Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Investigating the Diversity and Potential Applications of the Microbiome at AlLith Province Hot Springs

Abdulaziz M Kusa

AKUSA.stu@uj.edu.sa

University of Jeddah

Mohammed N. Baeshen

University of Jeddah

Roba M.S Attar

University of Jeddah

Hulayyil Al-Harbi

University of Jeddah

Research Article

Keywords: hot spring, metagenomics, extremophiles, thermophiles, community analysis, functional analysis

Posted Date: February 9th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8801073/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background: Extreme ecosystems, such as hot springs, are of great interest as a source of novel extremophilic biodiversity, enzymes, survival metabolic functions, and biotechno-logical products. In hot springs, physicochemical factors such as temperature, pH, and mineral content influence microbial communities. In this study, we aimed to investigate the microbial diversity in the water, spring sediments, and wet-shore sediments of the Al-Lith hot spring in Saudi Arabia.

Methods: To assess the microbial diversity and identify potential applications, we conducted metagenomic analyses.

Results: The phyla cyanobacteria, *Chloroflexi*, and *protobacteria* were dominant in samples collected from water, spring sediment, and wet-shore sediments, respectively. In all three samples, *Thaumarcheotal*was the most important phylum in the Archaea domain, whereas **Mucoromycete** was the most important phylum in the eukaryotic domain. Functional potential analysis of the Kyoto Encyclopedia of Genes and Genomes database revealed that metabolic activi-ties were the most prominent, followed by genetic information, environmental information, cellular and human diseases, and organismal processes. Carbohydrate-active enzymes annotation showed that six main groups of enzymes were present: glycoside hydrolases, glycosyl transferases, carbohydrate-binding modules, carbohydrate esterases, polysaccharide lyases, and auxiliary activities. Functional classification of eggNOG re-vealed unknown functions, followed by amino acid transport and metabolism, energy production, replication, recombination, repair, and conversion. Conclusion: The microbial diversity in the Al-Lith hot spring includes a variety of microorganism species that can potentially be used in biotechnology as a novel source of untapped functional potential for various biotechnological and environmental applications.

1. Introduction

Thermal springs are natural geological features formed by the expulsion of hot water (with temperatures > 36.7°C) from the Earth's interior [1]. This geothermal heat is primarily attributed to the radioactive decay of elements present in the Earth's mantle, which accounts for approximately 45–90% of the heat that escapes from the Earth [2, 3]. The most important radioactive elements include uranium-235, potassium-40, thorium-232, and uranium-238 [4]. In addition to hot springs, other geothermally heated habitats include alkaline, carbonate-containing hot springs, and acidic mud holes [5]. Minerals such as calcium, chloride, magnesium, sodium, silica, and sulfates are abundant in hot spring water as they dissolve into the water as it passes through the rocks to the Earth surface. Geothermal groundwater typically has a chemical composition derived from archaic meteorological characteristics or the primordial sea [6]. Extensive research has been conducted on hot springs because of their unique geothermal environments, which are characterized by distinct biogeography and physicochemical properties. The capacity of hot spring systems to preserve fossils, from microbes to trees and

macroscopic animals, has fundamentally transformed the comprehension of microbial diversity in hot springs [7].

Hot springs provide favorable conditions to promote the proliferation of a diverse range of thermotolerant and thermophilic organisms [8]. Recent study has also reported a high abundance of *Acidithiobacillus*, a hyperthermophilic bacteria, in hot springs. Thermophiles are microorganisms that thrive at temperatures > 50°C, whereas hyperthermophiles thrive at temperature > 80°C [12]. The abundance of organic substrates was a major contributor to microbial diversity at the study site [11]. The presence of thermophilic microorganisms in hot springs has a significant potential for bioprospecting, thermostability, and biomolecular production [10]. Hot springs contain an abundance of valuable minerals, various chemical compounds, and a diverse range of microorganisms [9]. Hot springs are also notable for their ability to preserve fossils, ranging from microbes to trees and macroscopic animals, which has fundamentally transformed our understanding of microbial diversity in hot springs [7]. They also serve as biodiversity hotspots as they support a diverse array of microorganisms, including archaea, fungi, algae, bacteria, protozoa, cyanobacteria, and viruses.

Consequently, culture-independent techniques, such as metagenomics and metatranscriptomics, have been employed [13, 14, 15]. In addition to hot springs, metagenomics has been used to improve our understanding of microbial environments located in extreme and remote habitats, such as glacial lakes and hydrothermal vents. Metagenomic analysis from these extreme environments has led to the discovery of new genes, which can be used to develop biocatalysts that have important uses in society [16,17,1819]. The genomic and metabolic adaptabilities of microbial communities in these ecosystems have evolved because of their adaptation to extreme environments. Microorganisms isolated from these environments display remarkable genomic plasticity and metabolic diversity [20]. The most influential factors affecting diversity in hot spring environments include water temperature, pH, and chemical composition. [21, 22, 23].

Recently, there has been a rise in the production and diversity of thermostable enzymes, driven by advancements in the methods used to extract these enzymes from heat-loving eukaryotes, bacteria, and archaea found in various environments,(24) such as thermophilic and hyperthermophilic enzymes, also known as thermozymes, that are resistant to extreme conditions. Thermostability of proteins is determined by their intrinsic characteristics and is directly related to their molecular structure, resulting in increased rigidity and resistance to unfolding. Furthermore, extrinsic factors, such as solutes, metal binding, and molecular chaperones, contribute to protein stabilization [25]. The enzymes from bacteria and archaea have evolved and developed structural and functional characteristics that exhibit at temperatures higher than 70°C, it is highly thermostable and optimally active [26]. Several industrial sectors such as the sugar, paper, and fruit industries rely heavily on these enzymes for processing, manufacturing, and production of food, lactose-free milk products, animal feed products, and laundry detergent products [27]. Although mesophilic enzymes are not suitable for high-temperature and extreme pH conditions, thermophilic proteins are distinguished by their shorter amino acid lengths compared to proteins found in non-thermophilic organisms. Moreover, thermophilic microorganisms

have been shown to be essential for the production and characterization of all enzymes obtained from hydrolysis because of their ability to resist proteolytic and denaturing agents such as salts, organic solvents, high temperatures, and metal ions. Therefore, these microorganisms can be used in filtration processes to remove heavy metals from waste.[28]

Geothermal ecosystems are influenced by the microbiomes of hot springs, which can be both beneficial and harmful. Although most microorganisms are symbiotic, a minority are pathogenic. Biological and environmental factors such as the composition of organic substrates, nutrients, and water chemistry influence the diversity and abundance of microorganisms. Thermophilic bacteria have been studied for decades owing to their biotechnological potential [29]. In recent years, several molecular biology techniques have been developed that allow for the identification and characterization of microorganisms at the species and subspecies levels, including rep-PCR profiling and 16S rRNA sequencing [30, 31, 32, 29, 33]. A notable example is the use of the Taq DNA polymerase for amplification. The hot springs at Yellowstone National Park are the source of Taq polymerase extracted from the bacterium Thermus aquaticus [34]. In addition to exopolysaccharides, thermophiles produce thermostable enzymes such as amylases, proteases, cellulases, xylanases, and lipases [35, 36, 37, 38]. In the present study, we aimed to evaluate the potential applications of several thermophilic microorganisms in the Al-Lith region of SA. The comprehensive data expected to be generated during the study can provide profound in-sights into the intricacies of thermophilic microbial diversity.

The present study explores the diversity of thermophilic microorganisms and their potential applications in the Al-Lith region of SA. The investigation aims to acquire comprehensive data, thereby providing profound insights into the intricacies of thermophilic microorganism diversity. This study holds significant promise for advancing biotechnological applications and making substantial contributions to the thermozymes industry, thereby aligning with the strategic objectives outlined in Saudi Vision 2030.

2. Materials and Methods

2.1. Study site

The Wadi Al-Lith is one of the most significant landmarks in the region and is considered one of the largest valleys in the Tihamah area of Hijaz in SA. The source of Wadi Al-Lith is located in the village of Al-Tifalan. Originating from the Sarawat Mountains, located south of Taif, Wadi Al-Lith flows westward towards the coastal plain. Its meandering course supplies water to wells and agricultural lands, eventually emptying into the Red Sea near the city of Al-Lith. The total length of Wadi Al-Lith is approximately 100 km, with an average gradient of 14 m/km.

2.2. Sample collection

Samples water, spring Sediment and wet-shore sediment were collected on June 26, 2022, from the Ain Al-Hara hot spring (20°27’43.9"N & 40°28’16.1"E), located in Gomaigh within the province of Al-Lith, in the western part of the SA (Fig. 1**).** Portable thermometers and pH meters were used to measure the

temperature and pH of the sample collection sites (Table 1). Ain Al-Hara hot spring is acidic (pH 4), and the temperature of the source water exceeds 80°C. Ain Al-Hara is located 165 m above sea level. Following collection, samples were transported to the laboratory under cool and sterile conditions at 4°C and stored at -20°C.

Table 1
Details of Ain Al-Hara water, sediment, and wet shore sediment samples.

Samples	Samples Code	Sample quantity and size	Number of samples	Temp (°C)	pH	Depth (cm)
Water	W	20 L filled in sterile bottles	1	55	8.5	All levels
Spring Sediment	SS	Filled in 50 ml sterile falcon tube	1	62	8.5	18
Wet shore Sediment	SW		1	37	8.3	100

2.3. DNA Extraction and Shotgun Sequencing

Using the Ph-Ch method, genomic DNA was extracted from water, sediment, and wet-shore sediment samples. Under vacuum, the water samples were filtered through 0.22 µm membrane filters (cellulose ester, Millipore, Billerica, MA, USA) and stored at-20°C. DNA was extracted from membrane filters, soil, and sediment, as described previously [39]. The quantity and integrity of the isolated DNA were determined using a 1% agarose gel electrophoresis and NanoDrop 2000™ (Thermo Fisher Scientific™).

2.4. Taxonomic profiling and functional annotation

The Illumina sequencing platform was pre-processed using Readfq (V8), resulting in clean data for analysis. The following steps were followed for this process: a) Low-quality reads exceeding a certain proportion (default length of 40 base pair) of low-quality bases were removed. B) Reads with N bases exceeding a certain threshold (default length, 10 bp) were discarded. In addition, reads that overlapped with adapters beyond a certain length (default of 15 bp) were eliminated. To filter out the reads originating from the hosts, clean data were subjected to BLAST against the host database. Bowtie2 software (version 2.2.4) was used with the following parameter settings: end-to-end, sensitive, I 200, and × 400 [40], [41].

For assembly analysis, MEGAHIT software (v1.0.4-beta) was used with --presets meta-large parameter settings (--end-to-end, --sensitive, -I 200, -× 400). Scaftigs without N were obtained by breaking the scaffolds at the N junction [42–44].

ORF prediction for the scaftigs of each sample was performed using MetaGeneMark (version 3.5) and default parameters. Predictions shorter than 100 nt were filtered [45], [46], [40], [41], [47], [48], [42], [44], [49], [50], [51]. CD-HIT software (V4.5.8) was used to eliminate redundancy and obtain the non-redundant initial gene catalogue with parameter settings of c 0.95, G 0, aS 0.9, g 1, d 0 [52], [53], [54], and [50].

