samples were analysed by amplicon sequencing using Ion Torrent sequencer. The metagenomic soil DNA was extracted using OMEGA EZNA. Soil DNA Kit, while for water DNA, OMEGA EZNA. Water DNA Kit was used, following the manufacturer's protocol. The quantification was done using Qubit 2.0 following the Qubit dsDNA HS Assay (Life Technologies, Germany). Purified DNA samples (5 ng/μl) were used as template to generate variable V2 region 16s rDNA amplicons. The primers used for amplification were forward 5′CCTACGGGNGGCWGCAG3′ and reverse 5′GACTACHVGGGTATCTAATCC3′. Using 0.5 μl each forward and reverse primer (10 pM) and 12.5 μl 2× KAPA HiFiHotStart Ready Mix (Kapa Biosystems, UK) PCR was performed at 95°C for 3 min initial denaturation, 25 cycles of 98°C for 20 s, 60°C for 10 s, 72°C for 12 s and final extension at 72°C for 40 s.

The amplified product was loaded in 1.5% agarose gel, and the desired bands were eluted and purified using MinElute PCR Purification Kit (Qiagen GmbH, Germany). The eluted DNA was quantified using Qubit and further utilized to prepare amplicon libraries using Ion Xpress™ Plus Fragment Library Kit, according to the manufacturer's protocol. The clonal amplification was performed using Xpress Template Kit V2.0 (Life Technologies, USA) and the quality of libraries was quantified using Agilent 2100 bioanalyzer (Agilent Biotechnologies, USA). The diluted and quantified libraries were further used for sequencing using Ion Sequencing 400 Kit (Life Technologies, USA) and 316v2 chip following the corresponding protocol for sequencing and final sequencing, which was carried out on an Ion Torrent Personal Genome Machine (PGM) system (Agilent Biotechnologies, USA).

Metagenomic analysis

All nine samples were analysed using Metagenome Rapid Annotation using Subsystem Technology (MG-RAST) version 4.0 (ref. 11). The Ion Torrent-generated reads were submitted to MG-RAST under deposition of project ID mgp81214. The low-quality regions were trimmed using SolexaQA with default parameters in MG-RAST. The analysed files were downloaded from RDP classifier database of MG-RAST server with 70% identity with a cut-off of 1×10^{-5} E-value, and further used to carry out diversity analysis. A range of diversity profiles was generated using the statistical software PAST[12]. As a part of beta diversity, the data were uploaded to on-line METAGENassist[13] and partial least square-discriminant analysis (PLS-DA), and cluster analysis was also done. As a step of data normalization, the samples were scaled to each other by sum and taxon within samples through autoscaling in METAGEN assist (mean centred divided by standard deviation of each variable).

Results

Physico-chemical analysis

The temperature of the Tuva Timba water samples in winter, monsoon and summer was 60°C, 60°C and 66°C respectively. Also observed pH value was 7.25, 7.07 and 7.65 in winter, monsoon and summer respectively. The sulphate value varied from 92.38 to 167.85 mg/l in Tuva Timba. On the other hand, the temperature of the Dholera water samples was 42°C, 44°C and 55°C in winter, monsoon and summer respectively. The pH value was 7.00, 7.31 and 7.54 in winter, monsoon and summer respectively. The sulphate value varied from 15.81 to 91.02 mg/l in Dholera.

PCA of the Tuva Timba and Dholera hot-water springs and physico-chemical parameters suggests evidence between the two locations and the first two principal components explained 96.10% of the total variance with eigenvalue 13,550.2 (Figure 1). The separate placement of data points in Figure 1 distinctly differentiates the place of sample collection as well as the seasonal variations observed at both the places in terms of physico-chemical parameters.

A total of 1,133,723 reads for amplicons were generated with an average of 125,968 reads per sample (six water
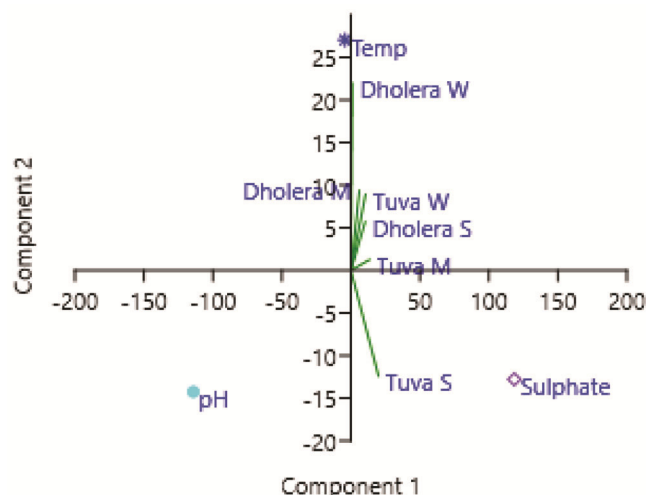
Figure 1. Biplot for the principle component analysis of physico-chemical parameters (S, summer; W, winter and M, monsoon).

