

Taxonomic and Functional Metagenomics Profiling of Tuwa and Unnai Hot Springs Microbial Communities

Disha Vora

Gujarat Biotechnology Research Centre

Satyamitra Shekh

Gujarat Biotechnology Research Centre

Madhvi Joshi

Gujarat Biotechnology Research Centre

Amrutlal Patel

Gujarat Biotechnology Research Centre

Chaitanya Joshi (✉ director@gbrc.res.in)

Gujarat Biotechnology Research Centre

Research Article

Keywords: Hot springs, Metagenomics, Next generation sequencing, KEGG, CAZy

Posted Date: November 10th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-1002337/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Hot springs are of great importance due to their unique physicochemical properties. Due to unique selection pressure in this habitat, a diverse microbial community is prevailing and can be analyzed by high throughput sequencing technology. Present study focuses on metagenomic sequencing of two hot springs from Gujarat, India namely Tuwa and Unnai through both, culturable and culture independent approach. Sequence analysis from both the water reservoirs depicted higher species richness and diversity based on various diversity indices. The microbial community structure at both the hot springs was distinct and dependent on physicochemical factors like temperature, pH, mineral content etc. Enrichment by cultivation before metagenome sequencing revealed the abundance of *Firmicutes* (up to 96%) representing cultivable organisms in hot springs. The bacterial phyla *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Thermotogae*, *Deinococcus-Thermus*, and *Chloroflexi* dominate the thermoalkaline spring at Unnai and Tuwa in different proportion. Economically important microorganisms belonging to genera *Thermus*, *Brevibacillus*, *Anoxybacillus*, *Bacillus*, *Pseudomonas*, and *Geobacillus* were prevalent in hot springs. The analysis of functional potential by KEGG revealed pathways for metabolism of carbohydrates, amino acids, vitamins, cofactors and xenobiotics. Annotation with Carbohydrate Active EnZymes (CAZy) revealed the presence of four major classes of enzymes: glycosyl transferase, glycoside hydrolase, polysaccharide lyase and carbohydrate-binding modules. The study provides insight into the microbial community structure and their untapped functional potential for various biotechnological and environmental applications.

Introduction

Extremophiles are organisms known to survive in environments with extremities of temperature, salinity, pH, pressure and radiation. Diverse thermophilic microorganisms specifically inhabit extreme environments such as hot springs, deep-sea hydrothermal vents, and volcanic island sediments [1]. Thermophilic microorganisms survive in hot environments with the help of thermostable enzymes that remain catalytically active at a higher temperature. Studying thermophiles provides a closer insight into the origin of life and evolutionary mechanisms as scientists believe that life has evolved at high temperatures during the prebiotic stage on the Earth [2]. Metagenomic analysis for the presence of microbial communities in several hot springs and their potential to produce useful metabolites has revealed the applications of thermophiles for industrial and biotechnological applications [3]. Geothermal hot springs are well documented for therapeutic significance and the chemical composition depends on the physicochemical properties of surrounding rock material. The microbial community structure of the hot springs is mainly shaped by the geochemical characteristics of the water reservoirs such as pH, temperature, ecological interactions, redox potential, dissolved minerals, and other chemicals [4].

Advancement in sequencing technology for the metagenomic profiling has served three major purposes: understanding the microbial community structural profiling, harnessing the functional potential of the existing microbiota, and discovery of novel natural biomolecules without culturing them. Several culture-independent studies have been done to explore microbial diversity and their functional potential to produce various beneficial metabolites from hot springs. The hot springs mainly harbor Gammaproteobacteria and Alphaproteobacteria including *Geobacillus*, *Anoxybacillus* [5], *Bacillus* [6], *Hydrogenobacter*, *Thermus* [7] *Clostridium* [8], *Opitutusterrae*, *Rhodococcus*, *Cellovibrio* [9], *Exiguobacterium*, *Pseudomonas*, *Paenisporosarcina*, *Acinetobacter*, *Stenotrophomonas*, *Thermanaerotherix* and *Thermoanaerobaculum* as major bacterial genera. Many studies have also reported the presence of the genes for catalysts of biotechnological potential for the metabolism of carbohydrate, nitrogen, sulfur, methane, xenobiotics and antibiotic biosynthesis as well as resistance genes [10–12]. Microbiota residing in the sediments of various habitats harbors genes of interest for biodegradation,

bioremediation, and stress response [13–15]. Fewer studies reported the metagenomics analysis of soils and sediments of hot water reservoirs which revealed the presence of novel genes for the biotechnological potential [16–17]. Hence it is essential to explore microbial community and their functional capabilities of still unexplored hot springs.

In the present work, we focused on analyzing the whole metagenome sequences derived from water and the soil of Tuwa and Unnai hot springs. This investigation aimed to study the microbial diversity of the hot springs and their functional profiling to create a metagenomic wealth of novel genes encoding enzymes of industrial potential. The study also focused on detecting the presence of important cultivable organisms and comparison for community structure between soil and water from Tuwa as well as between Tuwa and Unnai hot springs.

Materials And Methods

Site description and sampling

Two hot springs from Gujarat were selected for the current study namely Tuwa (22.799444°N and 73.4602778°E) and Unnai (20.8533333°N and 73.3344444°E). Both of the hot springs are less explored and situated in Panchmahal and Navsari District respectively in Gujarat, India. To understand the microbial community, physicochemical analysis of water samples from both the hot springs was carried out. Water samples and soil samples were collected in sterile 50 mL Falcon tubes following standard protocols to avoid direct exposure to the environment. Before sampling, the pH was recorded at the sampling site with a handheld pH meter (Whatman PHA 260) calibrated with warm buffer. Additionally, Tuwa water and soil samples were enriched in the medium containing 1 g L⁻¹ Yeast extract, 1.3 g L⁻¹ (NH₄)₂SO₄, 0.25 g L⁻¹ KH₂PO₄, 0.25 g L⁻¹ MgSO₄·7H₂O, 0.08 g L⁻¹ CaCl₂·2H₂O by culturing at 63°C for 24 h before metagenomics analysis.

Isolation of genomic DNA and NGS sequencing

Total four samples were processed for the sequencing: ES- enriched soil samples from Tuwa, EW- enriched water samples from Tuwa, Tuwa_S- soil samples from Tuwa, Unnai_S- soil samples from Unnai. The total DNA from soil, water and enriched samples was isolated using either DNeasy PowerWater or DNeasy PowerSoil Kits (QIAGEN). Libraries were prepared using Ion plus fragment library kit (Thermo Fisher Scientific) and the sequencing was performed using Ion 520 chip and 400-bp chemistry of Ion S5 plus system.

