e-ISSN:2582-7219

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH IN SCIENCE, ENGINEERING AND TECHNOLOGY

Volume 7, Issue 11, November 2024

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.521

© 2024 IJMRSET | Volume 7, Issue 11, November 2024| DOI: 10.15680/IJMRSET.2024.0711063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Data Mining and Bioinformatics Approaches to Elucidate Bacterial Communities in the Extreme Environment

Mrs.Deepika K S, Prajwal N, Ganesha R M

Assistant Professor, Department of Computer Science and Engineering, CIT, Gubbi, Tumkur, Karnataka, India

U.G. Student, Department of Computer Science and Engineering, CIT, Gubbi, Tumkur, Karnataka, India

U.G. Student, Department of Computer Science and Engineering, CIT, Gubbi, Tumkur, Karnataka, India

ABSTRACT: Numerous habitats are home to microscopic organisms called bacteria. They are plentiful and play a variety of roles in our lives. To determine which bacteria are required for several industrial uses, it is imperative that we study bacteria. The primary issue, though, is that most bacteria cannot be cultured, which makes it difficult to study microorganisms from various habitats. The metagenomics method of studying bacteria, which makes use of Next Generation Sequencing technology, may make use of the 16S rRNA marker gene. This study aims to demonstrate the use of data mining and bioinformatics approaches to analyze 16S rRNA sequencing data. The Biological database provided the raw sequencing data for 16S rRNA. The data was then clustered, denoised, and trimmed to produce Amplicon Sequence Variants (ASVs). Diversity (alpha and beta) and taxonomic analyses were then performed to elucidate the bacterial diversity and taxonomic profile of ASVs. The study's findings demonstrated that China's hot springs have higher Shannon and PD indices than those in Singapore and the United States. Additionally, a noteworthy variation was noted in the PD index. The bacterial populations at the three sites differed significantly, according to the unweighted UniFrac distance as well. Furthermore, the taxonomic study identified Proteobacteria, Chloroflexi, Cyanobacteria, and Crenarchaeota as the main bacterial groups in the ecosystem. The results may serve as a foundation for more bacterial metagenomic research, particularly that which focuses on the environment of hot springs.

KEYWORDS: Environmental factors, metagenomics, bacterial communities, data mining, 16S Rrna

I. INTRODUCTION

Bacteria are single-celled organisms that live in a variety of environments, from commonplace ones like soils and oceans to extreme ones like ice and hot springs [1].

Moreover, the human body and plants both contain them [1]. Bacteria belong to the two groups of organisms (Archaea and Bacteria) known as prokaryotes because they lack a nucleus and membrane-bound organelles. Broadly speaking, bacteria are essential to human health as well as a number of biogeochemical cycles, including the nitrogen and carbon cycles [2, 3].

Hot springs are formed when hot water rises to the surface as a result of tectonic or volcanic activity. The hot spring's temperature, which ranges from 40°C to the boiling point, creates an intense environment that encourages the growth of a specific kind of bacteria called thermophiles [4]. Apart from its high temperature, the hot spring often faces additional environmental stressors such fluctuations in pH between low and high, the presence of dissolved sulfide and other minerals, and insufficient supply of nutrients.

As a result of overcoming these obstacles, some thermophiles fall within the category of polyextremophiles, which are able to adjust to a range of harsh environments [5].

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Since bacteria significantly affect human lives, it is crucial to research germs from a variety of settings, particularly harsh ones. This allows us to clarify the diversity and makeup of certain bacteria at a given niche or environment [6]. Finding the bacteria that contribute to a certain ecological process in an environment is another advantage of determining the taxonomic structure [7]. We might even be able to extract new germs from harsh settings that could be used for a variety of things, including medicine [8]. microorganisms are typically isolated and grown in a growth medium as part of the culture-dependent approaches used in the traditional approach to studying microorganisms [9]. Even though around 90% of the bacterial species on Earth are still unidentified, only a small portion of bacteria can be isolated and grown in a laboratory [10]. This turns becomes the primary issue since it prevents the investigation of new, promising bacteria that could have industrial uses. Culture-neutral techniques were created to solve the issue. The genetic diversity and makeup of bacteria, particularly those that cannot be cultivated, can be determined using culture-independent techniques [11].