Table 1. Diversity indices based on metagenomics data of hot water spring samples

Index	Tuva Timba water samples			Tuva Timba soil samples			Dholera water samples		
	Winter	Monsoon	Summer	Winter	Monsoon	Summer	Winter	Monsoon	Summer
Taxa_S	159	154	165	97	**238**	228	163	146	199
Dominance	0.604	0.657	0.506	0.433	0.106	0.095	0.238	0.231	0.253
Simpson	0.395	0.343	0.4939	0.566	0.893	**0.904**	0.761	0.768	0.746
Shannon	1.133	1.025	1.489	1.505	3.133	**3.242**	2.027	2.047	2.078
Evenness	0.019	0.018	0.027	0.046	0.096	0.112	0.046	0.053	0.040
Berger–Parker	0.770	**0.806**	0.704	0.636	0.235	0.260	0.407	0.329	0.393
Chao-1	223.5	179.1	175	102.2	277	**262.2**	214.2	169.9	233.2
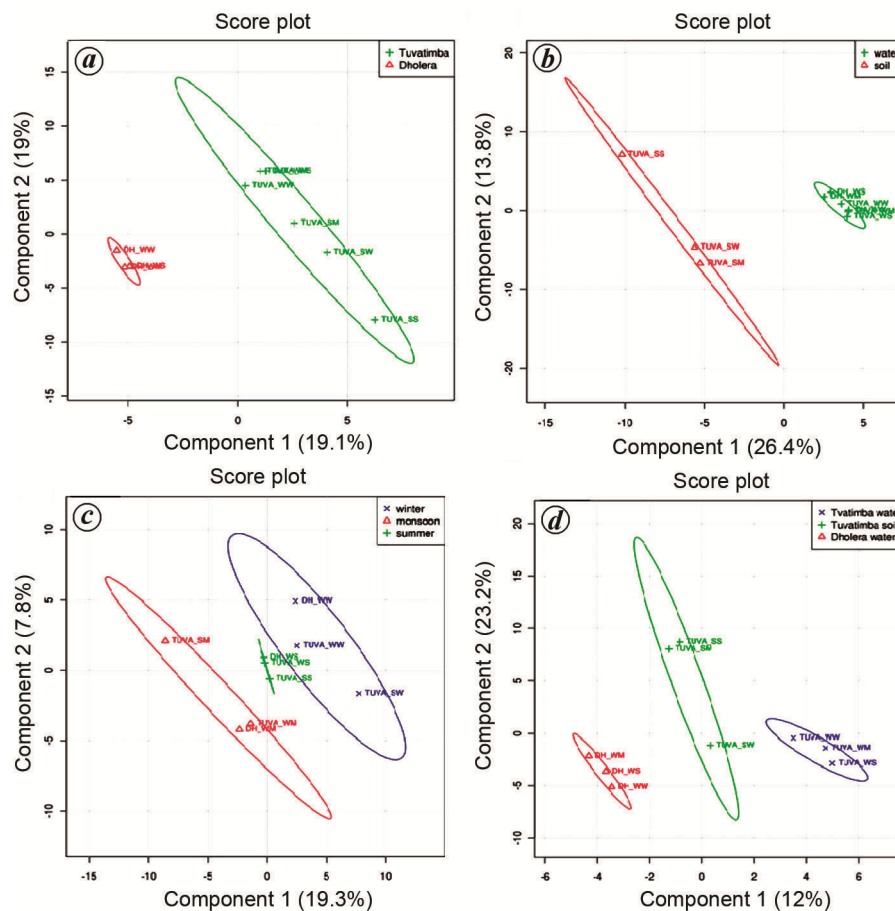
Figure 2. Score plot between (*a*) Tuva Timba and Dholera (location-wise), (*b*) Water and soil (sample-wise), (*c*) Winter, monsoon and summer (season-wise) and (*d*) Tuva Timba water and soil, and Dholera water (sample-wise three different samples).

and three soil samples) for the nine samples from Dholera and Tuva Timba hot-water springs using Ion Torrent sequencer. The generated reads were submitted to on-line metagenome server MG-RAST[14]. The adequacy of the reads generated for each sample was estimated through rarefaction curve in MG-RAST. Most of the samples revealed enough reads per sample, except for a few. A range of diversity indices was generated from the output of MG-RAST files to estimate microbial diversity among different samples of the hot-water springs.