Bioinformatics analysis of metagenomics data

Raw sequences were imported into a pipeline where reads were filtered using FastQC (Galaxy version 0.72) and processed with PRINSEQ version 0.20.4 lite. All reads having read length shorter than 100 bp, mean quality ≤20, containing any ambiguous base or incorrect primer sequences were excluded from the dataset. The taxonomic profiling of the obtained reads was done using the SILVA database with an e-value <1e-5 and a minimum identity of 90%, and analyzed up to genus level. The SILVA output files were further analyzed using STAMP to visualize the taxonomic composition for each sample from phylum to genus level. Rarefaction curves and various diversity indices were measured using PAleontological STatistics (PAST) v4.03. Principal coordinates analysis (PCoA) was carried out to complement the output of the cluster analyses.

Metabolic potential of hot spring samples

The SqueezeMeta pipelines [18] were employed for assembly and functional analysis of reads obtained. This pipeline used the co-assembly mode using Megahit and reads from all the samples were pooled before the assembly. The functional diversity of the samples from Tuwa and Unnai hot springs was analyzed following Kyoto Encyclopaedia of Genes and Genomics (KEGG) analysis with e-value < 1e-5. Functional classification in all the samples was performed up to the level-2 of KEGG annotations. The protein sequences of the predicted open reading frames (ORF) obtained were aligned against Carbohydrate Active Enzyme database (CAZy) database using BLASTp in DIAMOND [19]. The contigs were mapped and screened at 1e-5 e-value threshold and gene prediction was strengthened further by applying the criteria of $\geq 90\%$ identity.

Results And Discussion

In the present work, comparison of microbial diversity between two hot springs in the Gujarat area was carried out to observe location variation. Both the sites harbor unique physicochemical factors, which have shaped the existing microbial community. The temperature of Unnai hot springs usually ranges from 54 to 56°C and for Tuwa 63 to 64°C. pH ranges from 7.9 to 8.4 for Unnai and 8.2 to 9 for Tuwa. The pH of the water is slightly alkaline and has influence on the biological and chemical reactions. The rise in pH may have resulted due to the contamination of the water reservoirs by organic matter as inferred from the high BOD value in previous report [20]. pH is one of the factors that has an impact on the microbial diversity [21]. The abundance of *Gammaproteobacteria* and *Betaproteobacteria* is directly correlated with the pH of the environment [22]. Higher temperature of the water reservoir also has a critical influence on the existence and abundance of microbial communities as well as their metabolic activities. Vast numbers of phyla dominate the habitats with a range of temperatures. In this study, *Proteobacteria*, *Deinococcus-Thermus*, *Firmicutes* and *Bacteroides* account as the pivotal phyla in both the hot springs. Similar observations have been reported earlier in the metagenomic analysis of the various hot springs in India having temperature ranging from 43 to 98°C [10, 17, 23-25].

Sequence analysis

To characterize the microbial community structure of the environmental samples DNA was extracted from two different hot springs. The DNA extracted from the enriched as well as the raw samples were used for further downstream processing. A total of 22,517,835 reads were generated by high throughput shotgun WGS of 4 samples. Maximum reads were obtained in Tuwa soil sample, while the lowest was obtained in enriched water sample from Tuwa. After trimming and filtering the reads of each sample, average read length of about 276 - 296 and approximately 98% high quality reads at $\geq Q20$ were generated. The high-quality reads were assembled in 491,980 contigs, with N50 score of 1065, and the mean length of contigs were 836. The results of sequence assembly statistics are shown in Table 1. NGS is a powerful tool to identify the potential genes or enzymes from the DNA samples obtained from environment [26]. Metagenomic sequencing of both the hot spring samples with and without enrichment step provided plenty of sequence data. Thus, analysis of this data provides the information on the community structure and their functional potential using different tools. Such studies facilitate our knowledge about the role of culturable as well as uncultured microbes in the environment.

Diversity and species richness analysis

Rarefaction analysis and other diversity indices like Chao, Simpson, Shannon etc. were studied for species diversity. An overview of the alpha diversity of the community in both the hot springs is displayed in Table 2. The rarefaction curve (Fig. 1a) depict that generated reads are insufficient to capture the microbial diversity as the curve did not reach plateau. This upward trend in the soil samples without enrichment indicated higher bacterial diversity. However, for

enriched sample the curve reached to plateau which is obvious as only very few of the organisms may have grown under the growth conditions. Shannon index at phylum level for enriched samples was 0.25 - 0.35 and for soil samples 2.06 - 2.6. For genera level the Shannon index ranged from 2.1 - 3.74 for all analyzed samples. Same pattern of diversity in terms of evenness was obtained for Simpson index at phylum and genera level. Analysis of Shannon and inverse Simpson index at the phylum and genus level represented a higher diversity in terms of richness and evenness in soil samples without enrichment. PCoA analysis was performed using STAMP to check the cluster analysis of existing bacterial community (Fig. 1b). The two dimensions of the plot accounted for 81.1 and 14.4% of the diversity variations. Thus, PCoA1 and PCoA2 jointly expresses 95% variation among four samples. PCoA of the each metagenome sample revealed distinct and non-overlapping community structure with high variation at genus level between two hot springs.

The taxonomic profiles of both the sites identified microbes belonging to a total of 67 phyla, 158 classes, 326 families, and 491 genera. The metagenomic samples were found to be dominated by 6 bacterial phyla- *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Thermotogae*, *Deinococcus-Thermus*, and *Chloroflexi* (Fig. 2). Phyla present in less abundance were represented as others and listed in supplementary file 1. The phylum *Proteobacteria* exhibited the highest relative abundance (23%) in the soil samples from Tuwa. Similar abundance of *Proteobacteria* in sediment and water was observed from Soldhar and Polok hot springs respectively in previous reports in India [24, 27]. Hot springs at Soldhar and Polok show similar alkaline pH range but the temperature was higher (76 and 95°C respectively) than hot spring at Tuwa. *Deinococcus-Thermus* was the most abundant phylum (45%) in the Unnai soil samples. This finding resembles with the hot spring in Spain where the alkalinity was similar but the temperature was >76°C [28]. The microbial community structure signposted the alkaline pH and higher temperature as the prime selection factors favoring the abundance of *Proteobacteria* and *Deinococcus-Thermus* in Tuwa and Unnai soil respectively. In the enriched soil and water samples from the Tuwa, *Firmicutes* (94-96%) was found to be the largest phylum and comprises a vast diversity of bacterial strains. From the heatmap it is very clear that phylum *Firmicutes* got enriched while *Deinococcus-Thermus*, *Proteobacteria*, *Bacteroidetes*, *Thermotogae* and *Chloroflexi*, though abundant in the sample, are not enriched. This observation implies that the medium used for enrichment may be suboptimal for the growth of the organisms. The enrichment of *Firmicutes* was significant and it denotes the presence of cultivable organisms in Tuwa hot spring. Metagenome analysis of water samples from both the sites are reported earlier [29, 30]. Diversity at these sites matches with earlier studies nonetheless, the proportion of abundant phyla and genera varied in soil samples. Some distinct phyla such as *Deinococcus-Thermus*, *Bacteroidetes*, *Chloroflexi* and *Thermotogae* were found to be present in soil and water samples from both the sites in comparison to earlier reports from these sites for water samples (Supplementary file 2). This observation supports that each hot spring harbors unique microbial community due to different physicochemical properties as well as gradient of temperature and light. Habitat of thermoalkaline spring also differs in the form of sediment, thermal fluids or microbial mats depending on the geological profile of the surrounding rock. All these factors play crucial role in shaping the existing microflora [31]. The obtained results about dominating phyla of the samples analyzed in this study also coincides with the earlier reports for hot springs in Indian regions and other parts of the world [10, 17, 23-25, 32-34]. Cellulose, chitin, hemicellulose and starch degrading bacterial genes have been reported in the metagenomes where phyla *Firmicutes*, *Proteobacteria*, *Deinococcus-Thermus*, *Bacteroidetes* etc dominate the environment [35]. This is the first report of metagenomic analysis of enriched soil and water samples from the Tuwa as well as soil from both the hot springs. The enrichment approach depicts the wide variety and abundance of culturable organisms present in the hot spring environment.