Using Bowtie2 (Bowtie2.2.4), the clean data of each sample were aligned to the initial gene catalogue with the following parameter settings: --end-to-end, --sensitive, -I 200, and-x 400. Genes with reads ≤ 2 were filtered out of each sample to form the final gene catalogue (Unigenes) [42], [51], [54]. Gene abundance in each sample was calculated based on the number of aligned reads and the length of each gene [41], [49], [50], [55], [56], [57]. In addition to basic information statistics, core pan gene analysis, correlation analyses between samples, and Venn diagrams of gene numbers were constructed for each sample based on gene abundance.

DIAMOND software (V0.9.9.110) was used to align unigene sequences with those of bacteria, fungi, archaea, and viruses from the NCBI NR database (Version 2018-01-02), with the parameters blastp, e 1e-5. The result with a minimum evaluative value of 10 was selected for the alignment. An LCA algorithm in the MEGAN software was used to determine the species annotations [47], [59].

The abundance of each sample at each taxonomic level (kingdom, phylum, class, order, family, genus, and species) was estimated using LCA annotations and gene abundance tables. In a sample, the abundance of a species is equal to the sum of the abundances of annotated genes [40], [41], [49]. The number of genes of a species in a sample equals the number of genes with non-zero abundance among those annotated as belonging to that species. Based on the abundance tables at each taxonomic level, Krona analysis, relative abundance overview, and abundance clustering heatmaps were performed (60).

To align the unigenes with those in the functional databases, DIAMOND software (V 0.9.9) was used with the following parameters: blastp, -e 1e-5 [51], [61]. Three functional databases exist: KEGG (Version 2018-01-01), eggNOG (version 4.5), and CAZy (version 201801) [62], [63], [64]. The selection of the best blast-hit results was then analyzed [61], [51], and [65]. The relative abundances at different functional levels were calculated based on the alignment results [40], [51], [54]. Using the functional annotation and gene abundance tables, we derived gene number tables for each sample at each taxonomic level. Among the genes annotated with a particular function, the number of genes with non-zero abundance equals the number of genes with that function in the sample.

Annotated gene statistics, relative abundance overviews, and abundance clustering heat maps were generated from the abundance tables for each taxonomic level, combined with PCA and NMDS analysis for dimension reduction, ANOSIM analysis of inter-and intra-group differences based on functional abundance, metabolic pathway comparisons, and Metastat and LEfSe analyses of inter-group functional differences.

3. Results

Statistical analysis of raw data, including the total number of reads (sum of reads 1 and 2), base pairs (bp), GC content (%), and ratios of bases with Phred quality scores exceeding 20 and 30 (Q20 (%) and Q30 (%)), are summarized in Table 2 for all three samples. These samples were subjected to paired-end sequencing of the shotgun metagenome using an Illumina sequencing platform, which resulted in 37,738.5 raw reads. Each of these sequences had a length of 350 bp and the average number of reads

across the samples was 12,579.5 reads. The raw data displayed GC content for the 3 samples W, SS, and SW of 52.99%, 59.64%, and 58.46%. In contrast, Q20 and Q30 values were notably high, exceeding 96% and 91%, respectively. The Cont. samples exhibited the highest values and percentages across most of the analyzed parameters, except for the GC content, which was 52.99% for water, 59.64% for SS, and 58.46% for SW.

Table 2. Summary of sequencing, assembly, and pre-processing statistics.

Data	Sample ID	W	SS	SW
Raw Data	Total raw reads	12,844.95	12,422.66	12,470.89
	Clean Data	12,771.96	12,304.18	12,426.44
	GC (%)	52.99	59.64	58.46
	Q20	96.88	96.81	96.69
	Q30	91.92	92.03	91.68
	Effective (%)	99.432	99.046	99.644
Assembly Data	total len. (bp)	360,180,406	205,231,866	636,385,578
	Num.	259,165	124,012	542,578
	Average len. (bp)	1,389.77	1,654.94	1,172.89
	N50 Len.(bp)	1,802	2,463	1,272
	Max len.(bp)	357,393	373,944	236,462.00

3.1. Results of Assembly (by MEGAHIT)

MEGAHIT software was used to generate the assembled data (Table 2) by merging the raw paired-end sequences. A total of 1,201,797,850 sequences were assembled, with an average of 400,599,283.33 reads observed across the three metagenome samples. Sample SS displayed lower assembly metrics, with 205,231,866 assembled samples, which is a relatively smaller dataset than the other samples. Conversely, the SW sample exhibited the highest number of read bases, totaling 636,385,578 bp, and the number of reads in the W sample was 360,180,406 bp.

3.2. PreProcessing and Clustering (by Bowtie2 Software)

Following assembly, Bowtie2 software was used to perform the pre-processing raw reads in Table 2, which included trimming, filtering, and clustering of sequencing reads. Table 2 presents the results indicating the number of clean reads in the study samples. Clean reads for W, SS, and SW were obtained after preprocessing 12,771.96, 12,304.18, and 12,426.44, with an average of 12,500.86 reads.

3.4. Taxonomic profiling and functional annotation

Microbial communities were revealed by Krona analysis in the Al-Lith hot spring (Figure 2). These results represent three different samples from the Al-Lith hot springs. Water samples (W) were highly abundant in bacteria (90%), whereas Eukaryotes and Archaea accounted for 0.1% and 0.8%, respectively. The percentage of unclassified microbes was 0.3%, whereas that of unknown microorganisms was 9%. In the spring sediment sample (SS), bacteria were highly abundant at 90%, whereas Eukaryotes and Archaea accounted for 0.02% and 1%, respectively. The percentage of unclassified microbes was 0.3%, whereas that of unknown microorganisms was 9%. Wet shore sediment (SW) bacteria exhibited a high level of abundance, reaching 84%, whereas Eukaryotes and Archaea accounted for 0.06% and 2%, respectively. The percentage of unclassified microbes was 0.4%, whereas that of unknown microorganisms was 14%.

in krona **(Figure 3)** showed the microbial community, at water samples, the bacterial domain was dominated by the phyla *Cyanobacteria* (31%), *Proteobacteria* (26%), *Bacteroidetes* (11%), *Chloroflexi* (5%), *Bacteroidetes* (6%). and *Planctomycetes* (1%). At spring sediment sample *Chloroflexi* (46%), *Armatimonadetes* (11%), *Proteobacteria* (7%), *Planctomycetes* (7%), *Chlorobi* (3%), *Cyanobacteria* (2%), *Acidobacteria* (1%). and wetshore sediments *Proteobacteria* (25%), *Chloroflexi* (17%), *Bacteroidetes* (8%), *Verrucomicrobia* (8%), *Planctomycetes* (3%), *Cyanobacteria* (3%). Distance variations in the bacterial domain community at the phylum

level were detected in all three samples. *Cyanobacteria* were highly dominant in the water samples (31%), followed by *Chloroflexi* in the SS (46%) and *Proteobacteria* in the SW (25%).

The results in Krona Figures 4 show that the Archaea domain taxonomy was dominated by the phyla *Thaumarchaeota* (63%), *Euryarchaeota* (13%), *Candidatus_Bathyarchaeota* (3%), *Candidatus_Woesearchaeota* (3%), *Crenarchaeota* (2%), and *Candidatus_Heimdallarchaeota* (0.1%). The spring sediment samples *Thaumarchaeota* (65%), Euryarchaeota (15%), *Candidatus_Bathyarchaeota* (3%), *Crenarchaeota* (3%), *Candidatus_Micrarchaeota* (2%), and *Candidatus_Diapherotrites* (0.1%). and in wet shore sediment sample *Thaumarchaeota* (52%), *Euryarchaeota* (14%), *Candidatus_Woesearchaeota* (13%), *Candidatus_Bathyarchaeota* (1%), *Crenarchaeota* (0.4%), and *Candidatus_Heimdallarchaeota* (0.1%). There were differences in the distances between the three samples in the archaea domain at the phylum level. *Thaumarchaeota* was highly dominant in the SS sample (65%), followed by W (63%) and SW (52%). *Candidatus_Korarchaeota* was the most abundant in W (0.2%), *Candidatus_Diapherotrites* was most abundant in the SS sample (0.4%), and *Candidatus_Heimdallarchaeota* was most abundant in the SW sample (0,1%).

The results in krona (**Figure 5)** shows that the eukaryotic domain was dominated by the phyla *Mucoromycota* (28%), *Ascomycota* (10%), *Basidiomycota* (10%), Chytridiomycota (9%), *Zoopagomycota* (5%), *Microsporidia* (3%), *Blastocladiomycota* (0.9%), and *Cryptomycota* (0.8%). In the SS sample *Mucoromycota* (35%), *Ascomycota* (19%), *Chytridiomycota* (13%), *Basidiomycota* (4%), *Zoopagomycota* (4%), *Microsporidia* (2%), *Cryptomycota* (0.5%), and *Blastocladiomycota* (0.4%). In SW sample *Mucoromycota* (45%), *Ascomycota* (22%), *Basidiomycota* (15%), *Microsporidia* (6%), *Zoopagomycota*

(3%), *Chytridiomycota* (2%), *Blastocladiomycota* (0.01%). Distance variations in the eukaryotic domain community at the phylum level were detected in the three samples. *Mucoromycota* were highly dominant in the SW sample (45%), followed by SS (35%) and W (28%). *Blastocladiomycota* was the most abundant in the SW (0.1%) and SS (0.4%) samples, and *Cryptomycota* in the W sample (0,8).

Based on the KEGG database, various analyses were performed to predict the functional groups of the operating taxonomic units (**Figure 6 and 7**). The results showed that metabolic activities had the highest number of matched genes (261802, followed by genetic information processes (80761 matched genes, environmental information processes(62003 identified genes, cellular processes (49688 matched genes, human diseases (27223 matched genes, and organismal system (15223 matched genes.

In the determination of carbohydrate enzymes **(Figure 8)**, glycoside hydrolases had the highest number of matched genes in 19733, followed by glycosyl transferases, which had the number of matched genes in 19627. Carbohydrate-binding modules matched approximately 9724 genes, carbohydrate esterases matched approximately 3523 genes, polysaccharide lyases matched approximately 897 genes, and auxiliary activity matched approximately 475 genes.

In the orthologous groups (OGs) of proteins **(Figure 9)**, genes with unknown functions had the highest relative abundance of 287292 matched genes, followed by amino acid transport and metabolism, energy production, replication, recombination, repair, and conversion, whereas RNA processing and modification, extracellular structures, and nuclear structure had the lowest relative abundance.

4. Discussion

In this study, we described the biodiversity of biological communities found in the Al-Lith hot springs using modern taxonomic methods that involved studying and classifying the genetic material of living organisms within their environments.

In environmental research, shotgun metagenomic analysis is useful for identifying microbial assemblages at specific locations and understanding their interactions and taxonomic classifications. Microbial assemblages, which are essentially microbial in hot springs, play a significant role in biogeochemical processes that support hot spring ecosystems. A comprehensive metagenomic analysis of Al-Lith hot spring samples over time in the present study provides an understanding of bacterial, archaeal, and eukaryotic communities. The high diversity of this community contributes to its stability and functionality. However, multiclass pollutants and other ecological factors significantly affect microbial community abundance.

4.1. Diversity of Microbial Communities in Al-Lith hot spring

In the present study, we used metagenomic sequence analysis to determine the relative abundances of bacteria, archaea, and eukaryotic taxonomic groups in water, spring sediment, and wet-shore sediment samples collected from the Al-Lith hot spring, SA. Most of the sequences belonged to the bacterial

domain. Bacteria were the most dominant taxa of all the three samples. The phylum *Cyanobacteria* dominated the water sample (32%), whereas the phylum *Chloroflexi* (46%) dominated SS. The phylum *Proteobacteria* (25%) was dominated by the SW samples. A correlation between *cyanobacterial* species and physicochemical variables has previously been reported in the Al-Lith springs [66]. *Thaumarchaeota* dominates the Archaea domain taxonomy in water, spring sediments, and wet shore sediments, accounting for 63%, 65%, and 52%, respectively; these results aligh with those of previous studies [67, 68]. According to [69], *Thaumarchaeota* is a newly identified phylum. We found that *Mucoromycota* phyla were the most dominant in the eukaryotic domain in the water, spring sediment, and wet shore sediment (28%), (35%), and (45%, respectively).

