CrossMark

Comparative analysis of microbial diversity in two hot springs of Bakreshwar, West Bengal, India

Biswadeep Chaudhuri[1], Trinath Chowdhury[1], Brajadulal Chattopadhyay*

Department of Physics, Jadavpur University, Raja Subodh Chandra Mallick Road, Kolkata 700 032, West Bengal, India

ABSTRACT

Various aspects of hot springs at Bakreshwar (Lat. 23°52′48″N; Long. 87°22′40″E) in West Bengal, India have been investigated since the middle of 20th century, but comprehending the complete diversity and the complexity of the microbial population therein has been in the continuing process. Some of these microorganisms are found to have immense industrial importance. Microbes generally exist in milieus of varying complexities and diversities. Attempting the usually employed cultivation-based techniques in experimentation with those microbes had confronted various limitations. To overcome these limitations a strategy based on high-throughput sequencing of 16S rRNA gene amplicon analysis was employed for studying the differential diversity and the detailed nature of microbial population of the two hot springs of Bakreshwar (54 °C & 65 °C). Paired-end libraries of amplified V-3 hyper-variable 16S rDNA fragments from sets of samples that varied in their contents, ranging from a single bacterium to highly complex communities were sequenced. The comparison revealed the differential aspects in the two hot spring waters; the samples at 54 °C showed the bacterial phylum *Firmicutes* (65.85%) and *Synergistetes* (27.24%) predominating and those from hot spring water at 65 °C showed the abundance of the phyla *Firmicutes* (96.10%) and *Proteobacteria* (3.36%). The presence of Archaea in the hot springs could not be ascertained.

1. Introduction

Bakreshwar hot springs in the geothermal belt (35 °C to 81 °C), were formed as a result of volcanic or tectonic activities belong to the Chotanagpur gneissic complex within a Precambrian craton in the eastern part of Peninsular India [1]. The environmental conditions prevailing during the origin of a Precambrian craton included high temperature, poor amount of free oxygen, high concentration of sulfur and lots of reducing gases like methane, ammonia, carbon dioxide [2]. The environmental conditions of these hot springs suited for being the natural niches for some unicellular organisms, prokaryotes known as thermophiles (optimal growth temperature > 50 °C) and hyperthermophiles (OGT > 80 °C) limited in the domain of Archaea and bacterial viruses [3].

Places which were once perceived to be sterile abound with microorganisms with great diversity. The search for them has been intensified recently as the enzymes isolated from them are proven to be extremely valuable biocatalysts with potentially significant application for use in industry and biotechnology [2]. An important feature of the cluster of Bakreshwar hot springs is the high Helium content in their water with values ranging from 1.2% to 2% in air at 1 m above the ground, [4] which is far above the atmospheric background level, and the gas is released in periodic bursts [5]. The thermal water is alkaline with low to moderate sodium-HCO_3^- and SO_4^{++} as compared to chloride. The Cl^-, HCO_3^- and SO_4^{++} ternary plot of thermal water in Bakreshwar falls in HCO_3^- field indicating sodium-bicarbonate type water [6]. This has led to the isolation and identification of a number of metal-reducing bacteria. Three bacterial strains so isolated and partially characterized from water of the 65 °C hot spring consortium were found efficacious for enhancement of concrete strength, silica leaching activity and other industrial applications [7–12].

The utilization of the procedure of Metagenomics bypasses the requirement for obtaining pure cultures. It has now become a standard method for characterization of the biodiversity, genome contents and inferred functions of bacterial and Archaeal communities [13]. Metagenomic sequencing has also been applied to diverse microbial communities existing in surface soils, fresh and marine waters, where enormity of previously undetected biodiversity has been revealed [14–17].

Revelations obtained from the work have led to a markedly clear

---

* Corresponding author at: Jadavpur University, Kolkata 700032, West Bengal, India.
  E-mail address: bdc@phys.jdvu.ac.in (B. Chattopadhyay).
[1] These authors contributed equally to this work.

breakthrough in understanding the differences in diversity of the bacterial population of two hot springs in Bakreshwar. The use of 16S rRNA amplicon sequencing in MiSeq System, Illumina established the presence, besides a few other bacteria, of novel non-marine thermophilic relatives thereof. Further investigative studies can be directed towards metagenomic sequencing of microbes in other hot springs in the same region with probable results of interest. Sequencing can even be aimed at for other differently variable regions (V1–V9) as well.

2. Materials and methods

2.1. Reagents

All the analytical grade chemicals and instruments used in this investigation were purchased from the Sigma, USA, Merck, Germany and Amresco, USA, GE, USA, KAPA Biosystems, USA, Applied Biosystems, USA, Qiagen, USA, Agilent, USA, Thermo Scientific, USA.

2.2. Collection of sample

Samples were collected from two hot springs (65 °C and 54 °C) of Bakreshwar. The surface temperature of the sediment of both the hot springs varied between 2 °C and 3 °C. The pH of the water was measured varying from 7.8 to 8.0. The sediment contained 1.1 to 1.5% organic carbons and, interestingly, 280 to 422 mol of reducible Fe (III) and 280 to 600 mol of reduced iron per g of wet sediment [18]. The temperature of the water at the time was around 65 °C and 54 °C. 50–60 l of hot spring water was passed through membranes 6 times in the span of 4 years and subsequent DNA extraction was performed after mixing the different samples. Due to extremely low yield of DNA the process had to be carried out several times.

