Metagenomics revealing molecular profiles of microbial community structure and metabolic capacity in Bamucuo lake, Tibet

Cai Wei [a,b,c], Dan Sun [a,b,c], Wenliang Yuan [d], Lei Li [e], Chaoxu Dai [a,b,c], Zuozhou Chen [a,b,c], Xiaomin Zeng [f], Shihang Wang [a,b,c], Yuyang Zhang [a,b,c], Shouwen Jiang [a,b,c], Zhichao Wu [a,b,c], Dong Liu [a,b,c,**], Linhua Jiang [e,***], Sihua Peng [a,b,c,*]

[a] Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Shanghai Ocean University, Shanghai, 201306, PR China
[b] National Pathogen Collection Center for Aquatic Animals, Ministry of Agriculture of China, Shanghai, 201306, China
[c] International Research Center for Marine Biosciences at Shanghai Ocean University, Ministry of Science and Technology, Shanghai, 201306, PR China
[d] College of Mathematics and Information Engineering, Jiaxing University, Jiaxing, 314033, PR China
[e] Engineering Research Center of AI & Robotics, Ministry of Education, Academy for Engineering & Technology, Fudan University, Shanghai, 200433, PR China
[f] Central South University Xiangya Public Health School, Changsha, 410078, PR China

ARTICLE INFO

ABSTRACT

Microorganisms play critical ecological roles in the global biogeochemical cycles. However, extensive information on the microbial communities in Qinghai-Tibet Plateau (QTP), which is the highest plateau in the world, is still lacking, particularly in high elevation locations above 4500 m. Here, we performed a survey of th e soil and water microbial communities in Bamucuo Lake, Tibet, by using shotgun metagenomic methods. In the soil and water samples, we reconstructed 75 almost complete metagenomic assembly genomes, and 74 of the metagenomic assembly genomes from the water sample represented novel species. Proteobacteria and Actinobacteria were found to be the dominant bacterial phyla, while Euryarchaeota was the dominant archaeal phylum. The largest virus, Pandoravirus salinus, was found in the soil microbial community. We concluded that the microorganisms in Bamucuo Lake are most likely to fix carbon mainly through the 3-hydroxypropionic bi-cycle pathway. This study, for the first time, characterized the microbial community composition and metabolic capacity in QTP high-elevation locations with 4555 m, confirming that QTP is a vast and valuable resource pool, in which many microorganisms can be used to develop new bioactive substances and new antibiotics to which pathogenic microorganisms have not yet developed resistance.

1. Introduction

The Qinghai-Tibet Plateau (QTP) is the world's highest plateau and is known as the "third Pole", which is one of the most important water resources in East Asia and plays an important role in regulating global climate change (Chen et al., 2014; Kuang and Jiao, 2016). Via various metabolic pathways, the microorganisms in QTP form the basis of carbon, nitrogen and various nutrient cycles (Rui et al., 2015; Cavicchioli et al., 2019), and influence climate change to some extent, such as greenhouse gas emission (CO_2, CH_4, N_2O) (Raymond et al., 2013; Saunois et al., 2016; DelSontro et al., 2018). However, up to now, the diversity information of the microbial communities in QTP is still lacking.

To address this issue, we initiated this research by using metagenomics to increase the understanding of the microorganisms in this habitat.

With the advancement of second generation sequencing technology, it is now possible to collect information from previously uncultured microorganisms using metagenomics (Jousset et al., 2017). In recent years, the diversity of the microbial communities of the permafrost in QTP has become the focus of most research due to its unique and fragile characteristics, that is, permafrost microbial communities are susceptible to climate change (Hu et al., 2014, 2016; Wei et al., 2014; Chen et al., 2017). Han et al. revealed the structure and diversity of the microbial community at Keke Salt Lake (also in QTP) (Han et al., 2017). Xing's research on the soil microbial community in Qaidam Basin showed that environmental factors are the primary driving force for the establishment of bacterial community structure (Xing et al., 2019). However, at an elevation of above 4000 m, only a few environmental microbiology investigations have been conducted, most of them using the simpler 16S ribosomal RNA gene sequencing technique rather than shotgun metagenomic sequencing (Wei et al., 2014; Hu et al., 2016; Chen et al., 2017; Han et al., 2017; Xing et al., 2019; Ma et al., 2020). In comparison to shotgun metagenomic sequencing, 16S rRNA gene profiling reveals just the taxonomic composition of the microbial community but not its functional annotation (Sunagawa et al., 2013; Ranjan et al., 2016).

Shotgun metagenomic sequencing not only provides a stronger and more reliable assessment of microbial diversity, but also provides valuable information on the functional annotation including the metabolic potential of microbial communities (Sharpton, 2014; Zhou et al., 2015; Alneberg et al., 2018; Wiseschart et al., 2019).

Bamucuo Lake, at an elevation of 4555 m, is located in QTP and is primarily fed by surface runoff from the surrounding basin.

Because of the harsh environmental conditions, research on the microbial community of high-altitude lakes is nearly non-existent. We discovered that there are few high-altitude metagenomics research literatures in the world, and the majority of the existing literatures are based on 16S ribosomal RNA gene sequencing. Are the microbial community distributions at high and low altitudes very different? What are the different characteristics of the community structure, metabolic potential, material cycles in the ecosystems, and various specific gene clusters compared with those of microorganisms at low altitudes? These are issues that the international academic community may be concerned about. In this regard, there is an urgent need to initiate this research.

Here we investigated the microbial community in Bamucuo Lake using shotgun metagenomics. To the best of our knowledge, this study is the first shotgun metagenomics study of the environmental microbial community at an elevation of over 4000 m.

2. Materials and methods

2.1. Sampling site and sample collection

The soil and water samples involved in this study were collected on June 20, 2017 at an elevation of 4555 m in Bamucuo Lake (31°34′N, 90°6′E), Bango County, Nagqu Prefecture, Tibet, China (Fig. 1). To make the sampling more representative of the lake water, we collected multiple buckets of water every 10 m at a depth of 0.5m within a range of about 100 m, and a total of three replicate samples of 30L each were collected. Then, the water collected was prefiltered through a 3 μm pore size organic microporous membrane filter, followed by a 0.22 μm pore size mixed fiber membrane filter, and the water temperature was 13 °C at the time of sampling. The surrounding air temperature was 19 °C.

Similarly, to make the soil sampling representative of the lakeshore soil, we also adopted a multi-point sampling strategy. The three replicate soil samples were collected at about 15 m from the bank of the lake, each of which was a mixture of soil samples from three locations about 100 m apart. For detail, after removing the surface soil, roughly 10g of meadow soil was taken from 20 cm below the surface and placed in a 50 ml centrifuge tube, which was then stored in ice boxes before being transported to the laboratory.

After DNA extraction, we selected one sample with the best DNA extraction quality from three water samples and from three soil samples, respectively, for subsequent metagenome sequencing. In this paper, we refer to these two samples as s05-02 (soil) and s05-03 (water).