Proteobacteria is predominant phylum in moderate-temperature thermal springs across various global locations. This includes Siloam in South Africa, as reported by [70] Tatta Pani in Pakistan;[71], Dusun Tua in Malaysia; and Sungai Serai, Ayer Hangat,[22] Attri, and Yumthang in India [72, 73]

Identifying previously unknown microorganisms in harsh environments is a novel and important advancement in our understanding of the microbiome of hot springs. Our findings suggest that new thermophilic or hyperthermophilic microorganisms may be discovered in other hot springs and geothermal habitats, potentially offering valuable applications in biotechnology in the future.

4.2. Advanced Analysis Functional Prediction Profiling

Using shotgun metagenomics, we primarily focused on the functional profiles. There was no clear separation of functional gene structures between the water, spring sediment, and wet shore sediment metagenomes, as well as between the KEGG, EggNOG, and CAZy datasets. These findings suggested that metabolic activities, genetic information processes, environmental formation processes, cellular processes, human diseases, and organismal systems dominated KEGG functional categories. A previous study found that the expression of metabolic genes is higher [74]. Genes are replicated, transcribed, translated, folded, sorted, and degenerated in microbes due to the processing of genetic information [75]. A shotgun metagenomic analysis of the KEGG database predicted that environmental information-processing genes were the third most prevalent. Evolution and diversity depend heavily on interactions between organisms and their environments over time [76]. Among the four functional genes, cellular processes help microbes perform functions such as supporting the body's structure, absorbing nutrients from food, and converting them into energy [77]. The KEGG pathway and annotation revealed that pollutants may trigger genes associated with human diseases, such as cancer, metabolic disease, cardiovascular disease, endocrine substance dependence, and drug resistance.

Within microbial genomes, the CAZy pathway is a critical genetic sequence encoding a suite of carbohydrate enzymes. This pathway facilitates the development of a sophisticated enzymatic system capable of breaking down and utilizing complex carbohydrates, as detailed in [78]. Glycoside hydrolases were the most common genes. These molecules include glycosyltransferases, carbohydrate-binding modules, carbohydrate esterases, and polysaccharide lyases. Glycoside hydrolases (GH) are crucial for the metabolism of polysaccharides such as starch, chitin, cellulose, and xylan [79]. Glycosyltransferase

enzymes play a crucial role in saccharide synthesis and glycosylation. Enzymes that break down carbohydrates can be affected by changes in the types of microbes that live in certain community structures in the Al-Lith hot spring.

The BLAST database uses unsupervised orthologous functional annotation to match the predicted gene amino acid sequences with eggNOG. The metagenomic findings included amino acid transport and metabolism, energy production, replication, recombination, repair, conversion, RNA processing, and nuclear structure. Most functional activities in the collected Al-Lith hot-spring microbiome are unknown.

Using shotgun metagenomic analysis, we discovered a rich diversity of living microorganisms that could benefit the development of biotechnological applications and the thermozyme industry in the environment, which is one of the goals of Saudi Vision 2030. The findings of the study provide insights as instead of relying on microbial cultures, we employed a DNA-based method to classify them based on the abundance and diversity of living organisms. Furthermore, our findings build on the taxonomic, structural, and functional genomes of microorganisms in the Al-Lith hot springs.

Abbreviations

The following abbreviations are used in this manuscript:

3He	Helium-3
4He	Helium-4
Bp	base pairs
GC	Guanine-cytosine content
CAZy	carbohydrate-active enzymes
DNA	Deoxyribonucleic acid
EDTA	Ethylenediaminetetraacetic acid
eggNOG	evolutionary genealogy of genes: Non-supervised Orthologous Groups
gDNA	Genomic deoxyribonucleic acid
KEGG	Kyoto Encyclopedia of Genes and Genomes
NCBI	National Center for Biotechnology Information, U.S. National Library of Medicine.
NGS	Next-generation sequencing
ORF	open reading frames
PCA	Principal component analysis
PCR	Polymerase chain reaction
Ph-Ch	Phenol/Chloroform
Rep-PCR	Repetitive-element PCR
RNA	Ribonucleic acid
rRNA	Ribosomal RNA gene
RTA	Real Time Analysis
SA	Saudi Arabia
SAS	Saudi Arabia Shield
SDS	Sodium dodecyl sulfate
SS	Spring sediment
SW	Wet shore sediment
W	Water
WGS	Whole-genome-shotgun

Declarations

Author Contributions: Author Contributions: Abdulaziz M Kusa collected the study samples, achieved lab experiments, data analysis and wrote the manuscript. Mohammed N. Baeshen is the PI of the project and general manager. Roba M.S Atta worked on manuscript editing and reviewing. Hulayyil Al-Harbi worked on bioinformatics and data analysis.

Funding: This research was funded by Deputyship for Research & Innovation, Ministry of Education in Saudi Arabia, grant number MoE-IF-UJ-R2-22-04100088-1.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

References

1. Pentecost, A.; Jones, B.; Renaut, R.W. What is a hot spring?. *Can J Earth Sci* 2003, **40**, 1443–1446. doi: 10.1139/e03-083.
2. Wheeler, J. Geodynamics by D. L. Turcotte and G. Schubert. Geol. J. 2005, 40, 456. doi: https://doi.org/10.1002/gj.987.
3. van Middlesworth, P.E.; Wood, S.A. The aqueous geochemistry of the rare earth elements and yttrium. Part 7. REE, Th and U contents in thermal springs associated with the Idaho batholith. *Appl. Geochem.* 1998, **13**, 861–884.
4. Khoonphunnarai, P.; Yongsiriwit, P.; Seetapong, N.; Awae, M. Potential of geothermal energy resource from heat production in Thung Nui hot spring rocks area, Satun Province, Southern of Thailand. In Journal of Physics: Conference Series, Siam Physics Congress 2018–A Creative Path to Sustainable Innovation, Pitsanulok, Thailand, 21–23 May 2018. doi: 10.1088/1742-6596/1144/1/012039.
5. Morozkina, E.V.; Slutskaya, E.S.; Fedorova, T.V.; Tugay, T.I.; Golubeva, L.I.; Koroleva, O.V. Extremophilic microorganisms: Biochemical adaptation and biotechnological application (review). Appl. Biochem. Microbiol. 2010, 46, 1–14. doi: 10.1134/S0003683810010011. [6] Ármannsson, H.; Fridriksson, T. Application of geochemical methods in geothermal exploration. Presented at Short Course on

Surface Exploration for Geothermal Resources organized by UNU-GTP and LaGeo, Ahuachapan and Santa Tecla, El Salvador, 17–30 October 2009.

6. Des Marais, D. J.; Walter, M. R. Terrestrial hot spring systems: Introduction. *Astrobiology* **2019**, *19* (12) doi: 10.1089/ast.2018.1976.
7. Pace, N.R. A Molecular View of Microbial Diversity and the Biosphere. *Science* 1997, **276**, 734–740
8. Sharma, N.; Kumar, J.; Abedin, M.M.; Sahoo, D.; Pandey, A.; Rai, A.K.; Singh, S.P. Metagenomics revealing molecular profiling of community structure and metabolic pathways in natural hot springs of the Sikkim Himalaya. *BMC Microbiol.* 2020, **20**, 246. doi: 10.1186/s12866-020-01923-3. [10] Yasir, M.; Qureshi, A.K.; Khan, I.; Bibi, F.; Rehan, M.; Khan, S.B.; Azhar, E.I. Culturomics-Based Taxonomic Diversity of Bacterial Communities in the Hot Springs of Saudi Arabia. *OMICS: A J. Integr. Biol.* 2019, **23**, 1. doi: 10.1089/omi.2018.0176.
9. Menzel, P.; Gudbergsdóttir, S.R.; Rike, A.G.; Lin, L.; Zhang, Q.; Contursi, P.; Moracci, M.; Kristjansson, J.K.; Bolduc, B.; Gavrilov, S.; Ravin, N.; Mardanov, A.; Bonch-Osmolovskaya, E.; Young, M.; Krogh, A.; Peng, X. Comparative Metagenomics of Eight Geographically Remote Terrestrial Hot Springs. *Environ. Microbiol.* 2015, **70**, 411–424. doi: 10.1007/s00248-015-0576-9.
10. Kumar, S.; Dangi, A.K.; Shukla, P.; Baishya, D.; Khare, S.K. Thermozymes: Adaptive strategies and tools for their biotechnological applications. *Bioresour. Technol.* 2019, **278**, 372–382 doi: 10.1016/j.biortech.2019.01.088.
11. Bhaya, D.; Grossman, A.R.; Steunou, A.-S.; Khuri, N.; Cohan, F.M.; Hamamura, N.; Melendrez, M.C.; Bateson, M.M.; Ward, D.M.; Heidelberg, J.F. Population level functional diversity in a microbial community revealed by comparative genomic and metagenomic analyses. *ISME J.* 2007, **1**, 703–713
12. Liu, Z.; Klatt, C.G.; Wood, J.M.; Rusch, D.B.; Ludwig, M.; Wittekindt, N.; Tomsho, L.P.; Schuster, S.C.; Ward, D.M.; Bryant, D.A. Metatranscriptomic analyses of chlorophototrophs of a hot-spring microbial mat. *ISME J.* 2011, **5**, 1279–1290.
13. López-López, O.; Cerdán, M.E.; González-Siso, M.I. Hot spring metagenomics. *Life* 2013, **3**, 308–320.
14. Kaushal, G.; Kumar, J.; Sangwan, R.S.; Singh, S.P. Metagenomic analysis of geothermal water reservoir sites exploring carbohydrate-related thermozymes. *Int. J. Biol. Macromol.* 2018, **119**, 882–895. doi: 10.1016/j.ijbiomac.2018.07.196.
15. Agarwal, N.; Narnoliya, L.K.; Singh, S.P. Characterization of a novel amylosucrase gene from the metagenome of a thermal aquatic habitat, and its use in turanose production from sucrose biomass. *Enzyme Microb. Technol.* 2019, **131**, 109372. doi: 10.1016/j.enzmictec.2019.109372.
16. Patel, S.N.; Kaushal, G.; Singh, S.P. A novel d-allulose 3-epimerase gene from the metagenome of a thermal aquatic habitat and d-allulose production by *Bacillus subtilis* whole-cell catalysis. *Appl. Environ. Microbiol.* 2020, **86**, 5. doi: 10.1128/AEM.02605-19.
17. Joshi, N.; Sharma, M.; Singh, S.P. Characterization of a novel xylanase from an extreme temperature hot spring metagenome for xylooligosaccharide production. *Appl. Microbiol. Biotechnol.* 2020, **104**, 4889–4901. doi: 10.1007/s00253-020-10562-7.

