sustainability

MDPI

Article

A Comprehensive Metagenomic Analysis Framework Revealing Microbiome Profile and Potential for Hydrocarbon Degradation and Carbohydrate Metabolism in a Himalayan Artificial Lake

Girija Kaushal [1], Monika Thakur [1], Amit Kumar Rai [2,*] and Sudhir P. Singh [1,*]

[1] Center of Innovative and Applied Bioprocessing (DBT-CIAB), Sahibzada Ajit Singh Nagar 140306, India
[2] Institute of Bioresources and Sustainable Development (DBT-IBSD), Gangtok 737102, India
* Correspondence: amitrai.ibsd@gov.in (A.K.R.); sudhirsingh@ciab.res.in (S.P.S.)

Abstract: The high-altitude lakes are ecological habitats accommodating a vast diversity of microbial populations. These microbes are efficient sources for a variety of enzymes. The objective of this study is to perform in-depth metagenomic profiling of an artificial lake ecosystem located in the Sikkim Himalayan region, deciphering the hydrocarbon degradation potential of this site and mining biocatalysts of industrial importance. In the present study, metagenomic analysis of an artificial Himalayan lake, located in North Sikkim, India, was performed. A comprehensive taxonomic and functional profiling revealed gene mapped to pathways for degradation of hydrocarbons such as toluene, benzoate, ethylbenzene, etc. This site was rich in iron, and the metagenomic investigation revealed genomic signatures of the iron-reducing bacterium; *Geothrix fermentans*. The appraisal of the carbohydrate metabolic potential of this site divulged the predominance of β-galactosidase genes. The artificial lake metagenome was further compared to publicly available saline and freshwater lakes. At the taxonomic, as well as functional levels, it was found to be closer to freshwater lake metagenome, e.g., Medonta Lake, US, and freshwater Vanda Lake, Antarctica. The microbial community profiling and functional contribution of the artificial Himalayan lake would be beneficial for mining genes encoding various industrially relevant enzymes.

Keywords: high-altitude; artificial lake; Himalayan; metagenome; β-galactosidase; hydrocarbon degradation

Citation: Kaushal, G.; Thakur, M.; Rai, A.K.; Singh, S.P. A Comprehensive Metagenomic Analysis Framework Revealing Microbiome Profile and Potential for Hydrocarbon Degradation and Carbohydrate Metabolism in a Himalayan Artificial Lake. *Sustainability* **2022**, *14*, 11455. https://doi.org/10.3390/su141811455

Academic Editor: Dino Musmarra

Received: 18 August 2022
Accepted: 7 September 2022
Published: 13 September 2022

1. Introduction

The lakes of the Himalayan region are one of the most vulnerable ecosystems on earth [1]. The Himalayan state of the Sikkim plateau is surrounded by many mountain lakes, of which Gurudongmar and ChhoLamo are very famous (www.sikkim.gov.in (accessed on 2 June 2022)). The earth's fresh water system is a crucial biosphere inhabited by assorted organisms, including methane degrading, dissimilatory iron-reducing and nitrifying bacteria, which impart to the cycling of substances by different pathways [2]. Hence, it draws the scientific community's attention to an in-depth genomic inspection of the freshwater ecosystem. The highly efficient, next-generation sequencing platforms are awarding an opportunity to elaborate on the microbial potential of various environmental niches. The metagenomic investigations of diverse habitats provide a good plan of action for initiating projects associated with the activities on adaptation in the ecological niches of the Himalayan regions. In-depth genomic studies have divulged the microbial diversity and functional profiling of many aquatic habitats, such as Lough Neagh Lake in Ireland, Lake Soyang in South Korea, Soda Lake brines in Russia, Lonar Lake, Tsomgo Lake and Pangong Lake in India, Antarctic Lake and Salt Lake, China [3–10]. These evaluations boost our understanding of the vast range of indeterminate metabolic pathways and adaptation associated with hydrocarbon degradation and conversion of complex carbohydrates into value-added products, etc.

The natural incidences lead to the formation of new artificial water reservoirs in the Himalayan regions. The Landslide in Dzongu of North Sikkim (India) resulted in the emergence of an artificial lake, as a freshwater ecosystem, by the blockage of the Kanaka River (Arunachal24.in (accessed on 15 June 2022)). This river is the major tributary of the Teesta River, which is sourced by ChhoLamo (www.sikkim.gov.in (accessed on 2 June 2022)). The lakes represent independent ecosystems with unique characteristics [4]. The surrounding environmental conditions render micro-floral diversity in the Himalayan lakes, making them a rich source of the crucial microbial community that is yet to be explored and exploited for societal benefits [5].

The extreme niches often inhabit microbial populations with sturdy metabolic potential in harsh conditions. Therefore, these extremophiles could be a potential source of versatile biocatalysts. In recent reports, the high-altitude sites have reported a vast collection of enzymes with industrial applications such as lipase [11], pullulanase [12], alkaline pectate lyase [13], amylase [14], protease [14], etc. Therefore, the exploration of these sites could be highly beneficial in the mining of cold-active enzymes with applications in the processing of biomass. Therefore, intending to generate genomic-data resources for future strategies, we have explored the in-depth metagenome of an artificial lake in the Himalayan region, revealing molecular taxon structure and metabolic functions prominent at this site. Apart from gene mining, designing bioremediation techniques by studying microbial populations and functional pathways from various habitats is an important contribution of metagenomics. This study also elaborates a molecular description of the hydrocarbon degradation and carbohydrate metabolic-pathway genes that have prevailed at this site. In addition, a comparative metagenomic study with other low-temperature freshwater and saline aquatic habitats was implemented to discover the unique capabilities of this environmental niche.

2. Materials and Methods

2.1. Sample Collection and Extraction of Metagenomic DNA

The natural landslide near the Teesta River in Dzongu, North Sikkim Himalayas, created an artificial lake at a high altitude (N 27.53° and E 88.49°) of the Sikkim Himalayas. The sample collection was undertaken from the Himalayan artificial lake, Sikkim, India (Figure 1). The pH and temperature of the site were measured using a portable pH meter and a thermometer, respectively. A sterilized scoop was used to collect the muddy-water sample from the lake. The samples were taken from about 12 inches below the water surface. Five random samples were harvested in aseptic containers. The samples were carried to the laboratory in cool condition and stored at −20 °C. All the samples of the Himalayan artificial lake site were pooled before DNA extraction. The MP Bio-medicals DNA™ isolation kit was employed to extract metagenomic DNA from the environmental sample. In order to access the quality of isolated DNA, NanoDrop 2000™ (Thermo fisher Scientific™, Madison, WI, USA) was used.

2.2. Metagenome Sequencing, Assembly, and Elementary Analysis

TruSeq DNA Sample Preparation Kit was employed to prepare the DNA sequencing library as instructed in the manufacturer's protocol (Illumina, San Diego, CA, USA). Illumina HiSeq (Illumina, San Diego, CA, USA) was used to perform microbiome sequencing, generating about 150 nucleotides paired-end reads. These reads were submitted to the FASTQC server for evaluating their quality based on the Phred score (www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 28 May 2022)). The Cut adapt (version-1.8.1) was used for adapter trimming [15]. These reads were termed as high-quality reads and further processed for assembly. The MetaSPAdes (version 3.10.1) was used to assemble these high-quality reads at default parameters [16]. The assembly was analyzed by calculating its statistical parameters using the QUAST web-server [17]. The assembled contigs were submitted to MetaGene Annotator (MGA) for the prediction of ORFs [18]. These ORFs were taken as input for the taxonomic and functional distribution of the metagenome.

Figure 1. The Sikkim Himalayan artificial lake sampling site. (**a**) The geographical coordinates of the sampling site have been shown by using Google Maps, 2021 (https://www.google.com/maps, accessed on 12 June 2022). (**b**) Landslide view that leads to formation of lake (**c**) Sample collection site.

2.3. Taxonomic and Functional Profiling

The taxonomic classification was derived by aligning the predicted ORFs against the NCBI taxonomy data-set at the stringent e-value of 1×10^{-5}, using the Diamond BLASTx approach [19], and then the alignment file was imported to MEGAN [20]. The number of ORFs mapped on a particular taxon indicated the sample's taxonomic abundance. The microbial diversity indices were computed to check the species richness and evenness in the given metagenomic sample by using the PAST4.03 statistical tool [21].

To analyze the functional properties of the artificial lake metagenome, the predicted ORFs were aligned against the SEED and KEGG Orthology (KO) databases at the e-value cutoff of 1×10^{-5}. Additionally, antibiotic-resistance genes (ARGs) were predicted by aligning the ORFs against the CARD database [22]. The ORFs exhibiting at least 80% identity and 70% subject coverage were considered as antibiotic resistance genes (ARGs).

2.4. Exploration of Hydrocarbon Degradation and Carbohydrate Metabolism Pathways

The KEGG Orthology (KO)IDs [23] aligned with metagenomic ORFs were mapped over hydrocarbon degradation pathways to check the presence of enzymes involved in processing hydrocarbons in the artificial lake metagenome. The carbohydrate metabolism pathway was investigated by aligning the predicted ORFs over the Carbohydrate Active Enzyme (CAZy) database [24]. While predicting CAZymes, the stringent criteria of 50% to 70% identity with at least 50% subject coverage were considered.

2.5. Statistical Comparison with Other Lake Metagenomes

The metagenome of the Himalayan artificial lake was compared with publicly available lake metagenomes (Table 1) at IMG/M server [25]. The metagenomic data were statistically compared using STAMP (version 2.1.3) statistical software [26]. The phylum, COG and KO distribution were analyzed using the Principal Component Analysis (PCA). The clustering heat maps were plotted for COG and KO profiles to compare the functional profiles of the metagenomes. The taxonomic distribution was compared via two-variable analysis by applying the Fisher Exact test and Bonferroni correction, and plotting an extended bar plot.

Table 1. Metagenomic data obtained from IMG/M server for comparative analysis.