At the class level taxonomic profiling from enriched water and soil samples from Tuwa, *Bacilli* (78-82%), *Negativicutes* (7-9%), *Clostridia* (5-8%) were abundant and their dominance support the enrichment of culturable

group. *Flavobacteriia*, *Alphaproteobacteria*, *Gammaproteobacteria*, *Deinococci* and *Actinobacteria* were lesser than 2% (Fig. 3). For soil samples from Tuwa and Unnai (without enrichment) major abundant class were *Deinococci*, *Gammaproteobacteria*, *Bacilli*, *Clostridia*, *Flavobacteria*, *Deltaproteobacteria*, *Alphaproteobacteria*, *Negativicutes*, *Thermotogae*, *Anaerolineae*, and *Ignavibacteria*. The metagenomes contained diverse bacterial genera in comparison to earlier reports for water samples from both the hot springs [29, 30]. Analysis for bacterial genera prevailing in all four samples revealed the presence of total of 491 genera (Fig. 4a, 4b). Analysis of dominant genera in thermoalkaline spring revealed the presence of autotrophs, heterotrophs, chemoorganotrophs and chemolithotrophs; thermophilic, neutrophilic, and alkaliphilic organisms; as well as aerobes, microaerophilic, facultative anaerobes and obligate anaerobes. The most abundant chemorganotrophs and chemolithotrophs belonged to *Brevibacillus* (26-42.5%), *Staphylococcus* (13-15%), *Anaerosinus* (8-9%), *Anoxybacillus* (7-9%), *Streptococcus* (7-8%), and *Bacillus* (5-11%) which are carbohydrate decomposing cultivable genera significantly enriched in the water and soil samples from Tuwa [36]. The proportion of all the genera in samples with or without enrichment varied. The abundance of *Thermus* and *Brevibacillus* was significantly higher (up to 44%) in Unnai soil and enriched soil from Tuwa respectively. Other major genera in the soil samples were *Myroides* (4-6%), *Staphylococcus* (4-6%), *Fervidobacterium* (2-4%), and *Vibrio* (3-5%). Genus *Brevibacillus* was not detected in Tuwa soil but enriched in soil and water from Tuwa. In enriched soil samples, *Brevibacillus*, *Anaerosinus* and *Streptococcus* were abundant. Whereas in enriched water samples, *Staphylococcus*, *Anoxybacillus* and *Bacillus* were abundant. Human pathogens like *Staphylococcus* and *Streptococcus* are also present in the hot spring environment. Some important genera like *Bacillus*, *Brevibacillus*, *Anoxybacillus*, *Thermus*, *Geobacillus*, *Aeribacillus*, *Pseudomonas* etc. are identified as potential culturable organisms for the industrially important thermozyms for degradation of polysaccharides, proteins, lipids, xenobiotics and other components [37-39].

KEGG analysis

To reveal the functional potential of bacteria prevailing in the hot springs, the metagenomic data was analyzed using KEGG and a total of 20,321 genes/hits were annotated with KO IDs. The detailed analysis showed that KOs were assigned to pathways for metabolism (>4,000), genetic information processing, cellular processes, environmental information processing, organismal systems and Human diseases (Fig. 5a). Further analysis of the major pathways for metabolism at level-2 showed 18 sub-pathways (Fig. 5b). Pathways for carbohydrate metabolism, energy metabolism, metabolism of cofactors and vitamins, amino acid metabolism, xenobiotics degradation and metabolism and membrane transport were found to be dominant in the four metagenomes. The abundance of genes related to metabolic pathways may be responsible for survival of the organisms into the hot springs environment [10]. Soil samples of both the hot springs showed higher proportion of the reads assigned to different KO categories as compared to enriched samples. Soil samples represent total microbiome present in comparison to enriched samples where only cultivable group dominate the samples due to selective enrichment. Interestingly other KEGG subsystems were found to be associated with biosynthesis of glycan and other secondary metabolites, metabolism of terpenoids, and polyketides etc.

Mining for carbohydrate active enzymes

To have insight into carbohydrate related enzymes present in the metagenomes, predicted ORF reads obtained from the SqueezeMeta pipeline were matched against the CAZy database. The metagenome sequences were distributed among four major enzyme classes of CAZy database-glycosyl transferase (GT), glycoside hydrolase (GH), polysaccharide lyase (PL), and Carbohydrate-Binding Modules (CBMs) [25]. Proportion of GH family was higher (>80%) followed by GT, PL and CMBs (Fig. 6a). GH13 was the most abundant (33%) within the GH family. GH2, GH3,

GH57, and other families were comparatively less abundant than GH13 (Fig. 6b). The CAZy enzymes are reported to be involved in the catabolism of polysaccharides. GH and GT enzymes have an important role in energy metabolism, communication and structural support in the organisms [40]. GHs are involved in the catalysis of carbohydrates at *O*-glycosidic bond and efficiently hydrolyse cellulose, chitin, other abundant carbohydrates in the habitat etc. Amongst GH enzymes whole gene sequences for debranching enzymes (GH13, GH16, GH23, GH51, GH67, GH 94, GH97), hemicellulose degrading (GH2, GH3, GH4, GH5, GH10, GH26, GH31, GH35, GH38, GH43, GH53), and cellulases (GH3, GH5, GH9) are present (Fig. 6c). GT are responsible for the glycosylation of amino acids during cellular processes [41]

