

Article

Spatial Metagenomic Analysis in Understanding the Microbial Diversity of Thar Desert

Jagdish Parihar ¹, Suraj P. Parihar ², Prashanth Suravajhala ^{3,4,*} and Ashima Bagaria ^{1,*}

- ¹ Department of Physics, Manipal University Jaipur, Jaipur 303007, India; jagdish.parihar@muj.manipal.edu
- ² Wellcome Centre for Infectious Diseases Research in Africa (CIDRI-Africa), Institute of Infectious Diseases and Molecular Medicine (IDM), Division of Medical Microbiology, Faculty of Health Sciences, University of Cape Town, Private Bag X3, Rondebosch, Cape Town 7701, South Africa; suraj.parihar@uct.ac.za
- ³ Bioclues.org, Vivekananda Nagar, Kukatpally, Hyderabad 500072, India
- ⁴ Amrita School of Biotechnology, Amrita Vishwavidyapeetham, Amritapuri Campus, Clappana P.O., Kollam 690525, India
- * Correspondence: prash@bioclues.org (P.S.); ashima.bagaria@jaipur.manipal.edu (A.B.)

Simple Summary: We present a systematic investigation of the distribution of microbial communities in arid and semi-arid regions of Thar Desert Rajasthan, India. Their responses in multiple environmental stresses, including surface soil, surface water and underground water were evaluated. We further assess the biotechnological potential of native microorganisms and discover functional species with results providing a detailed understanding of the abundance of microbial communities in these regions, associated with various stress-related biogeochemical and biotechnological processes. We hope our work will facilitate the development of effective future strategies for the use of extremophiles in complex environments.

Abstract: The arid and semi-arid regions of Rajasthan are one of the most extreme biomes of India, possessing diverse microbial communities that exhibit immense biotechnological potential for industries. Herein, we sampled study sites from arid and semi-arid regions of Thar Desert, Rajasthan, India and subjected them to chemical, physical and metagenomics analysis. The microbial diversity was studied using V3–V4 amplicon sequencing of 16S rRNA gene by Illumina MiSeq. Our metagenomic analyses revealed that the sampled sites consist mainly of Proteobacteria (19–31%) followed by unclassified bacteria (5–21%), Actinobacteria (3–25%), Planctomycetes (5–13%), Chloroflexi (2–14%), Bacteroidetes (3–12%), Firmicutes (3–7%), Acidobacteria (1–4%) and Patescibacteria (1–4%). We have found Proteobacteria in abundance which is associated with a range of activities involved in biogeochemical cycles such as carbon, nitrogen, and sulphur. Our study is perhaps the first of its kind to explore soil bacteria from arid and semi-arid regions of Rajasthan, India. We believe that the new microbial candidates found can be further explored for various industrial and biotechnological applications.

Keywords: metagenomics; extremophiles; microbial diversity; soil bacteria; evolutionary taxonomy



Citation: Parihar, J.; Parihar, S.P.; Suravajhala, P.; Bagaria, A. Spatial Metagenomic Analysis in Understanding the Microbial Diversity of Thar Desert. *Biology* **2022**, *11*, 461. <https://doi.org/10.3390/biology11030461>

Academic Editor: Ian Blaby

Received: 11 January 2022

Accepted: 7 March 2022

Published: 17 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Extreme environments refer to the habitats that are detrimental to living organisms. These habitats include various physicochemical parameters such as temperature (>40 °C & <5 °C), pH (acidic pH < 5, alkaline pH 9), nutrient content, pressure (0.1 MPa), salinity (35%) and radiation (aw 0.8) [1]. The organisms flourishing under these extremities are called extremophiles. The great Indian Thar desert spans over the vast arid and semi-arid regions in the north-western part of India and offers temperature fluctuations varying from 5 to 55 °C. This extreme habitat poses a threat to most life forms but is also a haven for the flourishing diversity of thermophiles [2]. The optimal temperature range further breaks down the classification of these surviving microbes as mesophiles (37–40 °C); moderately

thermophiles (40–60 °C), thermophiles (60–80 °C) and hyperthermophiles for microbes sustaining above 80 °C [3,4]. Thermophilic bacteria thrive in habitats such as hot springs, geothermal vents and desert soil [5,6]. In terms of soil microbial community structure and function, it has been shown that desert biomes vary significantly from other biomes [7]. The most recent comparison of soil microbial population across different biomes using metagenomic sequencing depicts the lower values of phylogenetic and functional diversity crossing several main taxa [8]. Microbiologists have been studying the diversity of biomes across plants and animals for a long time and it is only in the recent past that the scientific community has started exploring soil microbiomes [9,10]. Microorganisms perform vital functions in environmental ecosystems and are sensitive to environmental stress [11,12]. Microbial diversity of soil and other extreme environments, however, is still limited to only 1–3% of cultivable soil microbes and is estimated to comprise approximately 4000–5000 different bacterial genomic units in 1 gm of soil [13,14]. The recent advancement in next generation sequencing (NGS) has heralded rigorous microbial studies, thereby providing new insights towards understanding microbial diversity, stability, and adaptability under extreme environmentally stressed conditions [15–17]. Microbes within the natural surroundings exist in complicated communities marked by intricate webs of different interactions [18]. The ecological interactions among microbial communities possess properties of positive cooperation or negative antagonism that have an impact on overall microorganism composition [19]. With soil samples possessing huge diversity across different environmental conditions [20,21], it is estimated that there would be approximately 3500–8800 genome equivalents in a pasture soil sample. Given the relative abundance of about 10,000 different species in the soil ecosystem, this immense diversity is a challenge for researchers trying to categorize microbial communities into distinct species [22]. Minor fluctuations in the soil environment tend to have a major impact on the diversity and community structure of the species [23]. Heavy metals, organic contaminants and pesticides are just a few environmental variables which affect the structure of the microbial community, with the parameters that influence microbial composition include particle size, permeability, porosity, water content, mineral composition, and plant cover.

The 16S and 18S rRNA gene sequencing gives insights into the bacterial and fungal diversity, respectively, whereas sequencing the inter transcribed sequence (ITS) region gives information about eukaryotic diversity [24]. With 16S rRNA gene sequencing being a commonly used tool for identifying bacteria, tracing phylogenetic relationships between bacteria, and identifying bacteria from various sources such as soil and water play an important role [25]. On the other hand, the 18S rRNA gene is a prevalent molecular marker for biodiversity research due to its highly conserved nature and slow evolution and thus assists in the analysis of species [26].

The bacterial 16S rRNA gene has nine hypervariable regions (V1–V9). We have selected the V3–V4 region because of the heterogeneity of the region and low evolution compared to other hypervariable regions, which can help better identify the organism. The sequences obtained from the Greengenes, NCBI, RDP, and SILVA databases representing the primers chosen for V1–V2 were significantly smaller (30.3%) than that of V3–V4 (90%), V4 (90.9%), and V4–V5 (87.8%) [27,28]. Whereas V3–V4 fragments do not allow distinguishing some species-level operational taxonomic units (OTU) merging into single species, the clusters could be determined at the level of genetic distances of V3–V4 fragments. The main reason why databases have a major proportion of 16S sequences submitted from the V3–V4 region is because they are produced by V3–V4 region primers (most used primers in microbial community analysis).