Sample Name	Sample ID in Given Analysis	Temperature	Reference
Saline Ace Lake, Antarctica	S1_3300005936	1 °C to 6 °C	https://link.springer.com/article/10.1007/s00300-014-1553-3 (accessed date 12 July 2022)
Freshwater Vanda Lake, Antarctica	S2_3300015360	4 °C to 10 °C	https://www.mdpi.com/2076-2607/3/3/391 (accessed date 12 July 2022)
Saline organic lake in Antarctica	S3_3300031218	−10 °C to −14 °C	https://eprints.utas.edu.au/14434/ (accessed date 15 July 2022)
Freshwater Towuti Lake, Indonesia	S4_3300031587	26 °C to 29 °C	https://www.frontiersin.org/articles/10.3389/fmicb.2016.01007/full (accessed date 15 July 2022)
Freshwater Mendota Lake, United States	S5_3300035663	21 °C to 23 °C	https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2427.2002.01011.x (accessed date 15 July 2022)
Freshwater artificial North Sikkim lake, India	S6_3300047516	12 °C to 14 °C	(This study)

2.6. ICP-MS Analysis

The concentration of different metals in the lake's sample was determined using inductively coupled plasma mass spectrometry (ICP-MS) (Agilent Scientific Instruments, Santa Clara, CA, USA). A multi-metal standard of different concentrations was used for reference.

3. Results and Discussion

3.1. Site Description and Physicochemical Analysis

At the time of the sample collection, the pH and temperature of the lake water were recorded to be 8.3, and 14 °C, respectively. The ICP-MS analysis of the lake samples revealed iron (126.69 ppm) as the prominent metal in the sampling site, followed by K (33.83 ppm), Mg (30.83 ppm), Al (22.89 ppm), and Ca (2.19 ppm) (Table 2). The concentration of heavy metals such as Cr, Cu, Co, Zn, and Ni was found within limits permitted by WHO (https://www.omicsonline.org/articles-images/2161$-$0525$-$5$-$334-t011.html (accessed on 20 July 2022)).

Table 2. Concentration of various elements in the Sikkim Himalayan artificial lake sample.

Element	Concentration (ppm)
Fe	126.69
K	33.83
Mg	30.83
Al	22.89
Ca	2.19
Mn	1.86
Na	1.0
Rb	0.64
Ba	0.48
Zn	0.29
V	0.23
Li	0.19
Cr	0.16
Cu	0.097
As	0.096
Ni	0.091
Cs	0.073
U	0.061
Co	0.047
Pb	0.046
Ga	0.044
Sr	0.038
Se	0.008
Be	0.004
Tl	0.0005
Cd	0.0003
Ag	0.000006

3.2. Assembly Statistics and Taxonomic Distribution

The Illumina sequencing of the artificial lake metagenome resulted in 19,302,650 high-quality paired-end reads with the Phred score of $\geq$30. The mean read length was 150 bp, contributing a total of ~6 GB data. These high-quality reads were assembled into 64,369 contigs, the average size of which was computed to be more than 500 bp at the N50 value of 789. A total of 23,18,091 ORFs were predicted from assembled

contigs (Table S1). These ORFs were used for functional and taxonomic profiling of the Himalayan site.

The taxonomic profiling of the lake metagenome revealed the predominance of Proteobacteria (52.09%) as the most abundant phylum, followed by Bacteroidetes (11.47%) and Verrucomicrobia (9.08%), as shown in Figure 2a (Table S2). A similar profile of abundant phyla was found in previously investigated high-altitude Pangong lake (Ladakh, India), Baikal Lake and a mesotrophic lake [5,27,28]. The abundance of proteobacterial members indicates an affluent microbial population for nitrogen fixation, methane metabolism, sulphur metabolism, etc. [29,30]. Bacteroidetes followed the pattern of abundance recorded in a freshwater lake, loaded with organic matter due to the stochastic disturbing events in the environment [31]. This states a positive correlation with the formation of the Himalayan artificial lake due to an unpredicted natural event, i.e., a landslide. At the species level, the iron-reducing bacterium, *Geothrix fermentans* (5.3%), was the most abundant (Figure 2b). *Geothrix fermentans*, known to reduce Fe(III) via a dissimilatory mechanism, is commonly found in aquifers [32,33]. This corroborates the high occurrence of iron in the Himalayan artificial lake sample (Table 2). The substantial representation of *Bacteroidetes bacterium* GWB2_41_8 (2.38%), *Nitrospira* (2.19%), and *Anaeromyxobacter* (2.15%) signify the nitrogen fixation and denitrification capability of the sampling site [31,34].

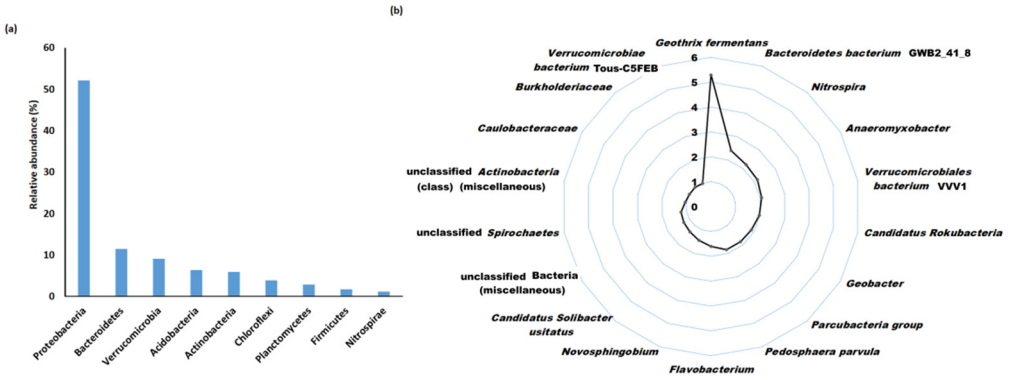


Figure 2. Taxonomic distribution in the Sikkim Himalayan artificial lake at (**a**) phylum level; (**b**) species level. The supportive details are given in Table S2.

The microbial diversity indices were computed, revealing the Simpson_1-D (0.98), Shannon_H (4.83), and Evenness (0.78) values (Table S3). On a scale of 0 to 1, the Simpson value towards 1 refers to a diverse microbial distribution, and the value of Evenness towards 1 relates to an evenly distributed species. A high Shannon_H value shows a rich species distribution in the sampling site. The species richness index was recorded to be higher than the samples collected from Prasher Lake (Shannon_H3.1–3.3) and Rewalsar lake (Shannon_H1.1–1.9) of Himachal Pradesh, India [35]. Thus, the statistical results signify high species diversity in the sample collected from the Himalayan artificial lake examined in the present study.

3.3. Functional Potential

The functional potential of the lake was predicted by aligning ORFs against the SEED database (Figure 3, Table S4). The carbohydrate metabolism (10.17%) was the most dominating among all functions in SEED classification (level 1), followed by Amino Acids and Derivatives (9.95%), Cofactors-Vitamins-Prosthetic Groups and Pigments (9.57%), and Protein Metabolism (6%). A similar pattern of functional distribution is reported in the metagenomic investigations of diverse aquatic habitats [5,36]. Interestingly, level 3 functional classification of the metagenomic data revealed the largest share of β-galactosidase (3.33%) in the carbohydrate metabolism. The contigs related to Chaperone protein DnaK (2.74%) were noted to be abundant in the category of Protein Metabolism. In the case of Amino Acids and Derivatives, the contribution of Cobalt-zinc-cadmium resistance protein

(8.54%) was prominent. Being a rich genetic resource of carbohydrate-metabolism-related genes, this site should be explored for industrially relevant enzymes for biomass processing applications. Level 3 SEED distribution revealed two predominant functions, Cofactors-Vitamins-Prosthetic Groups and Pigments, i.e., 5-FCL-like protein (17.25%) and Long-chain-fatty-acid-CoA ligase (12.73%). Although both enzymes are reported to serve fundamental metabolism functions, Long-chain-fatty-acid-CoA ligase helps in the pre-β-oxidation step of fatty acids with varying hydrocarbon chain lengths [37]. This draws attention towards exploring hydrocarbon metabolism in this lake's metagenome.