Additionally, the techniques concentrate on extracting genetic material from microorganisms, a process known as the metagenomics approach [11].

The study of unculturable bacteria has made extensive use of the metagenomics technique. The small subunit of ribosomal ribonucleic acid, or 16S rRNA, is used as a molecular marker in this method [12]. Since all bacterial species often contain the 16S rRNA, it is employed [13]. Additionally, due to moderate mutation rates, it includes areas that are comparatively preserved [13]. Because of this feature, the 16S rRNA is perfect for determining the taxonomic makeup and evolutionary relationships of bacterial populations [14]. The introduction of next-generation sequencing (NGS) technology has significantly improved the metagenomics technique by enabling the rapid and economical acquisition of high-throughput data [15]. Consequently, there has been a notable surge in the quantity of 16S rRNA sequencing data added to the biological database and GenBank in recent years [16]. To interpret the available sequencing data, data mining and computational biology/bioinformatics techniques must be used.

In this study, we illustrated how to use bioinformatics and data mining techniques to examine 16S rRNA sequencing data produced by NGS technology. The results of the study could serve as a guide to compare the variety of bacteria from different biological databases and to clarify the composition of bacteria.

II. RELATED WORKS

The environment around hot springs has been the subject of numerous studies. These mostly seek to identify and isolate thermophilic bacteria with the capacity to produce industrial enzymes. In 2019, 85 bacterial isolates from Turkish hot springs were examined by Gulmus and Gormez [17]. They discovered that over 50% of these isolates could produce the enzymes lipase, protease, and amylase. Abdollahi et al. 2020 carried out a similar investigation in which they characterized thermophilic bacteria from an Iranian hot spring using a culture-dependent methodology [18]. They identified two bacteria, Bacillus altitudinis and Thermomonas hydrothermalis, as possible sources of protease enzymes. Anoxybacillus flavithermus, a bacterium that produces amylase, was successfully identified from a Malaysian hot spring by Fazal et al. in 2022 [19]. Another study was carried out to identify the microorganisms discovered in the hot springs of the Republic of Korea. In 2022, Lee et al. identified 22 isolates that are capable of producing the enzymes lipase, protease, or amylase [20]. They also found other bacterial species that are common in these hot springs, including Aeribacillus, Bacillus, Caldibacillus, Geobacillus, and Thermoactinomyces [20]. Additionally, George et al. 2023 examined the diversity of bacteria at Singapore's Sembawang Hot Spring using the 16S rRNA gene sequencing technique [21]. The results showed that some bacterial taxa, including Roseiflexus sp., Thermosynechococcus elongatus, Oscillatoriales cyanobacterium, and Chloroflexus sp., predominate in the environment. According to earlier research, harsh settings like hot springs hold promise for the investigation of microorganisms with certain traits. Additionally, metagenomic research employing 16S rRNA gains popularity. Thus, we used data mining and bioinformatics tools to examine 16S rRNA sequencing data.

III. METHODOLOGY

Overall, the suggested approach employed in this work consists of the following steps (Fig. 1): taxonomic analysis, diversity analysis, data preprocessing, and data mining.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

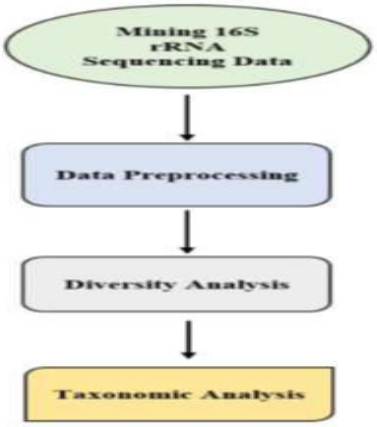

Fig. 1. The workflow of methods used in this study.