2.3. Isolation of genomic DNA

Bacterial genomic DNA was isolated by CTAB/NaCl method from the enrichment culture [19]. A 1.5 ml of the saturated bacterial culture was centrifuged at $6000 \times g$ for 5 min to collect the bacterial cell pellet. The cell pellet was then suspended in 567 µl of TE buffer (10 mm Tris, 1 mm EDTA, pH 8). Thereafter, 30 µl of 10% SDS and 3 µl of 20 mg/ml Proteinase-K (Amresco) were added to the same. The solution was mixed thoroughly and incubated at 37 °C for 1 h. 100 µl of 5 M NaCl and 80 µl of CTAB/NaCl solution (Sigma) was added and incubated at 65 °C for 10 min and centrifuged at $10000 \times g$ to remove cell wall debris, denatured proteins and polysaccharides. Extraction of total chromosomal DNA was done with an equal volume of a mixture of chloroform: isoamyl alcohol (24:1) at $10000 \times g$. The aqueous phase was again similarly extracted with phenol: chloroform: isoamyl alcohol mixture (25:24:1). DNA precipitation was done by adding 0.6 volume of 2-propanol and centrifuged at $12000 \times g$ for 10 min [20]. The DNA pellet was then washed with 70% ethanol, dried and dissolved in 50 µl autoclaved MilliQ water for sequencing purposes. The input DNA samples were quantitated using Qubit DNA quantitation assays (Thermo Scientific) that specifically quantitate dsDNA. The quality check of the DNA samples were done using Nano-Drop Spectrophotometer (Thermo Scientific) as well as using 0.8% Agarose TAE gel stained with Ethidium Bromide (Supplementary Fig. S1). Lower 260/230 ratio of the samples indicated presence of contaminants such as carbohydrates and phenols that absorb at 230 nm but this does not antagonize the sequencing procedure. Samples were taken that had a 260/280 ratio of 1.84 and 260/230 ratio of 1.79 for sequencing purpose (Supplementary Table S1). Five DNA samples each culled from both the hot springs were sequenced by mixing the five samples from each of the hot springs separately. Concentration of the DNA samples was found to be sufficient for library preparation after enriching the V3 hypervariable region of bacterial 16S rRNA gene. Both the samples were taken further for library preparation.

2.4. Designing amplification primers

2.4.1. 16S rRNA V3 enrichment and library preparation

The phylogenetic distribution of each primer sequence was analyzed using the RDP Probe Match [21]. The goal was to obtain broadly distributed primers that amplify regions of approximately 100–150 bp in length such that the paired end reads form an amplicon in some overlapping region. Sample-identifying barcode sequences were included at the 5′ ends on both the forward and the reverse primers. The barcodes were designed to be short, 3 or 4 bp, to minimize loss of sequence information; to have two or more differences from one another (such that a single sequencing error could not convert one into another); and to have the 3′ nucleotide of the barcode be a low frequency match of nucleotide upstream from the 5′ end of the priming site. Errors in the barcode and primer sequence of either of the paired end reads necessitated the removal of both reads in the pair; as in the majority of read-pairs that were culled; either only one of the reads perfectly matched the primer sequence or the paired reads lacked identical barcodes. Error frequencies were estimated for the barcode/primer regions of all reads with high sequence quality (Xq30, Q scores in Illumina are calculated using the following formula $Q = -10\log_{10}P$ where P stands for the probability value assigned to the base by the sequencer during the base call process. A quality of Q30 means that the base call accuracy is 99.9%.) for the first 25 nt with single nucleotide differences from the expected sequence.

The V3 hyper-variable region of bacterial 16S rRNA was enriched using modified HPLC purified 341F and 518R primers (Supplementary Table S2). In addition to V3 priming regions, these primers have complimentary sequences to Illumina forward, reverse and multiplex sequencing primers. The reverse primer contains a 6-bp index allowing for multiplexed sequencing.

Each PCR amplification reaction contained 25 pmol of each primer, a 200 M concentration of each of dNTPs, 1.5 mM $MgCl_2$ and 1 U High Fidelity DNA Polymerase (KAPA HiFi DNA Polymerase, KAPA Biosystems, USA). The PCR involved an initial denaturation step at 95 °C for 5 min followed by 20 cycles of 95 °C for 1 min, 50 °C for 1 min, and 72 °C for 1 min and ended with an extension step at 72 °C for 7 min in a Thermal Cycler (Veriti® Thermal Cycler, Applied Biosystems, USA). PCR amplicons were purified using QIAquick PCR purification kit (Qiagen, USA) and checked for desired fragment size and concentration using Agilent Bioanalyzer DNA 1000 assay. 50 ng of purified bacterial genomic DNA was used for enrichment of V3 region and library preparation. The flow chart representation of the sequencing work flow is depicted below (Fig. 1).

2.4.2. Sequencing library quality control (QC)

Sequencing libraries were QC and quantitated using Agilent Bioanalyzer DNA 1000 assay. Libraries were checked for right fragment size and required concentration for optimal cluster generation and sequencing.

Both libraries showed the expected fragment distribution around 300 nt and yielded sufficient concentration for cluster generation. The libraries were normalized to 10 nM and taken further for cluster generation and sequencing on MiSeq System, Illumina.

2.4.3. Sequence data QC

The libraries were sequenced using MiSeq 2×151 bp paired end format. The Illumina MiSeq generated sequence data was quality checked using FastQC (Version 0.11.3) and Rqc (Version 1.4.2) [22,23]. The data were checked for various parameters such as base call quality distribution, % bases above Q30, GC%, sequencing adapter contamination etc. (Supplementary Table S3).