Though thermophiles have been widely studied across the globe, an understanding of the biodiversity in desert biomes, specifically with respect to NGS, is scarce. The majority of protocols used for diversification are limited to morphological features. The advancement in next generation sequencing techniques have enabled researchers to characterize the microbial population utilizing these metagenomic techniques for both cultivable microorganisms as well as those that cannot be cultured. These culture independent tech-

niques based on DNA sequencing, do not depend on morphological parameters, and have filled the research gap to a large extent, but they have their own limitations. Asian countries harbour many hot springs which are good sources of thermophiles [29] and researchers have attempted to unlock the microbiome of various soil samples collected from different sites based on a non-redundant inter-transcribed sequence (ITS) region [30]. For example, the impact of agricultural management on soil microbial communities by generating amplicon and shotgun metagenome libraries using metagenomic approaches were studied [31,32]. The first culture independent study of microbial diversity within the hot spring hypersaline lakes Magadi and little Magadi in Kenya was carried out and their metagenomic study mainly focused on the comparison of total versus active microbial communities [33]. Another study showed the identification of the cellulolytic systems and addressed the strategies adopted for the discovery of new cellulases in the field of metagenomics, emphasizing on thermophilic microorganisms [3,34]. Furthermore, comparative metagenomic profiles not only help understand mechanisms for stress but also microbial adaptation to such impending stress associated with ecological patterns such as extreme hot/cold deserts and other niche communities [5]. Nevertheless, very small efforts have facilitated understanding microbes from varied samples in Indian conditions. For example, a study with soil and water samples could allow us to identify the genetic determinants between metagenomes of desert and non-desert soils. In bridging such gaps, our current metagenomic approach focuses on microbial diversity and their abundance in arid and semi-arid regions of the great Indian Thar Desert, Rajasthan, India, by inferring amplicon-based strategy. The goal of this study was to find the abundance of taxa present in the samples targeting the V3–V4 specific gene amplicon sequencing of 16S rRNA across six vivid samples (5 soil and 1 water sample). To the best of our knowledge, this study is perhaps the first of its kind, where thermophilic samples from extreme arid and semi-arid regions of Rajasthan, India have been analysed. We present the unique bacterial taxonomic clades obtained from our samples and discuss the challenges and implications of bacteria living under these extreme conditions.

2. Materials and Methods

2.1. Site Description and Sampling

The soil and water samples were collected maintaining heterogeneity from arid and semi-arid zones of the Thar during peak summer (June) from Jaisalmer, Barmer, Pachpadra and Pokhran, districts of Rajasthan, India (Table S1). The soil samples were collected by sterile spatula and bottle (1 kg for each soil sample and 1 litre water sample), marking the area to 1 × 1 × 1 foot (length, breadth and height) to remove the uppermost soil which may be influenced by environmental factors. The environmental parameters such as elevation, humidity, precipitation and temperature in the summer season were also measured during sampling. This belt is extremely dry and the temperature ranges from 40 °C to 60 °C. At the time of sampling, the average temperature of the soil sample measured was 55 °C. For temperature as an important factor, we observed temperature for three days and collected the samples at the highest temperature. There is no plant coverage in this study as all the samples were collected from an area that is barren or has low human and animal interference. A total of 5 soil samples were covered under the study and 1 water sample as negative control (water samples were collected from 6 sites from natural water reservoirs. At every single site 6 samples were collected and a total of 36 samples from 6 different sites were collected and pooled). In the case of a water sample (1 L each from 6 different sites was collected and pooled). The soil samples were collected and stored at 4 °C. These samples were transferred to the laboratory for further processing.

2.2. Physicochemical Analyses

The collected samples were filtered through 2 mm mesh and air dried to identify physicochemical properties. The chemical characteristics such as Total Organic Carbon (TOC) was determined by a rapid titration method. Available phosphorus and potassium

were measured by spectrophotometer and flame photometer and micronutrients content (iron, copper, zinc and manganese) were detected through atomic absorption spectroscopy (Agilent Technologies, 200 Series AA, Santa Clara, CA, USA) following content using standard operative methods. The physical characteristics of the samples such as pH and EC were evaluated by a multi parameter tester [35–38]. For all components analysed in this research, analytical uncertainties are <5 percent (Table S2).

2.3. DNA Extraction, PCR Amplification, and Illumina MiSeq Sequencing

The Fast DNA Spin Kit for Soil (Qiagen, La Jolla, CA, USA) was used to extract the genomic DNA from all field samples and incubated samples according to manufacturer instructions. A 25 ng of DNA was used to amplify hypervariable V3–V4 regions of total 16S rRNA genes from sediment and aqueous samples [39]. The reaction includes KAPA HiFi HotStart Ready Mix and 100 nm final concentrations of modified 341F and 785R primers [40]. The flowchart of the sequencing protocol followed is shown in Figure 1. The primers were designed to identify areas with amplicon length compatible with the current Illumina MiSeq sequencing platform (2×300 bp) by first examining the 16S rRNA gene in silico, and then targeting conserved and variable regions. The conserved regions were screened against the non-redundant database of bacterial model systems (*E. coli*) of NCBI BLAST and validated for their suitability in amplifying the 16S rRNA gene V3–V4 region. The full-length primers were designed using dual-index methods [41]. While designing full-length primers, each forward and reverse primers have an adaptor sequence (provided by NGS Company), four random nucleotide sequences followed by an eight-nucleotide indexing sequence, followed by a 16S rRNA V3–V4 unique forward priming sequence of 17 nucleotides (CCTACGGGNGGCWGCAG) and the reverse primer sequence containing 21 nucleotides (GACTACHVGGGTATCTAATCC). The purpose of the adapter sequence is to allow the DNA sequence attachment to the MiSeq flow cell and the connector sequence to increase the overall temperature of melting. These degenerate primers collections were further used with site-specific adaptor sequences to demonstrate the functional heterogeneity of soil and water samples. PCR was subsequently set with an initial denaturation of 95 °C for 5 min followed by 25 cycles of 95 °C for 30 s, 55 °C for 45 s and 72 °C for 30 s and a final extension at 72 °C for 7 min. The amplicons were purified using Ampure beads to remove unused primers. An additional 8 cycles of PCR were performed using Illumina barcoded adapters to prepare the sequencing libraries. The libraries were quantitated (Table S3), using Qubit DNA HS quantitation assay (Thermo Scientific, Waltham, MA, USA), which specifically quantitates dsDNA assay.

2.4. Technical Validation

The filtered dataset for the V3–V4 16S rRNA fragment included 1,362,076 reads with the average length of 301 base pairs. We obtained 6 clusters with a total of 617 species-level taxonomic clades wherein the V3–V4 dataset included 681,038 reads with an average length of 344 base pairs. For analysing the convergence, proportions of underestimated OTUs were checked for biological studies and validation. We removed all the duplicate, unwanted and unknown sequences and analysed 656 different sequences. Out of these, our analyses confirmed 64 unique sequences. Alpha diversity richness was demonstrated by heat maps and principal component analysis (PCA) plots while beta diversity richness was indicated by a phylogenetic tree.

The two sets of reads for every pattern (forward and reverse reads) were blended using (make.contigs) command. Sequences that failed to fulfil any one of the following standards were excluded, whereas an average length of 344 bases was analysed for the presence of ambiguities and nucleotide mismatch (>1) to the primer using (screen.seqs) command and duplicate sequences were further removed using (unique.seqs) command from the analyses. The pcr.seqs was used to infer V3–V4 vicinity of 16S rRNA genes wherein unique sequences were aligned against the SILVA reference database. By using (filter.seqs) command, gap characters had been pulled out without losing any facts in alignment of

columns. We then used the (pre.cluster) command to allow at most two variations between sequences followed by identifying chimeras using UCHIME algorithm (chimera.uchime). The sequences had been taxonomically classified using the Naive Bayesian classification with 80% confidence threshold via the use of (classify.seqs) command. The sequences that were not categorized to any one of the domains (unknown) or labelled in Chloroplast, Mitochondria, Eukaryota, and Archaea were eradicated using (remove.lineage) command, wherein each of these sequences are counted as a separate community.