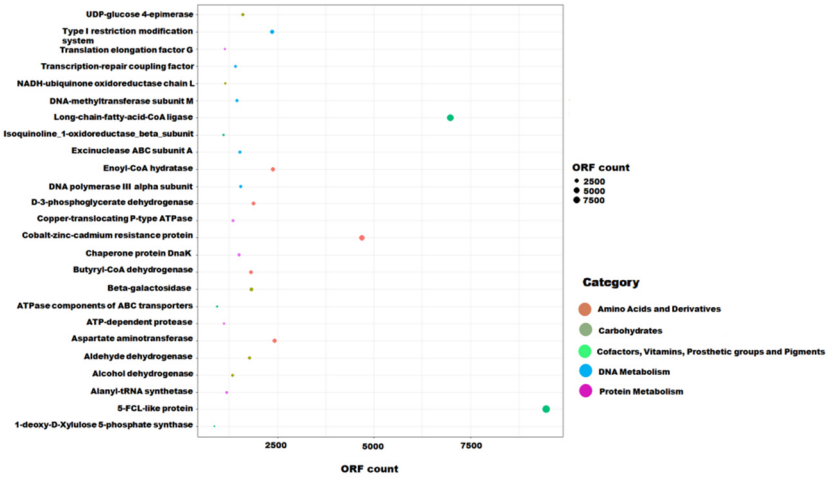
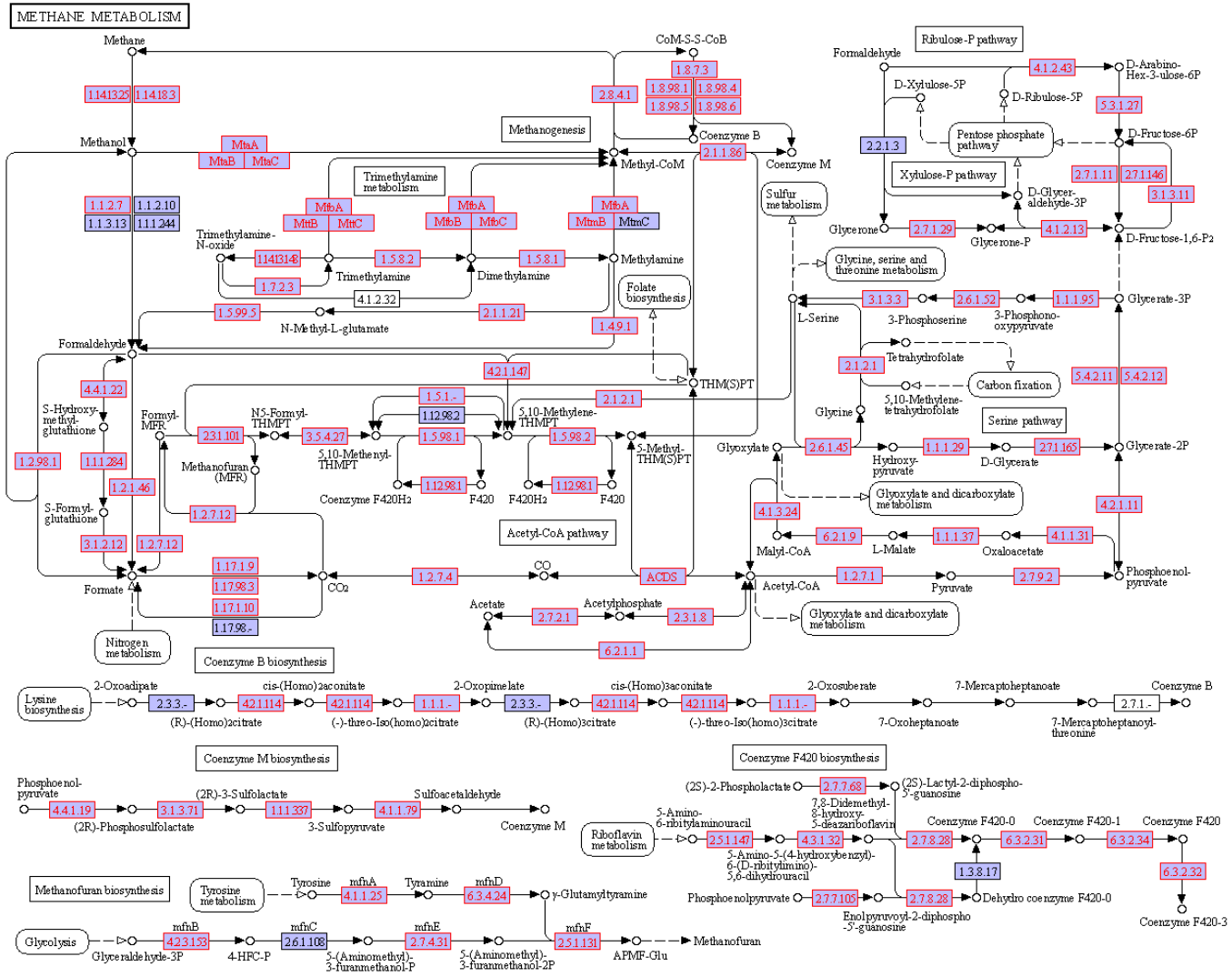


Figure 3. SEED function distribution in Sikkim Himalayan artificial lake. The supportive details are given in Table S4.

3.4. Hydrocarbon Degradation Pathway

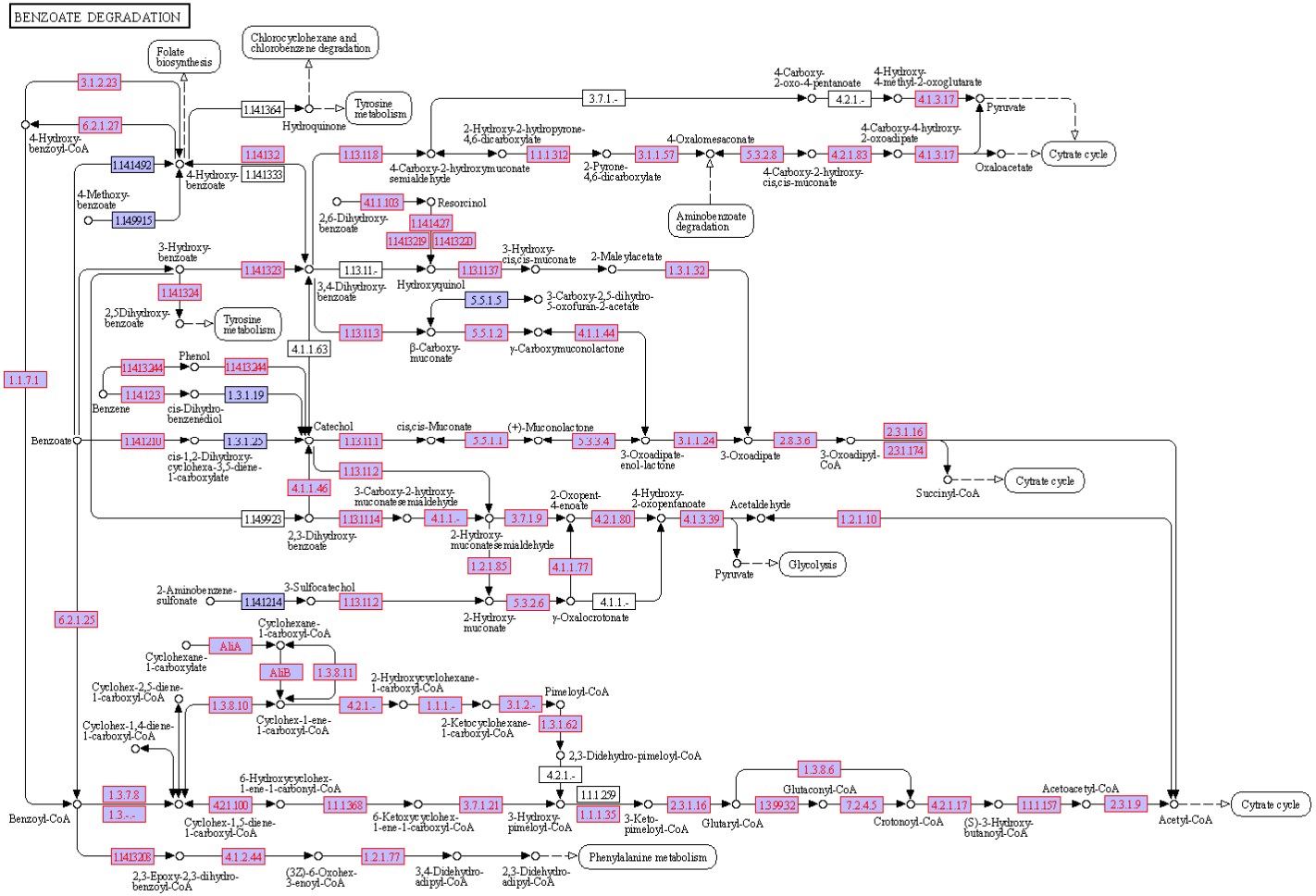
Hydrocarbons are known to be a severe threat to water bodies. The degradation of hydrocarbons via metabolic capabilities of various microbes could be a promising approach for treating hydrocarbon-polluted water [38]. In-depth metagenome sequence analysis revealed the occurrence of genes responsible for hydrocarbon degradation (Figure 4). The present sample was found enriched with methane metabolism enzymes. Methane contributes nearly 20% to the greenhouse effect, and lakes are the primary methane emission sources [39]. Thus, deciphering the genes mapped to the complete degradation pathway of methane would be helpful in designing strategies to diminish its harmful effect.

The other abundance pathways include Benzoate, Toluene, Benzene, and Xylene degradation pathways. Benzene, Toluene, and Xylene are classified as BETEX hydrocarbons, which constitute a single aromatic ring among other volatile organic compounds (VOCs) [40]. They are considered to be highly toxic and carcinogenic, and the predominance of their degradation pathways at the sampling site could exert a controlling effect on these harmful VOCs. The increased hydrocarbon-degrading enzymes in the sampling site could be due to the anthropogenic environmental pressure caused by land sliding [41]. The land-slide event possibly led to the degradation of plant-litter decomposition, resulting in the generation of VOCs [42]. In addition, being a tourist place, the human activities at the sampling site could lead to the accumulation of hydrocarbons. Thus, the microbial processes for the degradation of these hydrocarbons lead to an increase in hydrocarbon degrading enzymes in this region. These results are well related to East Taihu Lake (China), which showed an increase in hydrocarbon-degrading microbes due to disturbances caused by anthropogenic activities at the site [43]. This comprehensive metagenomic examination deciphered the genes involved in the degradation route of the hydrocarbons (Figure 4; Table S5).