a) Data collection

The sequence read archive (SRA) of the National Center for Biotechnology Information (NCBI) (https://www.ncbi.nlm.nih.gov) provided us with the sequencing data, which was generated by the Illumina Platform of NGS technology. We used the term "hot spring metagenome" to retrieve raw 16S rRNA amplicon sequencing data from China, Singapore, and the USA. The SRA Toolkit bioinformatics program was used to retrieve the fastq format containing the 16S rRNA sequencing data. Additionally, the NCBI provided the sequencing data's metadata. The data gathering technique yielded a total of 12 sequencing data.

b) Data pre-processing

The NCBI-provided raw sequencing data was analyzed using the QIIME 2 tool [22]. First, a QIIME extension file was created by importing the raw data using the import tool. Based on the quality score of the measurements, the imported data was trimmed and then denoised to remove contaminants. Both the denoising and trimming processes were carried out using the DADA2 program, which resulted in Amplicon Sequence Variants (ASVs). Following that, de novo clustering was used to group the ASVs at 99% sequence similarity. The data was clustered and viewed using the QIIME2 feature table tool.

c) Diversity analysis

The variety of microorganisms was examined in the R environment. Faith's Phylogenetic Diversity (PD) and Shannon's index (H) are two alpha diversity indices that were created using the vegan program. To ascertain whether alpha diversity varied significantly between places, the Kruskal-Wallis test was employed, and Dunn's test was employed as a post hoc analysis. Furthermore, the unweighted unifrac distance was used to estimate the community dissimilarity that accounts for evolutionary links (beta diversity).

d) Taxonomic analysis

The taxonomy analysis of the sequencing data was performed using the pre-trained Naïve Bayes classifier in the QIIME 2 classifier tool. The Greengenes 13_8 99% OTUs (Operational Taxonomic Units) were used to train the classifier.The QIIME 2 taxonomy barplot tool was then used to display the taxonomic makeup in bar plots. ASVs classified as "unassigned" were those whose taxonomic levels could not be ascertained.

IV. RESULTS AND DISCUSSION

Using the SRA Toolkit's fastq-dump function, we first obtained the raw data (fastq file) from the hot spring 16S rRNA amplicon sequencing. The Ilumina platform's paired-end library layout 16S rRNA sequencing data was the one we chose.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

This is because the sequencing data generated by paired-end sequencing is more comprehensive and of higher quality, which facilitates bioinformatics analysis [23]. The DADA2 tool was used to trim and denoize the received sequencing data. DADA2 employs a particular method to deduce the actual sample composition by modeling the errors produced during the sequencing process. Amplicon Sequence Variant (ASV) is produced by denoising amplicon reads to remove sequence mistakes and identify the right biological sequences, while trimming eliminates adapters and poor sequences [24]. Following denoising and trimming, 5,289 ASVs were produced. Following that, the ASVs were further grouped using a de novo clustering technique at 99% similarity, producing 4,440 ASVs with 650,353 features overall for the entire sample (Table I). A collection of sequences was grouped into pertinent taxonomic levels using the clustering technique [25]. Since the de novo clustering approach outperforms reference-based clustering in terms of assigning ASVs, we used it.

We next used the clustered ASVs to create a tree for phylogenetic diversity analyses. The phylogeny tool from QIIME2 was used for this. Bacterial diversity was then examined using the clustered ASVs and the phylogenetic tree. The analysis found that the Shannon index of China's hot springs was higher than that of Singapore's and the US's (Fig. 2). A higher value of the Shannon index, which measures the richness of the bacterial community, suggests that the environment contains a greater variety of bacterial species [26]. According to the statistical analysis, however, the three locations' Shannon indices did not differ much. Similar to the Shannon index, the PD index is higher in the Chinese hot spring than in other locations. Furthermore, the PD index's Kruskal-Wallis test revealed a significant difference across the three locations (Table II). Following that, we employed a post-hoc test called Dunn's Test to determine whether location pairs had significant differences. According to Table I, the results showed that the differences in bacterial diversity originated from China-USA marriages. By considering the phylogenetic relationships within the bacterial population, Faith's PD computes an index of bacterial diversity [27].