2.4.4. Read Quality Plot

Both the samples (Fig. 2) yielded > 400 K read pairs per sample, which was well above the minimum target of 100 K. The data had very
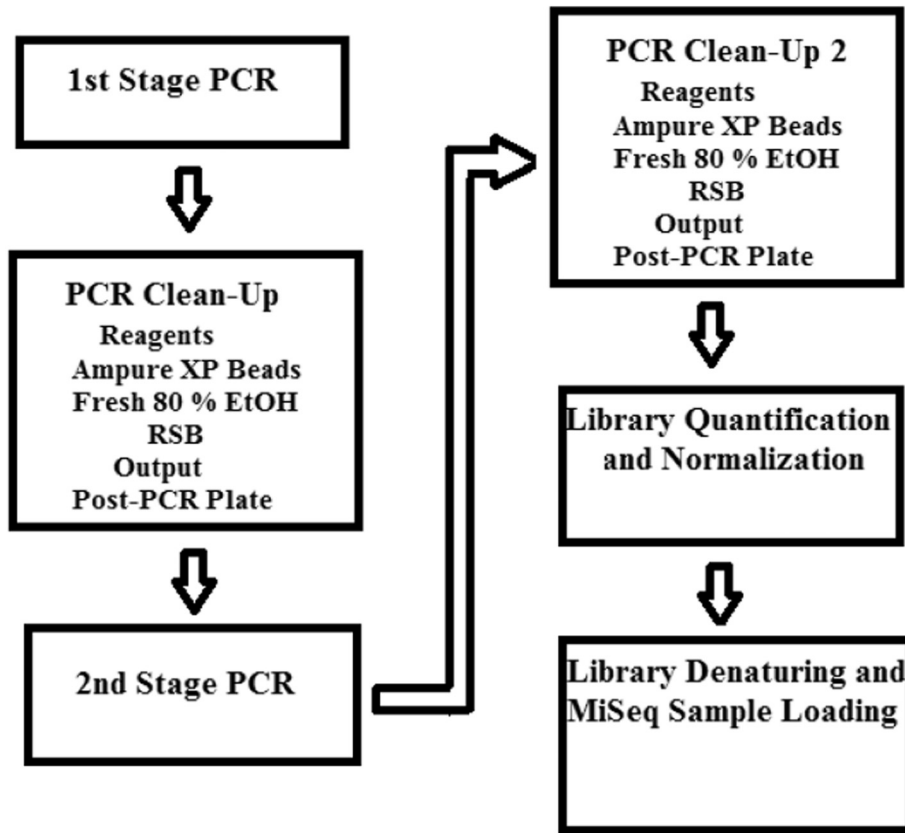
Fig. 1. Flowchart representation of 16S library preparation protocol in different steps.

high base call quality (High base call quality as on an average 97% of reads are above Q30.). The difference in quality between Read 1 - > forward pair, Read 2 - > reverse pair observed is normal as Illumina Read 2 shows slightly lesser quality as compared to Read 1. It was expected to have same GC% in Read 1 & 2 of a sample but the slight difference in this observed in B65 might be due to the quality difference between the read pairs. No quality trimming was done as the data was of good quality. There were no adapter sequence contaminations in the sequence data. Data sets were taken as input for Metagenome Analysis Pipeline. The data quality reported by FastQC and Rqc were observed to be above expectation hence no QC trimming

was carried.

2.4.5. Data analysis

For downstream analysis of the metagenome data, 'mothur' software [24] bundle was used. The quality filtered sequence reads were imported into mothur and the read pairs were aligned with each other to form contigs. The contigs were screened for errors and only those between 170 bp and 200 bp were taken for further analysis. Reads with ambiguous base calls were rejected; however filtering was not necessary as the read quality was extremely good. The high quality contigs were now checked for identical sequences and duplicates were merged.
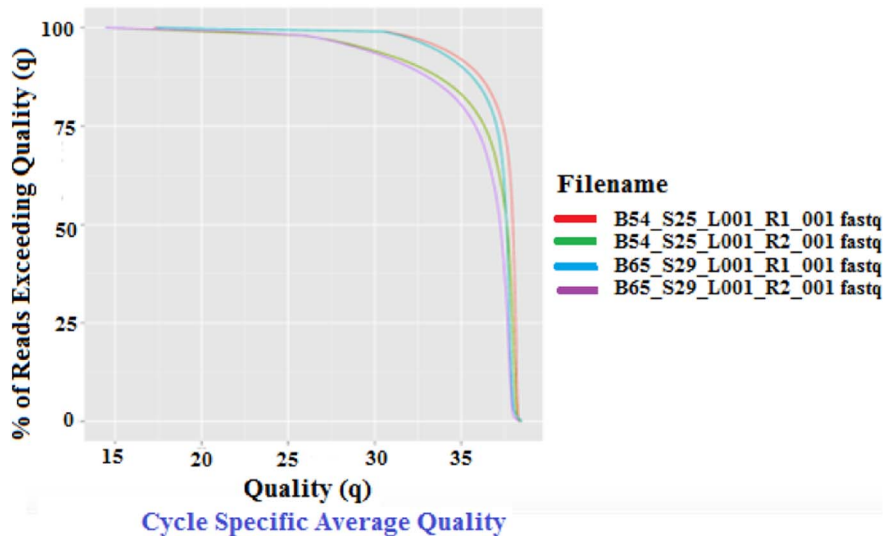

Fig. 2. Read Quality Plot in Illumina MiSeq. This plot describes the average quality pattern by showing on the X-axis quality thresholds and on the Y-axis the percentage of reads that exceed that quality level.

Although the primers for enrichment were designed for bacterial 16SrRNA V3 hyper variable region, there were chances of non specific amplification of other regions. In order to check for any such non specific amplification, the contigs were aligned to a known database for 16 s rRNA. Any ambiguous contigs aligning to untargeted regions were discarded. Gaps and overhangs introduced in the contigs during the alignment process described above were removed and the chimeras that might have resulted due to PCR errors were removed.