2.2. DNA extraction and sequencing

According to the manufacturer's instructions, genomic DNA was extracted using OMEGA's DNA extraction kit (http://www.omegabiotek.com.cn/, Guangzhou, China), and the experimental process was modified slightly, including the cell lysis step was increased by 5 min, and the centrifugal force in the DNA extraction step was increased from 12,000*g to 14,000*g, with centrifuge time increased by 20 s. The purity and concentration of the extracted genomic DNA were measured on NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Massachusetts, United States), and the quality of the extracted genomic DNA was characterized by 1% agarose gel electrophoresis. The total genomic
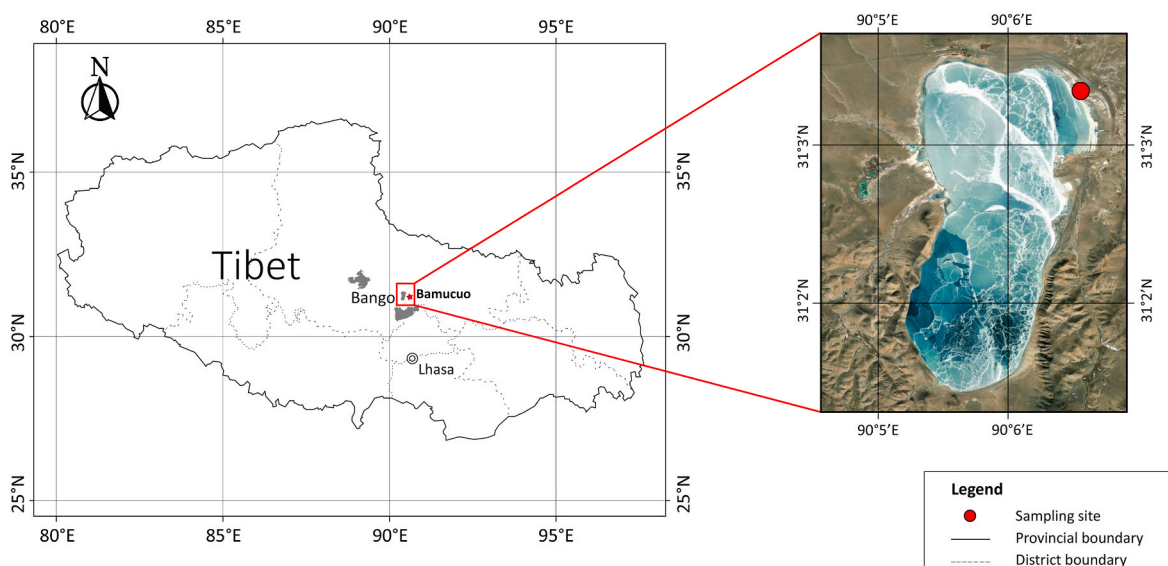

Fig. 1. The map indicating the sampling site in Bamucuo Lake, Tibet.

DNA was stored in a −20 °C refrigerator. Then, the genomic DNA was sent to Shanghai Meiji Biomedical Technology Co., Ltd for sequencing. The length of the DNA library was around 300bp, and Illumina HiSeq 4000 sequencer was used for sequencing with 150 bp paired-end sequencing.

2.3. Metagenomic assembly and gene abundance analysis

Fastp v. 0.19.5 was used to remove adaptor sequences at the 3′ and 5′ terminals from the raw reads (Chen et al., 2018a). The reads with lengths less than 50 bp after trimming, reads with average quality values (Phred values) less than 20, and reads containing N bases were all removed. FastQC v. 0.11.9 (Cock et al., 2010) was used to perform the sequence quality control. The assembly module of MetaWRAP v. 1.2.1 (Uritskiy et al., 2018) was then used for assembling, which include the assembly software MEGAHIT v 1.1.3 (Li et al., 2016), with the following parameters: minimum contig length of 1000 bp, and k-mer sizes of 21, 29, 39, 59, 79, 99, 119 and 141, respectively. BBMap v. 38.87 (https://sourceforge.net/projects/bbmap/) was used to map clean reads to each contig for calculating the coverage information. Prokka v. 1.14.6 (Seemann, 2014) was used to annotate the assembled sequence, with all the parameters set to their default values. Salmon v. 1.3.0 (Patro et al., 2017) was used to estimate the genetic abundances for each sample.

2.4. Species classification and functional annotation based on metagenomic data

Based on the metagenomic data, we integrated the annotation data acquired by Kraken2 v. 2.1.1 (Wood et al., 2019) and Bracken v. 2.6.0 (Lu et al., 2017), viewing the results of the species annotation by Krona v. 2.7.1 (Ondov et al., 2011).

Using Diamond v. 2.0.5 (Buchfink et al., 2015) as an aligner, alignments (e-value≤1E-5) were performed against the Non-Redundant Protein Sequence Database (NR), Carbohydrate-Active Enzyme Database (CAZy) (Lombard et al., 2014), Pathogen-Host Interactions Database (PHI) (Winnenburg et al., 2006), Virulence Factors Database (VFDB) (Chen et al., 2016), and Biocide and Metal Resistance Genes Database (BacMet) (Pal et al., 2014). The resistomes were predicted using Rgi v. 5.1.1, which is based on the Comprehensive Antibiotic Resistance Database (CARD). The Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa and Goto, 2000) annotation was performed using the KAAS online analysis website (Moriya et al., 2007). Clusters of Orthologous Groups of Proteins (COG) (Tatusov et al., 1997) and Gene Ontology (GO) (Ashburner et al., 2000; Carbon et al., 2017, 2019) functional annotation were performed using emapper v. 2.0.1 (Huerta-Cepas et al., 2017) based on the eggNOG orthology data (Huerta-Cepas et al., 2019).

2.5. Metagenomic binning

Three modules (MetaBAT2 (Kang et al., 2019), MaxBin (Wu et al., 2016), and CONCOCT (Alneberg et al., 2014)) from MetaWRAP (Uritskiy et al., 2018) were used to perform the binning operation for all the contigs, with the minimum contig length set to 1,500, yielding three different binning results. To calculate the completeness and contamination, CheckM v 1.0.12 was used (Creevey et al., 2011; Parks et al., 2015; Momper et al., 2017). Then, the bin_refinement module in MetaWRAP was used to obtain the bins with completeness of more than 70% and with contamination of less than 10% (parameter: c 70 -x 10), and these bins were called MAGs.

By using the MetaWRAP pipeline (Uritskiy et al., 2018), the clean reads from each sample were mapped to the assembled MAGs by using the quant_bins module; the relative abundance of all the MAGs was estimated; a preliminary gene annotation for each MAG was obtained by using the annotate_bins module; and the distribution of the contigs in each bin was visualized by the blobology module.

2.6. Taxonomic classification and phylogenetic analysis

FastANI v. 1.32 (Jain et al., 2018) was used to cluster the MAGs, and the MAGs with an average nucleotide identity (ANI) value less than 95% were considered to be different species of bacteria. GTDB-tk v. 1.3.0 (Chaumeil et al., 2020) "classify" workflow was used to conduct species classification and to calculate the ANI values of the MAGs based on the Genome Taxonomy Database (GTDB) (Parks et al., 2018, 2020). For phylogenetic analysis, we first searched the GTDB database to find the corresponding reference genomes which are closest to each MAG (Table S1), and then used GTDB-tk (Chaumeil et al., 2020) "align" method to perform multiple sequence alignments for the marker genes in the MAGs and the marker genes in the corresponding reference genomes. Importing the above alignment results, the "infer" method (FastTree v. 2.1.10) (Price et al., 2010) was used to construct a phylogenetic tree under the Whelan-And-Goldman (WAG) model, with the parameters set to "default". Finally, the phylogenetic tree was refined using ITOL v. 5.7 (Letunic and Bork, 2019).

2.7. Functional analysis of MAGs

The carbohydrate degradation enzymes were identified using dbCAN (Yin et al., 2012). We examined the KEGG (Kanehisa and Goto, 2000) biochemical maps for the six known carbon fixation pathways (reductive acetyl-CoA, 3-hydroxypropionate/4-hydroxybutylate cycle, 3-hydroxy-propionate bi-cycle, dicarboxylate/4-hydroxybutyrate cycle, reductive tricarboxylic acid cycle, and Calvin cycle (Momper et al., 2017; Lannes et al., 2019)) and identified the genes encoding the enzymes involved in methanogenesis and methane oxidation. The genes with functions in nitrogen and sulfur metabolism were investigated (Table S2). The completeness of annotated KEGG pathways was evaluated at three levels based on the percentage of annotated genes in the whole pathway gene set: "Full evidence" (All the genes involved in this pathway were present), "Strong evidence" (50%–99% genes present), and "Weak evidence" (20%–49% genes present). Diamond (Buchfink et al., 2015) was used to perform the alignments for all of the MAGs (e-value < 1E-5) against BacMet Database (Pal et al., 2014). The MAGs were submitted to the Anti-SMASH v. 5.2.0 (Edgar, 2004; Blin et al., 2019) website to screen out the secondary metabolite biosynthetic gene clusters (BGCs), with the model set to 'strict'. Rgi (Alcock et al., 2020) was used to predict the resistomes for the MAGs.

3. Results

3.1. Quality controlling and metagenome assembling

Shotgun sequencing yielded 29, 623, 109,936 bp and 24,284,563,524 bp of unassembled raw bases from the soil and water samples, respectively (Table 1). In total, 194,202,138 (98.99%) and 159,527,496 (99.19%) reads passed the quality control in the soil and water sample datasets, respectively (Table 1).