In general, there are differences in the alpha diversity indices among the three sites. More diversified bacteria may be able to survive and proliferate in the environment if the index value is greater.

TABLE I. AMPLICON SEQUENCE VARIANTS DATA GENERATED AFTER PRE-PROCESSING

Sample	Feature count	Number of ASVs	Total frequency count
SRR23559592	132476		
SRR23559593	128938		
SRR23559591	124276		
SRR23559594	120418		
SRR23783014	37123		
SRR23783011	32047	4,440	650,353
SRR23783010	27279		
SRR23783012	22422		
SRR23604208	12758		
SRR23604210	4546		
SRR23604209	4336		

© 2024 IJMRSET | Volume 7, Issue 11, November 2024| DOI: 10.15680/IJMRSET.2024.0711063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

TABLE II. STATISTICAL ANALYSIS OF THE ALPHA DIVERSITY INDICES

Statistical method	Alpha diversity index	P-value
Kruskal-Wallis	Shannon	0.215
	PD	**0.041***
Dunn's Test (Post-hoc)	PD (China-Singapore)	1.000
	PD (China-USA)	**0.041***
	PD (Singapore-USA)	0.182

*Significant at P-value < 0.05

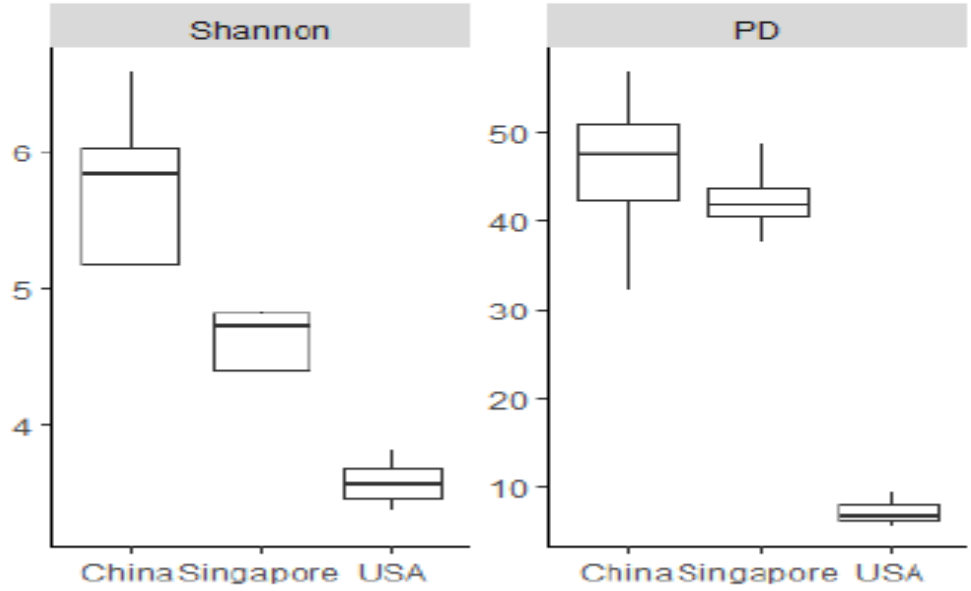

Fig. 2. Alpha diversity indices of hot spring environment from China, Singapore, and the USA.