Mothur Commands

```
make.contigs(file=input.files, processors=15)
screen.seqs(fasta=current, group=current, maxambig=0,minlength=170, maxlength=200)
unique.seqs()
count.seqs(name=current, group=current)
align.seqs(fasta=current, reference=silva.nr_v119.align)
summary.seqs()
screen.seqs(fasta=current, count=current, start=6332, end=13862, maxhomop=8)
filter.seqs(fasta=current, vertical=T, trump=.)
unique.seqs(fasta=current, count=current)
pre.cluster(fasta=current, count=current, diffs=2)
chimera.uchime(fasta=current, count=current, dereplicate=t)
remove.seqs(fasta=current, accnos=current)
classify.seqs(fasta=current, count=current, reference=silva.nr_v119.align,
taxonomy=silva.nr_v119.tax, cutoff=80)
remove.lineage(fasta=current, count=current, taxonomy=current, taxon=Chloroplast-
Mitochondria-unknown-Eukaryota)
dist.seqs(fasta=current, cutoff=0.15)
cluster(column=current, count=current, cutoff=0.03)
make.shared(list=current, count=current, label=0.03)
classify.otu(list=current, count=current, taxonomy=current, label=0.03)
quit()
```

UCHIME algorithm (Version 1.39.0) [25] was used to flag and remove chimeric contigs. A known reference of all the chimeric sequence was cross checked. The quality filtered contigs were classified into taxonomical outlines based on the RDP database [26].

The classified contigs were filtered for any undesirable lineage from the taxonomy file. Usually the undesirables are sequences from Mitochondria, Chloroplast, Fungi or Archaea. The contigs were then clustered into OTUs (Operational Taxonomic Unit) and the abundance of each OTU in the population was estimated (Fig. 3). The taxonomy

tables were then analyzed using phyloseq (Version 1.14.0) [27], a R based package (Version 3.2.2) [28] for downstream analysis and result visualization. Pie charts showing abundance were plotted using Krona Tools (Version 2.7) [29].

Phyloseq Commands

```
require(phyloseq)
input_mothur=import_mothur(
mothur_list_file =
"input.trim.contigs.good.unique.good.filter.unique.precluster.pick.silva.wang.pick.tx.list",
mothur_shared_file =
"input.trim.contigs.good.unique.good.filter.unique.precluster.pick.silva.wang.pick.tx.shared",
mothur_constaxonomy_file =
"input.trim.contigs.good.unique.good.filter.unique.precluster.pick.silva.wang.pick.tx.1.cons.taxon
omy")
plot_bar(input_mothur)
```

3. Results

From each of the two multiplexed Illumina sequencing runs over 40 million paired-reads were collected. The primer and barcodes sequences which were virtually not informative for sequence classification constituted over 25% of the sequencing reads. Inclusion of both these sequences together provided an essential filter for the removal of erroneous reads and estimation of error rates in the sequencing process.

3.1. Composition of bacterial communities in the hot spring water

The number of reads obtained by amplicon sequencing ranged from 416,074 to 442,064 in the two different samples (54 °C & 65 °C respectively) (Tables 1, 2) (NCBI Gene Bank Accession Number: BioProject ID:PRJNA305225, for 54 °C-SAMN0.4326631 & for 65 °C-SAMN0.4326630, 54 °C-SRR3025782 & 65 °C-SRR3025965). Based on the taxonomic classification of representative sequences from all OTUs, the composition of bacterial communities was revealed at phylum level (Fig. 4) and class level (Figs. 5 and 6). Among the analyzed sample (B 54-after pooling 5 samples from each of the 54 °C hot spring together) from the hot spring water at 54 °C, the bacterial phylum *Firmicutes* (65.85%) and *Synergistetes* (27.24%) predominated. Less abundant taxa
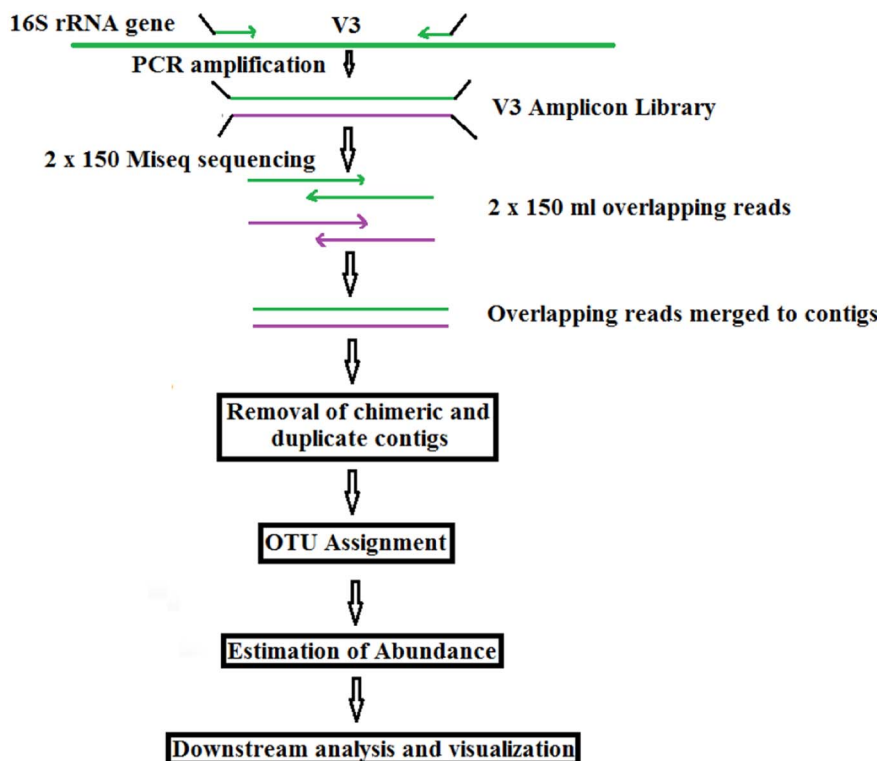

Fig. 3. Amplification using V3 primers, overlapping reads merged to contigs & downstream analysis and visualization.