Conclusion

The present study provide insight into the microbial composition of soil and enriched samples from hot springs at Tuwa and Unnai, Gujarat. Metagenomic analysis for taxonomic profiling of the hot springs depicts great microbial diversity in terms of alpha and beta diversity at the two hot water reservoirs. Due to unique environmental factors (pH and temperature) at the hot springs, the metagenomes show distinct community structure. *Deinococcus-Thermus* (45%) and *Proteobacteria* (23%) were the major abundant phyla at Unnai and Tuwa hot spring respectively. Enrichment dependent approach revealed presence and abundance of cultivable organisms from the phylum *Firmicutes* ($\geq 94\%$) in soil and water samples from Tuwa. In this study, some distinct cultivable and non cultivable organisms from *Firmicutes*, *Bacteroidetes*, *Proteobacteria*, *Deinococcus-Thermus*, *Thermotogae* and *Chloroflexi* were reported for the first time, when compared to earlier reports of water metagenomes of both the sites. The metagenomic analysis in comparison with public databases KEGG and CAZy was performed for the functional profiling. More than 20,000 reads were assigned to various pathways for genetic information processing, survival and human diseases. Organisms present in both the hot springs show potential for the production of important enzymes involved in complex carbohydrate metabolism. GHs were highly abundant in the metagenomes and contributed by various genera reported to survive at higher temperature and other extremities. These genera can be the important source of heat tolerant enzymes and secondary metabolites. This study provides an insight into unexplored metabolic potential of two hot springs in Gujarat that can be utilized for industrial application.

Declarations

Conflict of interest

Authors declare no conflict of interest

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and accession number

The datasets generated and/or analyzed during the current study are available in the NCBI repository [<https://www.ncbi.nlm.nih.gov/Traces/study/?acc=PRJNA755441>; Accession Number: PRJNA755441].

Competing Interest

Authors declare no competing interests

Funding

The research work is supported by Gujarat State Biotechnology Mission, DST (Grant number: GBRC/GoG/DST/JD-2/BS14/2017).

Authors' Contribution

D. Vora and S. Shekh contributed equally to this work as first authors for data analysis and manuscript writing. Dr. M. Joshi, Dr. A. Patel, and Prof. C. Joshi contributed to Data analysis, Interpretation, editing manuscript and overall supervision.

Acknowledgement

The authors would like to acknowledge the Gujarat State Biotechnology Mission (GSBTM) for the financial support.

References

1. Mirete, S., Morgante, V., & González-Pastor, J. E. (2016). Functional metagenomics of extreme environments. *Current opinion in biotechnology*, *38*, 143-149.
2. Burgess, E. A., Wagner, I. D., & Wiegel, J. (2007). Thermal environments and biodiversity. *Physiology and biochemistry of extremophiles*, 11-29.
3. Strazzulli, A., Iacono, R., Giglio, R., Moracci, M., & Cobucci-Ponzano, B. (2017). Metagenomics of hyperthermophilic environments: Biodiversity and biotechnology. In *Microbial Ecology of Extreme Environments* (pp. 103-135). Springer, Cham.
4. Siering, P. L., Clarke, J. M., & Wilson, M. S. (2006). Geochemical and biological diversity of acidic, hot springs in Lassen Volcanic National Park. *Geomicrobiology Journal*, *23*(2), 129-141.
5. Hussein, E. I., Jacob, J. H., Shakhathreh, M. A. K., Abd Al-razaq, M. A., Juhmani, A. S. F., & Cornelison, C. T. (2017). Exploring the microbial diversity in Jordanian hot springs by comparative metagenomic analysis. *MicrobiologyOpen*, *6*(6), e00521.
6. Mangrola, A., Dudhagara, P., Koringa, P., Joshi, C. G., Parmar, M., & Patel, R. (2015). Deciphering the microbiota of Tuwa hot spring, India using shotgun metagenomic sequencing approach. *Genomics data*, *4*, 153-155.
7. Bhatia, S., Batra, N., Pathak, A., Green, S. J., Joshi, A., & Chauhan, A. (2015). Metagenomic evaluation of bacterial and archaeal diversity in the geothermal hot springs of Manikaran, India. *Genome announcements*, *3*(1), e01544-14.
8. Ghelani, A., Patel, R., Mangrola, A., & Dudhagara, P. (2015). Cultivation-independent comprehensive survey of bacterial diversity in Tulsi Shyam Hot Springs, India. *Genomics Data*, *4*, 54-56.
9. Mehete, G. T., Paranjpe, A. S., Dastager, S. G., & Dharme, M. S. (2016). Complete metagenome sequencing based bacterial diversity and functional insights from basaltic hot spring of Unkeshwar, Maharashtra, India. *Genomics data*, *7*, 140-143.
10. Sharma, N., Kumar, J., Abedin, M. M., Sahoo, D., Pandey, A., Rai, A. K., & Singh, S. P. (2020). Metagenomics revealing molecular profiling of community structure and metabolic pathways in natural hot springs of the

Sikkim Himalaya. *BMC microbiology*, 20(1), 1-17.

11. Ferrandi, E. E., Previdi, A., Bassanini, I., Riva, S., Peng, X., & Monti, D. (2017). Novel thermostable amine transferases from hot spring metagenomes. *Applied microbiology and biotechnology*, 101(12), 4963-4979.
12. Littlechild, J. A. (2015). Enzymes from extreme environments and their industrial applications. *Frontiers in bioengineering and biotechnology*, 3, 161.
13. Pandit, A. S., Joshi, M. N., Bhargava, P., Shaikh, I., Ayachit, G. N., Raj, S. R., ... & Bagatharia, S. B. (2015). A snapshot of microbial communities from the Kutch: one of the largest salt deserts in the World. *Extremophiles*, 19(5), 973-987.
14. Kirchman, D. L., Hanson, T. E., Cottrell, M. T., & Hamdan, L. J. (2014). Metagenomic analysis of organic matter degradation in methane-rich Arctic Ocean sediments. *Limnology and oceanography*, 59(2), 548-559.
15. Behera, B. K., Chakraborty, H. J., Patra, B., Rout, A. K., Dehury, B., Das, B. K., ... & Mohapatra, T. (2020). Metagenomic analysis reveals bacterial and fungal diversity and their bioremediation potential from sediments of river Ganga and Yamuna in India. *Frontiers in microbiology*, 11, 2531.
16. Ranjan, R., Yadav, M. K., Suneja, G., & Sharma, R. (2018). Discovery of a diverse set of esterases from hot spring microbial mat and sea sediment metagenomes. *International journal of biological macromolecules*, 119, 572-581.
17. Najar, I. N., Sherpa, M. T., Das, S., Das, S., & Thakur, N. (2020). Diversity analysis and metagenomic insights into antibiotic and metal resistance among Himalayan hot spring bacteriobiome insinuating inherent environmental baseline levels of antibiotic and metal tolerance. *Journal of global antimicrobial resistance*, 21, 342-352.
18. Tamames, Javier, and Fernando Puente-Sánchez. "SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline." *Frontiers in microbiology* 9 (2019): 3349.
19. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., & Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic acids research*, 37(suppl_1), D233-D238.
20. Sodhi, K. K., Kumar, M., & Singh, D. K. (2021). Assessing the bacterial diversity and functional profiles of the River Yamuna using Illumina MiSeq sequencing. *Archives of Microbiology*, 203(1), 367-375.
21. Zhalnina, K., Dias, R., de Quadros, P. D., Davis-Richardson, A., Camargo, F. A., Clark, I. M., ... & Triplett, E. W. (2015). Soil pH determines microbial diversity and composition in the park grass experiment. *Microbial ecology*, 69(2), 395-406.
22. Zhou, J., Deng, Y., Shen, L., Wen, C., Yan, Q., Ning, D., ... & Brown, J. H. (2016). Temperature mediates continental-scale diversity of microbes in forest soils. *Nature communications*, 7(1), 1-10.
23. Chaudhuri, B., Chowdhury, T., & Chattopadhyay, B. (2017). Comparative analysis of microbial diversity in two hot springs of Bakreshwar, West Bengal, India. *Genomics data*, 12, 122-129.
24. Najar, I. N., Sherpa, M. T., Das, S., Das, S., & Thakur, N. (2018). Microbial ecology of two hot springs of Sikkim: predominate population and geochemistry. *Science of the Total Environment*, 637, 730-745.
25. Saxena, R., Dhakan, D. B., Mittal, P., Waiker, P., Chowdhury, A., Ghatak, A., & Sharma, V. K. (2017). Metagenomic analysis of hot springs in Central India reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Frontiers in microbiology*, 7, 2123.
26. Kimura, N. (2018). Novel biological resources screened from uncultured bacteria by a metagenomic method. In *Metagenomics* (pp. 273-288). Academic Press.