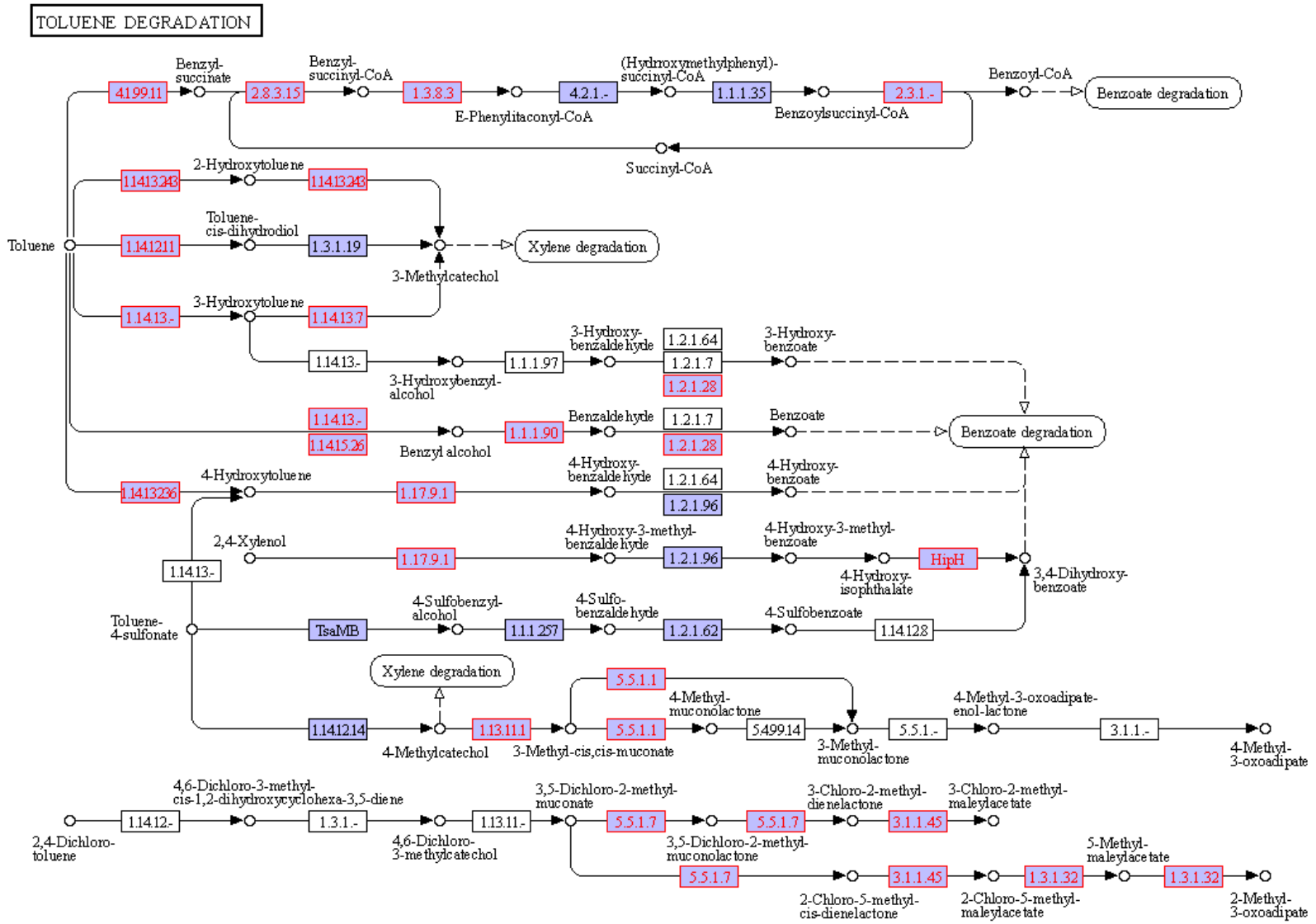
(**a**)

Figure 4. *Cont.*



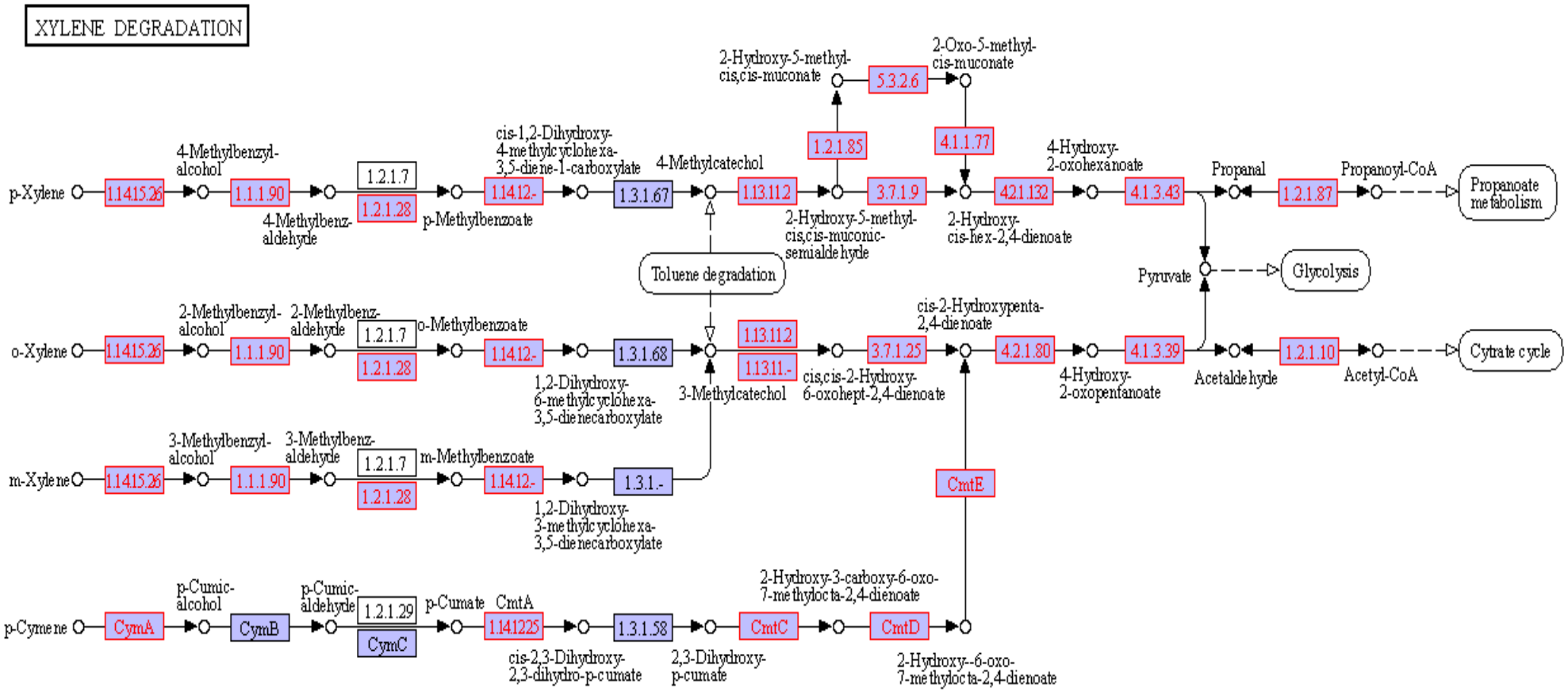
(**b**)

Figure 4. *Cont.*



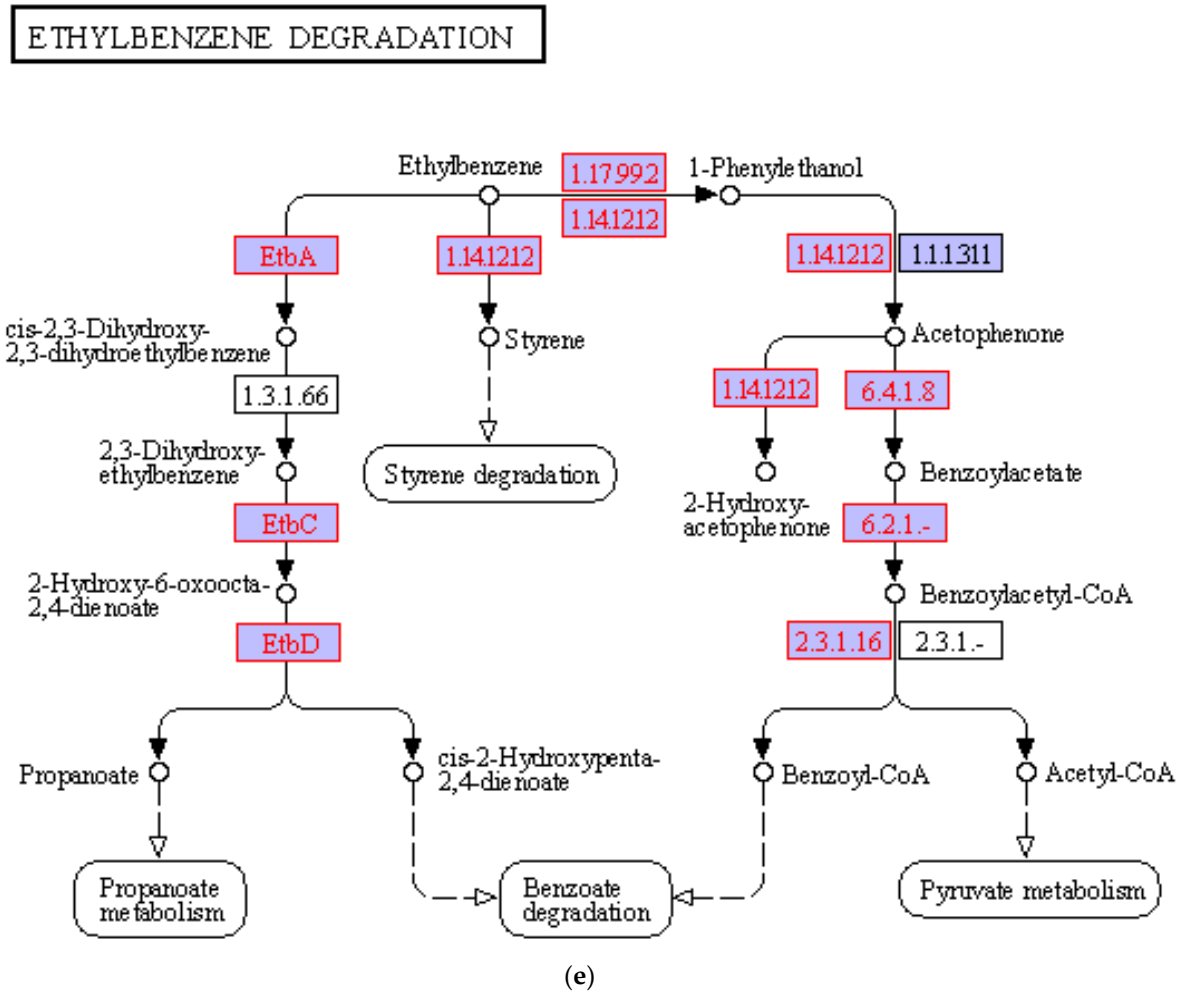
(**c**)

Figure 4. *Cont.*



(**d**)

Figure 4. *Cont*.



(**e**)

Figure 4. The metagenomic genes mapped to hydrocarbon degradation pathways (**a**) Methane degradation; (**b**) Benzoate degradation; (**c**) Toluene degradation; (**d**) Xylene degradation; (**e**) ethylbenzene degradation. The supportive details are given in Table S5.