18. Wemheuer, B.; Taube, R.; Akyol, P.; Wemheuer, F.; Daniel, R. Microbial diversity and biochemical potential encoded by thermal spring metagenomes derived from the Kamchatka peninsula. *Archaea* 2013, **2013**, 136714. doi: 10.1155/2013/136714.
19. Mahajan, M.M. Deciphering molecular mechanisms underlying salinity stress tolerance in Kharchia local, Ph.D. Degree, Division of Molecular Biology and Biotechnology National Research Centre on Plant Biotechnology Indian Agricultural Research Institute, New Delhi, 2017.
20. Chan, C.S.; Chan, K.-G.; Ee, R.; Hong, K.-W.; Urbieta, M.S.; Donati, E.R.; Shamsir, M.S.; Goh, K.M. Effects of Physiochemical Factors on Prokaryotic Biodiversity in Malaysian Circumneutral Hot Springs. *Front. Microbiol.* 2017, **8**, 1252. doi: 10.3389/fmicb.2017.01252.
21. Poddar, A.; Das, S. K. Microbiological studies of hot springs in India: a review. *Arch. Microbiol.* **2018**, *200*, 1–18. doi: 10.1007/s00203-017-1429-3.
22. Gomes, E.; de Souza, A. R.; Orjuela, G. L.; da Silva, R.; de Oliveira, T. B.; Rodrigues, A. Applications and benefits of thermophilic microorganisms and their enzymes for industrial biotechnology. In *Gene Expression Systems in Fungi: Advancements and Applications*; Gomes, E., Rodrigues de Souza, A., Ladino Orjuela, G., da Silva, R., Brito de Oliveira, T., Rodrigues, A., Eds.; Springer: Cham, 2016; pp 459–492. doi: 10.1007/978-3-319-27951-0_21.
23. Bruins, M. E.; Janssen, A. E. M.; Boom, R. M. Thermozymes and their applications. *Appl. Biochem. Biotechnol.* **2001**, *90*, 155–186.
24. Vieille, C.; Zeikus, G.J. Hyperthermophilic Enzymes: Sources, Uses, and Molecular Mechanisms for Thermostability. *Microbiol. Mol. Biol. Rev.* 2001, **65**, 1–43. doi: 10.1128/mmbr.65.1.1-43.2001.
25. Mehta, R.; Singhal, P.; Singh, H.; Damle, D.; Sharma, A. K. Insight into thermophiles and their wide-spectrum applications. *3 Biotech* **2016**, *6*, 81. doi: 10.1007/s13205-016-0368-z.
26. Ilyas, S.; Lee, J.C.; Kim, B.S. Bioremoval of heavy metals from recycling industry electronic waste by a consortium of moderate thermophiles: Process development and optimization. *J. Clean Prod.* 2014, **70**, 194–202. doi: 10.1016/j.jclepro.2014.02.019.
27. Adiguzel, A.; Ozkan, H.; Baris, O.; Inan, K.; Gulluce, M.; Sahin, F. Identification and characterization of thermophilic bacteria isolated from hot springs in Turkey. *J. Microbiol. Methods* 2009, **79**, 321–328. doi: 10.1016/j.mimet.2009.09.026.
28. J. L. W. Rademaker and F. J. De Bruijn, "Characterization and Classification of Microbes by REP_PCR Genomic Fingerprinting 52 PUBLICATIONS 4,307 CITATIONS SEE PROFILE," 1998. [Online]. Available: https://www.researchgate.net/publication/285185102
29. Perret, X.; Broughton, W.J. Rapid identification of *Rhizobium* strains by targeted PCR fingerprinting. *Plant Soil* 1998, **204**, 21–34 doi: 10.1023/A:1004370725605.
30. Nazina, T.N.; Sokolova, D.Sh.; Grigoryan, A.A.; Shestakova, N.M.; Mikhailova, E.M.; Poltaraus, A.B.; Tourova, T.P.; Lysenko, A.M.; Osipov, G.A.; Belyaev, S.S. *Geobacillus jurassicus* sp. nov., a new thermophilic bacterium isolated from a high-temperature petroleum reservoir, and the validation of the *Geobacillus* species. *Syst. Appl. Microbiol.* 2005, **28**, 43–53. doi: 10.1016/j.syapm.2004.09.001.

31. Rahman, R.; Leow, T.; Salleh, A.; Basri, M. *Geobacillus zalihae* sp. nov., a thermophilic lipolytic bacterium isolated from palm oil mill effluent in Malaysia. *BMC Microbiol.* 2007, **7**, 77. doi: 10.1186/1471-2180-7-77.
32. Chien, A.; Edgar, D.B.; Trela, J.M. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. *J. Bacteriol.* 1976, **127**, 1550–1557. doi: 10.1128/jb.127.3.1550-1557.1976.
33. Pinzón-Martínez, D.L.; Rodríguez-Gómez, C.; Miñana-Galbis, D.; Carrillo-Chávez, J.A.; Valerio-Alfaro, G.; Oliart-Ros, R. Thermophilic bacteria from Mexican thermal environments: isolation and potential applications. *Environ. Technol.* 2010, **31**, 957–966. doi: 10.1080/09593331003758797.
34. Singh, A.K.; Tripathi, B.M.; Sahay, H.; Singh, R.N.; Kaushik, R.; Saxena, A.K.; Arora, D.K. Biochemical and molecular characterization of thermo-alkali tolerant xylanase producing bacteria from thermal springs of Manikaran. *Indian J. Microbiol.* 2010, **50**, 2–9. doi: 10.1007/s12088-010-0071-4.
35. AL-Awsy, M.; Al Obiady, S.; Al Obaidi, A. Production of amylase enzyme from thermophilic bacteria using agricultural wastes as a substrate. *Aust. J. Basic & Appl. Sci.* **2017**, *11*, 158–164.
36. Suleiman, A.D.; Abdul Rahman, N.; Mohd Yusof, H.; Mohd Shariff, F.; Yasid, N.A. Effect of Cultural Conditions on Protease Production by a Thermophilic *Geobacillus thermoglucosidasius* SKF4 Isolated from Sungai Klah Hot Spring Park, Malaysia. *Molecules* 2020, **25**, 2609. doi: 10.3390/molecules25112609.
37. Djurhuus, A.; Port, J.; Closek, C.J.; Yamahara, K.M.; Romero-Maraccini, O.; Walz, K.R.; Goldsmith, D.B.; Michisaki, R.; Breitbart, M.; Boehm, A.B.; Chavez, F.P. Evaluation of Filtration and DNA Extraction Methods for Environmental DNA Biodiversity Assessments across Multiple Trophic Levels. *Front. Mar. Sci.* 2017, **4**, 314 doi: 10.3389/fmars.2017.00314.
38. Karlsson, F.H.; Fåk, F.; Nookaew, I.; Tremaroli, V.; Fagerberg, B.; Petranovic, D.; Bäckhed, F.; Nielsen, J. Symptomatic Atherosclerosis is Associated with an Altered Gut Metagenome. *Nat. Commun.* 2012, **3**, 1245 doi: 10.1038/ncomms2266.
39. Qin, J.; Li, R.; Raes, J.; Arumugam, M.; Burgdorf, K.S.; Manichanh, C.; Nielsen, T.; Pons, N.; Levenez, F.; Yamada, T.; Mende, D.R.; Li, J.; Xu, J.; Li, S.; Li, D.; Cao, J.; Wang, B.; Liang, H.; Zheng, H.; Xie, Y.; Tap, J.; Lepage, P.; Bertalan, M.; Batto, J.-M.; Hansen, T.; Le Paslier, D.; Linneberg, A.; Nielsen, H.B.; Pelletier, E.; Renault, P.; Sicheritz-Pontén, T.; Turner, K.; Zhu, H.; Yu, C.; Li, S.; Jian, M.; Zhou, Y.; Li, Y.; Zhang, X.; Li, S.; Qin, N.; Yang, H.; Wang, J.; Brunak, S.; Doré, J.; Guarner, F.; Kristiansen, K.; Pedersen, O.; Parkhill, J.; Weissenbach, J.; MetaHIT Consortium; Bork, P.; Ehrlich, S.D.; Wang, J. A Human Gut Microbial Gene Catalogue Established by Metagenomic Sequencing. *Nature* 2010, **464**, 59–65. doi: 10.1038/nature08821.
40. Feng, Q.; Liang, S.; Jia, H.; Stadlmayr, A.; Tang, L.; Lan, Z.; Zhang, D.; Xia, H.; Xu, X.; Jie, Z.; Su, L.; Li, X.; Li, X.; Li, J.; Xiao, L.; Huber-Schönauer, U.; Niederseer, D.; Xu, X.; Al-Aama, J.Y.; Yang, H.; Wang, J.; Kristiansen, K.; Arumugam, M.; Tilg, H.; Datz, C.; Wang, J. Gut Microbiome Development Along the Colorectal Adenoma–Carcinoma Sequence. *Nat. Commun.* 2015, **6**, 6528. doi: 10.1038/ncomms7528.

41. Nielsen, H.B.; Almeida, M.; Sierakowska Juncker, A.; Rasmussen, S.; Li, J.; Sunagawa, S.; Plichta, D.R.; Gautier, L.; Pedersen, A.G.; Le Chatelier, E.; Pelletier, E.; Bonde, I.; Nielsen, T.; Manichanh, C.; Arumugam, M.; Batto, J.-M.; Quintanilha dos Santos, M.B.; Blom, N.; Borruel, N.; Burgdorf, K.S.; Boumezbeur, F.; Casellas, F.; Doré, J.; Dworzynski, P.; Guarner, F.; Hansen, T.; Hildebrand, F.; Kaas, R.S.; Kennedy, S.; Kristiansen, K.; Kultima, J.R.; Léonard, P.; Levenez, F.; Lund, O.; Moumen, B.; Le Paslier, D.; Pons, N.; Pedersen, O.; Prifti, E.; Qin, J.; Raes, J.; Sørensen, S.; Tap, J.; Tims, S.; Ussery, D.W.; Yamada, T.; MetaHIT Consortium; Renault, P.; Sicheritz-Ponten, T.; Bork, P.; Wang, J.; Brunak, S.; Ehrlich, S.D. Identification and Assembly of Genomes and Genetic Elements in Complex Metagenomic Samples Without Using Reference Genomes. *Nat. Biotechnol.* 2014, **32**, 822–828. doi: 10.1038/nbt.2939.

42. Karlsson, F.H.; Tremaroli, V.; Nookaew, I.; Bergström, G.; Behre, C.J.; Fagerberg, B.; Nielsen, J.; Bäckhed, F. Gut Metagenome in European Women with Normal, Impaired, and Diabetic Glucose Control. *Nature* 2013, **498**, 99–103. doi: 10.1038/nature12198.

43. Handelsman, J.; Rondon, M.R.; Brady, S.F.; Clardy, J.; Goodman, R.M. Molecular Biological Access to the Chemistry of Unknown Soil Microbes: A New Frontier for Natural Products. *Chem. Biol.* 1998, **5**, R245–R249 doi: 10.1016/S1074-5521(98)90108-9.

44. Chen, K.; Pachter, L. Bioinformatics for Whole-Genome Shotgun Sequencing of Microbial Communities. *PLoS Comput. Biol.* 2005, **1**, e24. doi: 10.1371/journal.pcbi.0010024.

45. Zhu, W.; Lomsadze, A.; Borodovsky, M. Ab Initio Gene Identification in Metagenomic Sequences. *Nucleic Acids Res.* 2010, **38**, e132. doi: 10.1093/nar/gkq275.

46. Li, W.; Godzik, A. Cd-hit: A Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. *Bioinformatics* 2006, **22**, 1658–1659. doi: 10.1093/bioinformatics/btl158.