27. Sharma, A., Paul, D., Dhotre, D., Jani, K., Pandey, A., & Shouche, Y. S. (2017). Deep sequencing analysis of bacterial community structure of Soldhar hot spring, India. *Microbiology*, 86(1), 136-142.
28. López-López, O., Knapik, K., Cerdán, M. E., & González-Siso, M. I. (2015). Metagenomics of an alkaline hot spring in Galicia (Spain): microbial diversity analysis and screening for novel lipolytic enzymes. *Frontiers in microbiology*, 6, 1291.
29. Mangrola, A., Dudhagara, P., Koringa, P., Joshi, C. G., Parmar, M., & Patel, R. (2015). Deciphering the microbiota of Tuwa hot spring, India using shotgun metagenomic sequencing approach. *Genomics data*, 4, 153-155.
30. Mangrola, A., Dudhagara, P., Koringa, P., Joshi, C., & Patel, R. (2018). Metagenomic microbial community profiling of Unnai hot spring by Ion-Torrent based shotgun sequencing. *Microbiology (00262617)*, 87(1).
31. López-López, O., Cerdán, M. E., & González-Siso, M. I. (2013). Hot spring metagenomics. *Life*, 3(2), 308-320.
32. Iacono, R., Cobucci-Ponzano, B., De Lise, F., Curci, N., Maurelli, L., Moracci, M., & Strazzulli, A. (2020). Spatial metagenomics of three geothermal sites in Pisciarelli hot Spring focusing on the biochemical resources of the microbial consortia. *Molecules*, 25(17), 4023.
33. Rozanov, A. S., Bryanskaya, A. V., Ivanisenko, T. V., Malup, T. K., & Peltek, S. E. (2017). Biodiversity of the microbial mat of the Garga hot spring. *BMC evolutionary biology*, 17(2), 37-49.
34. Hussein, E. I., Jacob, J. H., Shakhathreh, M. A. K., Abd Al-razaq, M. A., Juhmani, A. S. F., & Cornelison, C. T. (2017). Exploring the microbial diversity in Jordanian hot springs by comparative metagenomic analysis. *MicrobiologyOpen*, 6(6), e00521.
35. Badhai, J., Ghosh, T. S., & Das, S. K. (2015). Taxonomic and functional characteristics of microbial communities and their correlation with physicochemical properties of four geothermal springs in Odisha, India. *Frontiers in microbiology*, 6, 1166.
36. Lewin, A., Wentzel, A., & Valla, S. (2013). Metagenomics of microbial life in extreme temperature environments. *Current opinion in biotechnology*, 24(3), 516-525.
37. Kambourova, M. (2018). Thermostable enzymes and polysaccharides produced by thermophilic bacteria isolated from Bulgarian hot springs. *Engineering in life sciences*, 18(11), 758-767.
38. González-Siso, M. I. (2019). Editorial for the special issue: thermophiles and thermozymes.
39. Pinzon-Martinez, D. L., Rodríguez-Gómez, C., Miñana-Galbis, D., Carrillo-Chávez, J. A., Valerio-Alfaro, G., & Oliart-Ros, R. (2010). Thermophilic bacteria from Mexican thermal environments: isolation and potential applications. *Environmental Technology*, 31(8-9), 957-966.
40. Chao, L., & Jongkees, S. (2019). High-Throughput Approaches in Carbohydrate-Active Enzymology: Glycosidase and Glycosyl Transferase Inhibitors, Evolution, and Discovery. *Angewandte Chemie*, 131(37), 12880-12890.
41. Singh, K. M., Reddy, B., Patel, D., Patel, A. K., Parmar, N., Patel, A., ... & Joshi, C. G. (2014). High potential source for biomass degradation enzyme discovery and environmental aspects revealed through metagenomics of Indian buffalo rumen. *BioMed research international*, 2014.

Tables

Table 1
Statistical summary of sequences analyzed

Sample ID	SRA ID	BEFORE PRINSEQ-LITE		AFTER PRINSEQ-LITE			
		INPUT SEQUENCES	INPUT MEAN LENGTH	GOOD SEQUENCES	GOOD MEAN LENGTH	BAD SEQUENCES	BAD MEAN LENGTH
ES	SAMN20822856 SAMN20822857	2650320	301.05	2595392	296.26	54928.5	108.9
EW	SAMN20822858	2057771	285.7	2018185	280.8	39586	107.39
Tuwa_S	SAMN20822859 SAMN20822860	3480148	299.03	3403106	293.59	77042.5	109.56
Unnai_S	SAMN20822861 SAMN20822862	3070679	281.53	2998875	276.33	71804	108.49

Table 2
Various diversity metrics for bacterial community in the hot spring samples at Phylum and Genus level

PHYLUM	ES	EW	Tuwa_S	Unai_S
Dominance_D	0.919	0.878	0.125	0.245
Simpson_1-D	0.080	0.121	0.875	0.754
Shannon_H	0.249	0.344	2.598	2.055
Menhinick	0.123	0.182	0.884	0.649
Margalef	1.705	2.275	7.089	6.173
Chao-1	18	23	68.5	62