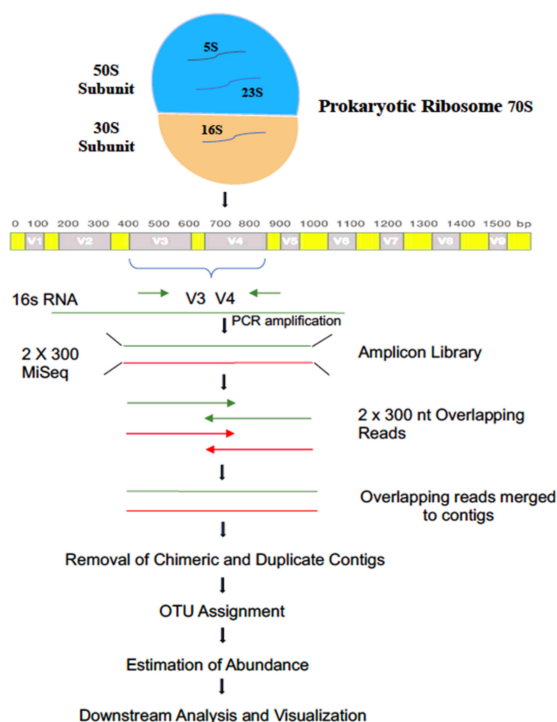


Figure 1. Flowchart depicting 16S rRNA metagenomic workflow.

2.5. Phylotype Analysis

The analysis of sequences into phylotypes as per taxonomic characterization was carried out using distance matrix sequences clustered into OTUs with cut-off 97% similarity (3% dissimilarity) using (dist.seqs) and (cluster) commands. The cut-off numbering of the phylotypes rises to four, which compared to the arranged level utilized in (cluster.split) command. In each OTU, the wide range of sequences was identified by using (make.shared) command and the unique OTUs were arranged using the (classify.otu) command.

2.6. Sequence Data Quality Control

The sequence data quality was checked using FastQC [42] and a consensus was reached using MultiQC [43]. The data was checked for base call quality distribution, bases% above Q20, Q30, GC%, and sequencing adapter contamination (Table S4 and Figure S1a,b). All the samples have passed QC threshold (Q20 > 95%) with the average GC percentage found to be 55% among all the samples.

2.7. NGS Analyses

The reads were trimmed (20 bp) from 5' end to remove the degenerate primers. These trimmed reads were then processed to remove adapter sequences and low-quality bases using Trim Galore [44]. The quality of raw reads was ensured using FastQC [42] and the analysis was carried out using Mothur pipeline [45]. The pairs were aligned with the contigs screened for errors and only those between 300 bp and 532 bp were retained. While contigs with ambiguous base calls were rejected, the high quality contigs were retained for

checking identical sequences, and duplicates were merged. The Contig length distribution was shown in (Figure S2) Although the primers for the experiment were designed for 16S bacterial rRNA, the chances for non-specific amplification of other regions cannot be neglected. To correct this, we aligned the contigs to a SILVA rRNA database for 16S rRNA. Depending on the variable region being amplified, most and contigs aligning to other regions on the database were discarded. Further the gaps and the overhangs from the contigs were removed and processed for removal of the artifacts or chimeras, which may have been formed due to PCR errors. UCHIME algorithm [46] was used to flag contigs with chimeric regions. A known reference of all the chimeric sequences was used to identify and remove possible chimeric sequences and shown in Figure 1.

The filtered contigs were processed and classified into taxonomic outlines based on the Silva v.132 database [47]. The contigs were then clustered into OTUs and the abundance was further estimated. Sample dissimilarities were calculated by Bray–Curtis matrices. The Mothur was used to generate curves of rarefaction (Figure S3) and Alpha-diversity [48,49]. With non-metric multidimensional scaling (NMDS) methodology [PAST software program, (v.3.16)] employed to display bacterial communities. The heatmaps were generated for top 20 phylum and genus normalized OTUs using heatmap (R package), while principal component analysis (PCA) was carried out using ClustVis (Beta) [50].

3. Results and Discussions

3.1. Microbial Structure, Diversity and Richness

Microbial community structures from different environmental samples were diverse, with relatively lower richness and diversity found in soil samples as compared to water samples. *Proteobacteria* were widespread and dominated microbial communities in all samples. The structure of microbial communities changed considerably during cultivation. In our soil and water samples, we found a microbiome consisting mainly of *Proteobacteria* (19–31%) in abundance, followed by unclassified bacteria (5–21%), then *Actinobacteria* (3–25%), *Planctomycetes* (5–13%), *Chloroflexi* (2–14%), *Bacteroidetes* (3–12%), *Firmicutes* (3–7%), *Acidobacteria* (1–4%) and *Patescibacteria* (1–4%), respectively, shown in (Tables S5 and S6). The heatmap shows distribution of top 20 Phyla with *proteobacteria* being abundant in Sample-6 and *Actinobacteria* abundant in Sample-2, compared to other samples shown in Figures 2a and 3a. Predominantly, we found *Proteobacteria* and *Acidobacteria* which have potential biotechnological applications such as paper and pulp industry, textiles industry, soap and detergents industry, leather processing industry, dairy industry, food and beverages, dairy industry, baking industry, meat industry, pharmacology and drug manufacturers, biofuel production, biomining, medicines, bioremediation, anti-inflammatory, antimicrobial, anti-cancer and also exploitation for different therapeutic applications.

The heatmap represents the distribution of the top 20 Genus, *Gemmataceae*; uncultured is more abundant in Sample-1. Similarly, *Curvibacter* is more abundant in sample-6 as compared to other samples shown in Figures 2b and 3b. In the heatmap, blue colour signifies less abundant, light blue and pale-yellow mean moderate and red colour signifies the most abundant of the species. This study was carried out to determine the overall similarity between the composition of microbial population in soil and water samples.

The stacked column bar graph represents the microbial community composition at the phylum level across all samples used in this study. Taxonomic identities that could not be shown to the respective level of resolution were considered as “unclassified” within their corresponding domain. Relative abundance data were analysed using MG-RAST against the SEED database and then visualized using Microsoft Excel. Sample names are included in the plot (sample-1 = soil; sample-2 = soil; sample-3 = soil; sample-4 = soil; sample-5 = soil; sample-6 = water).

Predicted community composition of each assembly from sample-1 to sample-6, coloured by genus, with proportion defined by summed length of contigs assigned to that genus. Deeper subsampling dramatically changes the community composition.