47. Zeller, G.; Tap, J.; Voigt, A.Y.; Sunagawa, S.; Kultima, J.R.; Costea, P.I.; Amiot, A.; Böhm, J.; Brunetti, F.; Habermann, N.; Hercog, R.; Koch, M.; Luciani, A.; Mende, D.R.; Schneider, M.A.; Schrotz-King, P.; Tournigand, C.; Tran Van Nhieu, J.; Yamada, T.; Zimmermann, J.; Benes, V.; Kloor, M.; Ulrich, C.M.; von Knebel Doeberitz, M.; Sobhani, I.; Bork, P. Potential of Fecal Microbiota for Early-Stage Detection of Colorectal Cancer. *Mol. Syst. Biol.* 2014, **10**, 766. doi: 10.15252/msb.20145645.

48. Sunagawa, S.; Coelho, L.P.; Chaffron, S.; Kultima, J.R.; Labadie, K.; Salazar, G.; Djahanschiri, B.; Zeller, G.; Mende, D.R.; Alberti, A.; Cornejo-Castillo, F.M.; Costea, P.I.; Cruaud, C.; d'Ovidio, F.; Engelen, S.; Ferrera, I.; Gasol, J.M.; Guidi, L.; Hildebrand, F.; Kokoszka, F.; Lepoivre, C.; Lima-Mendez, G.; Poulain, J.; Poulos, B.T.; Royo-Llonch, M.; Sarmento, H.; Vieira-Silva, S.; Dimier, C.; Picheral, M.; Searson, S.; Kandels-Lewis, S.; Tara Oceans coordinators; Bowler, C.; de Vargas, C.; Gorsky, G.; Grimsley, N.; Hingamp, P.; Iudicone, D.; Jaillon, O.; Not, F.; Ogata, H.; Pesant, S.; Speich, S.; Stemmann, L.; Sullivan, M.B.; Weissenbach, J.; Wincker, P.; Karsenti, E.; Raes, J.; Acinas, S.G.; Bork, P. Structure and Function of the Global Ocean Microbiome. *Science* **2015**, *348*, 6237. doi: 10.1126/science.1261359.

49. Oh, J.; Byrd, A.L.; Deming, C.; Conlan, S.; Kong, H.H.; Segre, J.A. Biogeography and individuality shape function in the human skin metagenome. *Nature* **2014**, *514*, 7520, 59-64. doi: 10.1038/nature13786.

50. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 23, 3150–3152. doi: 10.1093/bioinformatics/bts565.

51. Qin, J.; Li, Y.; Cai, Z.; Li, S.; Zhu, J.; Zhang, F.; Liang, S.; Zhang, W.; Guan, Y.; Shen, D.; Peng, Y.; Zhang, D.; Jie, Z.; Wu, W.; Qin, Y.; Xue, W.; Li, J.; Han, L.; Lu, D.; Wu, P.; Dai, Y.; Sun, X.; Li, Z.; Tang, A.; Zhong, S.; Li, W.; Chen, W.; Xu, R.; Wang, M.; Feng, Q.; Gong, M.; Yu, J.; Zhang, Y.; Zhang, M.; Hansen, T.; Sanchez, G.; Raes, J.; Falony, G.; Okuda, S.; Almeida, M.; LeChatelier, E.; Renault, P.; Pons, N.; Batto, J.-M.; Zhang, Z.; Chen, H.; Yang, R.; Zheng, W.; Li, S.; Yang, H.; Wang, J.; Ehrlich, S.D.; Nielsen, R.; Pedersen, O.; Kristiansen, K.; Wang, J. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **2012**, *490*, 55-60. doi: 10.1038/nature11450.

52. Li, J.; Jia, H.; Cai, X.; Zhong, H.; Feng, Q.; Sunagawa, S.; Arumugam, M.; Kultima, J. R.; Prifti, E.; Nielsen, T.; Juncker, A. S.; Manichanh, C.; Chen, B.; Zhang, W.; Levenez, F.; Wang, J.; Xu, X.; Xiao, L.; Liang, S.; Zhang, D.; Zhang, Z.; Chen, W.; Zhao, H.; Al-Aama, J. Y.; Edris, S.; Yang, H.; Wang, J.; Hansen, T.; Nielsen, H. B.; Brunak, S.; Kristiansen, K.; Guarner, F.; Pedersen, O.; Doré, J.; Ehrlich, S. D.; MetaHIT Consortium; Bork, P.; Wang, J. An integrated catalog of reference genes in the human gut microbiome. *Nature Biotechnology* **2014**, *32*, 834–841. doi: 10.1038/nbt.2942.

53. Villar, E.; Farrant, G. K.; Follows, M.; Garczarek, L.; Speich, S.; Audic, S.; Bittner, L.; Blanke, B.; Brum, J. R.; Brunet, C.; Casotti, R.; Chase, A.; Dolan, J. R.; d'Ortenzio, F.; Gattuso, J.-P.; Grima, N.; Guidi, L.; Hill, C. N.; Jahn, O.; Jamet, J.-L.; Le Goff, H.; Lepoivre, C.; Malviya, S.; Pelletier, E.; Romagnan, J.-B.; Roux, S.; Santini, S.; Scalco, E.; Schwenck, S. M.; Tanaka, A.; Testor, P.; Vannier, T.; Vincent, F.; Zingone, A.; Dimier, C.; Picheral, M.; Searson, S.; Kandels-Lewis, S.; Tara Oceans Coordinators; Acinas, S. G.; Bork, P.; Boss, E.; de Vargas, C.; Gorsky, G.; Ogata, H.; Pesant, S.; Sullivan, M. B.; Sunagawa, S.; Wincker, P.; Karsenti, E.; Bowler, C.; Not, F.; Hingamp, P.; Iudicone, D. Environmental characteristics of Agulhas rings affect interocean plankton transport. *Science* **2015**, *348*(6237), 1261447. doi: 10.1126/science.1261447.

54. Cotillard, A.; Kennedy, S. P.; Kong, L. C.; Prifti, E.; Pons, N.; Le Chatelier, E.; Almeida, M.; Quinquis, B.; Levenez, F.; Galleron, N.; Gougis, S.; Rizkalla, S.; Batto, J.-M.; Renault, P.; ANR MicroObes Consortium; Doré, J.; Zucker, J.-D.; Clément, K.; Ehrlich, S. D.; ANR MicroObes Consortium Members. Dietary intervention impact on gut microbial gene richness. *Nature* **2013**, *500*, 585–588. doi: 10.1038/nature12480.

55. Le Chatelier, E.; Nielsen, T.; Qin, J.; Prifti, E.; Hildebrand, F.; Falony, G.; Almeida, M.; Arumugam, M.; Batto, J.-M.; Kennedy, S.; Leonard, P.; Li, J.; Burgdorf, K.; Grarup, N.; Jørgensen, T.; Brandslund, I.; Nielsen, H. B.; Juncker, A. S.; Bertalan, M.; Levenez, F.; Pons, N.; Rasmussen, S.; Sunagawa, S.; Tap, J.; Tims, S.; Zoetendal, E. G.; Brunak, S.; Clément, K.; Doré, J.; Kleerebezem, M.; Kristiansen, K.; Renault, P.; Sicheritz-Ponten, T.; de Vos, W. M.; Zucker, J.-D.; Raes, J.; Hansen, T.; MetaHIT Consortium; Bork, P.; Wang, J.; Ehrlich, S. D.; Pedersen, O. Richness of human gut microbiome correlates with metabolic markers. *Nature* **2013**, *500*, 541–546. doi: 10.1038/nature12506.

56. Buchfink, B.; Xie, C.; Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **2015**, *12*, 59–60. doi: 10.1038/nmeth.3176.

57. Huson, D. H.; Mitra, S.; Ruscheweyh, H.-J.; Weber, N.; Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **2011**, *21*, 1552–1560. doi: 10.1101/gr.120618.111.

58. Ondov, B. D.; Bergman, N. H.; Phillippy, A. M. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics* **2011**, *12*, 385. doi: 10.1186/1471-2105-12-385.

59. Scher, J. U.; Sczesnak, A.; Longman, R. S.; Segata, N.; Ubeda, C.; Bielski, C.; Rostron, T.; Cerundolo, V.; Pamer, E. G.; Abramson, S. B.; Huttenhower, C.; Littman, D. R. Expansion of intestinal *Prevotella copri* correlates with enhanced susceptibility to arthritis. *eLife* **2013**, *2*, e01202. doi: 10.7554/eLife.01202.

60. Kanehisa, M. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **2006**, *34*, D354–D357. doi: 10.1093/nar/gkj102.

61. Powell, S.; Forslund, K.; Szklarczyk, D.; Trachana, K.; Roth, A.; Huerta-Cepas, J.; Gabaldón, T.; Rattei, T.; Creevey, C.; Kuhn, M.; et al. eggNOG v4.0: nested orthology inference across 3686 organisms. *Nucleic Acids Res* **2014**, *42*, D231–D239. doi: 10.1093/nar/gkt1253.

62. Cantarel, B. L.; Coutinho, P. M.; Rancurel, C.; Bernard, T.; Lombard, V.; Henrissat, B. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* **2009**, *37*, D233–D238. doi: 10.1093/nar/gkn663.

63. Bäckhed, F.; Roswall, J.; Peng, Y.; Feng, Q.; Jia, H.; Kovatcheva-Datchary, P.; Li, Y.; Xia, Y.; Xie, H.; Zhong, H.; Khan, M. T.; Zhang, J.; Li, J.; Xiao, L.; Al-Aama, J.; Zhang, D.; Lee, Y. S.; Kotowska, D.; Colding, C.; Tremaroli, V.; Yin, Y.; Bergman, S.; Xu, X.; Madsen, L.; Kristiansen, K.; Dahlgren, J.; Wang, J. Dynamics and Stabilization of the Human Gut Microbiome during the First Year of Life. *Cell* **2015**, *17*, 690–703. doi: 10.1016/j.chom.2015.04.004.

64. Abdelwahab, H.; Amin, A. Multivariate Analyses of Physicochemical Factors Controlling Cyanobacteria Biodiversity in Al-Lith Thermal Springs, KSA. *Egypt. Acad. J. Biol. Sci., H. Botany* **2019**, *10*, 45–58. doi: 10.21608/eajbsh.2019.47276.

65. Ghilamicael, A. M.; Budambula, N. L. M.; Anami, S. E.; Mehari, T.; Boga, H. I. Evaluation of Prokaryotic Diversity of Five Hot Springs in Eritrea. *BMC Microbiol.* **2017**, *17*, 1. doi: 10.1186/s12866-017-1113-4.

66. Huang, Q.; Dong, C. Z.; Dong, R. M.; Jiang, H.; Wang, S.; Wang, G.; Fang, B.; Ding, X.; Niu, L.; Li, X.; Zhang, C.; Dong, H. Archaeal and Bacterial Diversity in Hot Springs on the Tibetan Plateau, China. *Extremophiles* **2011**, *15*, 549. doi: 10.1007/s00792-011-0386-z.

67. Spang, A.; Hatzenpichler, R.; Brochier-Armanet, C.; Rattei, T.; Tischler, P.; Spieck, E.; Streit, W.; Stahl, D. A.; Wagner, M.; Schleper, C. Distinct Gene Set in Two Different Lineages of Ammonia-Oxidizing Archaea Supports the Phylum Thaumarchaeota. *Opinion* **2010**, *18*, 331–340. doi: 10.1016/j.tim.2010.06.003.

68. Tekere, M. Metagenomic Analysis of Bacterial Diversity of Siloam Hot Water Spring, Limpopo, South Africa. *Afr. J. Biotechnol.* **2011**, *10*, 78. doi: 10.5897/AJB11.899.