GENUS	ES	EW	Tuwa_S	Unai_S
Dominance_D	0.222	0.124	0.049	0.216
Simpson_1-D	0.777	0.875	0.950	0.783
Shannon_H	2.091	2.505	3.736	2.665
Menhinick	0.471	0.524	4.728	3.574
Margalef	6.819	6.723	38.47	34.5
Chao-1	70.5	67	385.9	339.5

Figures

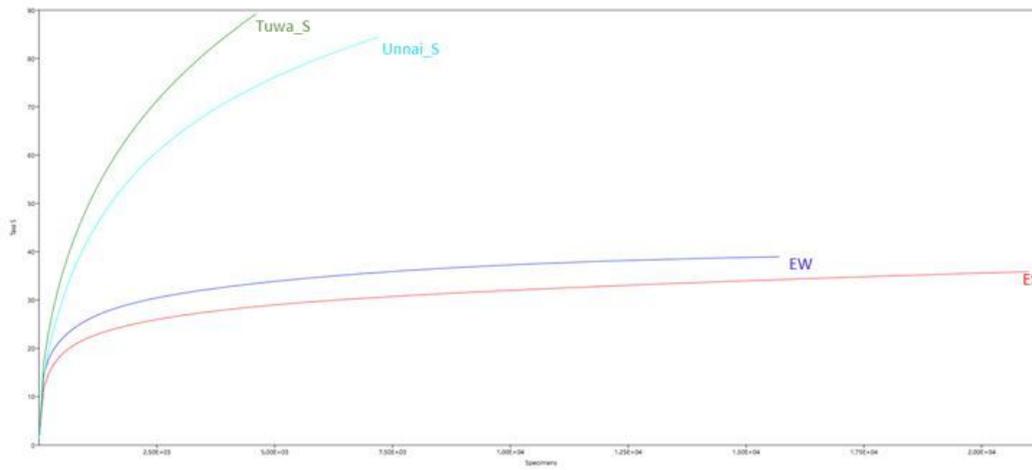


Figure 1a

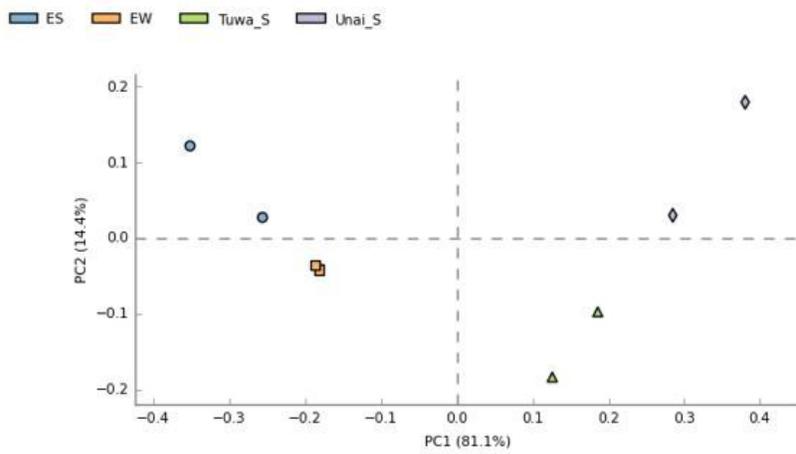


Figure 1b

Figure 1

a) Rarefaction analysis and b) PCoA of observed reads for sample EW, ES, Tuwa_S and Unnai_S

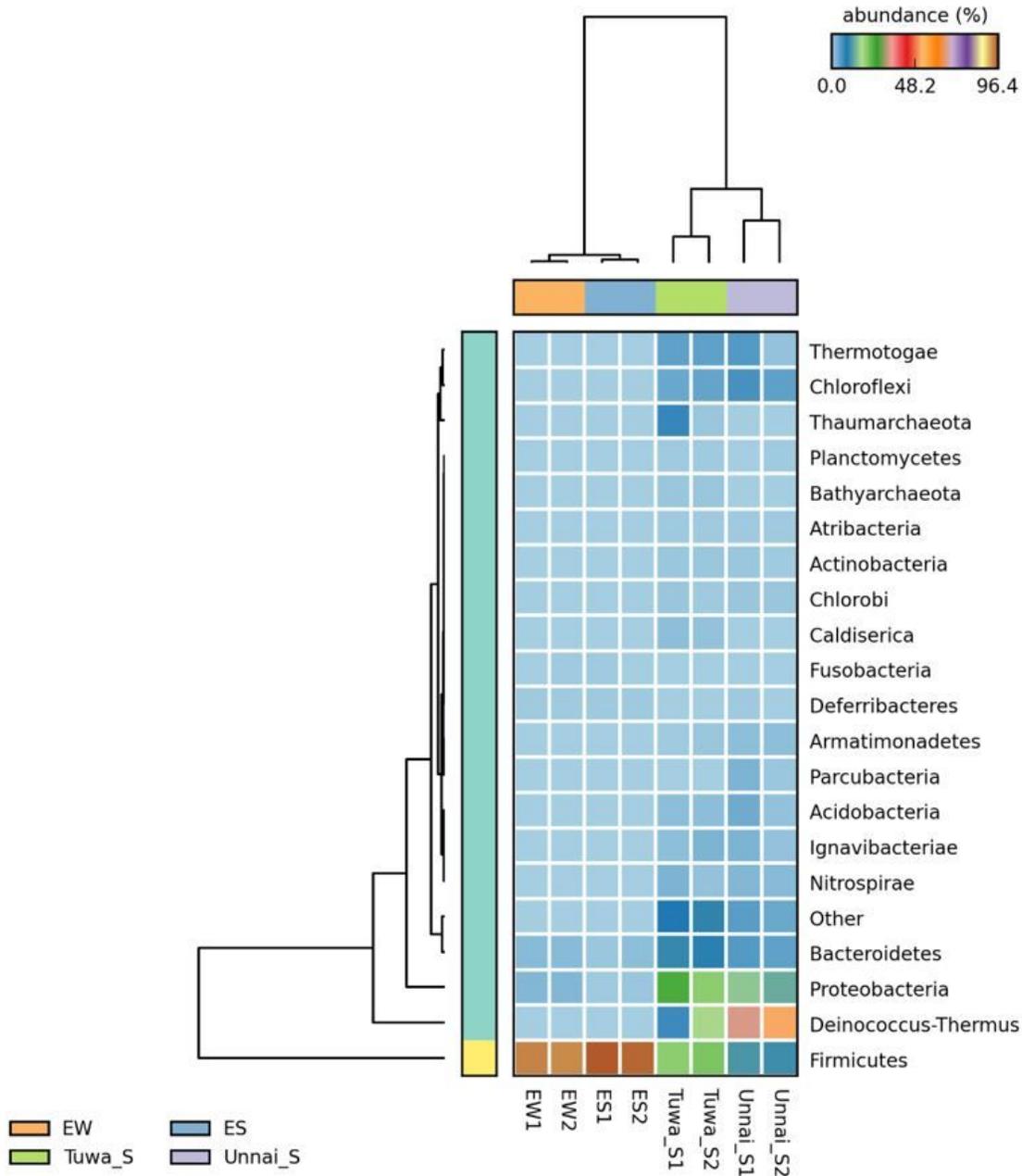


Figure 2

Heat map representing the taxonomic distribution of bacterial phyla in the soil and water samples from Tuwa and Unnai hot springs