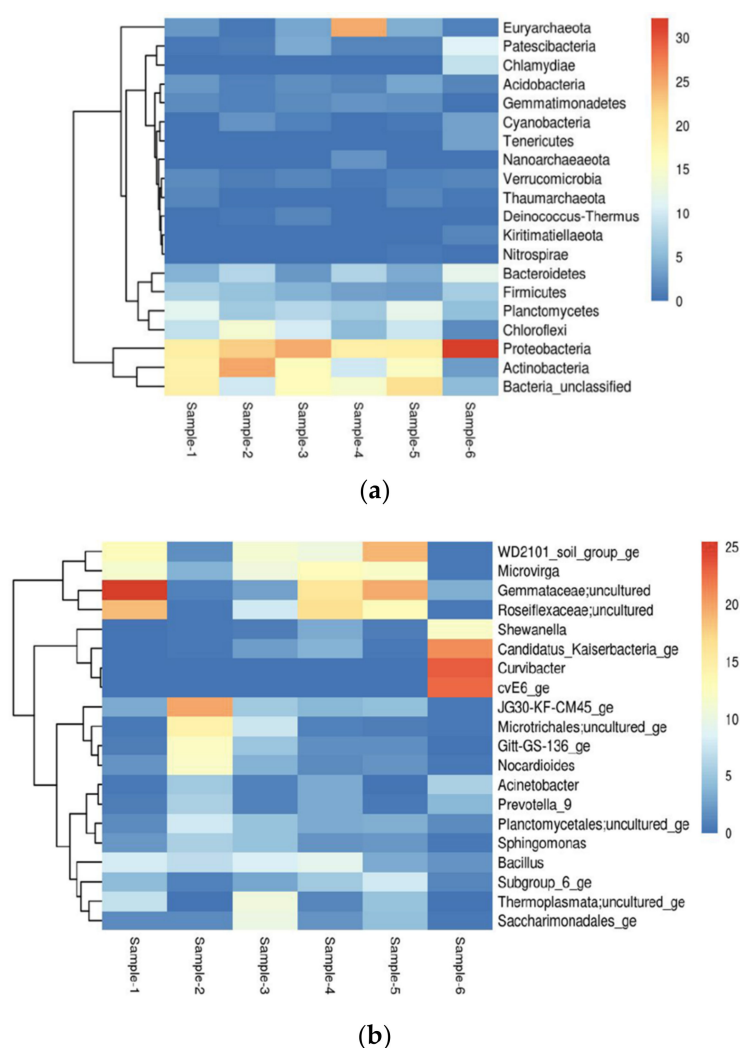
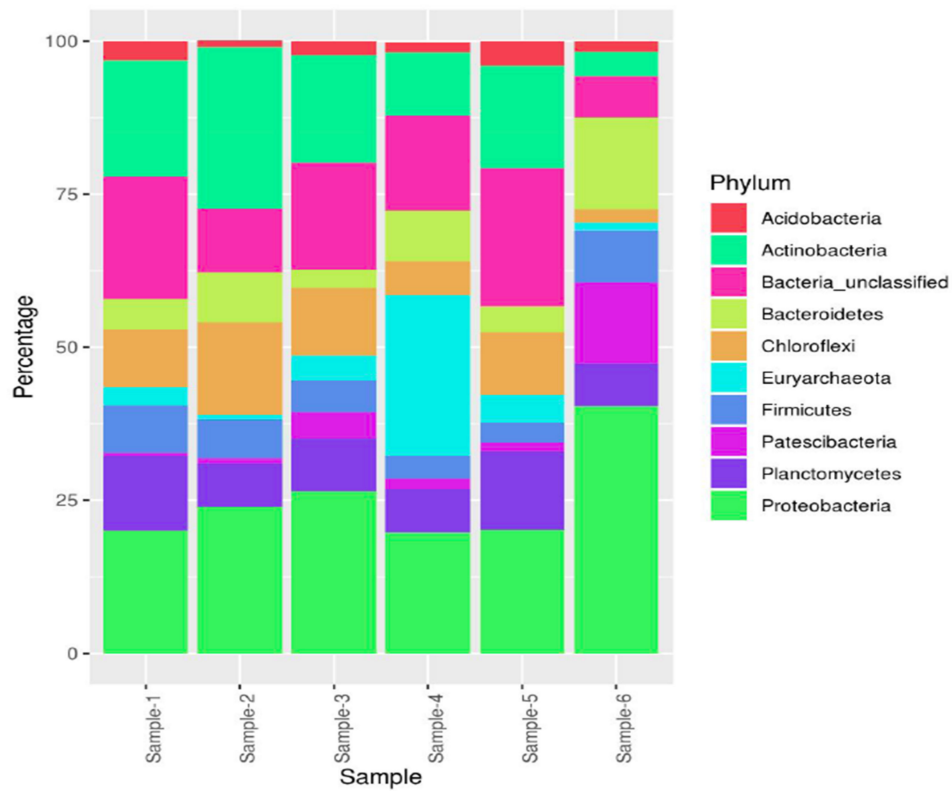


Figure 2. (a) Top 20 Phylum abundance distribution. (b) Heatmap representing distribution of top 20 Genus.

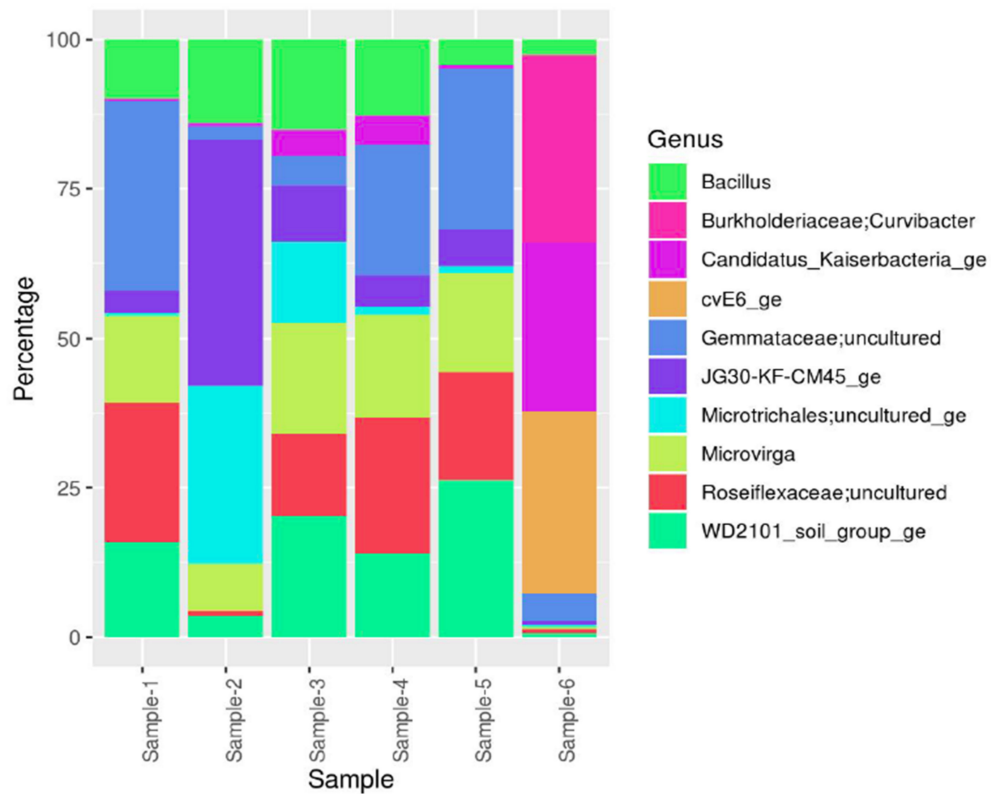
Microbial population data consisted of (percent) normalized OTUs. Graphical representations along the PC1 (33.8%) and PC2 (26.3%) axis had a 60.1 percent average variance. Sample 6 was widely distributed in different taxa as compared to sample 3, followed by sample 4, 5, 2 and 1, with sample 1 being the least abundant and shown in Figure 4. The Venn diagram showing the number of taxa shared by, or unique to, the different samples were built using the Venn and galaxy project (URL: usegalaxy.org, accessed on 27 February 2022) [51,52]. The most common species in all samples was *Wigglesworthia brevipalpis*, shown in (Figure S4a–o).

3.2. Alpha Diversity and Beta Diversity Analysis

Alpha diversity is a measurement of richness and relative abundance of bacteria within the sample. Chao1 and ACE indices represent the richness of the samples and Shannon, Simpson, InvSimpson and Fisher indices represent both richness and relative abundance shown in Figure 5a. The alpha diversity calculations were performed using Phyloseq an R package (vegan 2.4.4) for metagenome analysis. Using the alpha rarefaction.py script in QIIME, we calculated alpha OTU diversity by randomly sub-sampling (without substitute) each soil and water sample. In addition to the observed amount of OTUs, the Chao1, Shannon, phylogenetic diversity (PD) and Evenness indices of Faith were calculated.



(a)



(b)

Figure 3. (a) Top 10 Phyla abundance distribution. (b) Top 10 Genus abundance in coloured maps.