3.5. Carbohydrate Metabolism

SEED functional profiling of the lake sample suggested an abundance of carbohydrate metabolism, stipulating this site as a potential resource for carbohydrate transforming enzymes. The ORFs aligned against the CAZy database revealed the highest abundance of the Glycosyl hydrolase (GH) family, followed by the Glycosyltransferase (GT) family enzymes (Figure 5; Table S6). The GTs are involved in transferring essential regulatory proteins and metal ions required for microflora's survival [44,45]. Among all the CAZy classes, GT2 was the most contributing CAZymes class, which represents many enzymes associated with bacterial biofilm formation, and critical in bacterial structural functions [46,47]. Also, glycosyl transferases (GT) are known to transfer the sugar moiety on the acceptor molecules and contribute to the production of various oligosaccharides that are of enormous health benefits [48].

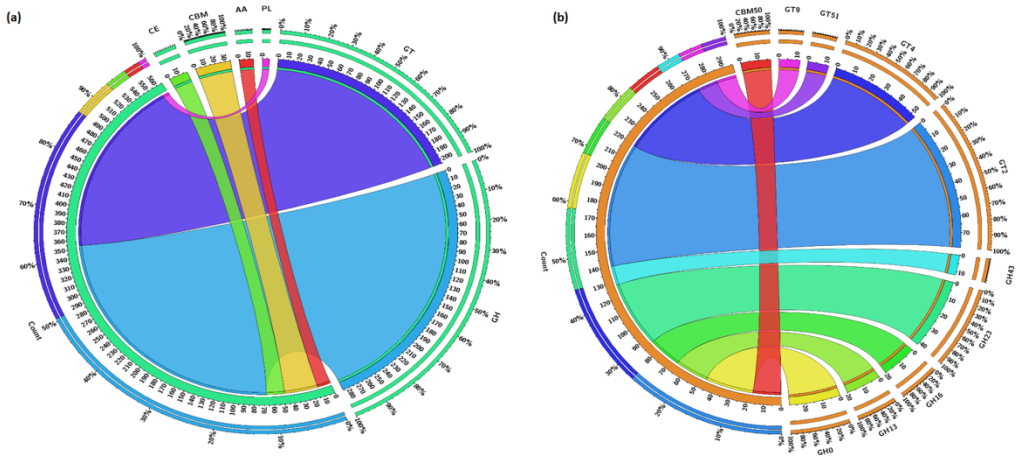


Figure 5. Distribution of carbohydrate active enzyme in metagenomic sample as (**a**) CAZy family; (**b**) top 10 CAZy classes. The supportive details are given in Table S6.

In the case of the GH family, GH23, GH0, and GH16 were the top three classes. The GH0 family of CAZymes majorly constituted proteins, which are not assigned to any glycosyl hydrolase family. The two proteins found under this category were lipoproteins and malate dehydrogenase. Lipoproteins are proteins anchored with the lipid bilayer of the bacterial cell wall and play a crucial role during membrane biogenesis [49]. Malate dehydrogenases are involved in the bacterial cell's oxidative and reductive TCA cycles. Besides converting malate to oxaloacetate, it also protects the cell from oxidative stress conditions [50]. In the case of GH23, lytic transglycosylases were one of the primary cell-machinery enzymes. These are mostly involved in the cleavage of the peptidoglycan backbone of bacteria by acting on glycosidic linkages and helping in the recycling of cell wall or formation of flagellar structures [51]. The metagenomic contigs mapped to GH16 enzyme family represented endo-β-1,4-galactosidase, β-glycosidase, endo-1,3-β-glucanase, etc. These enzymes are potential contributors to the processing of carbohydrate molecules [52]. Recently, we have characterized a novel β-galactosidase gene from the metagenome extracted from this high-altitude artificial lake site [53]. Thus, this site can be explored for mining various industrially relevant novel catalysts.

3.6. Antibiotic Resistance Genes

The antibiotic-resistance genes (ARGs) in the lake microbiome were assessed by performing BLASTX using ORFs against the CARD database. The results were analyzed at different alignment parameters, revealing a significant presence of tetracycline resistance gene, rpsJ (Table S8). These ARGs are widely found in aquatic habitats such as lakes [54,55]. Further, the metagenomic investigation detected the presence of genes associated to streptomycin resistance (rpsL), thiazolyl peptide resistance, and trimethoprim resistance at

this aquatic site. ARGs pose a potential threat to our environmental resources [56]. The Himalayan sites have continuous exposure to human activities that may cause a threat to their natural ecosystem leading towards the rise of antibiotic resistance in its microbial community [57]. Additionally, several microbial species tend to acquire the trait of antimicrobial resistance as a result of horizontal gene transfer when exposed to challenging environmental conditions [29,58]. The predominance of ARGs can be observed in a diverse set of environmental niches such as glaciers, hot springs, and lakes, across the globe [29]. Thus, the metagenomic assessment of ARGs' presence in a given lake would be helpful in defining the preventive measures for human activities in ecological niches.

3.7. Comparative Metagenomics

The metagenomic sample of the Himalayan artificial lake (present study) was compared with five other lake metagenomes reported from cold environments, i.e., Saline Ace Lake, Antarctica, Freshwater Vanda Lake, Antarctica, Saline organic lake in Antarctica, Freshwater Towuti Lake, Indonesia, and Freshwater Mendota Lake, United States. The temperature of these lakes ranges from 1 °C to 29 °C (Table 2). The metagenomic data of these sites were procured from IMG/M platform and analyzed using the server's built-in databases and algorithms. The taxonomic and functional distributions of all six metagenomic investigations were statistically compared. The PCA was performed by following STAMP multivariate analysis (Figure 6) at three levels, i.e., Phylum (Table S9), COG (Table S10), and KO (Table S11). A total of 78% to 88 % of the variance was covered in the PC1 and PC2 components of the PCA plot. Among all the six samples, taxonomic and functional classification grouped three samples together, Medonta Lake (US), Vanda lake (Antarctica), and the Himalayan artificial lake (India, present study). The three lakes shared the temperature range of 4 °C to 23 °C. All the freshwater samples were grouped in a single cluster except Freshwater Towuti Lake, Indonesia, the temperature of which was slightly higher (26–29 °C) than other freshwater samples. As expected, none of the saline lakes could cluster with any of these samples. This shows an association between the physicochemical characteristic of sampling sites and their taxonomic, as well as functional profiling.