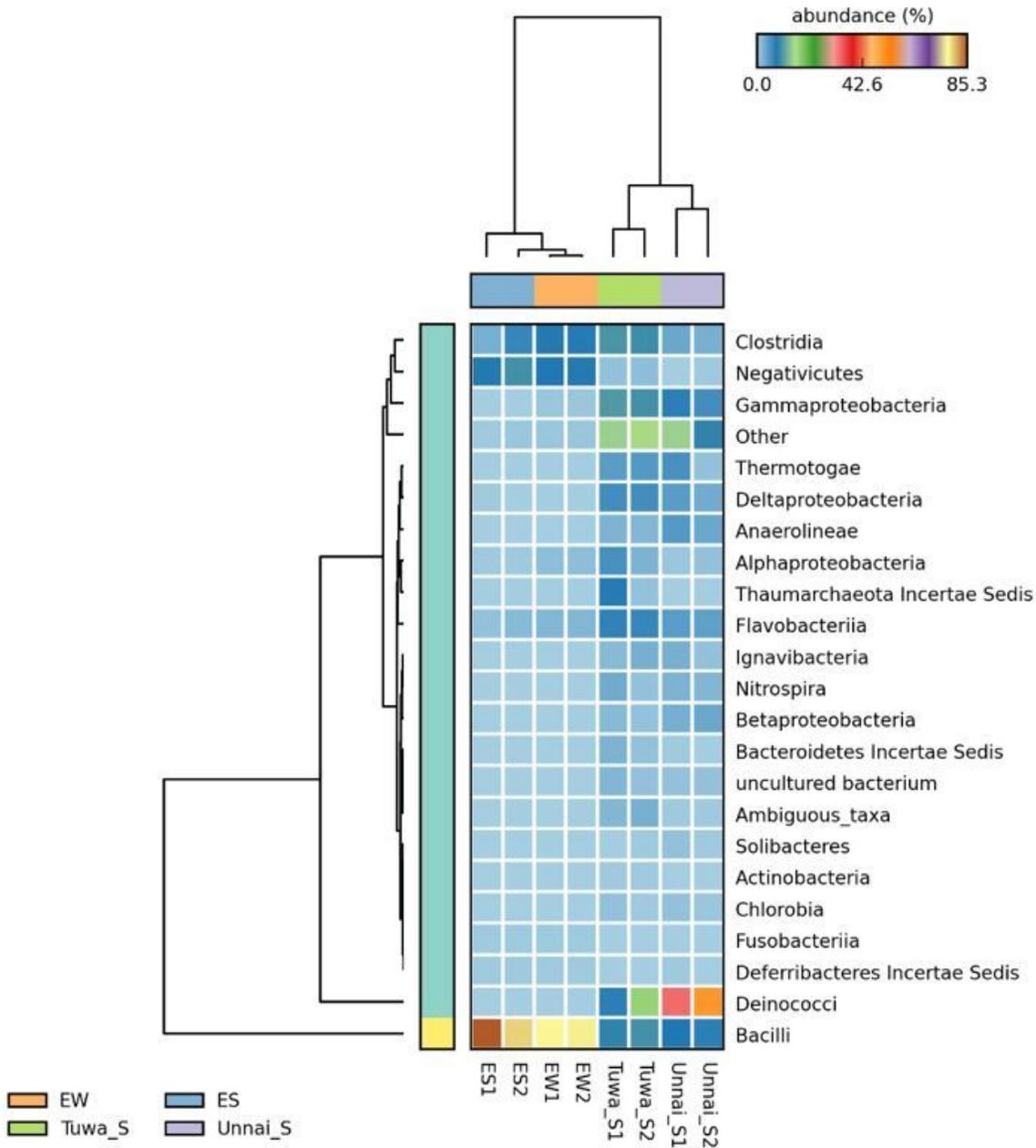


Figure 3

Heat map representing the microbial community profiling of all four samples at class level. Only dominant bacterial families present at relative abundance >1% were indicated in the graph.

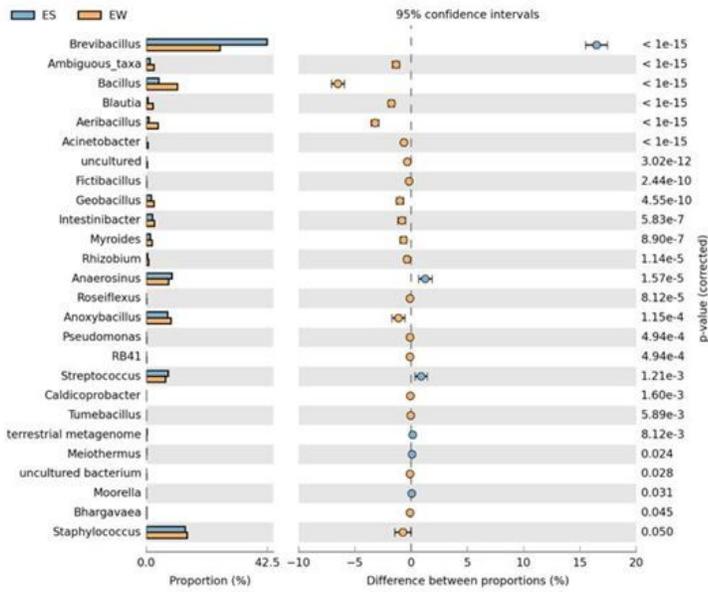


Figure 4a

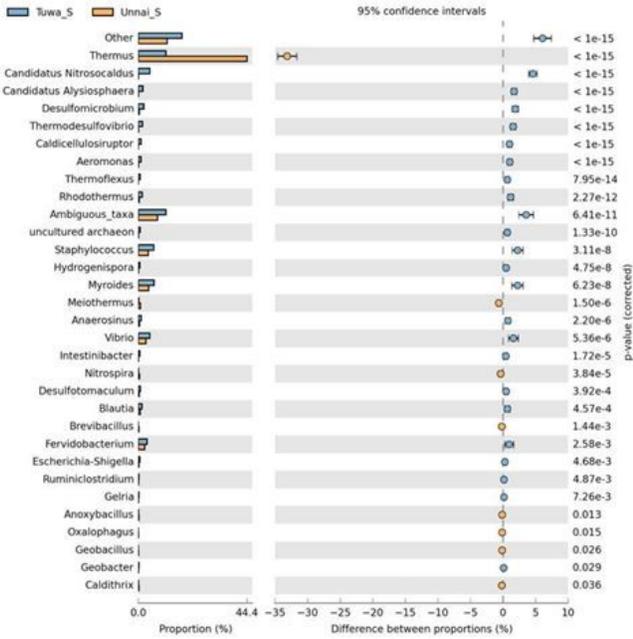


Figure 4b

Figure 4

Extended error bars displaying the mean proportion of bacterial genera present in a) enriched water and soil samples from Tuwa and b) soil samples from Tuwa and Unnai