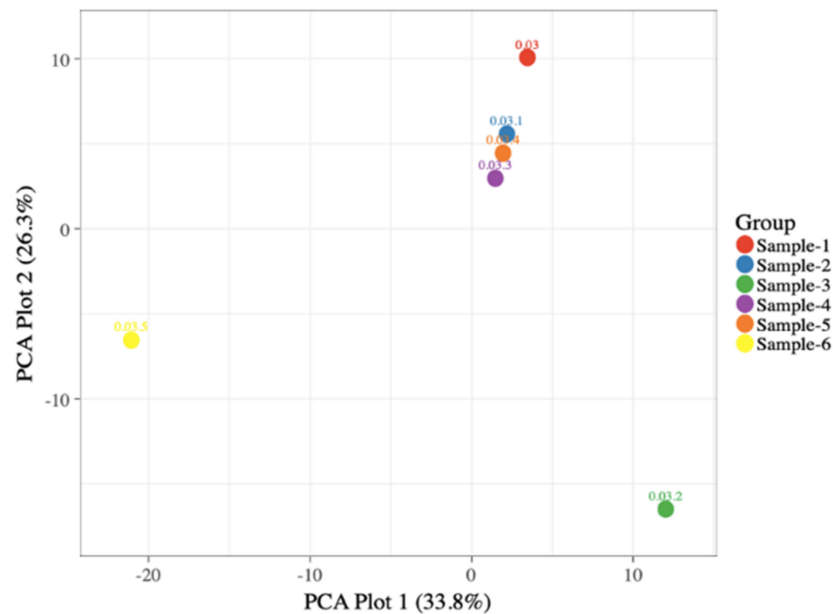


Figure 4. Principal component analysis (PCA). PCA plot illustrating correlation among six samples (five soil samples and one water soil sample) from arid and semi-arid regions of Rajasthan.

The rarefaction gives the diversity but not the rich coverage and, hence, the differences in coverage are expected after sampling a large number of individuals with select 1000 sequences from each sample, and so the average was calculated. While the quantitative estimation of coverage served as a basis for the adjustment of tests from the distance matrix/phylogeny tree, we ascertained the measures from binned fragments.

3.3. Comparative Analysis from Different Deserts

In the recent past, the investigation of soil microbiomes in different geographical regions of the world has been carried out by NGS. Some of the studies mentioned here show the presence of a diverse microbial population in desert soils. For example, the arid desert soils of Kazakhstan, Atacama Desert, Namib, Gobi, Thar, Negev, Sonoran, and Mojave deserts are highly diverse, dry and arid with low humidity and intense radiation, while these physical parameters allow special microbial communities to flourish in these extreme conditions. The dominant phyla present in these deserts are mainly Proteobacteria, Actinobacteria, Firmicutes, Chloroflexi, Verrucomicrobia, Acidobacteria, Planctomycetes, cyanobacteria and Bacteroidetes, respectively. The functional genes related to these taxa play an important role in biogeochemical cycles [53–60]. In our study we have also shown that Proteobacteria and Actinobacteria phyla are most abundant beside other phyla, respectively, among different soil samples. Therefore, we can conclude from above studies that our microbial diversity in the Thar desert is similar to the diversity of deserts in different geographical regions, though a thorough study is required to establish certain facts and figures with other desert biomes.

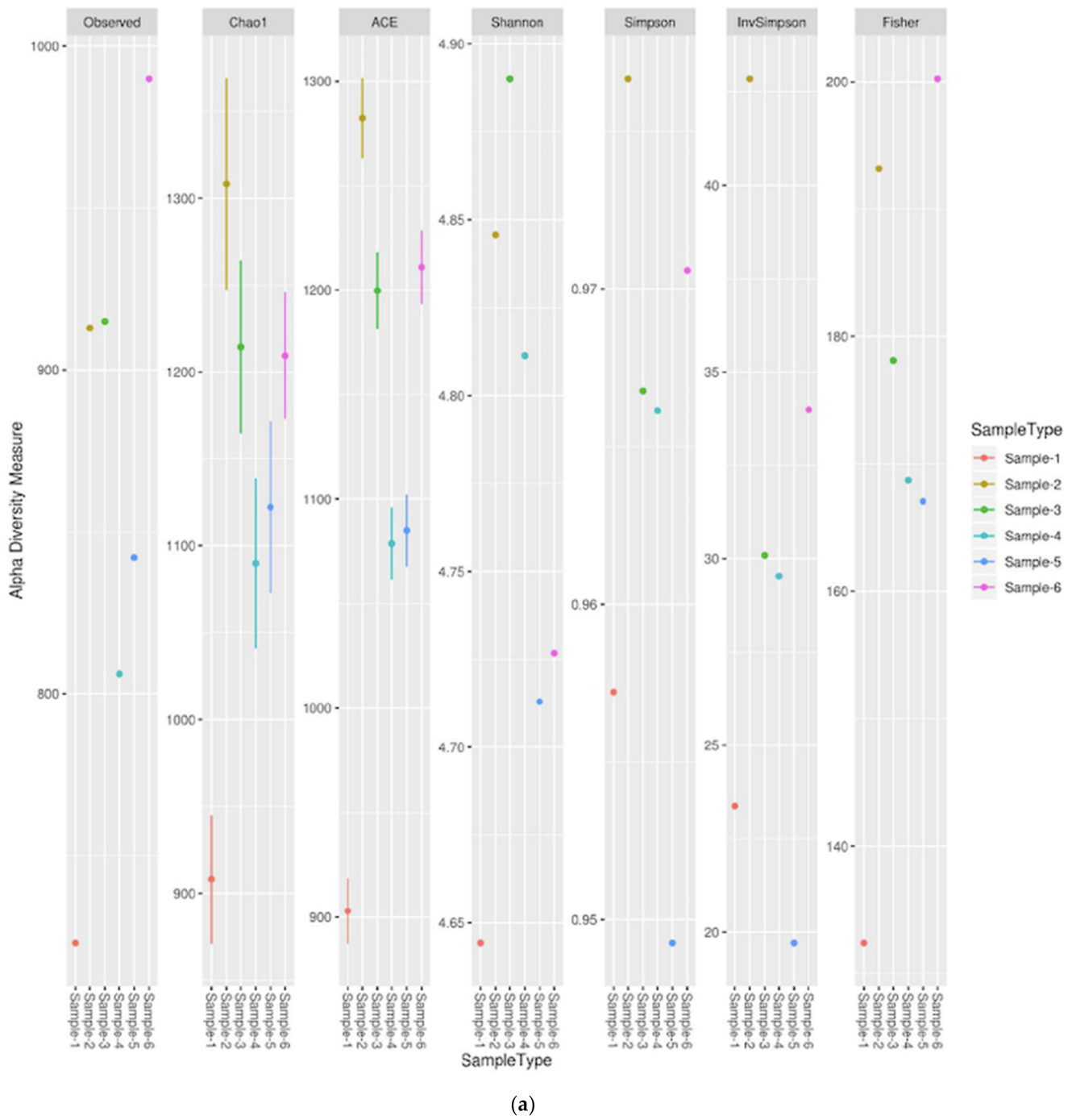


Figure 5. Cont.