After assembling, a total of 160,212 (soil) and 135,994 (water) contigs were generated. The longest contigs in the soil and water

Table 1
Summary of the metagenomic sequencing.

	Soil	Water
Raw reads	196,179,536	160,824,924
Raw bases (bp)	29,623,109,936	24,284,563,524
Clean reads (%)	194,202,138 (98.99%)	159,527,496 (99.19%)
Clean bases (%)	29,290,018,832 (98.88%)	24,062,066,782 (99.08%)
Contigs Num	160,212	135,994
Total length (bp)	356,933,003	397,377,554
Max length (bp)	833,193	286,311
N50 (bp)	2289	3841
GC content (%)	60.18	51.33

samples were 833,193 bp and 286,311 bp, respectively, with GC content of 60.18% and 51.33% and N50 lengths of 2289 bp and 3841 bp (Table 1).

After Prokka (Seemann, 2014) annotation, 351,242 genes from the soil sample and 390,943 from the water sample were discovered (Table 2).

NR: Non-Redundant Protein Sequence Database; KEGG: Kyoto Encyclopedia of Genes and Genomes; COG: Clusters of Orthologous Groups of Proteins; GO: Gene Ontology; CAZy: Carbohydrate-Active Enzyme Database; VFDB: Virulence Factors Database; BacMet: Biocide and Metal Resistance Genes Database; PHI: Pathogen-Host Interactions Database; CARD: Comprehensive Antibiotic Resistance Database.

3.2. Microbial community representation

To reveal the ecological diversity at high elevation locations, we characterized the microbial community structure in Bamucuo Lake.

For the soil sample, the results showed that bacteria accounted for the majority (99.5%). Only a small portion was archaea (0.3%), and the rest was viruses (0.08%) (Fig. 2). The three most abundant phyla in bacteria were Proteobacteria (54.0%), Actinobacteria (18.6%) and Firmicutes (9.8%), and another 29 phyla were identified (Table S3A). For the most abundant phylum, Proteobacteria, 112 families (Table S3B) and 408 genera (Table S3C) were identified, while for the Actinobacteria, 46 families (Table S3D) and 141 genera (Table S3E) were identified.

A total of 2400 bacterial species were identified in the whole dataset, including *Bacillus cereus* (5.1%), *Hymenobacter sedentarius* (4.5%), *Gemmatirosa kalamazoonesis* (3.0%).

In the archaea (Table S3F), Euryarchaeota (81.9%) accounted for the largest proportion, followed by Thaumarchaeota (14.7%) and Crenarchaeota (1.7%). Table S3 G, H, I, and J list the detailed information for the archaea at the class, order, family, and genus level, respectively.

In the soil sample, eighteen viral species were identified (Table S3K); Surprisingly, *Pandoravirus salinus* (15.3%), a giant virus, was the most abundant, followed by *Bacillus virus 250* (11.5%), *Mycobacterium phage Jobu08* (7.7%), etc.

For the water sample, the results showed that bacteria accounted for 99.1%, archaea accounted for 0.4%, and viruses accounted for 0.5% (Fig. 3). At the phylum level, Proteobacteria (54.4%), Bacteroidetes (17.9%) and Actinobacteria (15.1%) were the most abundant phyla, and another 30 phyla were identified (Table S3a). In the Proteobacteria, 124 families (Table S3b) and 464 genera (Table S3c) were identified, while in the Actinobacteria, 46 families (Table S3d) and 150 genera (Table S3e) were identified.

A total of 2714 bacterial species were identified, in which *Yoonia vestfoldensis* (12.2%), *Belliella baltica* (5.3%) and *Limnohabitans* sp. *63ED37-2* (3.6%) were dominant.

Among the archaea, Euryarchaeota (98.8%) accounted for the largest proportion (Table S3f).

In addition, a total of 52 species of viruses were identified

Table 2
Summary of functional gene annotations against different databases in the two samples.

Database	Soil	Water
Predicted genes	351,242	390,943
NR	232,679	294,919
KEGG	104,704	117,467
COG	211,757	222,635
GO	45,638	59,921
CAZy	52,427	51,940
VFDB	45,976	45,547
BacMet	21,273	18,305
PHI	59,516	60,782
CARD	11	12

(Table S3k), with *Chrysochromulina ericina virus* being the most abundant (12.7%), followed by *Phaeocystis globosa virus* (8.8%) and *Synechococcus phage S-PM2* (5.9%).

3.3. Reconstruction of the MAGs

Metagenomic binning yielded 75 reconstructed MAGs with completeness >70% and contamination <10%, including 18 from the soil sample and 57 from the water sample (Fig. 4). A total of 20 MAGs passed the high-quality threshold, with completeness >95% and contamination <5%, accounting for 26.67% of all the MAGs. The completeness, contamination, GC content, N50, genomic size, and GTDB (Parks et al., 2018, 2020) species categorizations are all listed in Table S4. The original reconstructed MAGs have been reassembled by using the reassemble_bins module in MetaWRAP (Uritskiy et al., 2018) and their genomic integrities have been significantly improved (Table S4). The MAGs from the soil sample had a median genome size of 3.58 Mb and a median GC content of 62.4%, which was similar to the MAGs from the water sample (median: 3.72 Mb; 60.9%).

3.4. Taxonomic identification

To identify the representative microorganisms in Bamucuo Lake, we conducted a comprehensive analysis of each reconstructed MAG. We found that the one-to-one ANI values of all the MAGs were lower than 95%, suggesting that the MAGs are different species, and then, we performed taxonomic annotation and phylogenetic analysis for the 75 MAGs by using GTDB-tk (Chaumeil et al., 2020), with the results showed that the 75 MAGs could be taxonomically assigned to 10 phyla (Fig. 5), including Proteobacteria (16 MAGs, 5 from the soil sample and 11 from the water sample), Bacteroidota (22 MAGs, 2 from the soil sample and the remaining 20 from the water sample), and Actinobacteriota (17 MAGs, 3 from the soil sample and 14 from the water sample). Furthermore, by searching the GTDB database, we found that 74 MAGs were not assigned species names, and only one MAG from the soil sample was assigned a species name (s05-02_bin.5) with an ANI of 98.06% and a 94% alignment fraction (AF), classified as Lactococcus piscium_C, and highly related to known Lactococcus_A piscium_C (GCF_000981525.1). The 74 unspecified MAGs from the soil and water samples were the draft genomes of potential new bacterial species, which deserve further isolation, culture and identification.

3.5. Functional analysis based on the MAGs

To further understand the functional potential of the representative microorganisms in Bamucuo Lake, we analyzed the carbon metabolism, methane metabolism, nitrogen metabolism, and sulfur metabolism for the 75 MAGs.

3.5.1. Carbohydrate degradation
In the 75 MAGs, 286 different types of hydrolases were retrieved, including 165 Glycoside Hydrolases (GHs), 40 Glycosyltransferases (GTs), 32 Polysaccharide Lyases (PLs), 23 Carbohydrate-Binding Modules (CBMs), 16 Carbohydrate Esterases (CEs), and 10 Auxiliary Activities (AAs) (Table S5). In the GH family, GH23 (57 of the 75 MAGs), GH3 (57 of the 75 MAGs), and GH109 (52 of the 75 MAGs) occurred most frequently. In ascending order of frequency, the GT4 family was found in all of the MAGs, followed by GT2, GT51, GT28, GT30, and GT19. In the CBM family, CBM48 was the most common, followed by CBM9, CBM67 and CBM6. PL10, P11, and P12 were the three most commonly occurring enzymes in the MAGs among the PLs. Sixteen types of CEs had been annotated from the MAGs, including CE1, CE11, CE14 and CE4. As for the AA family, 10 members were found, such as AA3, AA4, AA7 and AA1.