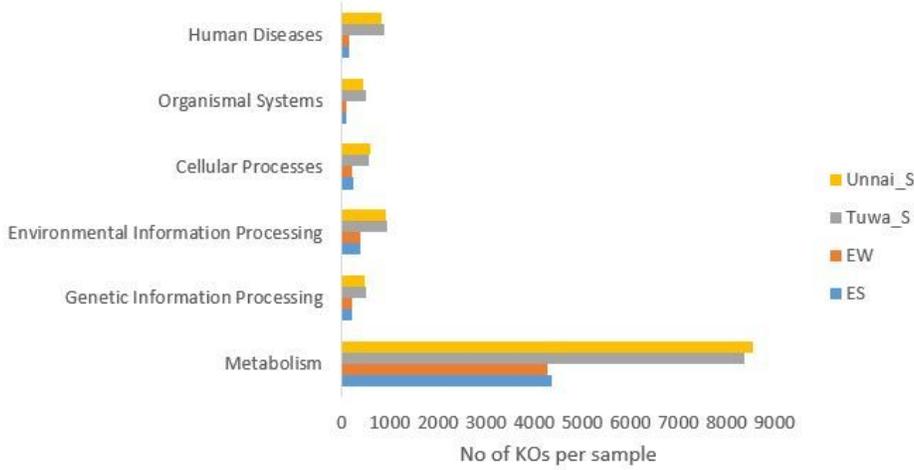


Figure 5a

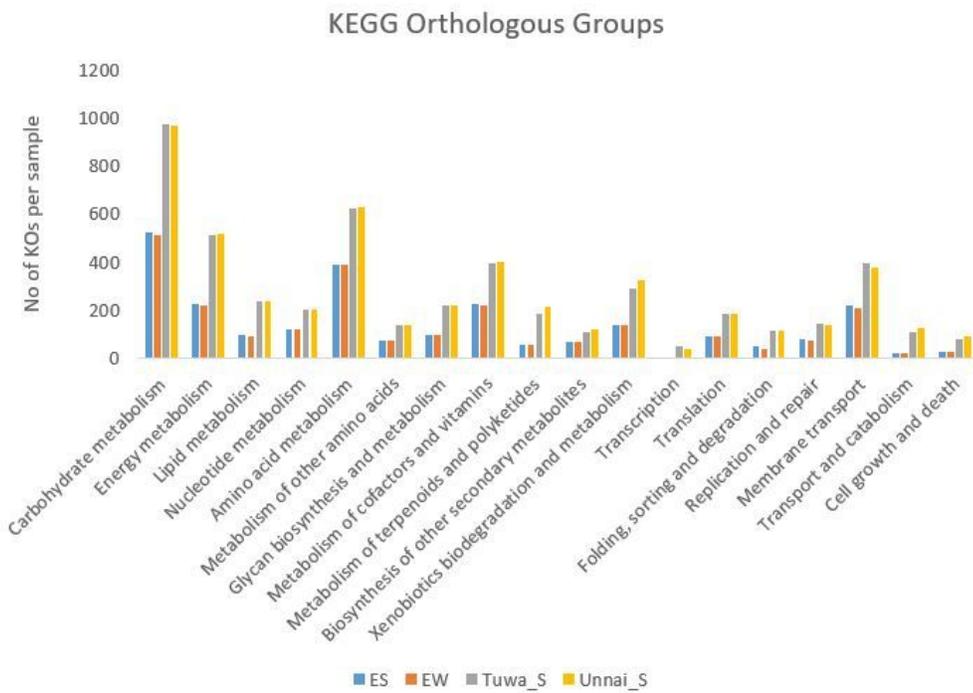


Figure 5b

Figure 5

Functional analysis showing relative abundance of a) KEGG pathways and b) subsystem based distribution for the both the hot spring sites

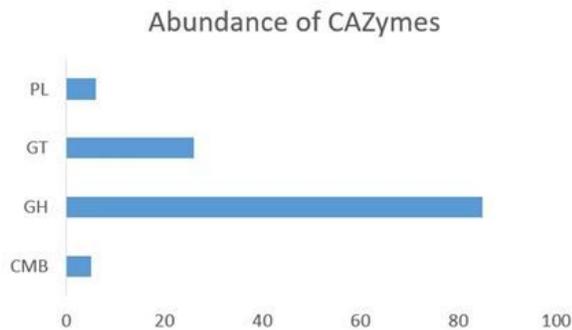


Figure 6a

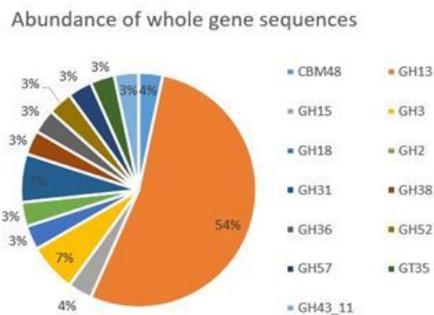


Figure 6b

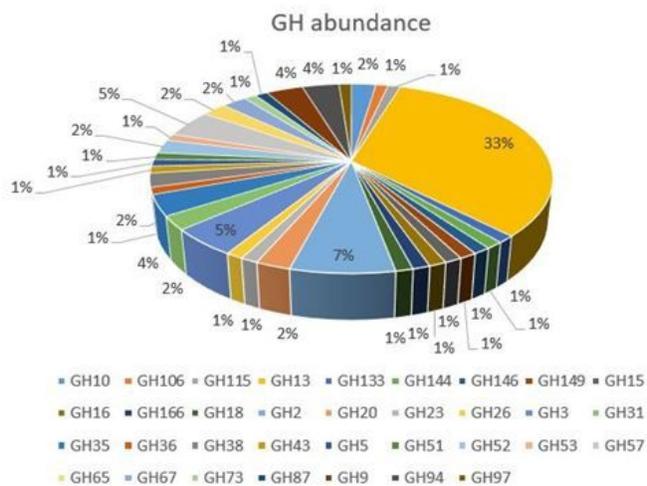


Figure 6c

Figure 6

a) Abundance of CAZy family enzymes (GT- Glycosyl Transferases; GH- Glycoside Hydrolases; CBM- Carbohydrate-Binding Modules; PL- Polysaccharide Lyases); b) whole gene sequences for three enzyme classes and c) glycoside hydrolase (GH) in the metagenomics data

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfile1.xlsx](#)

- [Supplementaryfile2.docx](#)