69. Amin, A.; Ahmed, I.; Salam, N.; Kim, B.-Y.; Singh, D.; Zhi, X.-Y.; Xiao, M.; Li, W.-J. Diversity and Distribution of Thermophilic Bacteria in Hot Springs of Pakistan. *Microb. Ecol.* **2017**, *74*, 116–127.

doi: 10.1007/s00248-017-0930-1.

70. Panda, A. K.; Bisht, S. S.; De Mandal, S.; Kumar, N. S. Bacterial and Archeal Community Composition in Hot Springs from Indo-Burma Region, North-east India. *AMB Express* **2016**, *6*, 1–12.
71. Tripathy, S.; Padhi, S. K.; Mohanty, S.; Samanta, M.; Maiti, N. K. Analysis of the Metatranscriptome of Microbial Communities of an Alkaline Hot Sulfur Spring Revealed Different Gene Encoding Pathway Enzymes Associated with Energy Metabolism. *Extremophiles* **2016**, *20*, 525–536.
72. Curran, J. F.; Zaggia, L.; Quero, G. M. Metagenomic Characterization of Microbial Pollutants and Antibiotic-and Metal-Resistance Genes in Sediments from the Canals of Venice. *Water (Basel)* **2022**, *14*, 1161.
73. Zhang, H.; Yuan, X.; Xiong, T.; Wang, H.; Jiang, L. Bioremediation of Co-contaminated Soil with Heavy Metals and Pesticides: Influence Factors, Mechanisms and Evaluation Methods. *Chem. Eng. J.* **2020**, *398*, 125657.
74. McFall-Ngai, M.; Bosch, T. C. G. Animal Development in the Microbial World: The Power of Experimental Model Systems. *Curr. Top. Dev. Biol.* **2021**, *141*, 371–397.
75. Li, S.; Show, P. L.; Ngo, H. H.; Ho, S.-H. Algae-Mediated Antibiotic Wastewater Treatment: A Critical Review. *Environ. Sci. Ecotechnol.* **2022**, *9*, 100145.
76. Cabral, L.; Persinoti, G. F.; Paixão, D. A. A.; Martins, M. P.; Morais, M. A. B.; Chinaglia, M.; Domingues, M. N.; Sforca, M. L.; Pirolla, R. A. S.; Generoso, W. C.; Santos, C. A.; Maciel, L. F.; Terrapon, N.; Lombard, V.; Henrissat, B.; Murakami, M. T. Gut Microbiome of the Largest Living Rodent Harbors Unprecedented Enzymatic Systems to Degrade Plant Polysaccharides. *Nat. Commun.* **2022**, *13*, 629.
77. McKee, L. S.; La Rosa, S. L.; Westereng, B.; Eijsink, V. G.; Pope, P. B.; Larsbrink, J. Polysaccharide Degradation by the Bacteroidetes: Mechanisms and Nomenclature. *Environ. Microbiol. Rep.* **2021**, *13*, 559–581.

Figures

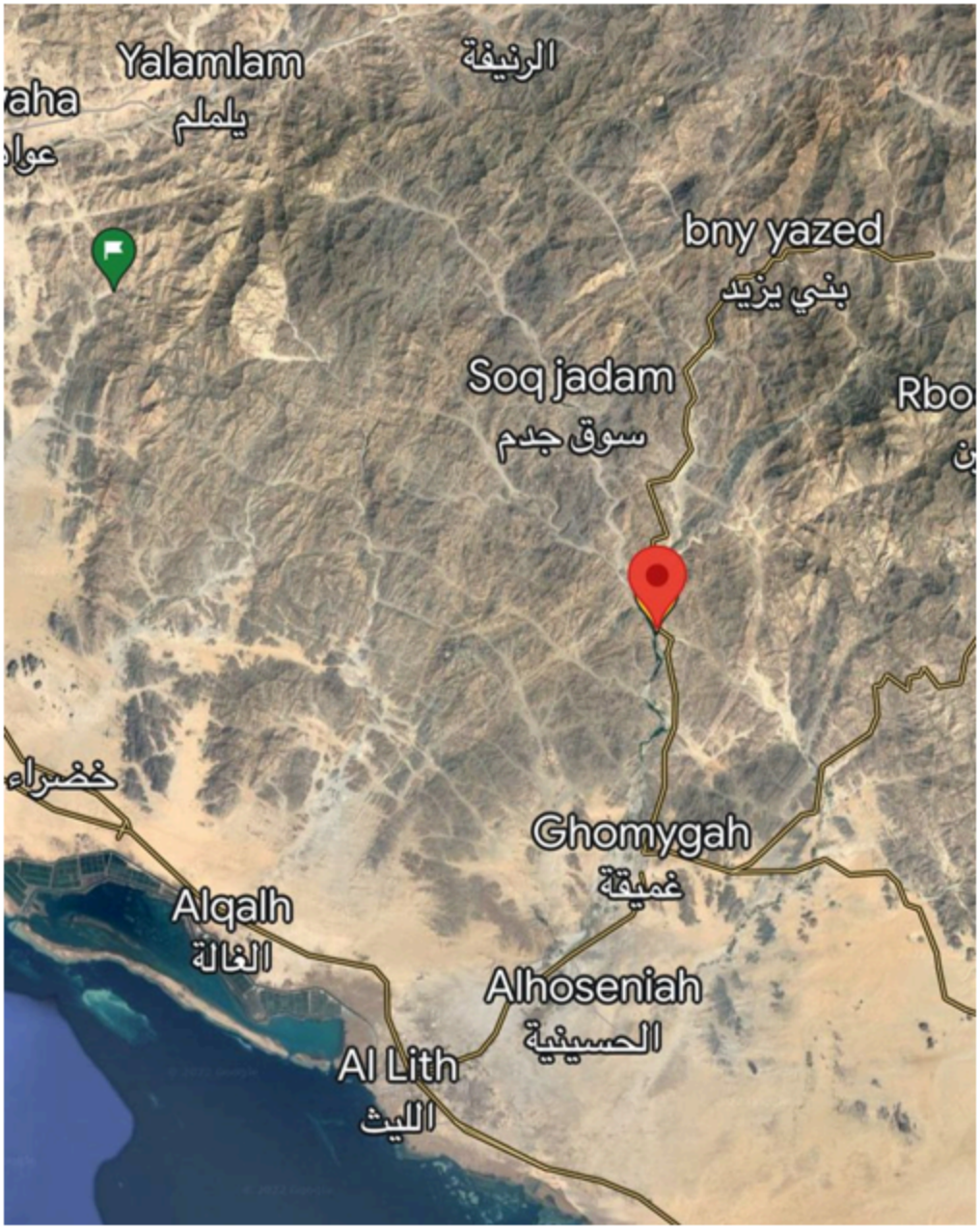

Figure 1

Location of the Ain Al-Hara hot spring in Gomaigh in the Al-Lith province.

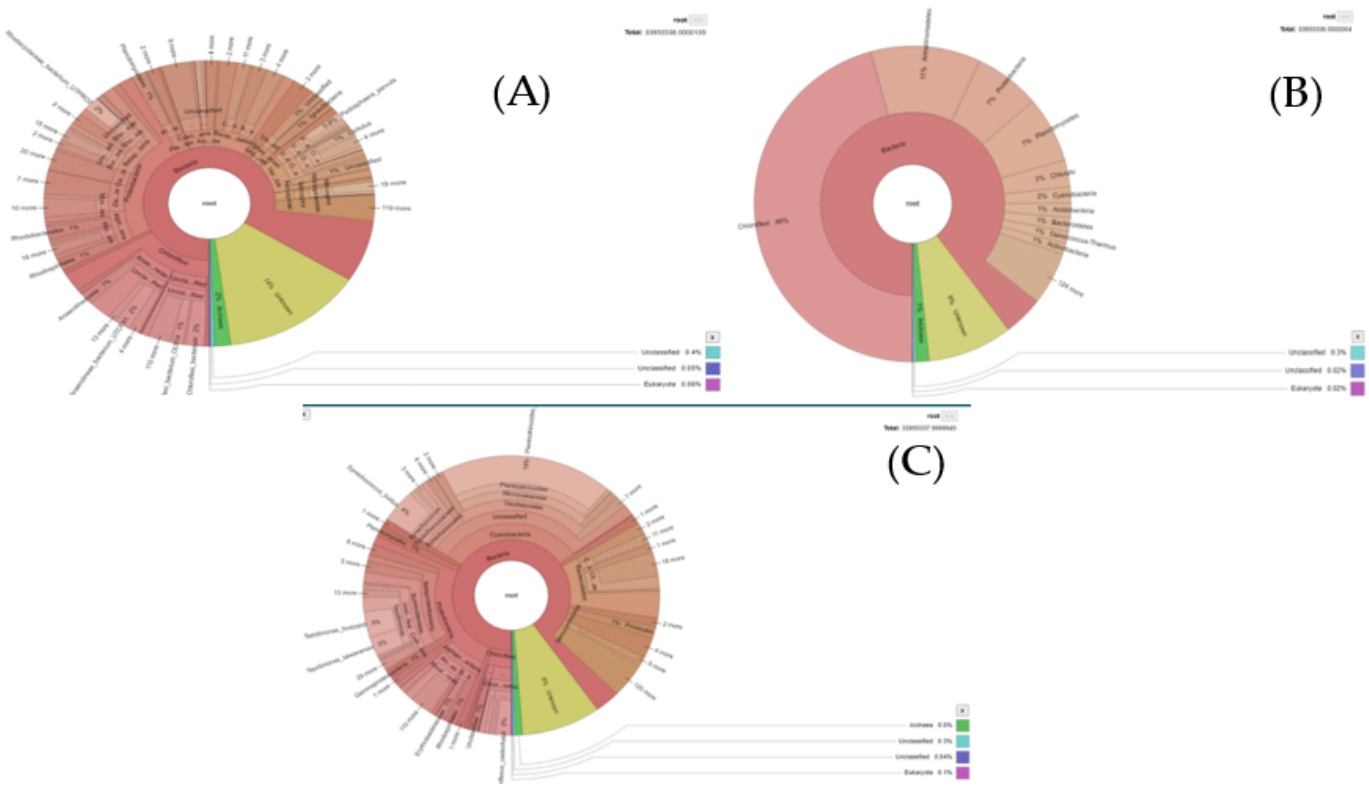

Figure 2

The Krona chart illustrates the taxonomic hierarchy and percentage composition of the microbiome, which includes phyla, classes, families, orders, and species. A) Water, B) spring sediments, and C) wet shore sediments.