The alpha diversity indices (total number of taxa, dominance, Simpson index, Shannon index, Berger–Parker index and Chao-1 index) were calculated using PAST[12]. The highest taxa diversity was found in Tuva Timba soil monsoon samples, while the lowest in Tuva Timba soil winter samples. The dominance of taxa was observed in Tuva Timba water monsoon samples compared to the other samples. The higher Simpson index in the Tuva Timba soil summer samples indicates an equal distribution of all taxa within the samples without predominance
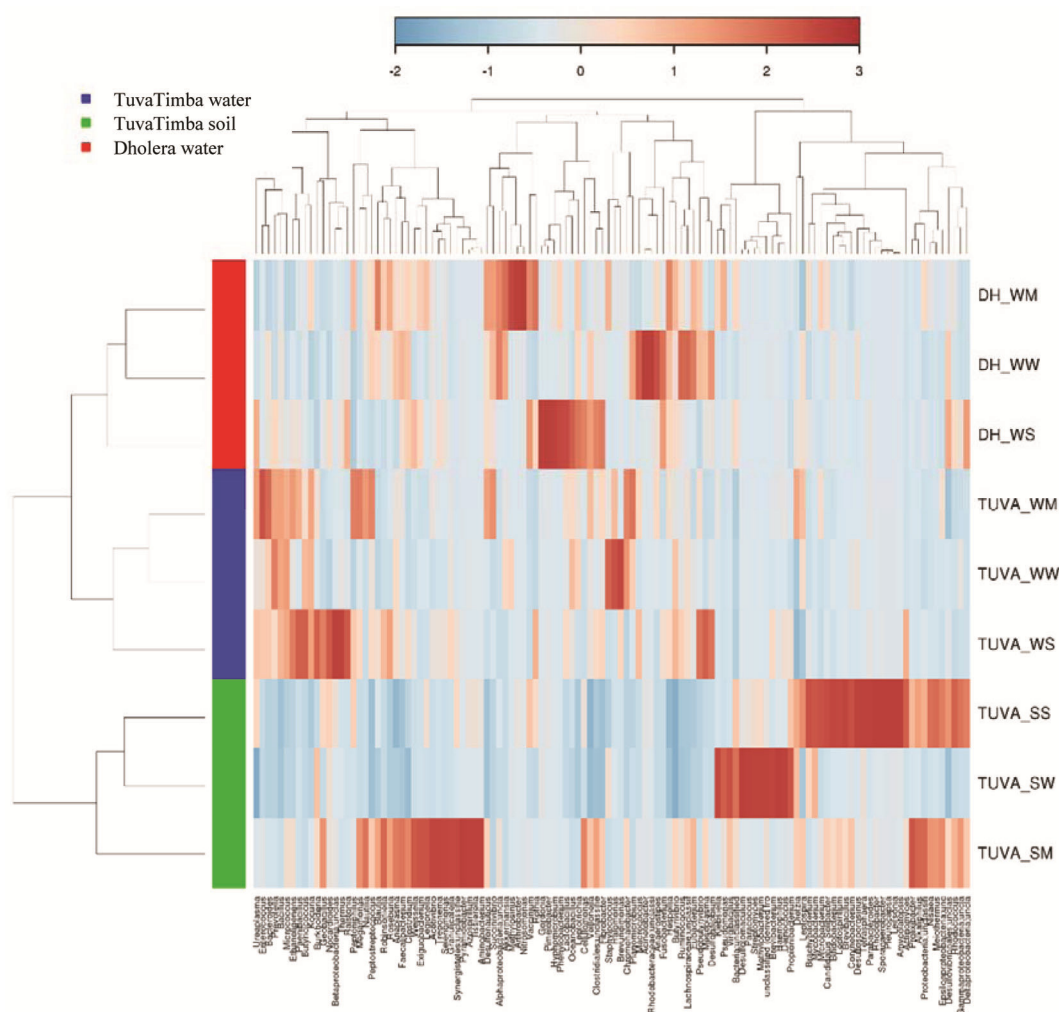
Figure 3. Heat map of samples according to locations using METAGENassist. Cluster generated from all seasonal samples. Total 418 bacterial species were obtained from all samples. Blue colour line indicates TUVA_WW (Tuva Timba water sample – winter season), TUVA_WM (Tuva Timba water sample – monsoon season) and TUVA_WS (Tuva Timba water sample – summer season). Green colour line indicate TUVA_SW (Tuva Timba soil sample – winter season), TUVA_SM (Tuva Timba soil sample – monsoon season) and TUVA_SS (Tuva Timba soil sample – summer season). Red colour line indicates DH_WW (Dholera water sample – winter season), DH_WM (Dholera water sample – monsoon season) and DH_WS (Dholera water sample – summer season).