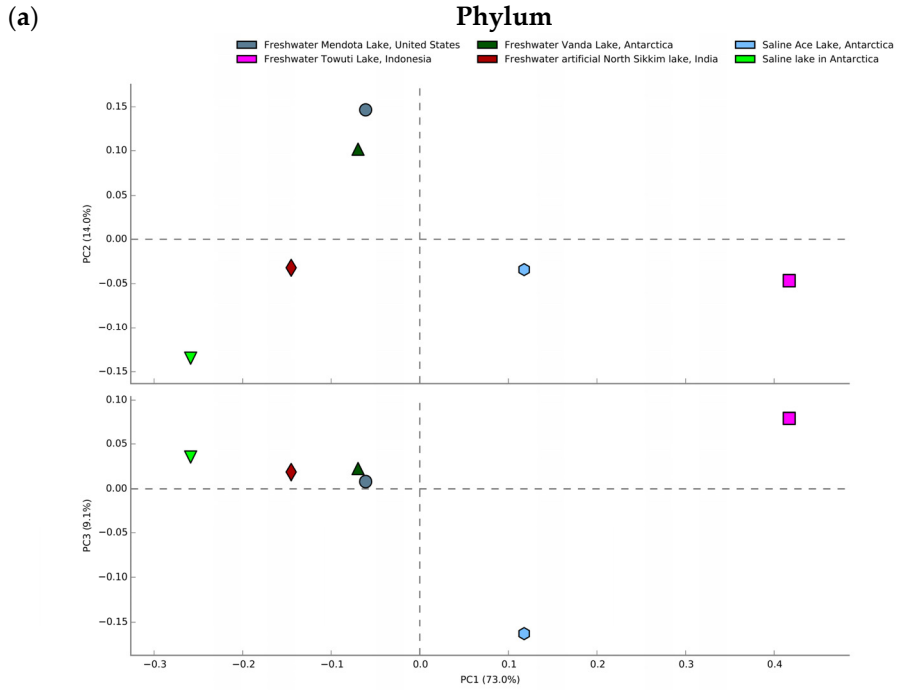


Figure 6. *Cont.*

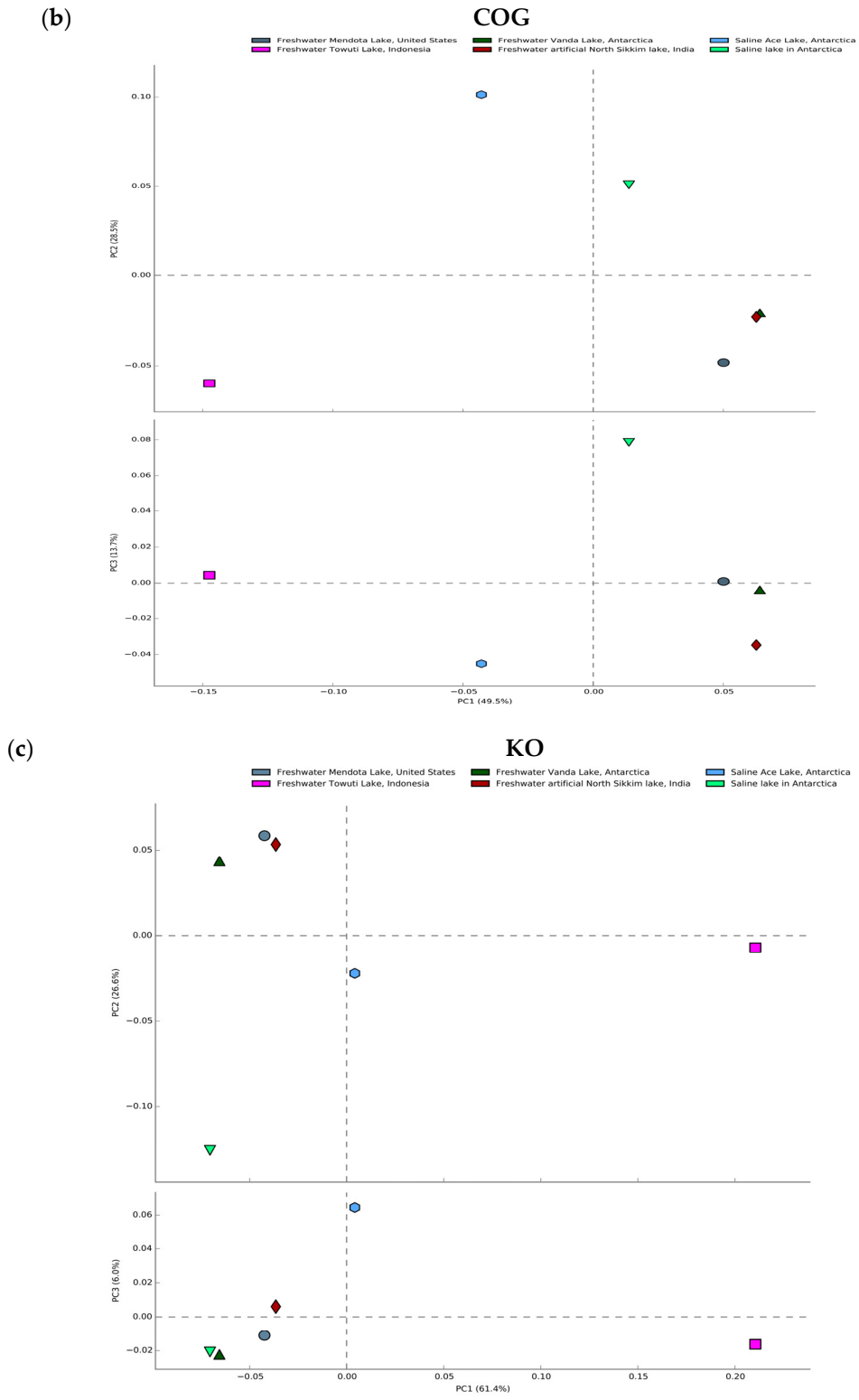


Figure 6. PCA plot showing diversity among the compared metagenomic samples on the basis of (**a**) Phylum; (**b**) COG; (**c**) KO.

The COG and KO functional profiles of all six lakes were compared by plotting heat maps (Figure 7). The heat-map clustering results were in accordance with the PCA, placing the functions from Medonta Lake (US), Vanda Lake (Antarctica), and the Himalayan artificial lake (India) more closely in comparison to other samples. Further, they shared a

comparable abundance of functions such as Glycosyltransferase (GTs), CheY chemotaxis protein, RNA polymerase sigma-70 factor, iron complex outer membrane receptor protein, ABC transport system permease protein, and serine/threonine-protein kinase. They contribute to the basic functional capabilities of the bacterial population. GTs are involved in the cell-wall biosynthesis phenomenon of microorganisms [59]. CheY protein contributes to chemotaxis in bacteria [60], and RNA polymerase sigma-70 factor is crucial in transcription initiation [61]. The genes related to the iron complex outer membrane receptor protein and the ABC transport-system permease protein are critical components of the iron transport system [62,63], and being rich in iron-reducing bacterial species, this functional potential is in agreement with the physicochemical and taxonomic contours of this site.

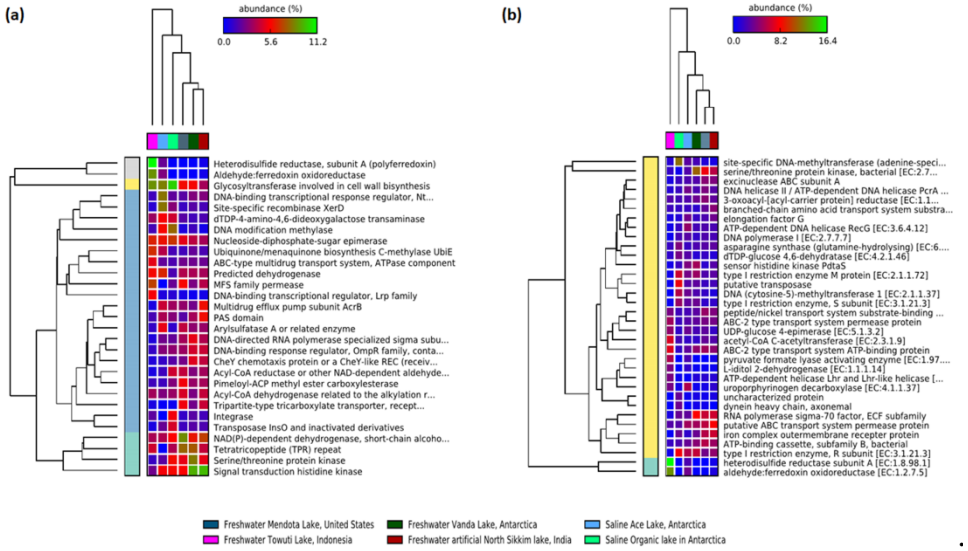


Figure 7. Heat maps representing functional profiles: (**a**) COG; (**b**) KO among the compared metagenomic samples.

Two-variable analysis was performed to determine the degree of similarity or difference among all the six lake samples at the phylum level (Figure 8). The error bars depicted that the Himalayan artificial lake exhibited a greater abundance of Proteobacteriaas compared to the Saline organic lake, Antarctica (p-value = 0.383). The Proteobacteria majorly include gram-negative bacterial families that usually hold great significance in industrial prospects. Although this phylum is reported to be widely distributed among a range of spatial-temporal conditions [6,8,64], few reports emphasized their higher abundance in freshwater lakes [31]. The phylum Atribacterota and Firmicutes were found abundant in the Saline Ace lake at a p-value of 0.086 and 0.211, respectively. Atribacterota was also detected in saline lake samples [65]. Firmicutes inhabits the saline lakes with higher abundance [8,66]. The abundance of Bacteriodetes in the Himalayan artificial lake (this study) was comparable to the other metagenomic samples, except Towuti Lake, which scored higher abundance at a p-value of 0.309. The distribution of phylum Bacteriodetes has been reported across a variety of ecological niches [5,27,30].

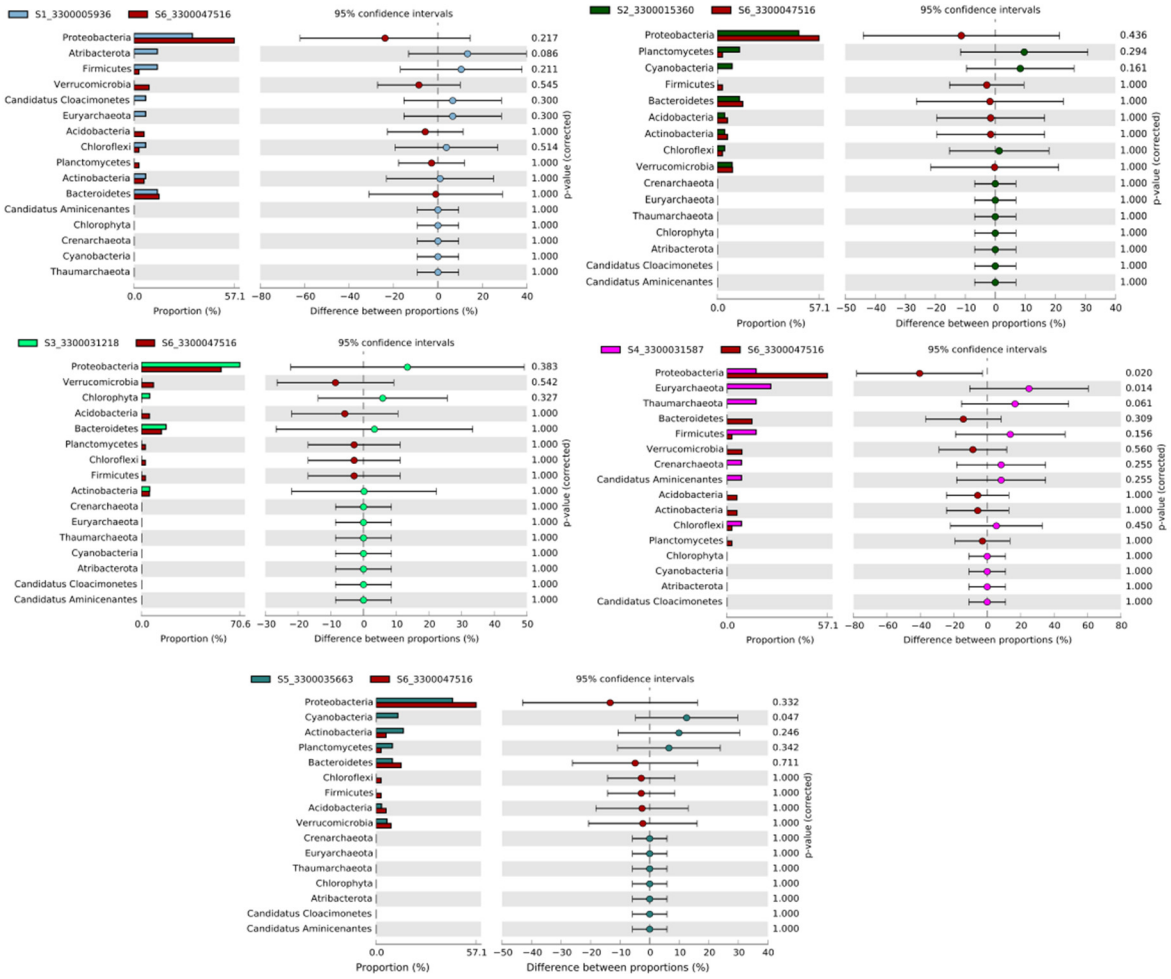


Figure 8. Error bar plots showing phylum distribution among the compared samples.