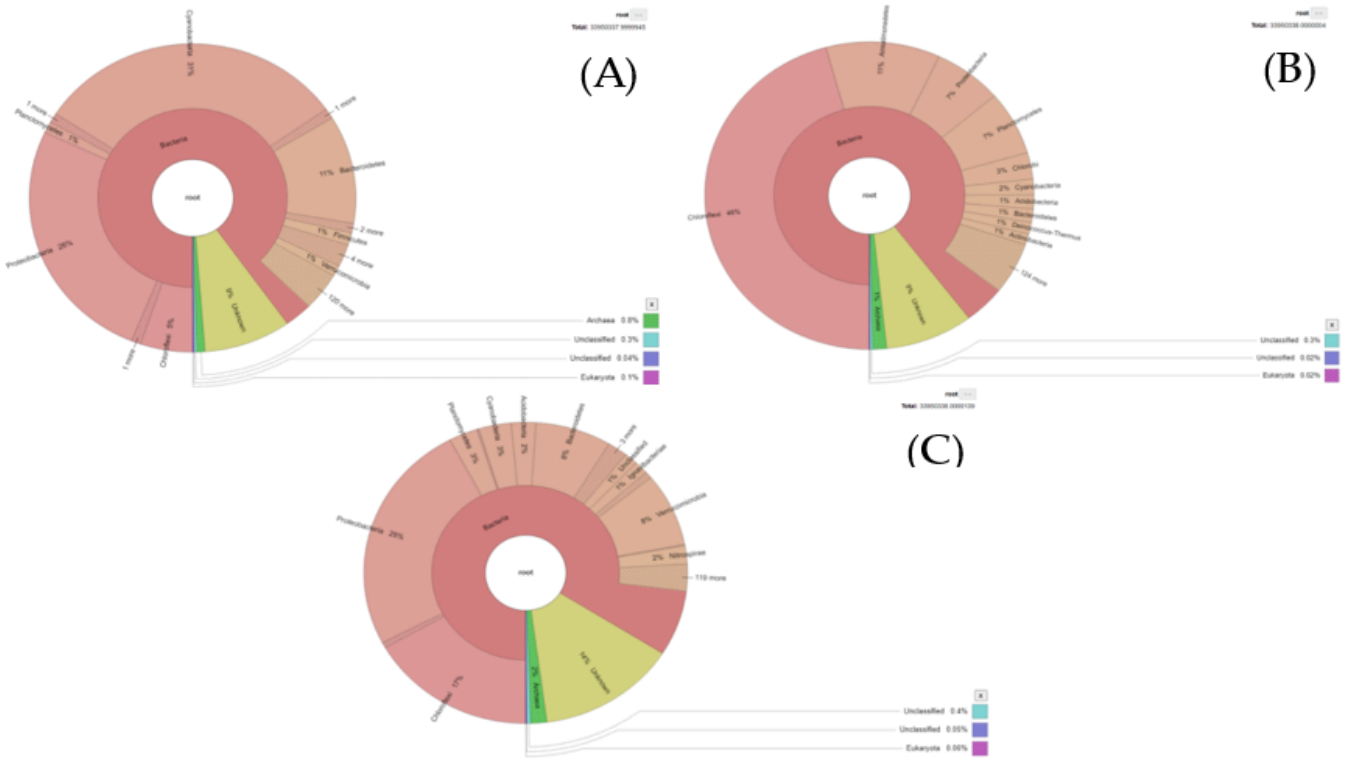

Figure 3

Krona abundance of bacteria domain taxonomic classification in hot-spring A) Water, B) spring sediments, and C) wet shore sediments.

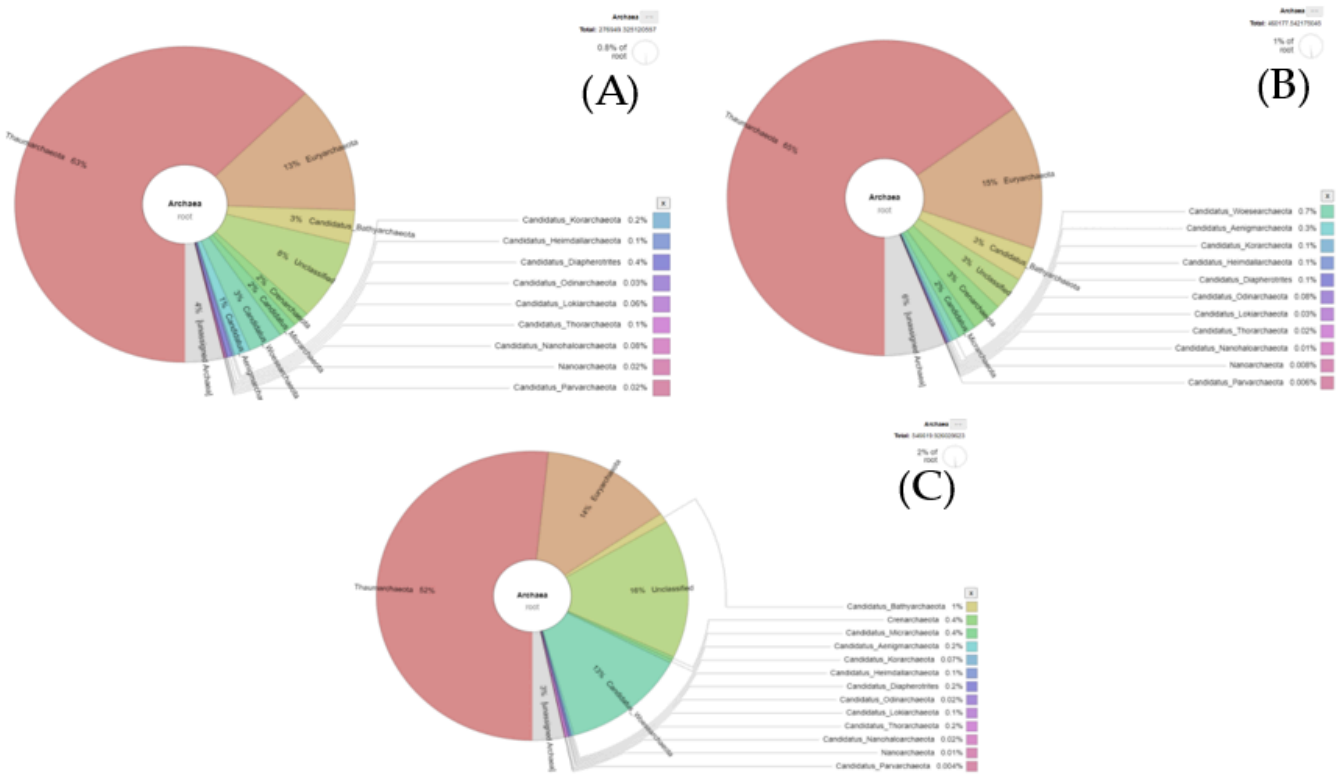

Figure 4

Krona abundance of archaea domain taxonomic classification in hot-spring A) Water, B) spring sediments, and C) wet shore sediments.

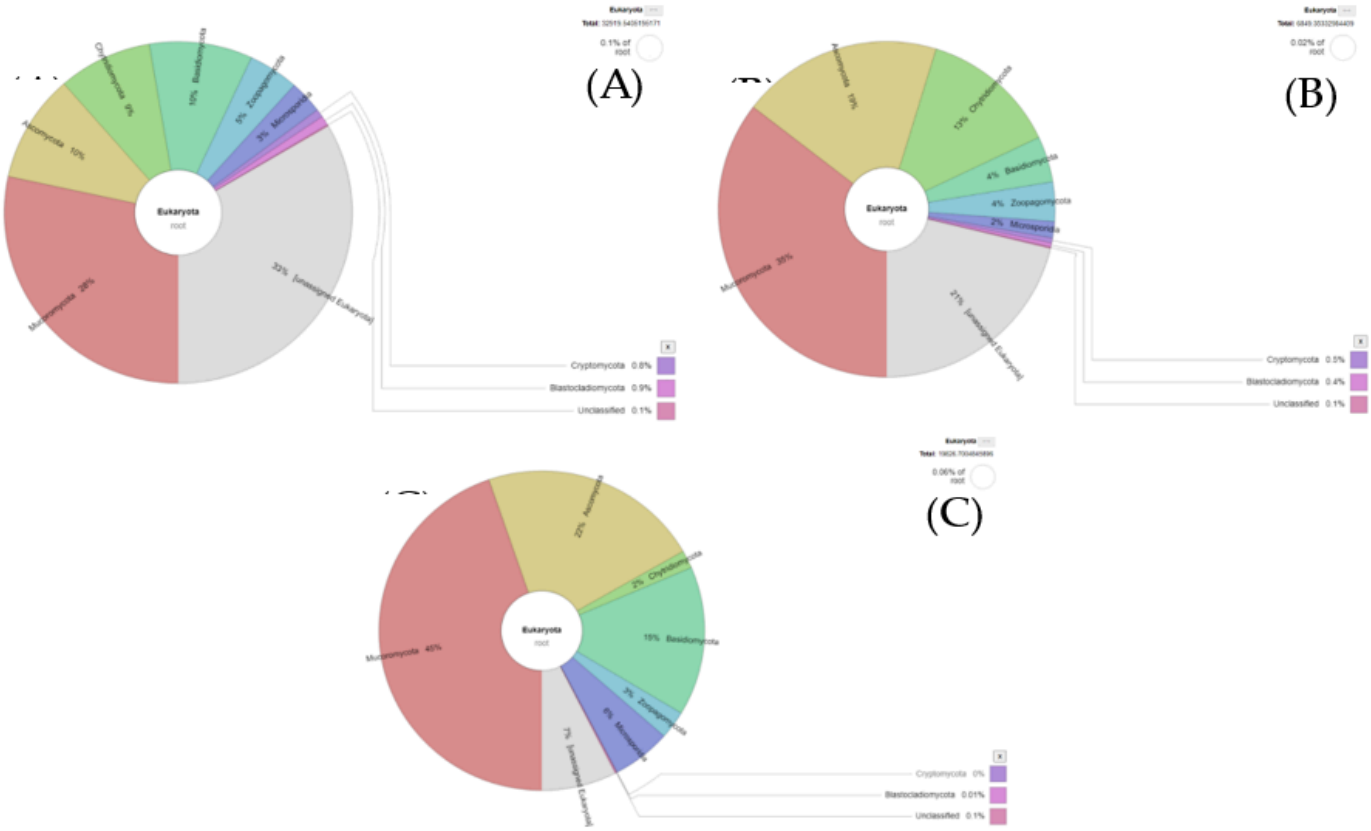

Figure 5

Krona abundance of Eukaryote domain taxonomic classification in hot-spring A) Water, B) spring sediments, and C) wet shore sediments.