Table 1
B54 top 10 & OUT distribution.

Rank	Kingdom	Phylum	Class	Order	Family	Genus	Abundance
1.	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae 1	*Caloramator*	3788
2.	Bacteria	Firmicutes	Clostridia	Clostridiales	Clostridiaceae 1	*Clostridium*_sensu_stricto	3501
3.	Bacteria	Synergistetes	Synergistia	Synergistales	Synergistaceae	Unclassified	3130
4.	Bacteria	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	447
5.	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_1	*Anoxybacillus*	99
6.	Bacteria	Firmicutes	Clostridia	Clostridiales	Unclassified	Unclassified	287
7.	Bacteria	Synergistetes	Synergistia	Synergistales	Synergistaceae	Unclassified	16
8.	Bacteria	Unclassified	Unclassified	Unclassified	Unclassified	Unclassified	78
9.	Bacteria	Firmicutes	Bacilli	Bacillales	Unclassified	Unclassified	20
10.	Bacteria	Firmicutes	Clostridia	Clostridiales	Syntrophomonadaceae	Unclassified	25

that were detected belonged to *Actinobacteria* (0.15%), *Proteobacteria* and *Bacteroidetes* (Fig. 5A & B, Supplementary Tables S4, S5). Around 5% of the sequences belonged to unclassified bacteria and could not be assigned to any bacterial phyla. The sample (B 65-after pooling 5 samples from each of the 65 °C hot spring together) from hot spring water at 65 °C showed the abundance of the phyla *Firmicutes* (96.10%) and *Proteobacteria* (3.36%). The less abundant taxa belonged to *Actinobacteria*, *Synergistetes* and *Bacteroidetes* (0.05%) respectively (Fig. 6A & B, Supplementary Tables S6, S7). At class level the microbiomes at 54 °C contained mostly representatives of *Clostridia* (65.85%) and *Synergistia* (27.23%). Less abundant representatives could be classified as Bacilli (1.2%), *Actinobacteria*, *Beta* and *Gamma-Proteobacteria* (< 0.2% each). In contrast the sample from 65 °C showed the abundance of the Class *Bacilli* (96.1%) and *Gammaproteobacteria* (3.3%). Some sequences could be classified as *Clostridia* (0.15%), *Actinobacteria* (0.067%), *Betaproteobacteria* (0.059%) and *Synergistia* (0.059%).

The putative core microbiome at 54 °C consisted of the genus *Caloramator* (33%) and *Clostridiumsensu_stricto* (30%). The sequences belonging to family Synergistaceae (27%) could not be assigned to any particular genera. Less abundant genera were represented by *Anoxybacillus* (Bacillaceae), *Pseudoxanthomonas*. The microbiome at 65 °C consisted predominantly of genus *Geobacillus* (96%), followed by *Pseudomonas* (3%). Rare taxa which could be identified belonged to *Caloramator*, *Halomonas*, *Pseudoxanthomonas*, *Acinetobacter*. The analysis of the microbial communities indicated a high degree of temperature specificity.

4. Discussion

Many unidentified classes and families in the respective consortium are still left to be investigated. Low-quality tags were then subsequently filtered out leaving only assembled reads that contained barcode and primer sequences at each end, which bounded a phylogenetically informative amplified region. Using RDP Classifier all the possible taxonomic identifications were done to assign order, family and genus.

A general trend of decrease in diversity of microbial community was

observed with increasing environmental temperature and this is in accordance with previous hot spring metagenomic analysis [30–33]. Multivariate analyses using PCoA (Principal Coordinate Analysis) plot deduced from the distance matrix calculated by the weighted normalized UniFrac algorithm using phylogenetic information can explain in a better way the differences in the two different microbiomes [26]. Previous studies showed that a high environmental temperature in combination with low pH values is associated with a high abundance of Archaea in the microbial community [34]. During preliminary DNA extraction and PCR amplification using normal Archaeal primers revealed no amplification in almost all the reactions. The hot springs of Bakreshwar are rather unique as the Archaea population which is one of the characteristic of any hot spring is significantly low or may even be absent in the consortium. It is tempting to speculate that fundamental differences in membrane and cell wall structure between bacteria and Archaea contribute to Archaea dominance in low pH, high temperature environments [17]. The reported numbers in diversity can be considered to be a lower bound on the richness of species; as rare species, especially in samples with few dominant species could only be captured by deep sequencing protocols [17].

A similarity pattern could be observed in the hot springs as both have high temperature and high pH as well as an influx of organic materials. Thus, community structure is largely determined by a combination of environmental parameters, rather than geographical distance. However, a very high abundance of a single species or members of the phylum will influence clustering of samples based on phylum level abundances. Previous studies identified differences in the amino acid composition between mesophiles and thermophiles [35,36], which originated at differences in genomic GC content as well as changes of relative frequencies of certain amino acids that affected the thermal stability of the protein [37,38]. The most variability of the amino acid profiles of the thermophiles was well justified by the genomic GC content, temperature having only little significant effect [39]. Nonetheless, in case of phyla-based clustering of organic-rich hot springs there is a negative correlation between Archaea/bacteria ratio and diversity [17].