4. Conclusions

This study presents a detailed metagenomic investigation of a high-altitude Himalayan artificial lake located in the Sikkim state of India. It enlightens in-depth information about the taxonomic distribution and functional potential of the aquatic site. The calculated microbial diversity indices depicted the differential distribution of various species. The presence of *Geothrix fermentans* microbial species correlates with the high iron content in the lake. The exploration of metabolic pathways revealed the richness of hydrocarbon related genes, constructing the complete hydrocarbon degradation pathway by gene mapping. The alignment against the CAZy database revealed this site as a wealthy resource for industrially relevant carbohydrate modifying enzymes. The comparative study with five publicly available metagenomes revealed its relatedness towards the freshwater-type habitat. The results propound this site as a potential metagenomic resource for mining genes for industrially critical enzymes that could be functional at cold to moderate temperatures.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/su141811455/s1, Table S1: Assembly statistics of lake metagenome; Table S2: Taxonomic distribution in artificial lake; Table S3: Diversity indices of artificial lake metagenome; Table S4: SEED function distribution in artificial lake metagenome; Table S5: Hydrocarbon degrading pathways derived from KEGG analysis in metagenome sample; Table S6: The CAZy distribution in artificial lake metagenome sample; Table S7: Enzymes encoded by GH and GT CAZY families; Table S8: Antibiotic resistant genes derived from CARD database in metagenome sample; Table S9: Taxonomic comparison on the basis of abundance (%) among six metagenomic

samples; Table S10: COG comparison on the basis of abundance (%) among six metagenomic samples; Table S11: KEGG comparison on the basis of abundance (%) among six metagenomic samples.

Author Contributions: Conceptualization, S.P.S. and A.K.R.; methodology G.K., M.T., S.P.S. and A.K.R.; formal analysis, G.K. and M.T.; investigation, S.P.S. and A.K.R.; resources, S.P.S. and A.K.R.; data curation, G.K. and M.T.; writing—original draft preparation, G.K., M.T. and S.P.S.; writing—review and editing, G.K., M.T., S.P.S. and A.K.R.; supervision, S.P.S. and A.K.R.; funding acquisition, S.P.S. and A.K.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The metagenomic data of the Himalayan artificial lake was submitted to the NCBI SRA database under BioProject ID PRJNA786703.

Acknowledgments: Authors acknowledge the Department of Biotechnology (DBT), Govt. of India, for supporting this research. DBT-CIAB and DBT-IBSD are acknowledged for facilitating the collaborative research work. The Sikkim Biodiversity Board is acknowledged for granting permission to perform research. DBT-NABI is acknowledged for sharing the High-Performance Computing facility. GK and MT acknowledge DST inspire and CSIR fellowships, respectively.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or; in the decision to publish the results.

References

1. Badar, B.; Romshoo, S.A.; Khan, M.A. Integrating biophysical and socioeconomic information for prioritizing watersheds in a Kashmir Himalayan lake: A remote sensing and GIS approach. *Environ. Monit. Assess.* **2013**, *185*, 6419–6445. [CrossRef] [PubMed]
2. Ahmad, T.; Gupta, G.; Sharma, A.; Kaur, B.; El-Sheikh, M.A.; Alyemeni, M.N. Metagenomic analysis exploring taxonomic and functional diversity of bacterial communities of a Himalayan urban fresh water lake. *PLoS ONE* **2021**, *16*, e0248116. [CrossRef] [PubMed]
3. Skvortsov, T.; De Leeuwe, C.; Quinn, J.P.; McGrath, J.W.; Allen, C.C.R.; McElarney, Y.; Watson, C.; Arkhipova, K.; Lavigne, R.; Kulakov, L.A. Metagenomic characterisation of the viral community of Lough Neagh, the largest freshwater lake in Ireland. *PLoS ONE* **2016**, *11*, e0150361. [CrossRef] [PubMed]
4. Moon, K.; Jeon, J.H.; Kang, I.; Park, K.S.; Lee, K.; Cha, C.-J.; Lee, S.H.; Cho, J.-C. Freshwater viral metagenome reveals novel and functional phage-borne antibiotic resistance genes. *Microbiome* **2020**, *8*, 1–15. [CrossRef] [PubMed]
5. Rathour, R.; Gupta, J.; Mishra, A.; Rajeev, A.C.; Dupont, C.L.; Thakur, I.S. A comparative metagenomic study reveals microbial diversity and their role in the biogeochemical cycling of Pangong lake. *Sci. Total Environ.* **2020**, *731*, 139074. [CrossRef] [PubMed]
6. Rai, A.; Bhattacharjee, A. Molecular profiling of microbial community structure and their CAZymes via metagenomics, from Tsomgo lake in the Eastern Himalayas. *Arch. Microbiol.* **2021**, *203*, 3135–3146. [CrossRef]
7. Vavourakis, C.D.; Ghai, R.; Rodriguez-Valera, F.; Sorokin, D.Y.; Tringe, S.G.; Hugenholtz, P.; Muyzer, G. Metagenomic insights into the uncultured diversity and physiology of microbes in four hypersaline soda lake brines. *Front. Microbiol.* **2016**, *7*, 211. [CrossRef]
8. Chakraborty, J.; Sapkale, V.; Shah, M.; Rajput, V.; Mehetre, G.; Agawane, S.; Kamble, S.; Dharne, M. Metagenome sequencing to unveil microbial community composition and prevalence of antibiotic and metal resistance genes in hypersaline and hyperalkaline Lonar Lake, India. *Ecol. Indic.* **2020**, *110*, 105827. [CrossRef]
9. Panwar, P.; Allen, M.A.; Williams, T.J.; Hancock, A.M.; Brazendale, S.; Bevington, J.; Roux, S.; Páez-Espino, D.; Nayfach, S.; Berg, M. Influence of the polar light cycle on seasonal dynamics of an Antarctic lake microbial community. *Microbiome* **2020**, *8*, 1–24. [CrossRef]
10. Liang, H.; Wang, F.; Mu, R.; Huang, J.; Zhao, R.; Li, X.; Yu, K.; Li, B. Metagenomics analysis revealing the occurrence of antibiotic resistome in salt lakes. *Sci. Total Environ.* **2021**, *790*, 148262. [CrossRef]
11. Phukon, L.C.; Chourasia, R.; Kumari, M.; Godan, T.K.; Sahoo, D.; Parameswaran, B.; Rai, A.K. Production and characterisation of lipase for application in detergent industry from a novel *Pseudomonas helmanticensis* HS6. *Bioresour. Technol.* **2020**, *309*, 123352. [CrossRef] [PubMed]
12. Thakur, M.; Sharma, N.; Rai, A.K.; Singh, S.P. A novel cold-active type I pullulanase from a hot-spring metagenome for effective debranching and production of resistant starch. *Bioresour. Technol.* **2021**, *320*, 124288. [CrossRef] [PubMed]
13. Sharma, N.; Sahoo, D.; Rai, A.K.; Singh, S.P. A highly alkaline pectate lyase from the Himalayan hot spring metagenome and its bioscouring applications. *Process Biochem.* **2022**, *115*, 100–109. [CrossRef]

14. Singh, A.K.; Kumari, M.; Sharma, N.; Rai, A.K.; Singh, S.P. Metagenomic views on taxonomic and functional profiles of the Himalayan Tsomgo cold lake and unveiling its deterzome potential. *Curr. Genet.* **2022**, 1–15. [CrossRef]
15. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* **2011**, *17*, 10–12. [CrossRef]
16. Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. metaSPAdes: A new versatile metagenomic assembler. *Genome Res.* **2017**, *27*, 824–834. [CrossRef]
17. Mikheenko, A.; Prjibelski, A.; Saveliev, V.; Antipov, D.; Gurevich, A. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* **2018**, *34*, i142–i150. [CrossRef]
18. Noguchi, H.; Taniguchi, T.; Itoh, T. MetaGeneAnnotator: Detecting species-specific patterns of ribosomal binding site for precise gene prediction in anonymous prokaryotic and phage genomes. *DNA Res.* **2008**, *15*, 387–396. [CrossRef]
19. Mount, D.W. Using the basic local alignment search tool (BLAST). *Cold Spring Harb. Protoc.* **2007**, *2007*, 17. [CrossRef]
20. Huson, D.H.; Beier, S.; Flade, I.; Górska, A.; El-Hadidi, M.; Mitra, S.; Ruscheweyh, H.-J.; Tappu, R. MEGAN community edition-interactive exploration and analysis of large-scale microbiome sequencing data. *PLoS Comput. Biol.* **2016**, *12*, e1004957. [CrossRef]
21. Hammer, Ø.; Harper, D.A.; Ryan, P.D. PAST: Paleontological statistics software package for education and data analysis. *Palaeontol. Electron.* **2001**, *4*, 9.
22. Jia, B.; Raphenya, A.R.; Alcock, B.; Waglechner, N.; Guo, P.; Tsang, K.K.; Lago, B.A.; Dave, B.M.; Pereira, S.; Sharma, A.N. CARD 2017: Expansion and model-centric curation of the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **2016**, *45*, 566–573. [CrossRef]
23. Kanehisa, M.; Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30. [CrossRef] [PubMed]
24. 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. [CrossRef] [PubMed]
25. Chen, I.M.A.; Chu, K.; Palaniappan, K.; Pillay, M.; Ratner, A.; Huang, J.; Huntemann, M.; Varghese, N.; White, J.R.; Seshadri, R. IMG/M v. 5.0: An integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res.* **2019**, *47*, D666–D677. [CrossRef] [PubMed]
26. Parks, D.H.; Tyson, G.W.; Hugenholtz, P.; Beiko, R.G. STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics* **2014**, *30*, 3123–3124. [CrossRef]
27. Coutinho, F.H.; Cabello-Yeves, P.J.; Gonzalez-Serrano, R.; Rosselli, R.; López-Pérez, M.; Zemskaya, T.I.; Zakharenko, A.S.; Ivanov, V.G.; Rodriguez-Valera, F. New viral biogeochemical roles revealed through metagenomic analysis of Lake Baikal. *Microbiome* **2020**, *8*, 163. [CrossRef]
28. Debroas, D.; Humbert, J.-F.; Enault, F.; Bronner, G.; Faubladier, M.; Cornillot, E. Metagenomic approach studying the taxonomic and functional diversity of the bacterial community in a mesotrophic lake (Lac du Bourget–France). *Environ. Microbiol.* **2009**, *11*, 2412–2424. [CrossRef]
29. 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*, 1–17. [CrossRef] [PubMed]
30. Zhou, Z.; Tran, P.Q.; Kieft, K.; Anantharaman, K. Genome diversification in globally distributed novel marine Proteobacteria is linked to environmental adaptation. *ISME J.* **2020**, *14*, 2060–2077. [CrossRef] [PubMed]
31. Newton, R.J.; Jones, S.E.; Eiler, A.; McMahon, K.D.; Bertilsson, S. A guide to the natural history of freshwater lake bacteria. *Microbiol. Mol. Biol. Rev.* **2011**, *75*, 14–49. [CrossRef]
32. Nevin, K.P.; Lovley, D.R. Mechanisms for accessing insoluble Fe(III) oxide during dissimilatory Fe(III) reduction by *Geothrix fermentans*. *Appl. Environ. Microbiol.* **2002**, *68*, 2294–2299. [CrossRef] [PubMed]
33. Coates, J.D.; Ellis, D.J.; Gaw, C.V.; Lovley, D.R. *Geothrix fermentans* gen. nov., sp. nov., a novel Fe(III)-reducing bacterium from a hydrocarbon-contaminated aquifer. *Int. J. Syst. Bacteriol.* **1999**, *49 Pt 4*, 1615–1622. [CrossRef] [PubMed]
34. Onley, J.R.; Ahsan, S.; Sanford, R.A.; Löffler, F.E. Denitrification by *Anaeromyxobacter dehalogenans*, a Common Soil Bacterium Lacking the Nitrite Reductase Genes nirS and nirK. *Appl. Environ. Microbiol.* **2018**, *84*, e01985-17. [CrossRef] [PubMed]
35. Thakur, R.K.; Jindal, R.; Singh, U.B.; Ahluwalia, A.S. Plankton diversity and water quality assessment of three freshwater lakes of Mandi (Himachal Pradesh, India) with special reference to planktonic indicators. *Environ. Monit. Assess.* **2013**, *185*, 8355–8373. [CrossRef]
36. Zhang, Y.; Xu, H.; Wang, L.; Liu, R.; Fu, L.; Lin, K. Unique bacterial communities and potential function along the vertical gradient in the deepest marine blue hole. *Environ. Microbiol. Rep.* **2021**, *13*, 911–927. [CrossRef]
37. Kumar, S.; Zhou, J.; Li, M.; Xiang, H.; Zhao, D. Insights into the metabolism pathway and functional genes of long-chain aliphatic alkane degradation in haloarchaea. *Extremophiles* **2020**, *24*, 475–483. [CrossRef]
38. Das, N.; Chandran, P. Microbial Degradation of Petroleum Hydrocarbon Contaminants: An Overview. *Biotechnol. Res. Int.* **2011**, *2011*, 941810. [CrossRef]
39. Bastviken, D.; Cole, J.; Pace, M.; Tranvik, L. Methane emissions from lakes: Dependence of lake characteristics, two regional assessments, and a global estimate. *Global Biogeochem. Cycles* **2004**, *18*, 4. [CrossRef]
40. Masekameni, M.D.; Moolla, R.; Gulumian, M.; Brouwer, D. Risk Assessment of Benzene, Toluene, Ethyl Benzene, and Xylene Concentrations from the Combustion of Coal in a Controlled Laboratory Environment. *Int. J. Environ. Res. Public Health* **2018**, *16*, 95. [CrossRef]