of any particular taxon. The Shannon index diversity indicates both the aspects of diversity like richness and evenness of taxa within the Tuva Timba soil samples in summer compared to rest of the samples. The highest Berger–Parker index in Tuva Timba water samples during monsoon confirms the dominancy of a few genera over others. Chao-1 index with the highest number of Tuva Timba soil samples in summer depicts the dominance of singletons and doubletons, indicating rare OTUs (operational taxonomic units) within the samples compared to other samples (Table 1).

As a part of beta diversity to reveal the variation between samples, PCA was carried out using METAGE-Nassist[13]. For this, the unassigned reads, that is, reads with greater than 50% zeros were removed along with the elimination of near-constant variables through interquar-

tile range filtering which was achieved using METAGEN assist.

Principal coordinates analysis (PCoA) was carried out according to the sample collection season, sample type, sample collection location as well as both sample type and location. PCoA showed clear-cut clusters of samples for all three groups. Figure 2 a reveals sharing of common taxa within samples at the two different locations, i.e. Tuva Timba and Dholera. Figure 2 b reveals sharing of some common taxa within samples for soil and water. Figure 2 c shows a grouping of three clusters for monsoon, winter and summer seasons. Figure 2 d reveals a tight clusters of all soil and water samples collected in a location and season wise pattern. Cluster analysis using Ward method generated clear three groups of samples according to Tuva Timba water, Tuva Timba soil and
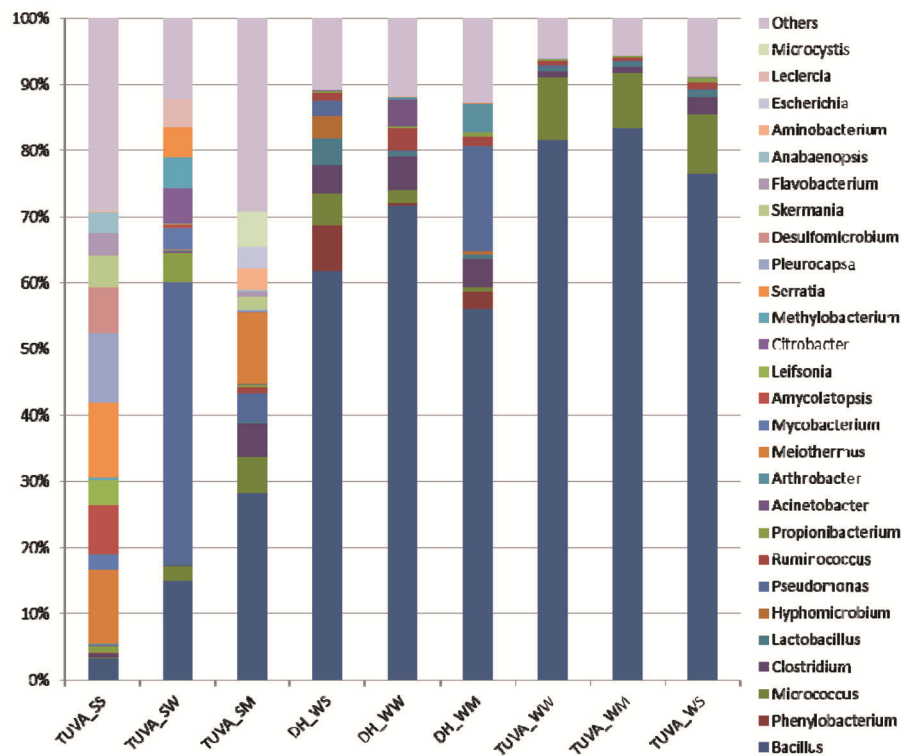
Figure 4. Comparison of genera from all seasonal samples (abundance of genera more than 3%).