Table 2
B65 top 10 & OUT distribution.

| Rank	Kingdom	Phylum	Class	Order	Family	Genus	Abundance
1.	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_1	*Geobacillus*	8558
2.	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_1	*Geobacillus*	2072
3.	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	319
4.	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_1	*Geobacillus*	523
5.	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_1	*Geobacillus*	40
6.	Bacteria	Firmicutes	Bacilli	Bacillales	Unclassified	Unclassified	33
7.	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_1	*Geobacillus*	11
8.	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	16
9.	Bacteria	Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	*Pseudomonas*	18
10.	Bacteria	Firmicutes	Bacilli	Bacillales	Bacillaceae_1	*Geobacillus*	6
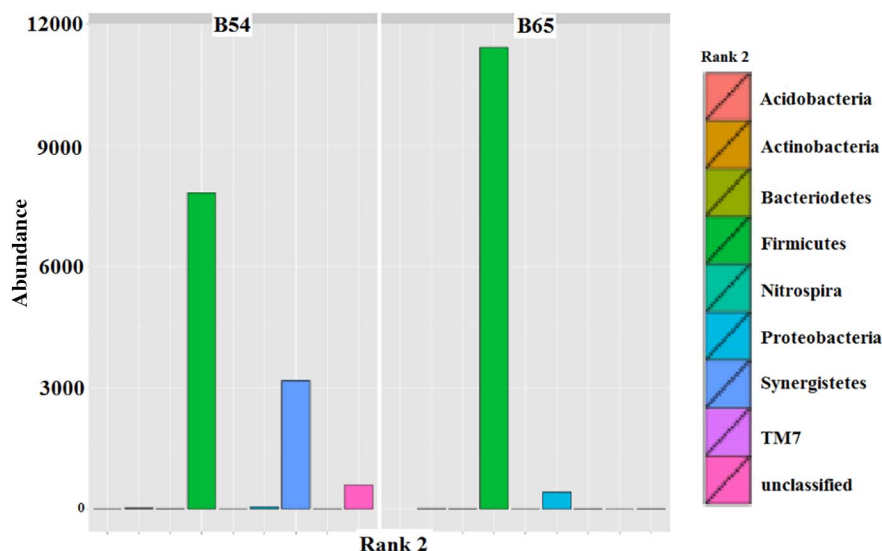
Fig. 4. Phylum level abundance in B54 and B65. (X-axis represents samples B54 and B65, Y-axis represents abundance of phylum, legend list of the phylums being represented on right side).

5. Conclusion

Comparative analysis of the microbiomes revealed primary variation in the *Firmicutes* population in the two different hot springs and absence of Archaea which is quite intriguing. Complete data obtained from single run at times cannot be fully conclusive hence primarily multivariate analyses using PCoA (Principal Coordinate Analysis) and other statistics may be taken by running the same sample in triplicates and geochemical processes of the hot springs should be aimed at in future. The Plethora of thermophilic organisms with great industrial significance resides in these hot springs. Novel sequencing strategies may probably open up new paradigm shift in understanding molecular ecology and especially microbial diversity of the other hot springs of the region. More studies can be directed towards understanding diversity in the microbial population of the other eight hot springs of Bakreshwar; as hot springs with varying temperatures in the same place is rather uncommon. Identification of special thermophilic populations from these hot springs and subsequent study thereof can be looked for. Understanding the microbial diversity in detail can also provide clue for

betterment of knowledge in the treatment of Silicosis as some of these bacteria do have some silica leaching ability. The population diversity revealed in the two hot springs are quite interesting inasmuch as it may help obtaining comprehensively comparative metagenomic analysis with other hot springs in different parts of the world.
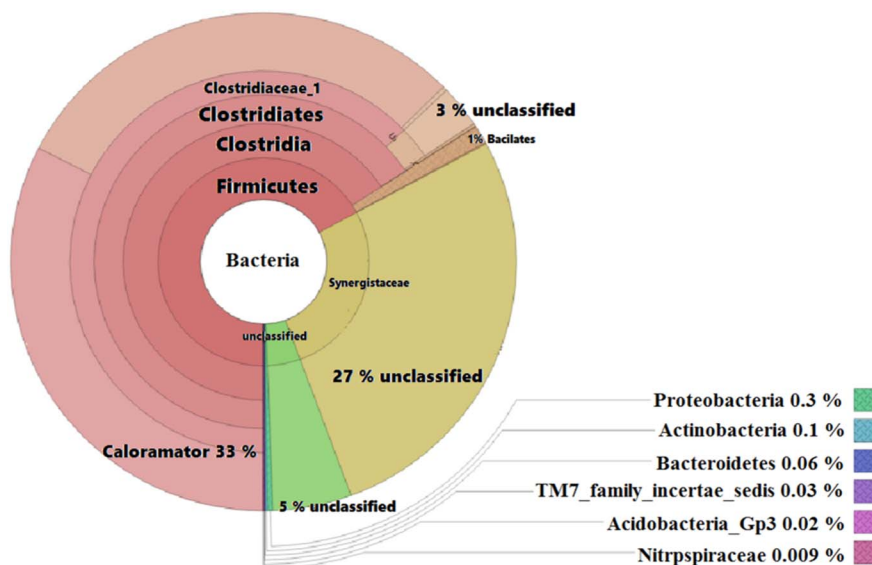
Fig. 5. A: Class distribution & abundance of B54 (hot spring 54 °C), distribution of class in R54, legend shows the different class in different representative colours.B: Same as Panel A but image generated from Krona. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
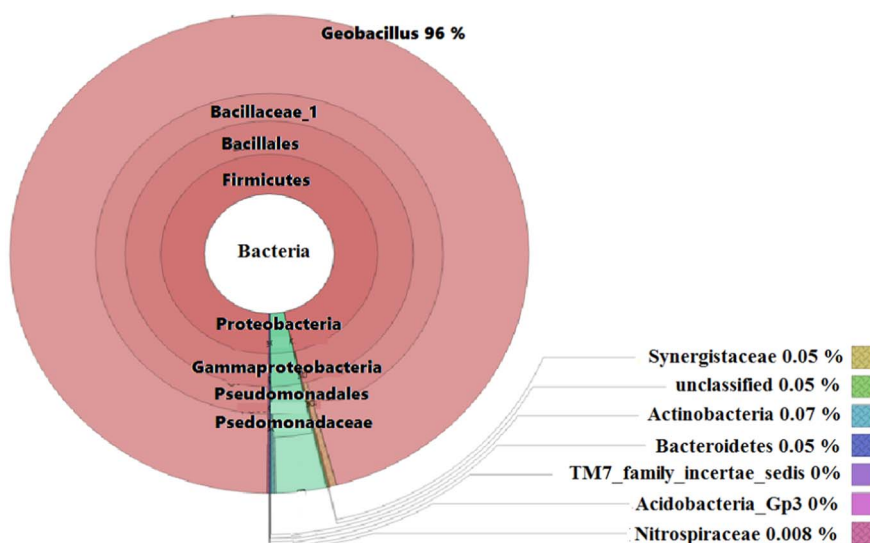
Fig. 6. A: Class distribution & abundance of B65 (hot spring 65 °C), distribution of class in R65, legend shows the different class in different representative colours. B: Same as Panel A but image generated from Krona. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.gdata.2017.04.001.