We conducted a beta diversity analysis of the hot spring habitat in addition to an alpha diversity analysis. The differences in bacterial species between two or more environments/ecosystems are reflected in the beta diversity. Here, we measured the differences in bacterial populations across the three hot spring locations using the unweighted UniFrac distance. For beta diversity analysis, the unweighted UniFrac distance has been used extensively since it calculates the fraction of branch length that is unique while accounting for species presence and absence data. The best method for detecting shifts in numbers in rare lineages is to use this distance [28]. According to the unweighted UniFrac distance result, the bacterial communities in the USA, Singapore, and China separated into three separate clusters (Fig. 3).

The bacterial species found in each site were relatively different, as the statistical test also showed that the bacterial communities of the three sites differed significantly (PERMANOVA, P-value <0.001).

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

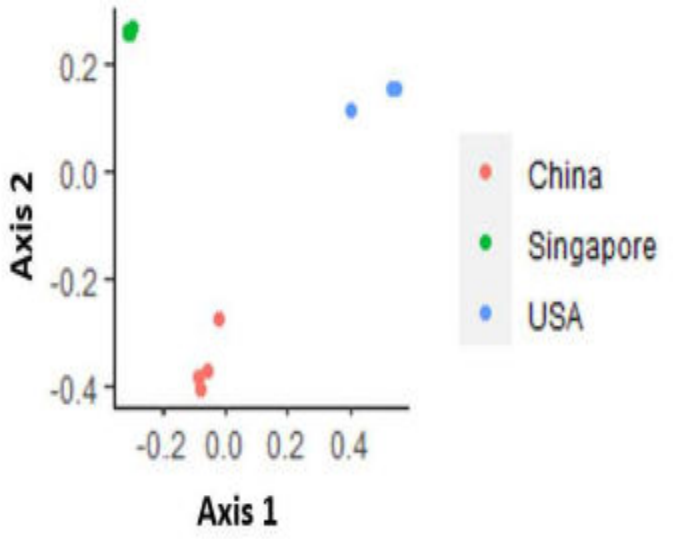

Fig. 3. Beta diversity of hot spring environment from China, Singapore, and the USA

We conducted a beta diversity analysis of the hot spring habitat in addition to an alpha diversity analysis. The differences in bacterial species between two or more environments/ecosystems are reflected in the beta diversity. Here, we measured the differences in bacterial populations across the three hot spring locations using the unweighted UniFrac distance. The unweighted UniFrac distance has been commonly used for beta diversity analysis since it incorporates species presence and absence data and calculates the unique proportion of branch length. This distance is the most effective way to identify changes in numbers in uncommon lineages [28]. Three separate clusters that were segregated from each other were created by the bacterial communities in China, Singapore, and the USA, according to the unweighted UniFrac distance result (Fig 3,).Additionally, the statistical test revealed a significant difference between the bacterial communities of the three sites (PERMANOVA, P-value <0.001), suggesting that the bacterial species found in each site were relatively different.

Taxonomic analysis was performed to calculate the number of species from the clustered ASVs using the 16S rRNA database from Greengenes (99% OTUs). It is demonstrated that each place has a different taxonomic profile. Chloroflexi, Cyanobacteria, and Bacteria (unknown species) dominated the bacterial population in the hot spring in China (Fig. 4). Proteobacteria, Chloroflexi, and Cyanobacteria were the three types of bacteria that were most prevalent in Singapore's hot spring (Fig. 4). Proteobacteria, Chloroflexi, and Cyanobacteria dominated the hot spring habitat in the USA, just like they did in Singapore (Fig. 4). It's noteworthy that while Crenarchaeota was not as common in the other locations, it was in the USA.

Despite being considered a harsh habitat, the hot spring may nonetheless be able to support a variety of animals.Hot springs are known to harbor Cyanobacteria and Chloroflexi [29]. They were able to adapt to difficult conditions, such as the high temperatures and sulfur concentrations of hot springs, which is why. A hot spring is probably home to Crenarchaeota, an archaeon that loves environments with high temperatures and sulfur [29]. Furthermore, a study conducted in 2021 by Choure et al. [30] showed that India's hot springs were rich in Proteobacteria. This implies that Proteobacteria can thrive in the hot spring environment.