Dholera water samples, suggesting sharing of common taxa within samples of different seasons, locations of hot springs (Figure 2).

The taxonomic distribution and its correlation to each other according to type of sample and location of hot springs were found to be strongly correlated to each other when heat map was generated. This finding supports the cluster analysis as well as PCoA depicting a close similarity of the same taxa according to location, season and type of samples (Figure 3).

Taxonomic composition

The taxonomic profiling of the hot-water spring samples was done using DGP database in MG-RAST with a maximum *E*-value of 1e-5 and a minimum identity of 70% and a maximum alignment of 25 bp, and phylotypes were analysed up to genus level. The MG-RAST output file was further used with STAMP software to visualize the taxonomic composition for each sample at phylum and genus level[15].

At phylum level, Firmicutes were highest up to 88% and 70% for Tuva Timba water and Dholera water samples respectively, for each season. While lowest Firmicutes were found in Tuva Timba soil summer sample depicting their overall higher abundance within water samples compared to soil samples. The Proteobacteria

phylum was highest in Tuva Timba soil winter samples (67%) followed by Dholera water samples, and lowest in Tuva Timba water samples (3–4%) of each season. The third most abundant phylum identified was Actinobacteria, highest in Tuva Timba soil samples (22%) and lowest in Dholera water winter samples (1.9%). Bacteroidetes, Deinococcus–Thermus, Cyanobacteria and Nitrospirae were prominently found in Tuva Timba soil samples of summer and monsoon season with higher abundance compared to other samples. Similarly, Deinococcus–Thermus, Cyanobacteria and Nitrospirae were predominant in Tuva Timba soil summer (18.2%) and monsoon (8.2%) samples, while synergists and planctomycetes were found dominant (1%) only in Tuva Timba soil monsoon samples. The other phylum (<1%) included Verrucomicrobia, Aquificae, Chlorobi, Gemmatimonadetes, Acidobacteria, Spirochaetes, Chloroflexi, Thermodesulfobacteria, Fusobacteria, Tenericutes, Thermotogae, Fibrobacteres, Ascomycota, Streptophyta, Chlorophyta and Basidiomycota.