References

[1] R.K. Sinharay, S. Srivastava, B.B. Bhattacharya, Audiomagnetotelluric Studies to Trace the Hydrological System of Thermal Fluid Flow of Bakreswar Hot Spring, Eastern India: A Case History, Geophysics 75 (5) (2010) B187–B195.
[2] M. Debnath, N.C. Mandal, S. Ray, The study of cyanobacterial flora from geothermal springs of Bakreswar, West Bengal, India, Algae 24 (2009) 185–193.
[3] M.T. Madigan, B.L. Marrs, Extremophiles, Scientific American, 1997.
[4] R.K. Nagar, G. Vishwanathan, S. Sagar, S. Sankaranarayanan, U.L. Pitale, R.N.Ž. Padhi, Geological, geophysical and geochemical investigations in Bakreshwar-Tantloi thermal field, Birbhum and Santhal Parganas districts, West Bengal and Bihar, India, Geol. Surv. India — Spec. Pub, 45 1996, pp. 349–360.
[5] D. Ghose, D.P. Chowdhury, B. Sinha, Large-scale helium escape from earth surface around Bakreswar-Tantloi geothermal area in Birbhum district, West Bengal and Dumka district, Jharkhand, India, Curr. Sci. 82 (2002) 993–996.
[6] D.K. Mukhopadhyay, P.B. Sarolkar, Geochemical appraisal of Bakreshwar–Tantloi hot springs West Bengal and Jharkhand India, Proceedings 37th Workshop on Geothermal Reservoir Engineering, Stanford University, 2012, pp. 1–5.
[7] M. Biswas, S. Majumdar, T. Chowdhury, B.D. Chattopadhyay, S. Mandal, U. Halder, S. Yamasaki, Bioremediase a unique protein from a novel bacterium BKH1, ushering a new hope in concrete technology, Enzym. Microb. Technol. 46 (2010) 581–587.
[8] T. Chowdhury, M. Sarkar, B. Chaudhuri, B.D. Chattopadhyay, U. Halder, Participatory role of zinc in structural and functional characterization of bioremediase: a unique thermostable microbial protein, J. Biol. Inorg. Chem. 20 (2015) 791–803.
[9] S. Majumdar, M. Sarkar, T. Chowdhury, B.D. Chattopadhyay, S. Mandal, Use of bacterial protein powder in commercial fly ash pozzolana cements for high performance construction materials, Open J. Civ. Eng. 2 (2012) 218–228.
[10] M. Sarkar, T. Chowdhury, B.D. Chattopadhyay, R. Gachhui, S. Mandal, Autonomous bioremediation of a microbial protein (bioremediase) in pozzolana cementitious composite, J. Mater. Sci. 49 (2014) 4461–4468.
[11] M. Sarkar, N. Alam, B. Chaudhuri, B.D. Chattopadhyay, S. Mandal, Development of an improved E. coli bacterial strain for green and sustainable concrete technology, RSC Adv. 5 (2015) 32175–32182.