© 2024 IJMRSET | Volume 7, Issue 11, November 2024| DOI: 10.15680/IJMRSET.2024.0711063

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

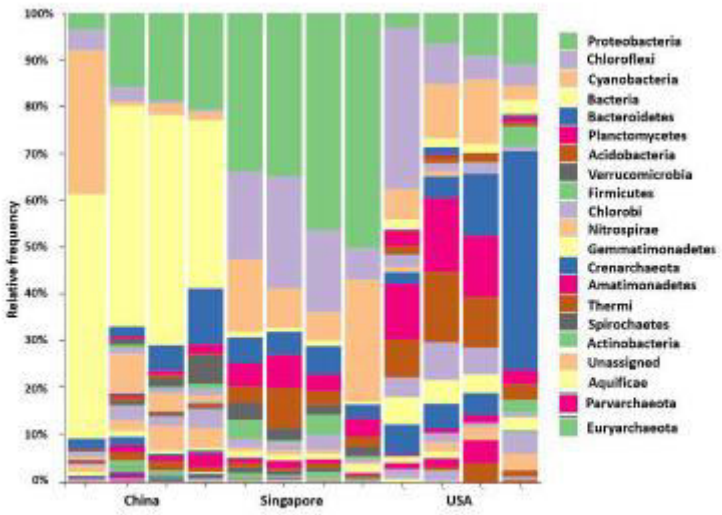

Fig. 4. Taxonomic profile of bacterial community from China, Singapore, and the USA

V. CONCLUSION

Utilizing bioinformatics and data mining techniques, we analyzed 16S rRNA sequencing data from three different hot spring settings. The results of this investigation showed that the alpha indices (Shannon and PD) of hot springs in China were superior to those in Singapore and the US.Additionally, the statistical test confirmed that the PD indices at the three locations varied significantly from one another. Significant variations in the bacterial communities in China, Singapore, and the USA were revealed by the unweighted UniFrac distance for the beta diversity analysis. The taxonomic analysis also found some taxa that dominated the environment, such as Proteobacteria, Chloroflexi, Cyanobacteria, and Crenarchaeota. Further bacterial metagenomic research, especially in the setting of hot springs, could be guided by the study's findings.

REFERENCES

[1] Turnbaugh PJ, Soen Y, Amir A, Zeisel A, Fuks G, Elgart M, et al.High-resolution profiling of microbial communities is made possible by combining variable sections of the 16S rRNA gene. 2018 Microbiome 6:1–13.It can be found at https://doi.org/https://doi.org/10.1186/s40168-017-0396-x.

[2] Zhang XD, Cao M, Liu F, Qian Y, Zhang SW, Wei ZG. Comparison of Amplicon Sequence-Based Operational Taxonomic Unit Selection Techniques. 2021; Front Microbiol 12:1–16.FMICB.2021.644012, https://doi.org/10.3389.

[3] Trinugroho JP, Asadi F, Pardamean B. Genome Data Mining Approach to Identify Potential Protein in Crop Plants. Proceeding 7th Int. Conf. Sci. Technol. (ICST-2022), 2022.

[4] Microbial diversity in harsh settings, Shu WS, Huang LN. Microbiol Nat Rev 2022;20:219–35. 1038/s41579-021-00648-y, https://doi.org.

[5] Wakerley GLJ, Evans DW, Button M, Power JF, Carere CR, Lee CK, et al. The microbial biogeography of New Zealand's 925 geothermal springs. 2018, Nat Commun, 9:2876. s41467-018-05020-y (https://doi.org/10.1038/article).

[6] Zhang Z, Ren J, Michail S, Sun F. MicroPro: utilizing metagenomic unmapped reads to provide light on the relationship between human microbiota and disease. 2019;20:1–13; Genome Biol.