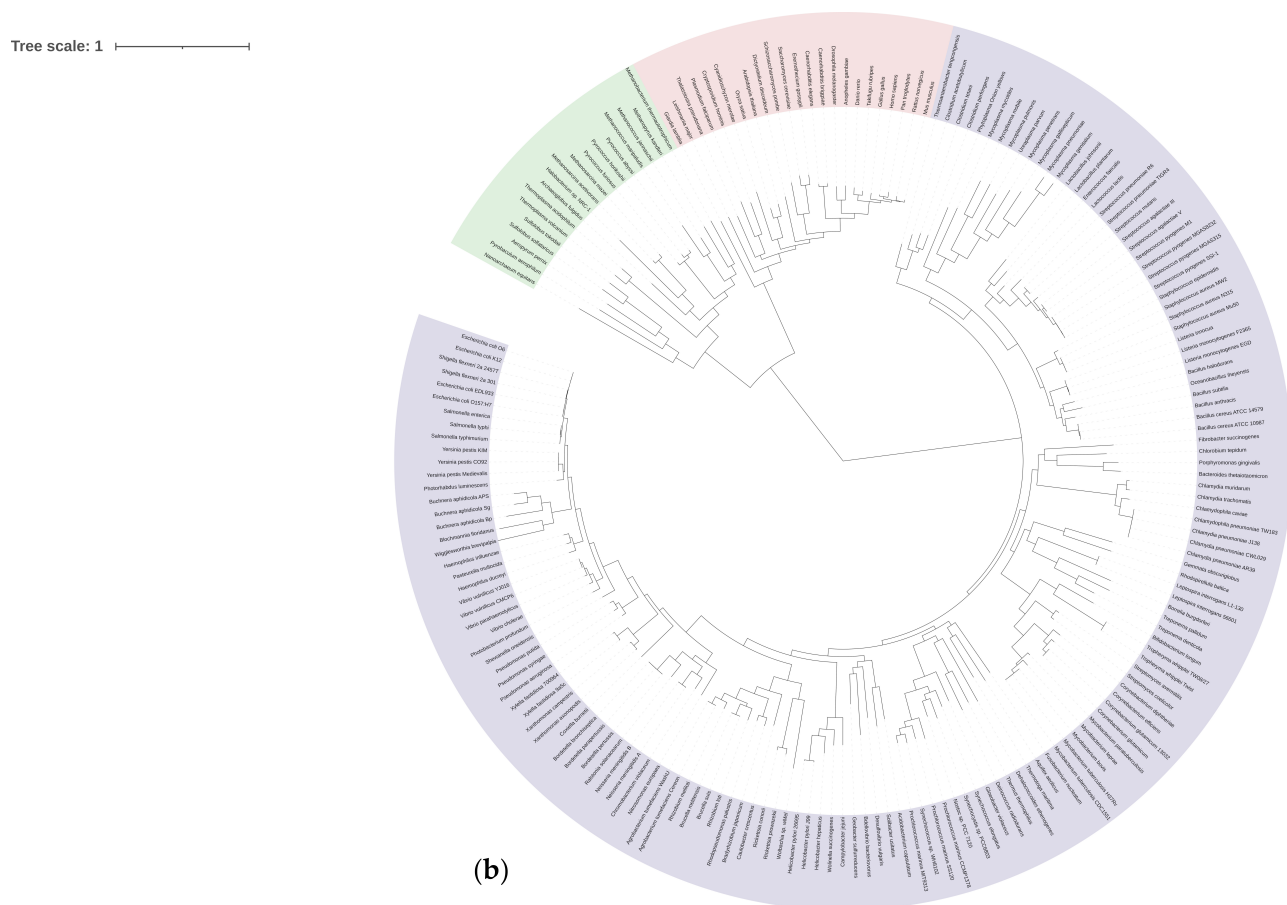


Figure 5. (a) Alpha diversity measurements relative to various statistical thresholds. Beta diversity refers to the ratio between local or alpha diversity and regional diversity. This is the diversity of species between two habitats or regions. The relationship between species and genomes is identified by the phylogenetic tree of life. The centre displays the last universal ancestor of all life on earth. The three domains of life are shown by different colours: pink represents eukaryota (animals, plants and fungi); blue represents bacteria; and green represents archaea. Taxonomic clades of six samples spanning from arid and semi-arid habitats showing the root clade. These figures are produced by iTOL and shown in Figure 5b [61]. (b) Taxonomic clades of six samples spanning from arid and semi-arid habitats showing the root clade [Color code: Bacteria-purple; Pink-Eukaryotes; Green-Archaea].

4. Conclusions and Future Aspects

The present study thus presents a systematic investigation of the distribution of microbial communities in arid and semi-arid regions of Thar Desert Rajasthan and their responses in multiple environmental stresses, including surface soil, surface water and underground water. It also deals with high temperature, salt concentration, and pH stress to assess the biotechnological potential of native microorganisms and to discover functional species. Our results provide a detailed understanding of the abundance of microbial communities in these regions, associated with various stress-related biogeochemical and biotechnological processes that will facilitate the development of effective future strategies for the use of extremophiles in complex environments.

The results obtained are among the few metagenomic diversity studies available for this region. The study displays the diversity of largely unknown soil microbiomes by preliminary analysis and further exploration is required to quantify the presence of microbes and their mechanism of function and stability. In this pilot study, we have majorly focused on how different environmental conditions affect the microbiome of any region. We have covered the maximum area for the study by collecting samples from different geographical locations which consist of different microbial populations. Hence, we have

shown the microbial composition of various sites. The characteristics of soils under various lands are usually different. The proportion of habitat disturbance produced by variations in land-use management methods, which impacted soil characteristics, was found to influence microbial diversity. This research offers up new possibilities for metagenomics research in Rajasthan's semi-arid habitats, and it contributes to the discovery of soil bacteria that are important for ecosystem processes.

Further evaluation of our samples will help us to enhance our knowledge of microbial composition and features in soil and learn about their one-of-a-kind pathways and metabolism. Future efforts will be directed towards identifying species and metabolisms associated with it in each geographic region. In addition, the datasets can be used in future for worldwide soil metagenomic projects for comparative purposes. Additional experimental and sequencing efforts will help in understanding soil microbial dynamics and the specific interacting microbes that would lead to enhancement in cutting-edge agricultural and soil sustainability practices.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/biology11030461/s1>, Figure S1: Histogram representing contig length distribution; Figure S1: Histogram representing contig length distribution; Figure S2. Rarefaction curve shows the measure of diversity that has been captured by a given number of reads in a six sample; Figure S3 (a). Histogram of reads with average sequence quality scores; (b). Histogram of reads with average base sequence quality scores; Figure S4 (a–h). Venn Diagram: Selective enrichment of microbial features at arid and semi-arid habitat. The Venn diagrams illustrate analysis of dominant features for arid and semi-arid habitats Table S1: Site information and Sample type; Table S2. Soil profiling of soil and water samples; Table S3. Library quantity estimate using Qubit HS kit; Table S4. Summary of Raw sequence data and quality; Table S5. Top 10 OTU's abundance among the samples; Table S6. Top 10 Phyla abundance among the samples.

Author Contributions: J.P. did the sampling and processed the samples in the lab; J.P. performed the bioinformatic processing of sequence data; Conceptualization, J.P., P.S. and A.B.; Data curation, J.P. and P.S.; Formal analysis, J.P. and P.S.; Investigation, J.P., S.P.P., P.S. and A.B.; Methodology, J.P. and P.S.; Supervision, P.S. and A.B.; Validation, J.P., S.P.P., P.S. and A.B.; Writing—original draft, J.P.; Writing—review and editing, J.P., S.P.P., P.S. and A.B. All authors participated in the experimental design, discussions about data interpretation and manuscript supervision. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Department of Science and Technology, Rajasthan, India Grant Reference No.: P.7(3) DST/B T R & D/E A C/2018/3145. SPP is supported by Wellcome Trust Grant Number 203135Z/16/Z.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The nucleotide paired-end sequences have been deposited in NCBI SRA database with accession ids: SRR14334654, SRR14334655, SRR14334656, SRR14334657, SRR14334658, and SRR14334659 for 6 samples for 16S rRNA (V3-V4 region) amplicon reads under the Bio Project PRJNA725519.

Acknowledgments: Jagdish Parihar thanks Manipal University Jaipur for Enhanced PhD Scholarship Ref: EF/2017-18/QE 02-11. We thank Birla Institute of Scientific Research, Jaipur, Rajasthan for providing the facilities to analyse the data.