At the genus level, the highest abundance of *Bacillus* was found in all three Tuva Timba water samples (76–81%) followed by Dholera water samples (56–71%), while it was lowest for Tuva Timba soil samples (3–28%). *Micrococcus* was highest in all three Tuva Timba water samples (8–9.5%), while *Clostridium* was higher in Tuva Timba soil monsoon samples and in all three Dholera water samples (4–5%). *Lactobacillus* and
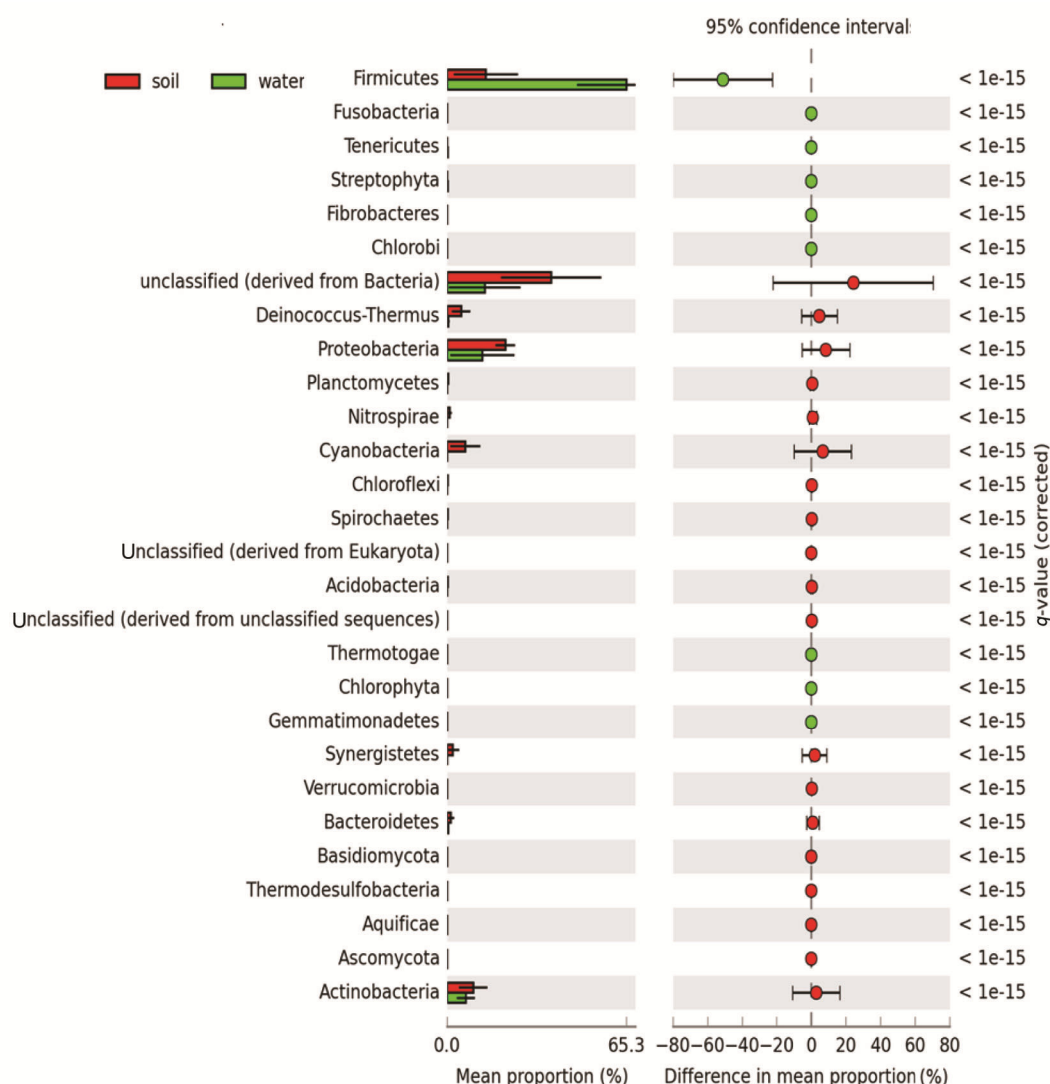
Figure 5. Mean proportion of various phyla between water and soil samples. 95% Confidence interval between soil and water samples is also shown.

Hyphomicrobium were abundant (3–4%) in only the Dholera water summer samples. Likewise, *Pseudomonas* was highest in Tuva Timba soil winter samples (42%), followed by Dhorela water monsoon samples (15%). *Meiothermus* was highest (10–11%) in only the Tuva Timba soil samples of summer and monsoon. *Mycobacterium*, *Citrobacter*, *Methylobacterium* and *Leclercia* were found only in Tuva Timba soil winter samples, while *Serratia* was found in both Tuva Timba soil samples in winter and monsoon. *Pleurocapsa*, *Desulfomicrobium*, *Skermania*, *Flavobacterium* and *Anabaenopsis* were present only in Tuva Timba soil summer samples. The overall abundance reveals that Tuva Timba water samples have less diverse microbes, while Tuva Timba soil samples contain wide diversity of microbes, specifically the Tuva Timba soil summer samples (Figure 4). The variation within sample groups according to type, soil

and water showed a significant difference for Firmicutes, Proteobacteria, Cyanobacteria, Actinobacteria and Deinococcus–Thermus genera (Figure 5). While no major significant difference was found between the three seasons.

Nucleotide sequence accession number

Metagenome sequence data are available on MG-RAST under the accession Tuva Timba soil sample (monsoon season) – mgm4754307, Tuva Timba soil sample (winter season) – mgm4754306, Tuva Timba soil sample (summer season) – mgm4754301, Tuva Timba water sample (monsoon season) – mgm4754308, Tuva Timba water sample (winter season) – mgm4754302, Tuva Timba water sample (summer season) – mgm4754305, Dholera

water sample (monsoon season) – mgm4754304, Dholera water sample (winter season) – mgm4754300, Dholera water sample (summer season) – mgm4754303.

Discussion

The study of an ecosystem like hot-water springs is of great interest as a source of extremophiles. Diversity studies based on cultivable organisms conducted at Tuva Timba and Dholera hot-water springs, which revealed purification of only 66 isolates; this is very low in terms of the variety of organisms present[16,17]. The 16S rDNA amplicon sequencing of the two hot-water springs helped uncover the presence of unique taxonomies and diversities of thermophile microbial communities. Season-wise water and soil samples from Tuva Timba and only water samples from Dholera were included. The clustering of samples by community structure of sequence composition clearly shows the effect of geographic location on microbial structure, as community structure is largely determined by a combination of environmental parameters as well as geographical distance. However, very high abundance of a single genus or member of a phylum will influence the clustering of samples based on phylum abundance[18].