[12] B. Chaudhuri, N. Alam, M. Sarkar, T. Chowdhury, B.D. Chattopadhyay, Phylogenetic characterization of BKH3 bacterium isolated from a hot spring consortium of Bakreshwar (India) and its application, Adv. Microbiol. 6 (2016) 453–461.
[13] C. Riesenfeld, P. Schloss, J. Handelsman, Metagenomics: genomic analysis of microbial communities, Annu. Rev. Genet. 38 (2004) 525–552.
[14] N. Fierer, R. Jackson, The diversity and biogeography of soil bacterial communities, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 626–631.
[15] D. Rusch, A. Halpern, G. Sutton, K. Heidelberg, S. Williamson, S. Yooseph, D. Wu, J. Eisen, J. Hoffman, K. Remington, K. Beeson, B. Tran, H. Smith, H. Baden-Tillson, C. Stewart, J. Thorpe, J. Freeman, C. Andrews-Pfannkoch, J. Venter, K. Li, S. Kravitz, J. Heidelberg, T. Utterback, Y. Rogers, L. Falc'on, V. Souza, G. Bonilla-Rosso, L. Eguiarte, D. Karl, S. Sathyendranath, T. Platt, E. Bermingham, V. Gallardo, G. Tamayo-Castillo, M. Ferrari, R. Strausberg, K. Nealson, R. Friedman, M. Frazier, J. Venter, The sorcerer II global ocean sampling expedition: northwest Atlantic through eastern tropical Pacific, PLoS Biol. e77 (2007) 5.
[16] C. Simon, R. Daniel, Achievements and new knowledge unraveled by metagenomic approaches, Appl. Microbiol. Biotechnol. 85 (2011) 265–276.
[17] P. Menzel, S.R. Gudbergsd'ottir, A.G. Rike, L. Lin, Q. Zhang, P. Contursi, M. Moracci, J.K. Kristjansson, B. Bolduc, S. Gavrilov, N. Ravin, A. Mardanov, E. Bonch-Osmolovskaya, M. Young, A. Krogh, X. XuPeng, Comparative metagenomics of eight geographically remote terrestrial hot springs, Microb. Ecol. 70 (2015) 411–424.
[18] D. Ghosh, B. Bal, V.K. Kashyap, S. Pal, Molecular phylogenetic exploration of bacterial diversity in a Bakreshwar (India) hot spring and culture of Shewanella-related thermophiles, Appl. Environ. Microbiol. 69 (2003) 4332–4336.
[19] F.M. Ausubel, R. Brent, R.E. Kingston, D.D. Moore, J.G. Seidman, J.A. Smith, K. Struhl, Curr. Protocols in Mol. Biol, Wiley, New York, 1994.
[20] K. Mukherjee, P. Tribedi, A. Chowdhury, T. Ray, A. Joardar, S. Giri, A.K. Sil, Isolation of a Pseudomonas aeruginosa strain from soil that can degrade polyurethane diol, Biodegradation 22 (2011) 377–388.
[21] J.R. Cole, Q. Wang, E. Cardenas, J. Fish, B. Chai, The ribosomal database project: improved alignments and new tools for rRNA analysis, Nucleic Acids Res. 37 (2009) 141–145.
[22] W. Souza, B. Carvalho, Rqc: quality control tool for high-throughput sequencing data.", Bioconductor (2017) [Online]. Available: http://bioconductor.org/packages/Rqc/. (Accessed: 09-May-2016).
[23] A. Simon, FastQC: a quality control tool for high throughput sequence data, Reference Source, 2010 www.bioinformatics.babraham.ac.uk.
[24] P.D. Schloss, Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities, Appl. Environ. Microbiol. 75 (2009) 7537–7541.
[25] R.C. Edgar, B.J. Haas, J.C. Clemente, C. Quince, R. Knight, UCHIME improves sensitivity and speed of chimera detection, Bioinformatics 10 (1093) (2011).
[26] P.H. Degnan, H. Ochman, Illumina-based analysis of microbial community diversity, ISME J. 6 (2012) 183–194.
[27] P.J. McMurdie, S. Holmes, phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data, PLoS ONE 8 (4) (2013) e61217 http://dx.plos.org/10.1371/journal.pone.0061217.
[28] R. C. Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2016 URL https://www.R-project.org/.
[29] B.D. Ondov, N.H. Bergman, A.M. Phillippy, Interactive metagenomic visualization in a web browser, BMC Bioinf. 12 (1) (2011 Sep 30) 385.
[30] A. Lewin, A. Wentzel, S. Valla, Metagenomics of microbial life in extreme temperature environments, Curr. Opin. Biotechnol. 10 (1016) (2012).

[31] S. Miller, A. Strong, K. Jones, M. Ungerer, Bar-coded pyrosequencing reveals shared bacterial community properties along the temperature gradients of two alkaline hot springs in Yellowstone national park, Appl. Environ. Microbiol. 75 (2009) 4565–4572.

[32] A. Valverde, M. Tuffin, D. Cowan, Biogeography of bacterial communities in hot springs: a focus on the actinobacteria, Extremophiles 16 (2012) 669–679.

[33] E. Mathur, A. Ortmann, M. Bateson, G. Geesey, M. Frazier, Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function, PLoS One 5 (2010) 10.1371.

[34] W. Inskeep, D. Rusch, Z. Jay, M. Herrgard, M. Kozubal, T. Richardson, R. Macur, N. Hamamura, R. Jennings, B. Fouke, A. Reysenbach, F. Roberto, M. Young, A. Schwartz, E. Boyd, J. Badger, Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function, PLoS One 9773 (5) (2010) (10.1371).

[35] C. Cambillau, J. Claverie, Structural and genomic correlates of hyperthermostability, J. Biol. Chem. 275 (2000) 383–386.

[36] K. Suhre, J. Claverie, Genomic correlates of hyperthermostability, an update, J. Biol. Chem. 278 (2003) 198–202.

[37] G. Saelensminde, J. Halskau, R. Helland, N. Willassen, I. Jonassen, Structure-dependent relationships between growth temperature of prokaryotes and the amino acid frequency in their proteins, Extremophiles 11 (2007) 585–596.

[38] G. Saelensminde, J. Halskau, I. Jonassen, Amino acid contacts in proteins adapted to different temperatures: hydrophobic interactions and surface charges play a key role, Extremophiles 13 (2009) 11–20.

[39] D. Kreil, C. Ouzounis, Identification of thermophilic species by the amino acid compositions deduced from their genomes, Nucleic Acids Res. 29 (2001) 1608–1615.