[7] Lapierre N, Ju CJ-T, Zhou G, Wang W. MetaPheno: A critical evaluation of deep learning and machine learning in metagenome based disease prediction. Methods 2020;166:74-82.https://doi.org/10.1016/j.ymeth.2019.03.003.MetaPheno.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[8] Mahesworo B, Pardamean B, Hidayat A., Trinugroho JP, Nirwantono R, and Sudigyo D. promising enzyme from Chelatococcus composti genome mining for sustainable D-Allulose synthesis. Earth Environ Sci IOP Conf Ser 2023;1169:012083. https://doi.org/10.1088/1755-1315/1169/1/012083.

[9] Bodor A, Bounedjoum N, Vincze GE, Erdeiné Kis Á, Laczi K, Bende G, et al. Challenges of unculturable bacteria: environmental perspectives. Rev Environ Sci Biotechnol 2020;19:1–22.https://doi.org/10.1007/s11157-020-09522-4.

[10] Hofer U. The majority is uncultured. Nat Rev Microbiol 2018;16:716–7. https://doi.org/10.1038/s41579-018-0097-x.

[11] Andharia K, Koringa P, Joshi A, Dumadiya A, Kotadiya R, Hinsu A, et al. Culture-dependent and culture-independent bacterial diversity from the peanut rhizosphere: a snapshot of whether to cultivate or not. e12035 in PeerJ 2021. https://doi.org/10.7717/peerj.12035.

[12] Laudadio I, Fulci V, Stronati L, and Carissimi C. Methodological Opportunities and Challenges in Next-Generation Metagenomics. In 2019, Omi A. J. Integr Biol. 23:327–33.OMI.2019.0073 (https://doi.org/10.1089/omi).

[13] Using 16S rRNA gene profiling and genome reconstruction, Peterson CT, Sharma V, Iablokov SN, Albayrak L, Khanipov K, Uchitel S, et al. uncover the prebiotic potential and community metabolic interactions of medicinal herbs utilized as nootropics and in neurological diseases. PLoS One, 14:1–25, 2019.10.1371/journal.pone.0213869 https://doi.org.

[14] 16S rRNA sequence embeddings: Convenient quantitative feature representations of nucleotide sequences for downstream analysis (Woloszynek S, Zhao Z, Chen J, Rosen GL). In 2019, PLoS Comput Biol 15:1–25. Journal.pcbi.1006721, https://doi.org/10.1371/journal.

[15] Trinugroho JP, Asadi F, Pardamean B. Identifying Cyanobacteria through Next-Generation Sequencing Technology for Modern Agriculture. E3S Web Conf 2023;388:01007. https://doi.org/10.1051/e3sconf/202338801007.

[16] Trinugroho JP, Perbangsa AS, Pardamean B. The Technology behind Genomic Database. 2022 4th Int. Conf. Cybern. Intell. Syst. ICORIS 2022, 2022. https://doi.org/10.1109/ICORIS56080.2022.10031455.

[17] Identification and Description of New Thermophilic Bacteria from Hot Springs in Erzurum, Turkey by Oztas Gulmus E and Gormez A. In 2020, Curr Microbiol 77:979–87. 10.1007/s00284-020-01880-0 is the URL.

[18] Isolation and Characterization of Thermophilic Bacteria from Gavmesh Goli Hot Spring in Sabalan Geothermal Field, Iran: Thermomonas hydrothermalis and Bacillus altitudinis Isolates as a Possible Source of Thermostable Protease By Abdollahi P, Ghane M, and Babaeekhou L. 87–95 in Geomicrobiol J 2020; 38.18.12774 https://doi.org/10.1080/01490451.2020.

Budiman C, Amin Z, Fazal BZ, and Ling CMWV. Amylase-producing bacteria from Poring Hot Spring in Sabah, Malaysia were screened, isolated, and described. 23:2807–15 in Biodiversitas 2022.Biodiv/d230604 (https://doi.org/10.13057).