41. Yan, Z.; Song, N.; Wang, C.; Jiang, H. Functional potential and assembly of microbes from sediments in a lake bay and adjoining river ecosystem for polycyclic aromatic hydrocarbon biodegradation. *Environ. Microbiol.* **2021**, *23*, 628–640. [CrossRef] [PubMed]
42. Derendorp, L.; Holzinger, R.; Röckmann, T. UV-induced emissions of C2–C5 hydrocarbons from leaf litter. *Environ. Chem.* **2011**, *8*, 602–611. [CrossRef]
43. Guo, J.; Fang, J.; Cao, J. Characteristics of petroleum contaminants and their distribution in Lake Taihu, China. *Chem. Cent. J.* **2012**, *6*, 92. [CrossRef] [PubMed]
44. Zhu, F.; Wu, R.; Zhang, H.; Wu, H. Structural and biochemical analysis of a bacterial glycosyltransferase. *Glycosyltransferases* **2013**, *1022*, 29–39.
45. Schmid, J.; Heider, D.; Wendel, N.J.; Sperl, N.; Sieber, V. Bacterial Glycosyltransferases: Challenges and Opportunities of a Highly Diverse Enzyme Class Toward Tailoring Natural Products. *Front. Microbiol.* **2016**, *7*, 182. [CrossRef]
46. Yakovlieva, L.; Walvoort, M.T.C. Processivity in Bacterial Glycosyltransferases. *ACS Chem. Biol.* **2020**, *15*, 3–16. [CrossRef]
47. Oehme, D.P.; Shafee, T.; Downton, M.T.; Bacic, A.; Doblin, M.S. Differences in protein structural regions that impact functional specificity in GT2 family β-glucan synthases. *PLoS ONE* **2019**, *14*, e0224442. [CrossRef]
48. Palcic, M.M. Glycosyltransferases as biocatalysts. *Curr. Opin. Chem. Biol.* **2011**, *15*, 226–233. [CrossRef]
49. Narita, S.I.; Tokuda, H. Bacterial lipoproteins; biogenesis, sorting and quality control. *Biochim. Biophys. Acta-Mol. Cell Biol. Lipids* **2017**, *1862*, 1414–1423. [CrossRef]
50. Takahashi-Íñiguez, T.; Aburto-Rodríguez, N.; Vilchis-González, A.L.; Flores, M.E. Function, kinetic properties, crystallization, and regulation of microbial malate dehydrogenase. *J. Zhejiang Univ. Sci. B* **2016**, *17*, 247–261. [CrossRef]
51. Jenkins, C.H.; Wallis, R.; Allcock, N.; Barnes, K.B.; Richards, M.I.; Auty, J.M.; Galyov, E.E.; Harding, S.V.; Mukamolova, G.V. The lytic transglycosylase, LtgG, controls cell morphology and virulence in Burkholderia pseudomallei. *Sci. Rep.* **2019**, *9*, 11060. [CrossRef] [PubMed]
52. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P.M.; Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2014**, *42*, D490–D495. [CrossRef] [PubMed]
53. Thakur, M.; Kumar Rai, A.; Singh, S.P. An acid-tolerant and cold-active β-galactosidase potentially suitable to process milk and whey samples. *Appl. Microbiol. Biotechnol.* **2022**, *106*, 3599–3610. [CrossRef] [PubMed]
54. Yang, Y.; Song, W.; Lin, H.; Wang, W.; Du, L.; Xing, W. Antibiotics and antibiotic resistance genes in global lakes: A review and meta-analysis. *Environ. Int.* **2018**, *116*, 60–73. [CrossRef] [PubMed]
55. Ohore, O.E.; Addo, F.G.; Han, N.; Li, X.; Zhang, S. Profiles of ARGs and their relationships with antibiotics, metals and environmental parameters in vertical sediment layers of three lakes in China. *J. Environ. Manag.* **2020**, *255*, 109583. [CrossRef]
56. Bengtsson-Palme, J.; Boulund, F.; Fick, J.; Kristiansson, E.; Larsson, D.G.J. Shotgun metagenomics reveals a wide array of antibiotic resistance genes and mobile elements in a polluted lake in India. *Front. Microbiol.* **2014**, *5*, 648. [CrossRef]
57. Sharma, N.; Kumari, R.; Thakur, M.; Rai, A.K.; Singh, S.P. Molecular dissemination of emerging antibiotic, biocide, and metal co-resistomes in the Himalayan hot springs. *J. Environ. Manag.* **2022**, *307*, 114569. [CrossRef]
58. Demanèche, S.; Kay, E.; Gourbière, F.; Simonet, P. Natural transformation of Pseudomonas fluorescens and Agrobacterium tumefaciens in soil. *Appl. Environ. Microbiol.* **2001**, *67*, 2617–2621. [CrossRef]
59. Mesleh, M.F.; Rajaratnam, P.; Conrad, M.; Chandrasekaran, V.; Liu, C.M.; Pandya, B.A.; Hwang, Y.S.; Rye, P.T.; Muldoon, C.; Becker, B.; et al. Targeting Bacterial Cell Wall Peptidoglycan Synthesis by Inhibition of Glycosyltransferase Activity. *Chem. Biol. Drug Des.* **2016**, *87*, 190–199. [CrossRef]
60. Sarkar, M.K.; Paul, K.; Blair, D. Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 9370. [CrossRef]
61. Paget, M.S.; Helmann, J.D. The sigma70 family of sigma factors. *Genome Biol.* **2003**, *4*, 203. [CrossRef] [PubMed]
62. Mosbahi, K.; Wojnowska, M.; Albalat, A.; Walker, D. Bacterial iron acquisition mediated by outer membrane translocation and cleavage of a host protein. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 6840. [CrossRef]
63. Lau, C.K.Y.; Krewulak, K.D.; Vogel, H.J. Bacterial ferrous iron transport: The Feo system. *FEMS Microbiol. Rev.* **2016**, *40*, 273–298. [CrossRef] [PubMed]
64. Eiler, A.; Zaremba-Niedzwiedzka, K.; Martínez-García, M.; McMahon, K.D.; Stepanauskas, R.; Andersson, S.G.E.; Bertilsson, S. Productivity and salinity structuring of the microplankton revealed by comparative freshwater metagenomics. *Environ. Microbiol.* **2014**, *16*, 2682–2698. [CrossRef] [PubMed]
65. Katayama, T.; Nobu, M.K.; Kusada, H.; Meng, X.-Y.; Hosogi, N.; Uematsu, K.; Yoshioka, H.; Kamagata, Y.; Tamaki, H. Isolation of a member of the candidate phylum 'Atribacteria' reveals a unique cell membrane structure. *Nat. Commun.* **2020**, *11*, 6381. [CrossRef] [PubMed]
66. Genderjahn, S.; Alawi, M.; Mangelsdorf, K.; Horn, F.; Wagner, D. Desiccation- and saline-tolerant bacteria and archaea in Kalahari pan sediments. *Front. Microbiol.* **2018**, *9*, 2082. [CrossRef] [PubMed]