Many other essential enzymes, such as starch-degrading enzymes (alpha-amylase, beta-amylase, glucoamylase, alpha-glucosidase,
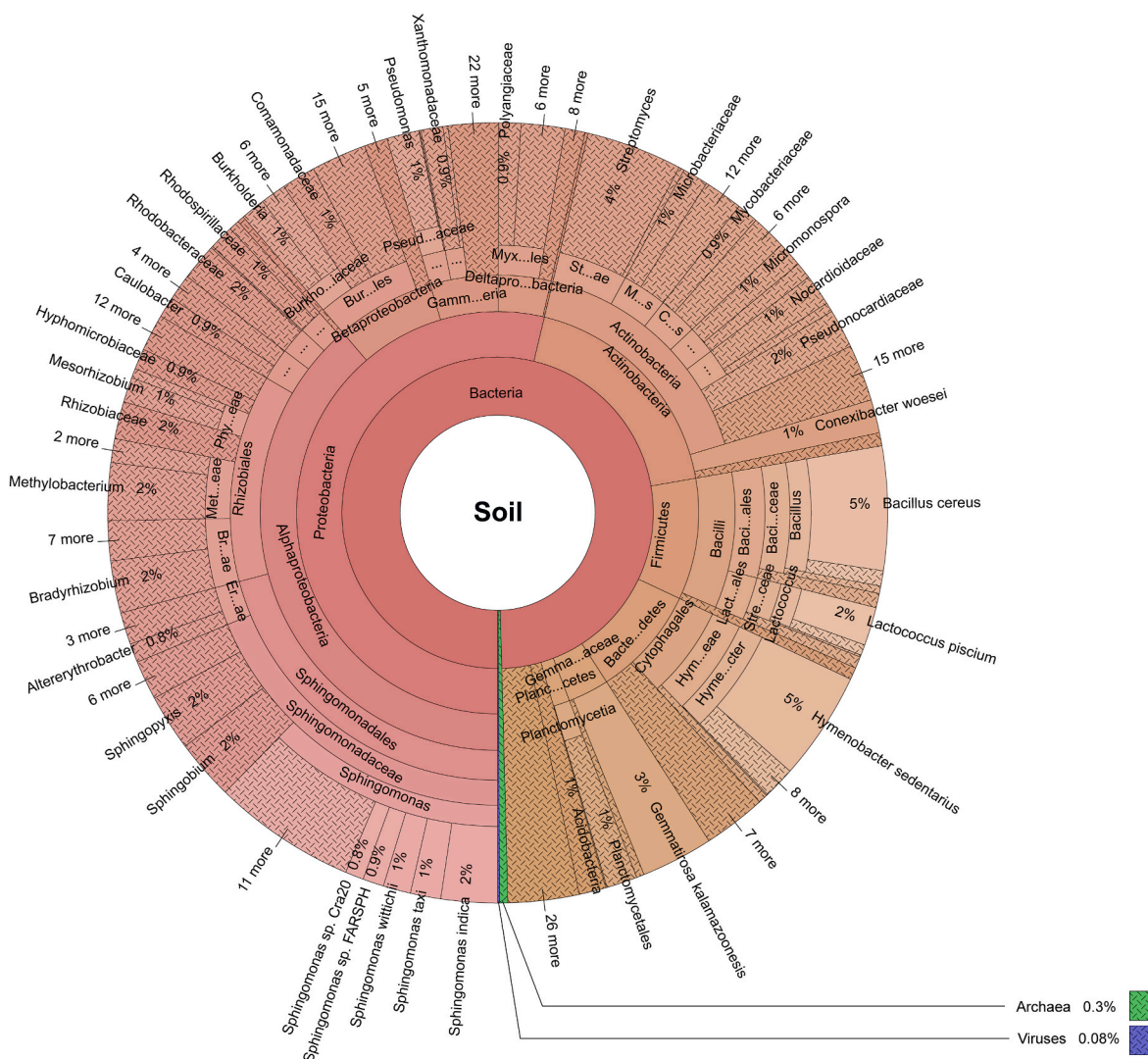
Fig. 2. The taxonomic composition of soil microbial communities. The Figure shows the classification levels of microbial community. The outermost to inner circles represent species, genus, family, orders, and phylum levels, respectively.

isoamylase and pullulanase), cellulose-degrading enzymes (endo-1,4-glucanase, exo-β-1,4-glucanases and β-1, 4-glucosidases), and other carbohydrate degrading enzymes, have also been annotated.

3.5.2. Carbon fixation

Genus *QHXQ01* (s05-02_bin.2) and genus *ZDH117* (s05-02_bin.18) were found to possess genes that participated in nearly all steps of the 3-hydroxypropionate bi-cycle pathway (Table S6).

On the other hand, we also investigated the four other important carbon fixation related pathways. Firstly, we found that although one or more genes encoding the enzymes involved in the reductive tricarboxylic acid cycle pathway were found in all of the MAGs, we cannot confirm that carbon fixation is through this pathway in Bamucuo Lake. Secondly, we investigated the Calvin cycle pathway and discovered that the genus *Cyanobium_A* (s05-03_bin.15) contained nearly complete gene sets for this pathway, the genes, encoding the key enzyme ribulose-1,5-bisphosphate carboxylase-oxygenase (RuBisCO), were only identified in the two MAGs (family AC-14 s05-02_bin.9 and genus *UBA11400* s05-03_bin.32), and another gene encoding the key enzyme phosphor-ibulokinase (PRK) was identified in the two MAGs (genus *Cyanobium_A* s05-03_bin.15 and genus *UBA6144* s05-03_bin.44). Thirdly, with the exception of two MAGs (family AC-14 s05-02 bin.9 and genus Cyanobium A s05-03 bin.15), 73 MAGs contained the genes encoding

enzymes involved in the reductive acetyl-CoA pathway. Finally, with the exception of the genus UBA8290 s05-03 bin.12, we found that the 74 MAGs contained the genes related to the dicarboxylate/4-hydroxybutyrate cycle pathway.

In conclusion, we can confirm that the microorganisms in Bamucuo Lake have the potential to fix carbon primarily through the 3-hydroxy-propionic bi-cycle pathway, and we also found strong evidence of the existence of the Calvin cycle in the MAG (s05-03_bin.15 genus *Cyanobium_A*).

3.5.3. Methane metabolism

Microbial methanogenesis and methane oxidation play essential roles in the earth's biogeochemical cycles. We found that, with the exception of family UBA7662 s05-03 bin.54, the other 74 MAGs contained genes related to the methanogenesis pathway (Table S6); however, the gene methyl-coenzyme M reductase (MCR), encoding the key enzyme in the methanogenesis pathway, was not detected in these 74 MAGs. Furthermore, no gene encoding the enzyme involved in the methane oxidation pathway was identified in all of the MAGs in the two samples (Fig. 6).