[19] Oh DK, Kim HW, Jeong GE, Cha IT, Lee YJ, Ganbat D, et al. Thermophilic bacteria were isolated and characterized from hot springs in the Republic of Korea. Microbes 10:0–9, 2022.10122375 (https://doi.org/10.3390/microorganisms)

[20] George C, Lim CXQ, Tong Y, Pointing SB. Community structure of thermophilic photosynthetic microbial mats and flocs at Sembawang Hot Spring, Singapore. Front Microbiol 2023;14:1–13. https://doi.org/10.3389/fmicb.2023.1189468.

[21] Al-Ghalith GA, Abnet CC, Bolyen E, Rideout JR, Dillon MR, and Bokulich NA, et al. Interactive, scalable, extendable, and reproducible microbiome data research using QIIME 2. Nat Biotechnol 2019;37:852–7. s41587-019-0209-9 (https://doi.org/10.1038/7).

[22] Zhang Y, Zhang H, Qu Z, Zhang X, Cui J, Wang C, et al. Comprehensive analysis of the molecular characterization of GM rice G6H1 using a paired-end sequencing approach. Food Chem 2020;309:125760.https://doi.org/https://doi.org/10.1016/j.foodchem.2019.125760.

[23] A AJ, Callahan BJ, Mcmurdie PJ, Rosen MJ, and Han AW. DADA2: Inference of high-resolution samples from Illumina amplicon data. 2016;13:581–3, Nat Methods. https://doi.org/10.1038/nmeth.3869.DADA2Antich A, Palacin C, Wangensteen OS, Turon X. To denoise or to cluster, that is not the question: optimizing pipelines for COI metabarcoding and metaphylogeography. BMC Bioinformatics 2021;22:1–24. https://doi.org/10.1186/s12859-021-04115-6.

[24] Yin L, Wan Y-D, Pan X-T, Zhou C-Y, Lin N, Ma C-T, et al. Association Between Gut Bacterial Diversity and Mortality in Septic Shock Patients : A Cohort Study. Med Sci Monit 2019;25:7376–82.https://doi.org/10.12659/MSM.916808.

ISSN: 2582-7219 | www.ijmrset.com | Impact Factor: 7.521| ESTD Year: 2018|

International Journal of Multidisciplinary Research in Science, Engineering and Technology (IJMRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

[25] Cantrell K, Huang S, Mcdonald D, Haiminen N, Armstrong G, Carrieri AP, et al. Faith's phylogenetic diversity may be computed efficiently and used to characterize microbiomes. https://doi.org/10.1101/gr.275777.121 Genome Res 2021;31:2131-7.

[26] Chen J, Bittinger K, Charlson ES, Hoffmann C, Lewis J, Wu GD, et al. Associating microbiome composition with environmental covariates using generalized UniFrac distances. Bioinformatics 2012;28:2106–13. https://doi.org/10.1093/bioinformatics/bts342.

[27] Björnsdóttir SH, Podar M, Yang Z, and Podar PT. Microbial Diversity in Icelandic and Yellowstone Hot Springs: A Comparative Study Across Temperature Gradients. Microbiol Front, 2020, 11:1–16.fmicb.2020.01625 (https://doi.org/10.3389/).

[28] Choure K, Parsai S, Kotoky R, Srivastava A, Tilwari A, Rai PK, et al. Comparative Metagenomic Analysis of Two Alkaline Hot Springs of Madhya Pradesh , India and Deciphering the Extremophiles for Industrial Enzymes. Front Genet 2021;12:1–11. https://doi.org/10.3389/fgene.2021.643423.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

IN SCIENCE, ENGINEERING AND TECHNOLOGY

| Mobile No: +91-6381907438 | Whatsapp: +91-6381907438 | ijmrset@gmail.com |

www.ijmrset.com