The study revealed that there was vast difference in taxonomic abundance compared to water and soil, as well as season-wise. Although comparatively Tuva Timba water samples (season-wise) depicted less diversity, while soil samples exhibited additional diverse taxa. The phylum Firmicutes was the most abundant for Tuva Timba water samples, while Proteobacteria was the most abundant for Dholera water and Tuva Timba soil samples. The third most abundant phylum was Actinobacteria, followed by Bacteroidetes and Deinacoccus–Thermus. A similar finding was observed for Tulsi Shyam hot-water spring in Gujarat using metagenome analysis[6]. The authors[6] also identified the most abundant phylum as Firmicutes followed by Proteobacteria, similar to our findings.

Amplicon sequencing analysis carried out by Tattapni geothermal fields reported a phylum Proteobacteria in their studies[2]. This phylum has also been reported in various studies carried out using 16S rRNA analysis of hot springs at various geographical locations of India and appears to be the indigenous phylum of these regions[6,19]. The microbes belonging to Proteobacteria are found to be abundant at this site, and also known to be highly abundant in petroleum-contaminated terrestrial and aquatic environments[20,21]. The methane-producing and sulphur-utilizing thermophilic bacteria belonging to genera such as *Acinetobacter*, *Pseudomonas*, *Methylobacterium* and *Desulfomicrobium* have been found in most of the samples. Similar bacteria belonging to these genera were also found in the study carried out at California, USA and

their association was found with high-temperature petroleum reservoirs[22,23]. The 16S rDNA-based study carried out at the Gulf of California, USA, also found sulphate-reducing bacterial populations, such as *Desulfobacter*, *Desulforhabdus*, *Thermodesulforhabdus*, etc.[24]. The presence of sulphate-reducing microbes is important in degrading organic matters under anoxic environments. They also play an important role in energy production as well as in the maintenance of the microbial community. They depend on sulphate and elemental sulphur as the thermal electron acceptor during anaerobic metabolism[25]. The presence of *Methylobacterium* shows the importance of this site for the isolation of novel methanotrophic species as few of them are known to convert Acetyl-CoA for subsequent energy-generation pathways, which was evident in many studies[2,26].

Conclusion

This culture-independent study has given an important insight into the potentially novel microbial diversity and community composition of two hot-water springs in Gujarat. These hot springs were abundant in phyla like Firmicutes, Proteobacteria and Actinobacteria. These phyla were found to be specifically correlated with physicochemical factors. Thus phylum Firmicutes and genus *Meiothermus* correlated with temperature, Proteobacteria with sulphide and *Mycobacterium* with methane. The presence of sulphate-reducing microbes indicates an active sulphur cycle in these hot springs. The identified thermotolerant and thermophilic microbes confirm the significant correlation between geographical and environmental parameters of hot-water springs.

Conflict of interest: The authors declare that no conflict of interest.

1. Mirete, S., Morgante, V. and Gonzalez-Pastor, J. E., Functional metagenomics of extreme environments. *Curr. Opin. Biotechnol.*, 2016, **38**, 143–149.
2. Saxena, R. *et al.*, Metagenomic analysis of hot springs in Central India reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Front. Microbiol.*, 2016, **7**, 2123.
3. Mangrola, A. *et al.*, Deciphering the microbiota of Tuwa hot spring, India using shotgun metagenomic sequencing approach. *Genom. Data*, 2015, **4**, 153–155.
4. Lopez-Lopez O., Cerdan, M. E. and Gonzalez-Siso, M. I., Hot spring metagenomics. *Life* (*Basel*), 2013, **3**, 308–320.
5. Mehetre, G. T., Paranjpe, A., Dastager, S. G. and Dharne, M. S., Investigation of microbial diversity in geothermal hot springs in Unkeshwar, India, based on 16S rRNA amplicon metagenome sequencing. *Genome Announc.*, 2016, **4**(1), e01766-15.
6. Ghelani, A., Patel, R., Mangrola, A. and Dudhagara, P., Cultivation-independent comprehensive survey of bacterial diversity in Tulsi Shyam Hot Springs, India. *Genom. Data*, 2015, **4**, 54–56.