3.5.4. Nitrogen metabolism

The genes involved in the assimilatory nitrate reduction pathway
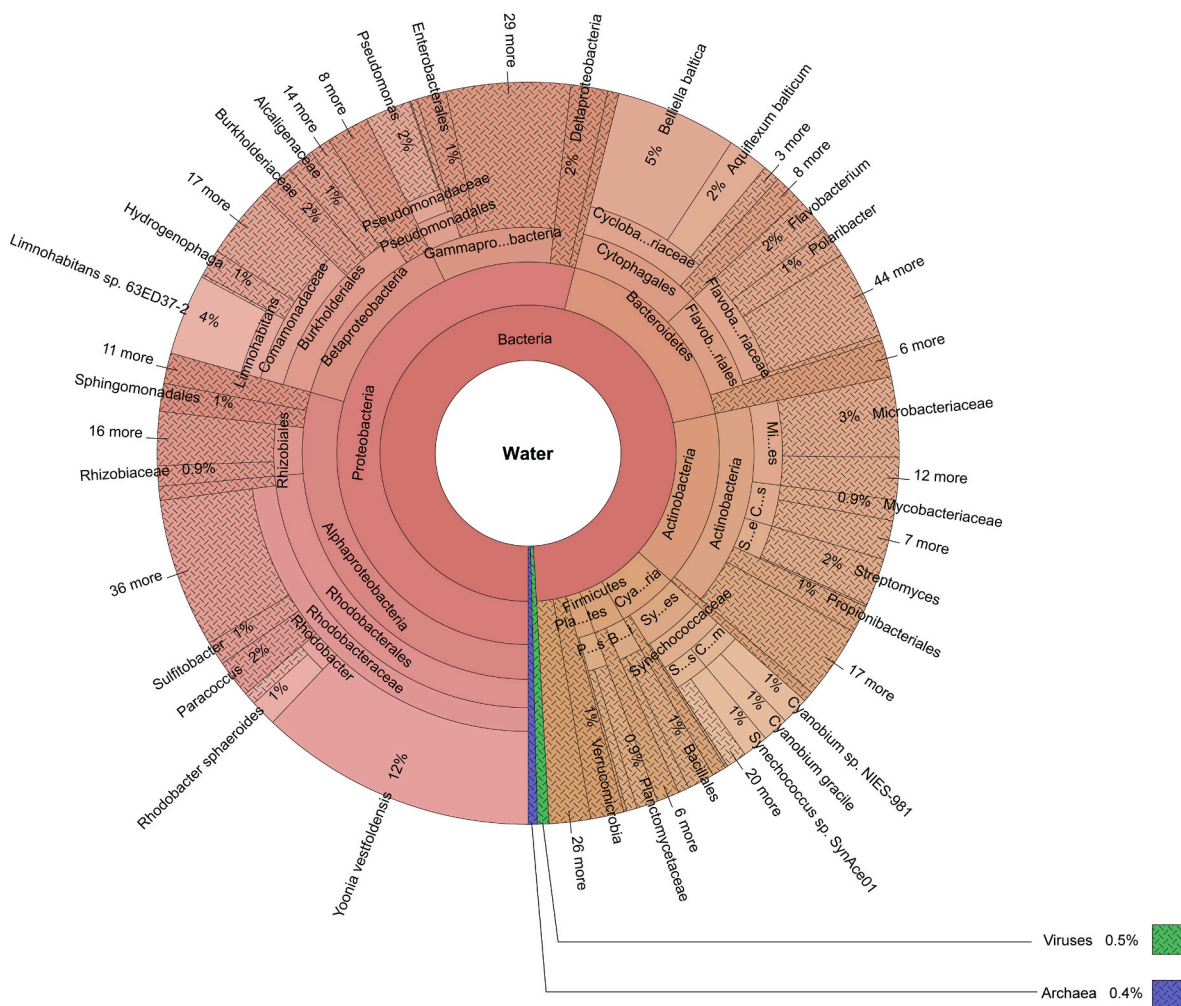
Fig. 3. Classification composition of the water microbial community. The Figure shows the classification level of the microbial community. The outermost to the inner circles represent species, genus, family, orders, and phylum levels, respectively.

were identified in eight MAGs (Fig. 6). Weak evidence of the nitrogen fixation pathway was found in three MAGs (genus *Hyphomonas* s05-03_bin.4, genus *Oceanicaulis* s05-03_bin.31 and genus *EhC02* s05-03_bin.35). The nitrogenase complex is composed of two parts: ferritin encoded by *nifH* (also known as nitrogenase reductase) and molybdenum ferritin encoded by *nifD* and *nifK* genes (also known as nitrogenase). The *nifH* gene was identified in three MAGs (genus *Hyphomonas* s05-03_bin.4, genus *Oceanicaulis* s05-03_bin.31 and genus *EhC02* s05-03_bin.35), while the gene *nifD* and *nifK* were not found.

In terms of denitrification and nitrification pathways, we found that four MAGs (Fig. 6) contained the genes linked to the denitrification pathway, while no MAG contained the genes linked to the nitrification pathway.

3.5.5. Sulfur metabolism

For sulfur metabolism, 52 MAGs contained the genes related to the assimilatory sulfate reduction pathway; eight MAGs contained the genes related to the dissimilatory sulfate reduction pathway; and 12 MAGs contained the genes related to the thiosulfate oxidation by sulfur-oxidation (SOX) pathway (Fig. 6). Among these 12 MAGs, the marker gene *soxB* related to SOX pathway was found in the ten MAGs (genus *QHVT01* s05-02_bin.1, family UBA7662 s05-03_bin.1, genus *UBA996* s05-03_bin.2, genus *EhC02* s05-03_bin.35, genus *Rhodonellum* s05-03_bin.41, genus *UBA6144* s05-03_bin.44, genus *Belliella*

s05-03_bin.46, family UBA955 s05-03_bin.51, family UBA7662 s05-03_bin.54, and family GCA-2696645 s05-03_bin.55).

3.6. Heavy metal resistance gene identification

To reveal the heavy metal resistance of the microorganisms in Bamucuo Lake, we investigated the common heavy metal resistance genes. Annotation data for all the genes with resistance to 19 types of common heavy metals were retrieved. (Table S7). The results showed that more than 1064 heavy metal resistance genes were identified in all of the MAGs (Fig. 7), among which the heavy metal resistance genes conferring resistance to multi-metals were abundant. Furthermore, the genes conferring resistance to antimony, arsenic, molybdenum and tungsten were identified in all of the MAGs.

3.7. Biosynthetic gene cluster identification

We identified the gene clusters and the annotated secondary metabolites in all of the MAGs. A total of 270 potential BGC clusters of secondary metabolites were collected, with 25 different types of BGCs found in the 75 MAGs (Table S8). Of the 40 Non-ribosomal peptide synthetases (NRPSs), 25 existed in the Acidobacteriota MAGs. Terpenes were found in the 61 MAGs, type III polyketide synthases (T3PKSs) in 17 MAGs, NRPSs in the Acidobacteriota MAGs and Bacteroidota MAGs, polyketide synthases (PKSs) in the Bacteroidota MAGs and Actinobacteria MAGs, and 11 bacteriocins in the Cyanobacteria MAG.
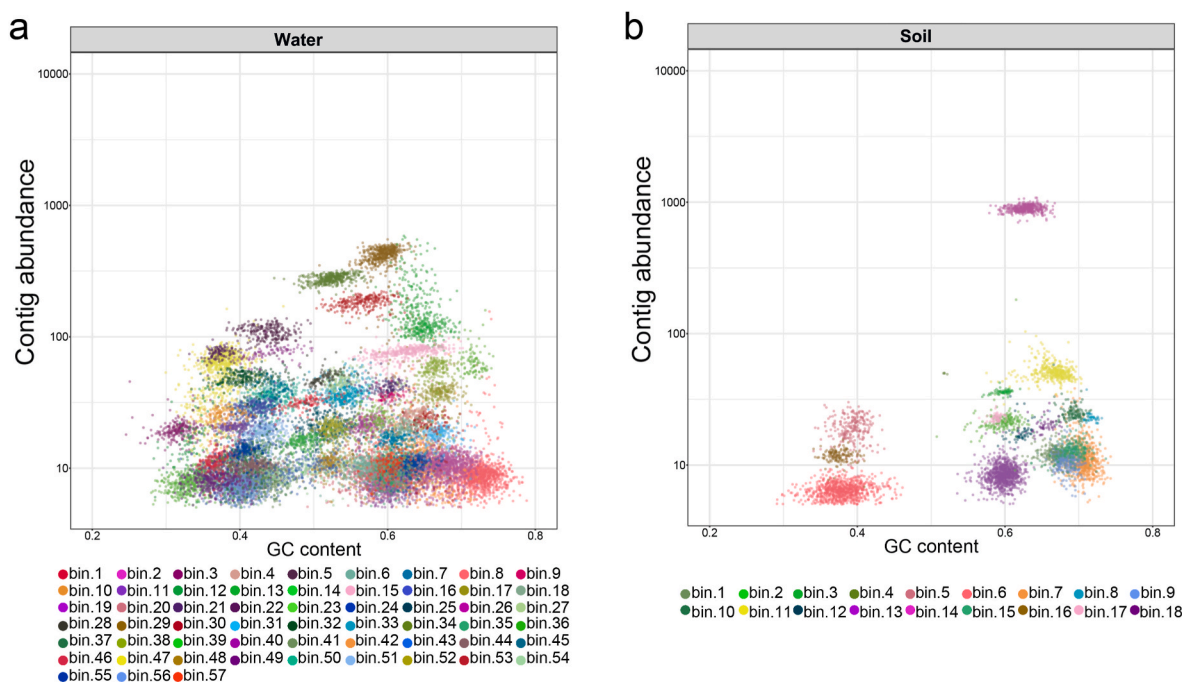
Fig. 4. Refined bin evaluation. (a) (Water) and (b) (Soil): The distribution of the contigs in each bin. The abscissa represents the GC content of the contigs, and the ordinate represents the contig abundance, with one dot representing one contig. The contigs in the same color belong to the same bin.