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

References

1. Kumar, S.; Suyal, D.C.; Yadav, A.; Shouche, Y.; Goel, R. Microbial diversity and soil physiochemical characteristic of higher altitude. *PLoS ONE* **2019**, *14*, e0213844. [[CrossRef](#)] [[PubMed](#)]
2. Ghadikolaei, K.K.; Gharechahi, J.; Haghbeen, K.; Akbari, N.K.; Hosseini, S.G.; Shahbani, Z.H. A cold-adapted endoglucanase from camel rumen with high catalytic activity at moderate and low temperatures: An anomaly of truly cold-adapted evolution in a mesophilic environment. *Extremophiles* **2018**, *22*, 315–326. [[CrossRef](#)]

3. Escuder-Rodríguez, J.-J.; DeCastro, M.E.; Cerdán, M.E.; Rodríguez-Belmonte, E.; Becerra, M.; González-Siso, M.I. Cellulases from thermophiles found by metagenomics. *Microorganisms* **2018**, *6*, 66. [[CrossRef](#)] [[PubMed](#)]
4. Bagaria, A.; Parihar, J. The Extremes of Life and Extremozymes: Diversity and Perspectives. *Acta Sci. Microbiol.* **2019**, *3*, 13.
5. Le, P.T.; Makhalanyane, T.P.; Guerrero, L.D.; Vikram, S.; Van de Peer, Y.; Cowan, D.A. Comparative metagenomic analysis reveals mechanisms for stress response in hypoliths from extreme hyperarid deserts. *Genome Biol. Evol.* **2016**, *8*, 2737–2747. [[CrossRef](#)] [[PubMed](#)]
6. Parihar, J.; Bagaria, A. Isolation, Identification and Characterization of a Thermo-halotolerant *Aeribacillus Pallidus* SHJP4, from the Arid Region of Rajasthan. *IOP Conf. Ser. Earth Environ. Sci.* **2021**, *796*, 012050. [[CrossRef](#)]
7. Koo, H.; Hakim, J.A.; Morrow, C.D.; Crowley, M.R.; Andersen, D.T.; Bej, A.K. Metagenomic analysis of microbial community compositions and cold-responsive stress genes in selected Antarctic lacustrine and soil ecosystems. *Life* **2018**, *8*, 29. [[CrossRef](#)]
8. Cowan, D.A.; Ramond, J.B.; Makhalanyane, T.P.; De Maayer, P. Metagenomics of extreme environments. *Curr. Opin. Microbiol.* **2015**, *25*, 97–102. [[CrossRef](#)]
9. Delattre, H.; Quémener, D.L.; Duquennoi, C.; Filali, A.; Bouchez, T. Consistent microbial dynamics and functional community patterns derived from first principles. *ISME J.* **2019**, *13*, 263–276. [[CrossRef](#)]
10. Jagdish, P.; Ashok, K.; Kamakhya, P.M.; Ashima, B. Extracellular Synthesis of Zinc Oxide Nanoparticles Using Thermo-Halotolerant *Aeribacillus pallidus* strain SJP 27: Characterization and Antibacterial Potential. *J. Nano Electron. Phys.* **2021**, *13*, 02007.
11. Batson, J.; Noe, G.B.; Hupp, C.R.; Krauss, K.W.; Rybicki, N.B.; Schenk, E.R. Soil greenhouse gas emissions and carbon budgeting in a short-hydroperiod floodplain wetland. *J. Geophys. Res. Biogeosci.* **2015**, *120*, 77–95. [[CrossRef](#)]
12. Li, C.; Fultz, L.M.; Moore-Kucera, J.; Acosta-Martínez, V.; Kakarla, M.; Weindorf, D.C. Soil microbial community restoration in Conservation Reserve Program semi-arid grasslands. *Soil Biol. Biochem.* **2018**, *118*, 166–177. [[CrossRef](#)]
13. Delmont, T.O.; Robe, P.; Cecillon, S.; Clark, I.M.; Constancias, F.; Simonet, P.; Hirsch, P.R.; Vogel, T.M. Accessing microbial diversity for soil metagenomic studies. *Appl. Environ. Microbiol.* **2011**, *77*, 1315–1324. [[CrossRef](#)] [[PubMed](#)]
14. Mocali, S.; Benedetti, A. Exploring research frontiers in microbiology: The challenge of metagenomics in soil microbiology. *Res. Microbiol.* **2010**, *161*, 497–505. [[CrossRef](#)]
15. Rondon, M.R.; August, P.R.; Bettermann, A.D.; Brady, S.F.; Grossman, T.H.; Liles, M.R.; Loiacono, K.A.; Lynch, B.A.; MacNeil, I.A.; Minor, C.; et al. Cloning the soil metagenome: A strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **2000**, *66*, 2541–2547. [[CrossRef](#)]
16. Zhang, J.; Chiodini, R.; Badr, A.; Zhang, G. The impact of next-generation sequencing on genomics. *J. Genet. Genom.* **2011**, *38*, 95–109. [[CrossRef](#)]
17. Knight, R.; Vrbanac, A.; Taylor, B.C.; Aksenov, A.; Callewaert, C.; Debelius, J.; Gonzalez, A.; Kosciolek, T.; McCall, L.I.; McDonald, D.; et al. Best practices for analysing microbiomes. *Nat. Rev. Microbiol.* **2018**, *16*, 410–422. [[CrossRef](#)]
18. Faust, K.; Raes, J. Microbial interactions: From networks to models. *Nat. Rev. Microbiol.* **2012**, *10*, 538–550. [[CrossRef](#)]
19. Berry, D.; Widder, S. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* **2014**, *5*, 219. [[CrossRef](#)]
20. Torsvik, V.; Øvreås, L. Microbial diversity and function in soil: From genes to ecosystems. *Curr. Opin. Microbiol.* **2002**, *5*, 240–245. [[CrossRef](#)]
21. Wang, Y.; Osman, J.R.; DuBow, M.S. Bacterial communities on the surface of the mineral sandy soil from the Desert of Maine (USA). *Curr. Microbiol.* **2020**, *77*, 1429–1437. [[CrossRef](#)] [[PubMed](#)]
22. Xu, Z.; Hansen, M.A.; Hansen, L.H.; Jacquioid, S.; Sørensen, S.J. Bioinformatic approaches reveal metagenomic characterization of soil microbial community. *PLoS ONE* **2014**, *9*, e93445.
23. McHugh, T.A.; McHugh, T.A.; Compson, Z.; van Gestel, N.; Hayer, M.; Ballard, L.; Haverty, M.; Hines, J.; Irvine, N.; Krassner, D.; et al. Climate controls prokaryotic community composition in desert soils of the southwestern United States. *FEMS Microbiol. Ecol.* **2017**, *93*. [[CrossRef](#)] [[PubMed](#)]
24. Buse, H.Y.; Lu, J.; Lu, X.; Mou, X.; Ashbolt, N.J. Microbial diversities (16S and 18S rRNA gene pyrosequencing) and environmental pathogens within drinking water biofilms grown on the common premise plumbing materials unplasticized polyvinylchloride and copper. *FEMS Microbiol. Ecol.* **2014**, *88*, 280–295. [[CrossRef](#)] [[PubMed](#)]
25. Kim, M.; Morrison, M.; Yu, Z. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. *J. Microbiol. Methods* **2011**, *84*, 81–87. [[CrossRef](#)]
26. Wu, S.; Xiong, J.; Yu, Y. Taxonomic resolutions based on 18S rRNA genes: A case study of subclass Copepoda. *PLoS ONE* **2015**, *10*, e0131498. [[CrossRef](#)]
27. Balvočiūtė, M.; Huson, D.H. OTT: How do these taxonomies compare. *BMC Genom.* **2017**, *18*, 114. [[CrossRef](#)]
28. Yang, B.; Wang, Y.; Qian, P.-Y. Sensitivity and correlation of hypervariable regions in 16S rRNA genes in phylogenetic analysis. *BMC Bioinform.* **2016**, *17*, 135. [[CrossRef](#)]
29. Saxena, R.; Dhakan, D.B.; Mittal, P.; Waiker, P.; Chowdhury, A.; Ghatak, A.; Sharma, V.K. Metagenomic analysis of hot springs in Central India reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Front. Microbiol.* **2017**, *7*, 2123. [[CrossRef](#)]
30. Coller, E.; Cestaro, A.; Zanzotti, R.; Bertoldi, D.; Pindo, M.; Larger, S.; Albanese, D.; Mescalchin, E.; Donati, C. Microbiome of vineyard soils is shaped by geography and management. *Microbiome* **2019**, *7*, 140. [[CrossRef](#)]