7. Mangrola, A. V., Dudhagara, P., Koringa, P., Joshi, C. G. and Patel, R. K., Shotgun metagenomic sequencing based microbial diversity assessment of Lasundra hot spring, India. *Genom. Data*, 2015, **4**, 73–75.

8. Panda, A. K., Bisht, S. S., Mandal, S. and Kumar, N. S., Bacterial and archeal community composition in hot springs from Indo-Burma region, North-east India. *AMB Express*, 2016, **6**(1), 111.

9. Bhatia, S., Batra, N., Pathak, A., Green, S. J., Joshi, A. and Chauhan, A., Metagenomic evaluation of bacterial and archaeal diversity in the geothermal hot springs of Manikaran, India. *Genome Annouc.*, 2015, **3**(1), e01544-e0154414.

10. Reaffirmed, 2003, IS:3025 Methods of Sampling and Test (Physical and Chemical) for Water and Wastewater Part 24 Sulphates.

11. Meyer, F. *et al.*, The metagenomics RAST server – a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform.*, 2008, **9**, 386.

12. Hammer, R., Harper, D. and Ryan, P., PAST: paleontological statistics software package for education and data analysis. *Palaeont Electro*, 2001, **4**, 9.

13. Arndt, D. *et al.*, METAGENassist: a comprehensive web server for comparative metagenomics. *Nucleic Acids Res.*, 2012, **40**, W88–W95.

14. Meyer, F. *et al.*, MG-RAST version 4 – lessons learned from a decade of low-budget ultra-high-throughput metagenome analysis. *Brief. Bioinform*, 2017, **20**(5), 1–9.

15. Parks, D. H., Tyson, G. W., Hugenholtz, P. and Beiko, R. G., STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*, 2014, **30**, 3123–3124.

16. Patel, D. N., Dave, S. R., Braganza, V. J. and Modi, H. A., Seasonal variation in bacterial diversity of Tuva Timba thermal springs of Gujarat, India. *Int. Res. J. Biol. Sci.*, 2019, **8**(2), 16–14.

17. Patel, D. N., Dave, S. R., Braganza, V. J. and Modi, H. A., Assessment of culturable microbial diversity of Dholera thermal springs of Gujarat, India. *J. Exp. Biol. Agric. Sci.*, 2019, **7**(1), 57–64.

18. Menzel, P. *et al.*, Comparative metagenomics of eight geographically remote terrestrial hot springs. *Microb. Ecol.*, 2015, **70**, 411–424.

19. Bowen De Leon, K., Gerlach, R., Peyton, B. M. and Fields, M. W., Archaeal and bacterial communities in three alkaline hot springs in Heart Lake Geyser Basin, Yellowstone National Park. *Front. Microbiol.*, 2013, **4**, 330.

20. Kimes, N. E. *et al.*, Metagenomic analysis and metabolite profiling of deep-sea sediments from the Gulf of Mexico following the Deepwater Horizon oil spill. *Front. Microbiol.*, 2013, **4**, 50.

21. Bao, Y. J. *et al.*, High-throughput metagenomic analysis of petroleum-contaminated soil microbiome reveals the versatility in xenobiotic aromatics metabolism. *J. Environ. Sci.* (*China*), 2017, **56**, 25–35.

22. Orphan, V. J., Taylor, L. T., Hafenbradl, D. and Delong, E. F., Culture-dependent and culture-independent characterization of microbial assemblages associated with high-temperature petroleum reservoirs. *Appl. Environ. Microbiol.*, 2000, **66**, 700–711.

23. Steven, B. *et al.*, Characterization of the microbial diversity in a permafrost sample from the Canadian high Arctic using culture-dependent and culture-independent methods. *FEMS Microbiol. Ecol.*, 2007, **59**, 513–523.

24. Dhillon, A., Teske, A., Dillon, J., Stahl, D. A. and Sogin, M. L., Molecular characterization of sulfate-reducing bacteria in the Guaymas Basin. *Appl. Environ. Microbiol.*, 2003, **69**, 2765–2772.

25. Elshahed, M. S. *et al.*, Bacterial diversity and sulfur cycling in a mesophilic sulfide-rich spring. *Appl. Environ. Microbiol.*, 2003, **69**, 5609–5621.

26. Ward, N. *et al.*, Genomic insights into methanotrophy: the complete genome sequence of *Methylococcus capsulatus* (Bath). *PLoS Biol.*, 2004, 2, e303.