3.8. Antibiotics resistance gene (ARG) identification

To investigate the distribution of ARGs in high elevation locations, we performed ARG analysis based on the MAGs in the two samples. The findings revealed that ARGs were discovered in 15 of the 75 MAGs, including adeF and rpsL (mycobacterium tuberculosis rpsL mutations conferring resistance to streptomycin). Among these 15 MAGs, adeF was found solely in the soil sample MAGs, while ampS was found in genus Microvirga s05-02_bin.10. In the water sample, except for adeF and ampS associated with antibiotic efflux, mycobacterium tuberculosis rpsL mutations conferring resistance to streptomycin accounted for the majority (5 out of 8). The detailed ARGs annotation information with perfect or strict hits predicted by Rgi is shown in Fig. 8.

4. Discussion

4.1. Microbial community structure

The QTP is the world's highest plateau, and microbial communities in such extreme environments play an important role in microbial ecology. We investigated the relationship between the distribution and function of the microbial community in Bamucuo Lake on the Qinghai-Tibet Plateau at an elevation of 4555 m, revealing the composition of the microbial community. We reconstructed 75 MAGs, including 20 high-quality MAGs (completeness >95% and contamination <5%). The median genome size and GC content of the reconstructed MAGs were 2.60 Mb and 56.5%, respectively. However, there were still numerous reads that were not assigned to any bins due to the presence of the sequence data from fungi, eukaryotes, and viruses, which adversely affected our reconstruction of the bacterial MAGs (Nayfach et al., 2021).

As for Archaea, Euryarchaeota is commonly considered as the most abundant Archaea in many ecosystems. This conclusion was also confirmed in our study. However, Thaumarchaeota and Crenarchaeota were only identified in the soil sample in Bamucuo Lake, demonstrating that the soil and water microbial communities differed. Pandora virus is the largest known virus reported by Philip et al., isolated from sediments in the central Chile estuary and freshwater ponds near Melbourne

(Australia) (Philippe et al., 2013); however, in our research, a type of Pandoravirus, Pandoravirus salinus was detected in the nearshore soil of Bamucuo Lake. The discovery of Pandoravirus salinus at such a high altitude is the first in the world, and why it inhabits here remains a mystery. Isolating, culturing and obtaining the complete genome of Pandoravirus salinus inhabiting the Qinghai-Tibet Plateau remains a significant challenge.

4.2. Carbon metabolism

Carbohydrate-Active Enzymes (CAZymes) are enzymes that degrade, alter, and generate glycosidic bonds. The Bacteroidota MAGs contained the most abundant CAZymes, followed by the Proteobacteria and Actinobacteriota MAGs (Table S5), showing that these bacterial phyla have high hydrolytic potential. The number of CAZymes types detected in Bacteroides MAGs (genus Segetibacter s05-02_bin.16, family Sphingo-bacteriaceae s05-02_bin.6 and genus UBA11400 s05-03_bin.32, etc.) accounted for up to 33.63% of the total number of CAZymes types detected in our investigation, which was consistent with previous reports that the Bacteroidota members may degrade complex carbohydrate-based biomass (McKee et al., 2021). An intriguing finding was the presence of GH23, a family of enzymes widely used in bioengineering and medicine (Nahar et al., 2018; Hukic et al., 2018), in the majority of the MAGs (57 of the 75, accounting for 76%).

As auxiliary enzymes, the GH2 and GH3 were also identified in the majority of the MAGs. The most common GH families involved in cellulose degradation were identified in 14 MAGs (genus *QHVT01* s05-02_bin.1, genus *Microvirga* s05-02_bin.10, genus *Sphingomonas_A* s05-02_bin.12, genus *Segetibacter* s05-02_bin.16, family T3Sed10-336 s05-03_bin.20, and genus *Rhodonellum* s05-03_bin.41, etc.), including GH5, GH9 and GH44 (Table S5).

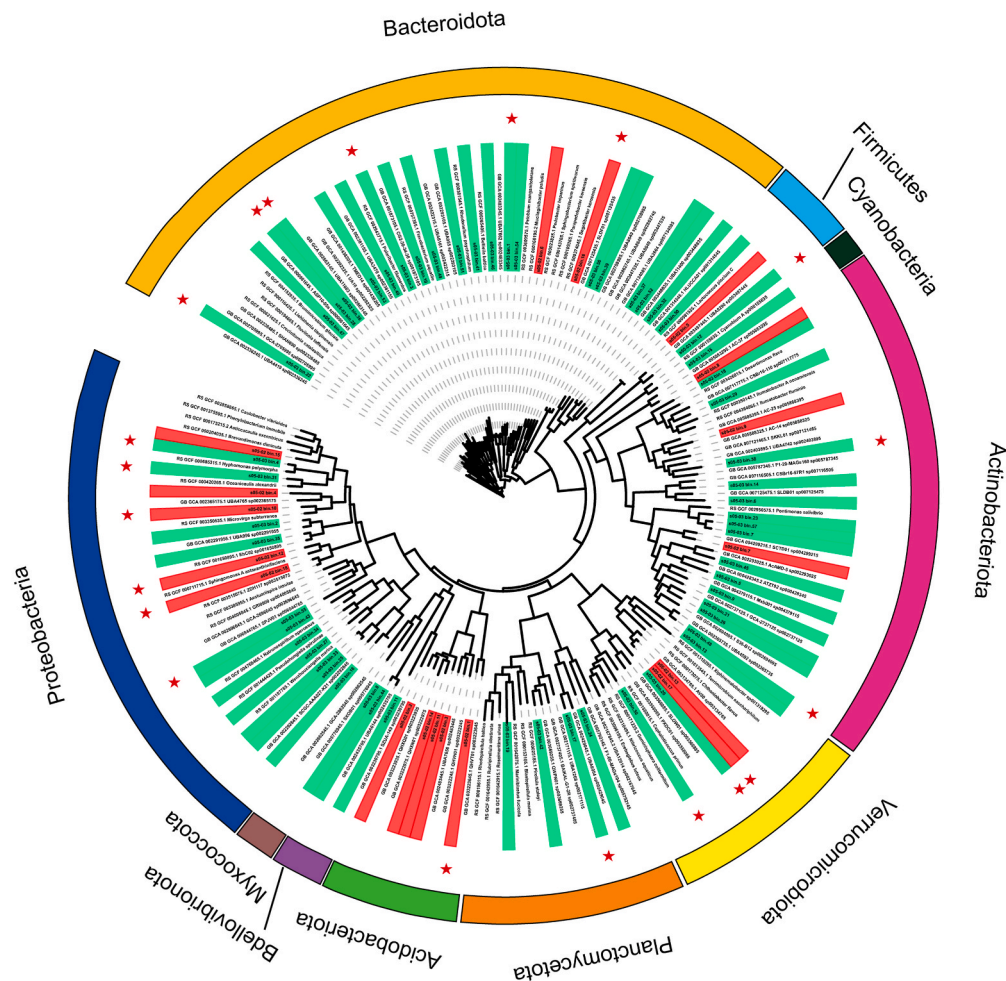The enzymes involved in xylan degradation were detected in 56 MAGs (genus *QHVT01* s05-02_bin.1, genus *QHXQ01* s05-02_bin.2, family UBA7656 s05-02_bin.3, genus *Microvirga* s05-02_bin.10, genus *QHXN01* s05-02_bin.11, genus *Sphingomonas_A* s05-02_bin.12, family UBA7662 s05-03_bin.1, genus *UBA996* s05-03_bin.2, genus *Polaribacter* s05-03_bin.3, etc.), including GH8, GH10, GH43, GH51, GH67, GH115,

Fig. 5. Phylogenetic tree of the 75 MAGs. The maximum likelihood evolutionary tree is reconstructed based on all the 75 MAGs and their nearest neighbor reference genomes. The MAGs of s05-02 (soil) were represented by the red background and the MAGs of s05-03 (water) by the green background. High quality MAGs, namely MAGs with completeness >95% and contamination <5%, are marked with a star sign, and the phyla to which they belong were distinguished in outer rings in different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

GH127, CE1, and CE2.