31. Orellana, L.H.; Chee-Sanford, J.C.; Sanford, R.A.; Löffler, F.E.; Konstantinidis, K.T. Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. *Appl. Environ. Microbiol.* **2018**, *84*, e01646–e17. [[CrossRef](#)] [[PubMed](#)]
32. Alsharif, W.; Saad, M.M.; Hirt, H. Desert microbes for boosting sustainable agriculture in extreme environments. *Front. Microbiol.* **2020**, *11*, 1666. [[CrossRef](#)] [[PubMed](#)]
33. Kambura, A.K.; Mwirichia, R.K.; Kasili, R.W.; Karanja, E.N.; Makonde, H.M.; Boga, H.I. Bacteria and Archaea diversity within the hot springs of Lake Magadi and Little Magadi in Kenya. *BMC Microbiol.* **2016**, *16*, 136. [[CrossRef](#)] [[PubMed](#)]
34. López-López, O.; Cerdan, M.E.; Siso, M.I.G. New extremophilic lipases and esterases from metagenomics. *Curr. Protein Pept. Sci.* **2014**, *15*, 445–455. [[CrossRef](#)]
35. Cota-Ruiz, K.; Hernández-Viezcas, J.A.; Varela-Ramírez, A.; Valdés, C.; Núñez-Gastélum, J.A.; Martínez-Martínez, A.; Delgado-Rios, M.; Peralta-Videa, J.R.; Gardea-Torresdey, J.L. Toxicity of copper hydroxide nanoparticles, bulk copper hydroxide, and ionic copper to alfalfa plants: A spectroscopic and gene expression study. *Environ. Pollut.* **2018**, *243*, 703–712. [[CrossRef](#)]
36. Bandyopadhyay, K.K.; Aggarwal, P.; Chakraborty, D.; Pradhan, S.; Garg, R.N.; Bhattacharyya, R.; Pramanik, P.; Singh, R. *Practical Manual on Soil and Water Erosion*; Division of Agricultural Physics, Indian Agricultural Research Institute: New Delhi, India, 2013; 46p.
37. Lindsay, W.L.; Norvell, W. Development of a DTPA soil test for zinc, iron, manganese, and copper 1. *Soil Sci. Soc. Am. J.* **1978**, *42*, 421–428. [[CrossRef](#)]
38. Sivakala, K.K.; Jose, P.A.; Anandham, R.; Thinesh, T.; Jebakumar, S.R.D.; Samaddar, S.; Chatterjee, P.; Sivakumar, N.; Sa, T. Spatial physicochemical and metagenomic analysis of desert environment. *J. Microbiol. Biotechnol.* **2018**, *28*, 1517–1526. [[CrossRef](#)]
39. Li, D.; Luo, R.; Liu, C.M.; Leung, C.M.; Ting, H.F.; Sadakane, K.; Yamashita, H.; Lam, T.W. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. *Methods* **2016**, *102*, 3–11. [[CrossRef](#)]
40. Klindworth, A.; Pruesse, E.; Schweer, T.; Peplies, J.; Quast, C.; Horn, M.; Glöckner, F.O. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* **2013**, *41*, e1. [[CrossRef](#)]
41. Kozich, J.J.; Westcott, S.L.; Baxter, N.T.; Highlander, S.K.; Schloss, P.D. Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Appl. Environ. Microbiol.* **2013**, *79*, 5112–5120. [[CrossRef](#)]
42. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. 2010. Available online: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc> (accessed on 1 January 2022).
43. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **2016**, *32*, 3047–3048. [[CrossRef](#)] [[PubMed](#)]
44. Bioinformatics, B. Trim Galore! 2017. Available online: <https://github.com/FelixKrueger/TrimGalore> (accessed on 1 January 2022).
45. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* **2009**, *75*, 7537–7541. [[CrossRef](#)] [[PubMed](#)]
46. Edgar, R.C.; Haas, B.J.; Clemente, J.C.; Quince, C.; Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **2011**, *27*, 2194–2200. [[CrossRef](#)] [[PubMed](#)]
47. Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **2012**, *41*, D590–D596. [[CrossRef](#)]
48. Amato, K.R.; Yeoman, C.J.; Kent, A.; Righini, N.; Carbonero, F.; Estrada, A.; Rex Gaskins, H.; Stumpf, R.M.; Yildirim, S.; Torralba, M. Habitat degradation impacts black howler monkey (*Alouatta pigra*) gastrointestinal microbiomes. *ISME J.* **2013**, *7*, 1344–1353. [[CrossRef](#)] [[PubMed](#)]
49. Hadizadeh, F.; Walter, S.; Belheouane, M.; Bonfiglio, F.; Heinsen, F.A.; Andreasson, A.; Agreus, L.; Engstrand, L.; Baines, J.F.; Rafter, J.; et al. Stool frequency is associated with gut microbiota composition. *Gut* **2017**, *66*, 559–560. [[CrossRef](#)]
50. Metsalu, T.; Vilo, J. ClustVis: A web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res.* **2015**, *43*, W566–W570. [[CrossRef](#)]
51. Afgan, E.; Baker, D.; Beek, M.; Blankenberg, D.; Bouvier, D.; Čech, M.; Chilton, J.; Clements, D.; Coraor, N.; Eberhard, C.; et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **2018**, *46*, W537–W544. [[CrossRef](#)]
52. Wood, D.E.; Salzberg, S.L. Kraken: Ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **2014**, *15*, R46. [[CrossRef](#)]
53. Kutovaya, O.V.; Lebedeva, M.P.; Tkachkova, A.K.; Ivanova, E.A.; Andronov, E.E. Metagenomic characterization of biodiversity in the extremely arid desert soils of Kazakhstan. *Eurasian Soil Sci.* **2015**, *48*, 493–500. [[CrossRef](#)]
54. Vikram, S.; Guerrero, L.D.; Makhalanyane, T.P.; Le, P.T.; Seely, M.; Cowan, D.A. Metagenomic analysis provides insights into functional capacity in a hyperarid desert soil niche community. *Environ. Microbiol.* **2016**, *18*, 1875–1888. [[CrossRef](#)]
55. Bull, A.T.; Asenjo, J.A.; Goodfellow, M.; Gomez-Silva, B. The Atacama Desert: Technical resources and the growing importance of novel microbial diversity. *Annu. Rev. Microbiol.* **2016**, *70*, 215–234. [[CrossRef](#)] [[PubMed](#)]
56. Andrew, D.R.; Fitak, R.R.; Munguia-Vega, A.; Racolta, A.; Martinson, V.G.; Dontsova, K. Abiotic factors shape microbial diversity in Sonoran Desert soils. *Appl. Environ. Microbiol.* **2012**, *78*, 7527–7537. [[CrossRef](#)] [[PubMed](#)]

57. Stomeo, F.; Valverde, A.; Pointing, S.B.; McKay, C.P.; Warren-Rhodes, K.A.; Tuffin, M.I.; Seely, M.; Cowan, D.A. Hypolithic and soil microbial community assembly along an aridity gradient in the Namib Desert. *Extremophiles* **2013**, *17*, 329–337. [[CrossRef](#)] [[PubMed](#)]
58. Rao, S.; Chan, Y.; Bugler-Lacap, D.C.; Bhatnagar, A.; Bhatnagar, M.; Pointing, S.B. Microbial diversity in soil, sand dune and rock substrates of the Thar Monsoon Desert, India. *Indian J. Microbiol.* **2016**, *56*, 35–45. [[CrossRef](#)] [[PubMed](#)]
59. Goodfellow, M.; Nouioui, I.; Sanderson, R.; Xie, F.; Bull, A.T. Rare taxa and dark microbial matter: Novel bioactive actinobacteria abound in Atacama Desert soils. *Antonie Van Leeuwenhoek* **2018**, *111*, 1315–1332. [[CrossRef](#)]
60. Mhete, M.; Eze, P.N.; Rahube, T.O.; Akinyemi, F.O. Soil properties influence bacterial abundance and diversity under different land-use regimes in semi-arid environments. *Sci. Afr.* **2020**, *7*, e00246. [[CrossRef](#)]
61. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **2019**, *47*, W256–W259. [[CrossRef](#)]