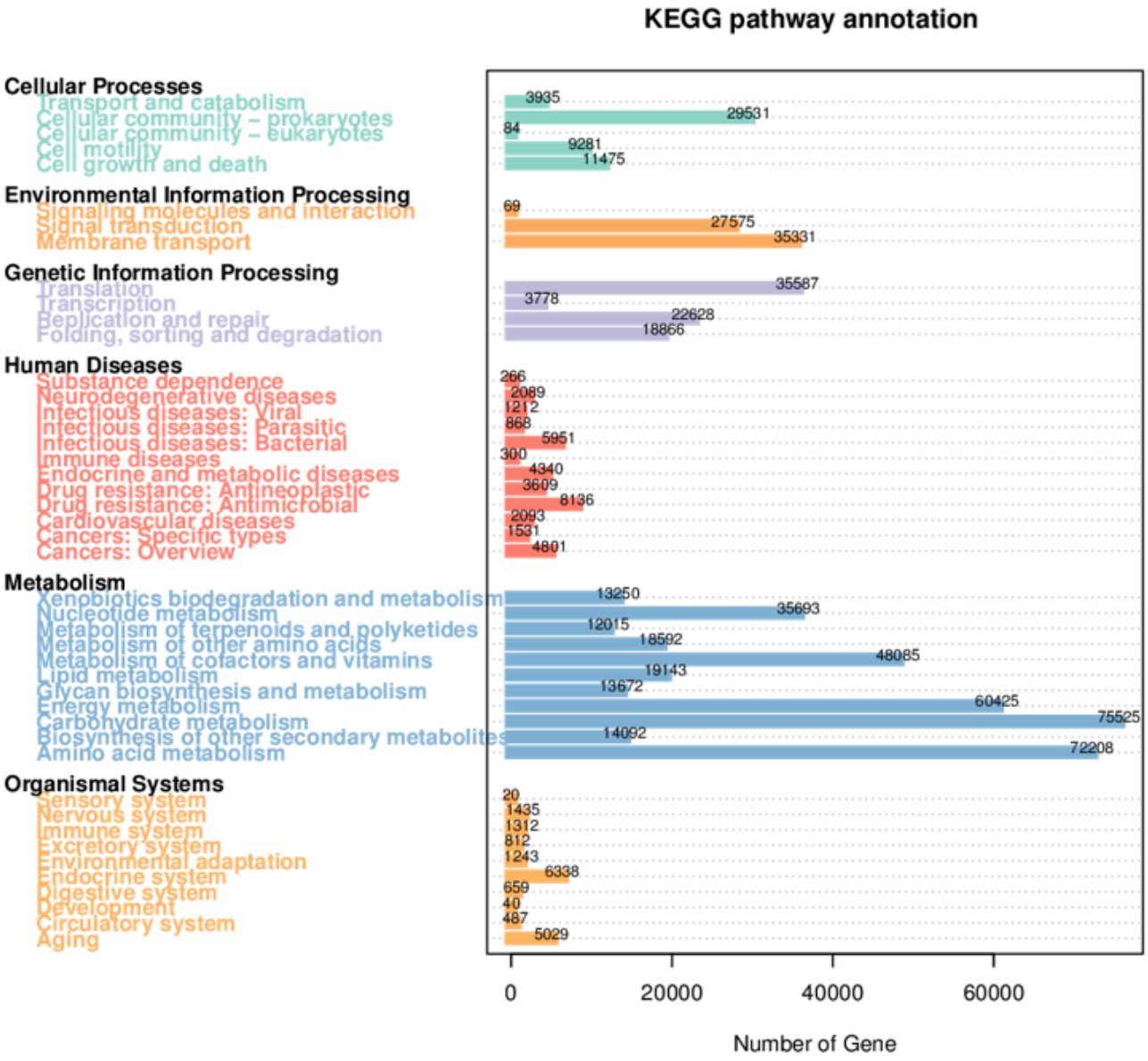

Figure 6

Relative abundance of the KEGG-level functional categories based on shotgun metagenomics data set for Water, spring sediment, and wet shore sediment from Al-Lith hot spring. The highest relative abundance being the metabolism and the lowest organismal systems.

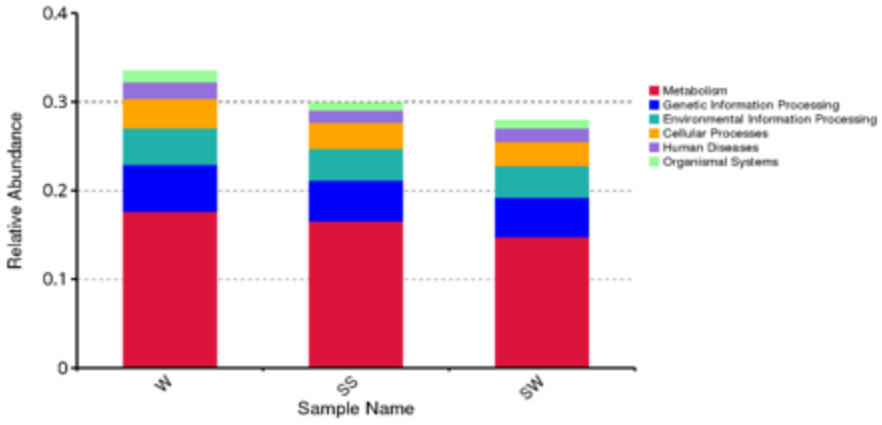

Figure 7

KEGG-level functional categories from Al-Lith hot spring based on shotgun metagenomic data. In terms of relative abundance, metabolism is the highest and organisms are the least abundant.

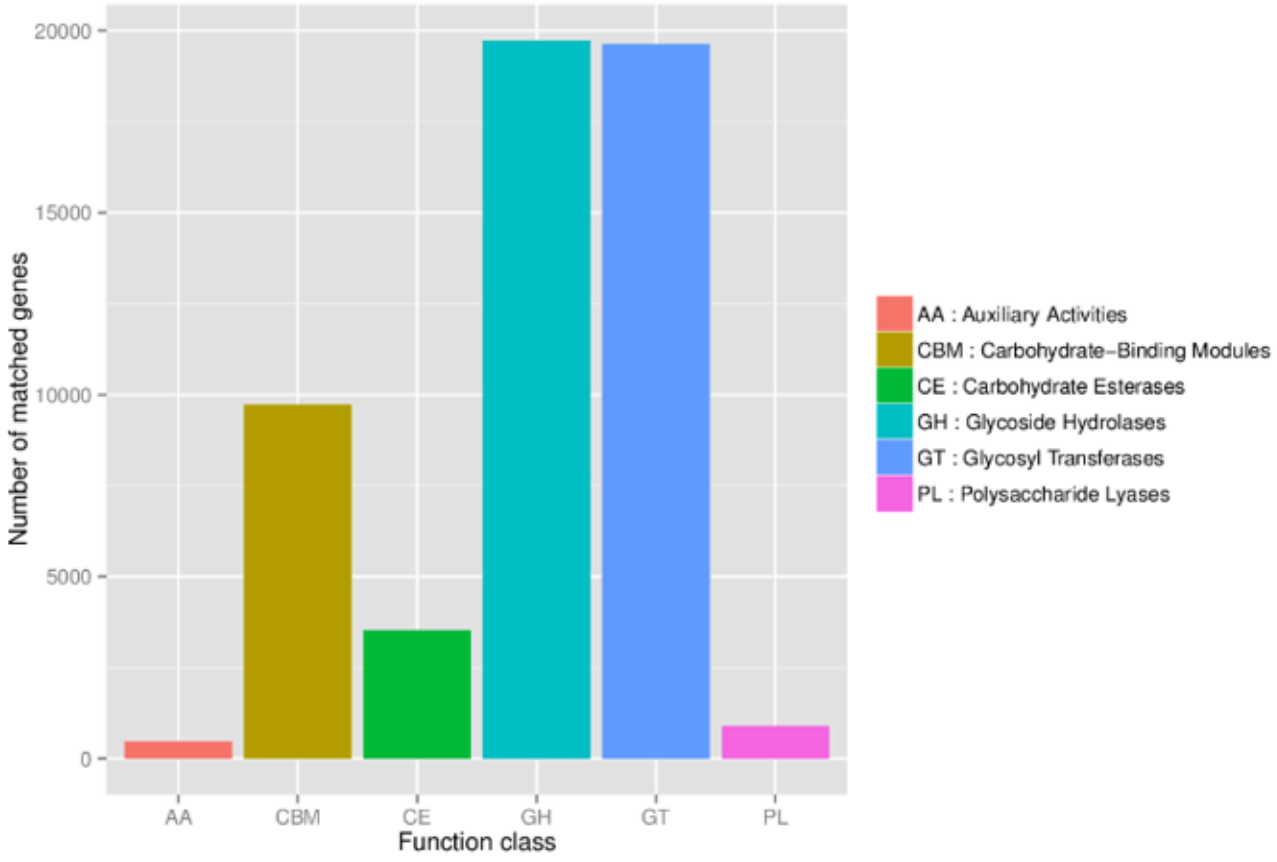

Figure 8

Based on a shotgun metagenomics data set for water, spring sediment, and wet shore sediment, functional groups of operating taxonomic units were predicted. Among the genes with the highest relative abundance were glycoside hydrolases and glycosyl transferases.

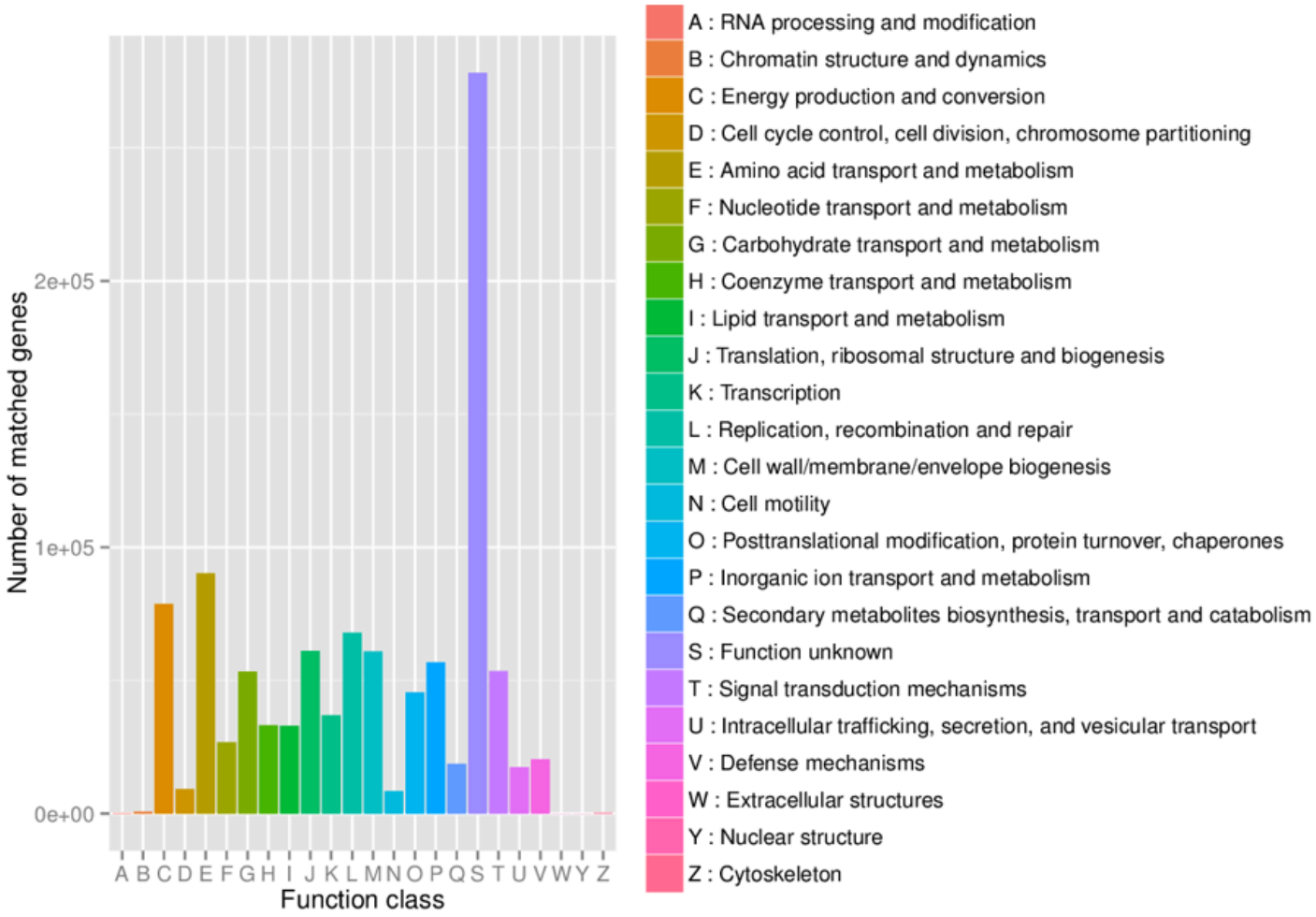

Figure 9

EggNOG and orthologous groups (OGs) of proteins at different taxonomic levels; functional categories based on shotgun metagenomics database for water and spring sediments, and wet shore sediments. The highest genes are pictured along with RNA processing and modification, extracellular structures, and nuclear structures.