The pectin-degrading families were identified in a total of 17 MAGs (family Sphingobacteriaceae s05-02_bin.6, genus *AC-37* s05-02_bin.8, genus *Sphingomonas_A* s05-02_bin.12, genus *UBA996* s05-03_bin.2, family T3Sed10-336 s05-03_bin.20, etc.), including GH28, PL1, PL9, PL11, CE8, and CE12, suggesting that the microorganisms in Bamucuo Lake have the ability of pectin degradation.

In addition, the starch-degrading enzymes were identified in 49 MAGs (genus *QHVT01* s05-02_bin.1, genus *QHXQ01* s05-02_bin.2, family UBA7656 s05-02_bin.3, family UBA7662 s05-03_bin.1, genus *Polaribacter* s05-03_bin.3, genus *Pontimonas* s05-03_bin.7, etc.), including GH13, GH77 and GH97. These observations suggest that the majority of microorganisms in Bamucuo Lake are involved in a variety of complex carbohydrate degradation reactions for processing organic matter from decaying plants and manure from various animals.

Carbon fixation is the process by which inorganic carbon from the atmosphere is assimilated and converted into organic compounds for further energy storage and biomolecular synthesis (Berg, 2011). Atmospheric carbon fixed by microorganisms and plants can be converted to dioxide via the chemosynthesis or photosynthesis processes (La Cono et al., 2018). Currently, the six known carbon fixation pathways are reductive acetyl-CoA, 3-hydroxypropionate/4-hydroxybutylate cycle, 3-hydroxypropionate bi-cycle, dicarboxylate/4-hydroxybutyrate cycle,

reductive tricarboxylic acid cycle, and Calvin cycle (Momper et al., 2017; Lannes et al., 2019).

The 3-hydroxypropionate bi-cycle (9.52%–52.38%) and Calvin cycle (11.76%–52.94%) were most common in the 75 MAGs (Fig. 6), despite the fact that the 3-hydroxypropionate bi-cycle pathway is generally rarely found in most environments because of its high energy cost (Berg, 2011). Our findings revealed that the whole gene set related to the 3-hydroxypropionate bi-cycle was found in two MAGs (genus *QHXQ01* s05-02_bin.2 and genus *ZDH117* s05-02_bin.18; Fig. 6) and the abundance of genes identified to encode enzymes in this pathway was also high in the remaining MAGs, suggesting that the 3-hydroxypropionate bi-cycle pathway is the predominant carbon fixation pathway in Bamucuo Lake.

In a review article, Berg concluded that the 3-hydroxypropionate bi-cycle pathway does not contain oxygen-sensitive steps, which is most suitable for mixotrophy and especially advantageous under neutrophilic and alkaliphilic conditions (Berg, 2011). However, Bamuco Lake is a saline lake with typical alkaline conditions, providing the most suitable environmental conditions for the 3-hydroxypropionate bi-cycle pathway. This may be a reasonable explanation that the 3-hydroxypropionate bi-cycle pathway is dominant in Bamuco Lake.

The Calvin cycle pathway is the main CO_2 fixation pathway, which is proved to be widely present in many autotrophic bacteria in most
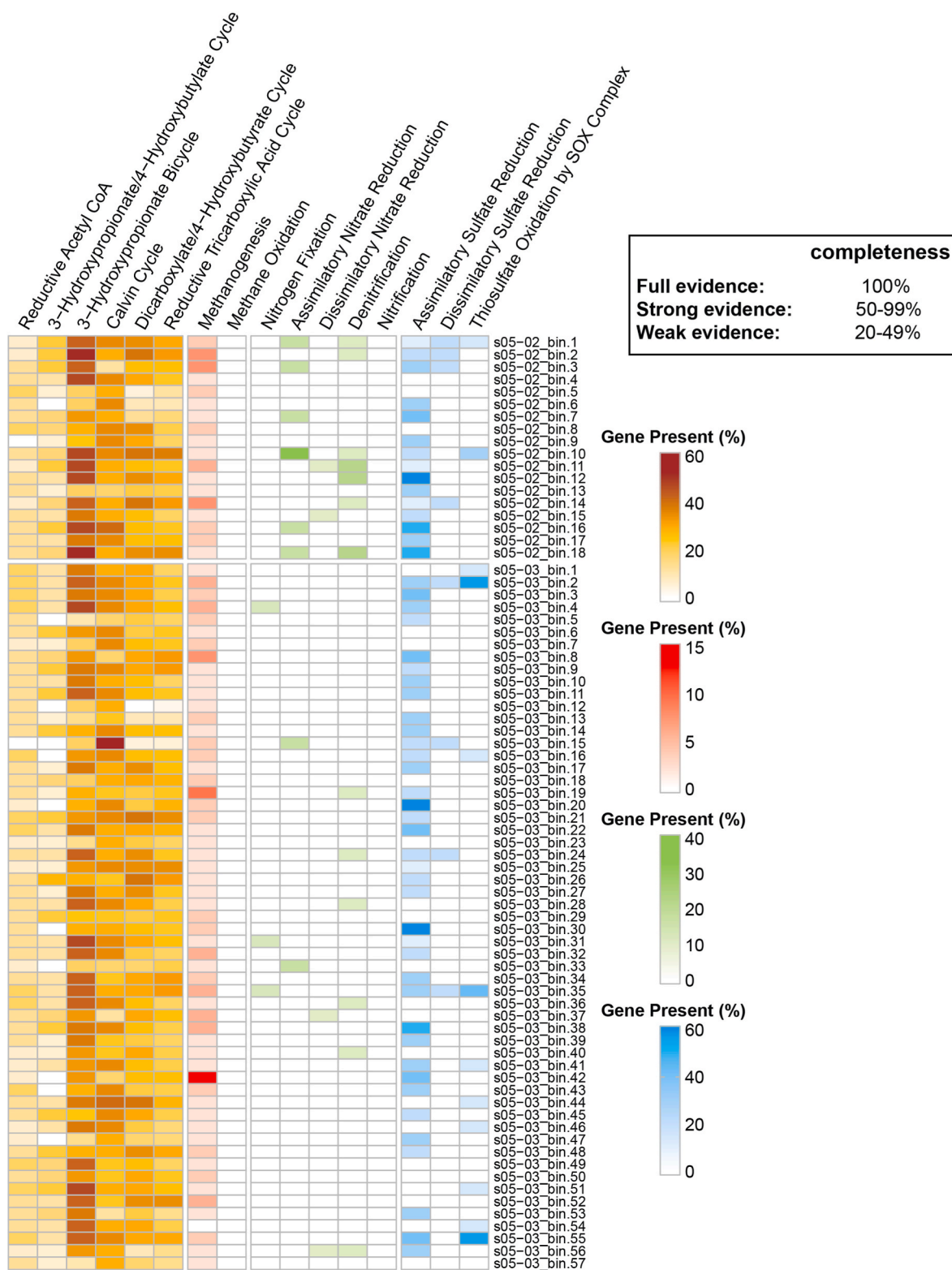
Fig. 6. KEGG pathway annotation of energy metabolism for the two samples. The heatmap shows the percentage of the associated genes. The orange boxes represent carbon fixation; the red boxes represent methane metabolism; the green boxes represent nitrogen metabolism; and the blue boxes represent sulfur metabolism. Here, we investigated the pathways of carbon fixation, methanogenesis, methane oxidation, nitrogen metabolism, and sulfur metabolism.

environments, such as Yap Trench (Zhang et al., 2018). However, even if most genes related to this pathway were detected, only the genus *Cyanobium_A* s05-03_bin.15 contained the nearly whole gene set related to the Calvin cycle in our data. Furthermore, the identified key enzymes revealed that the genes encoding RuBisCO were exclusively identified in

the two MAGs (family AC-14 s05-02_bin.9 and genus *UBA11400* s05-03_bin.32). Thus, the Calvin cycle pathway may not be the dominant carbon fixation pathway in this environment.

Compared with the Calvin cycle pathway, the reductive tricarboxylic acid pathway has lower energy consumption, which was found in
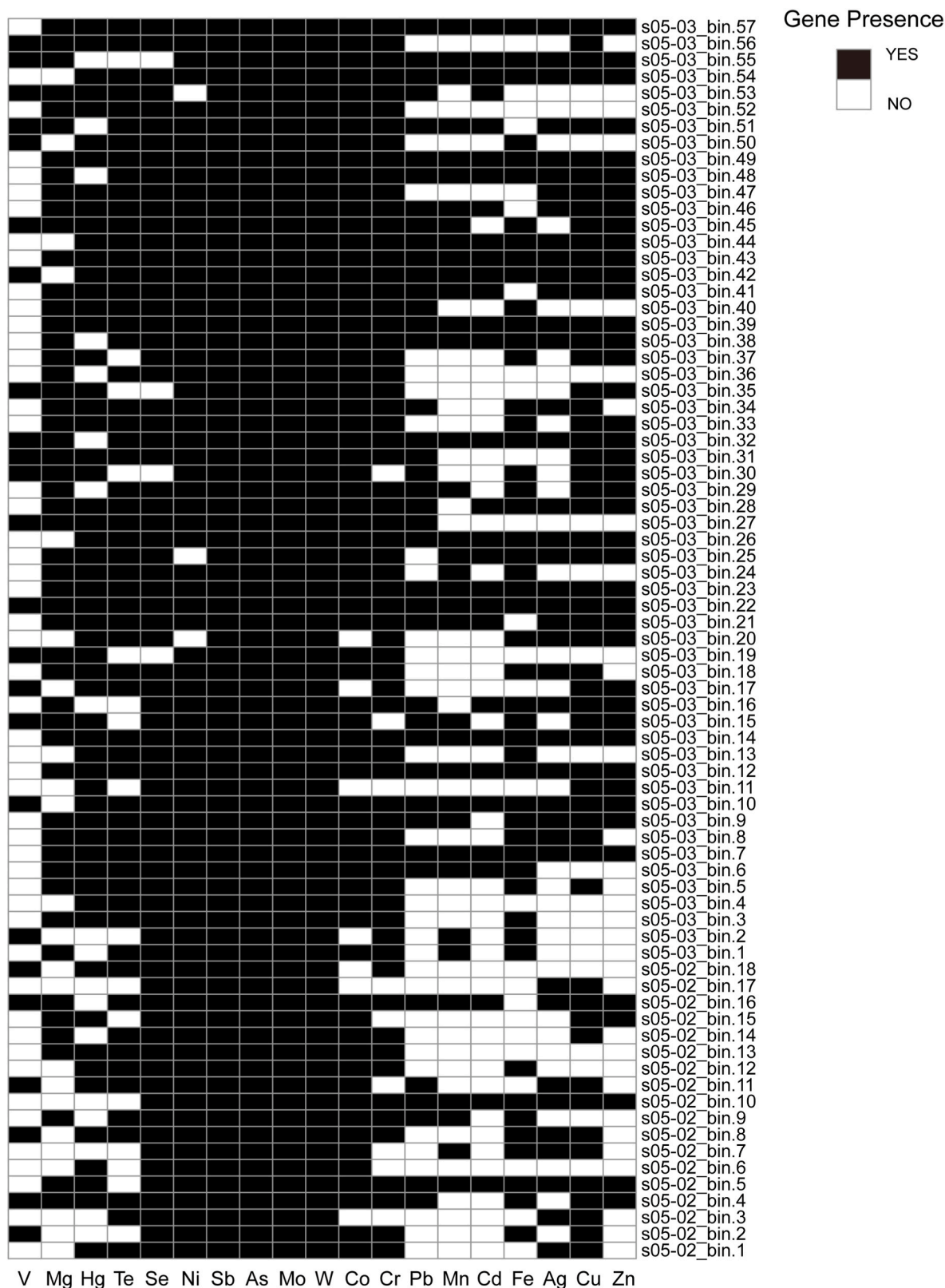
Fig. 7. Heavy metal resistance genes identified in the two samples. The black squares represent the presence of the resistance genes, and the blank squares represent the absence of the resistance genes.

certain autotrophic mesophilic bacteria and archaea (Berg, 2011), and was often used for biosynthetic purposes. Our results showed that the annotated gene set related to the reductive tricarboxylic acid pathway is incomplete (2.33%–37.21%), indicating that the evidence for the

presence of the carbon fixation pathway in Bamucuo Lake is weak.

In addition to the most common three carbon fixation pathways mentioned above, the three less common pathways were the reductive acetyl-CoA (0–18.75%), dicarboxylate/4-hydroxybutyrate (0–39.13%),
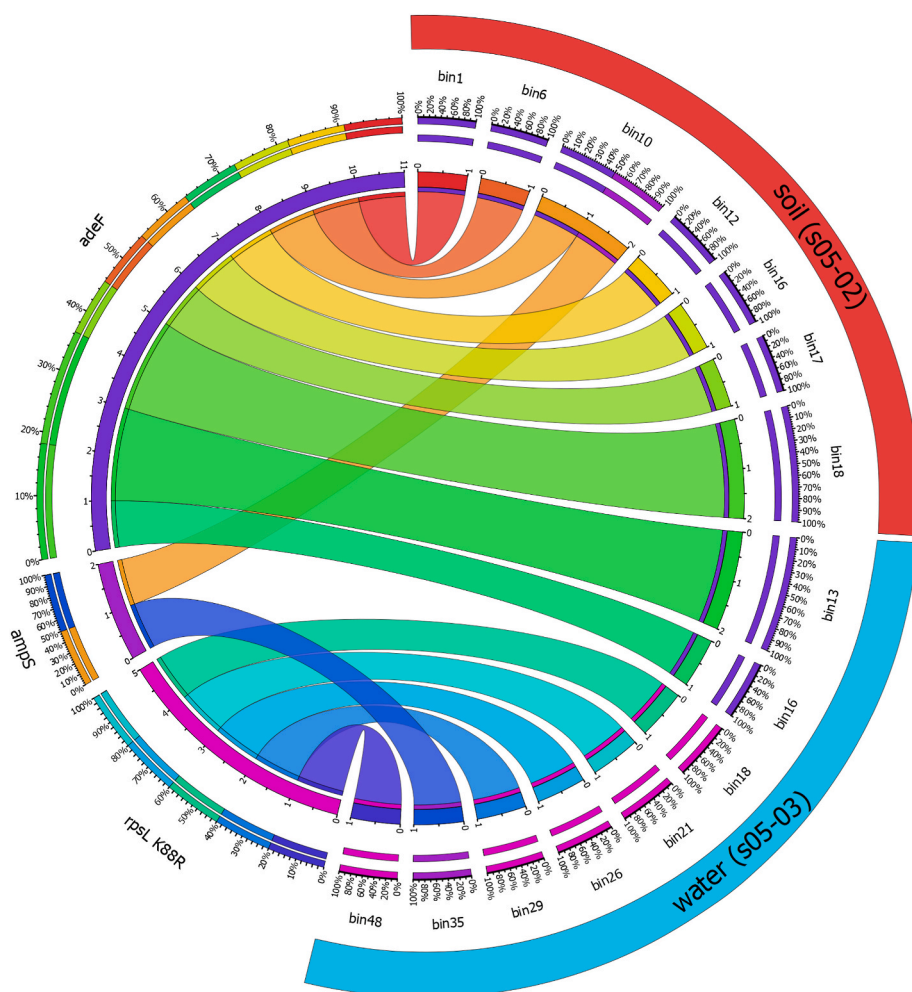
Fig. 8. The circos of the annotation of MAGs by using CARD (Alcock et al., 2020). The identified ARGs are on the left side of the peripheral ring, and the MAGs information is on the right side. The inner ring shows different MAGs and ARGs in different colors, and the scale represents the abundance information. The width of the band is proportional to the abundance of a particular ARG in each MAG. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

and 3-hydroxypropionate/4-hydroxybutylate cycle (0–27.78%) pathways. The dicarboxylate/4-hydroxybutyrate and reductive acetyl-CoA pathways are found exclusively in anaerobic organisms (Loder et al., 2016). Certain enzymes of the dicarboxylate/4-hydroxybutyrate pathway, which is strictly anaerobic, are sensitive to oxygen (Berg, 2011). In our investigation, weak evidence for the dicarboxylate/4-hydroxybutyrate pathway was found.

The reductive acetyl-CoA pathway (or the Wood-Ljungdahl pathway) is the main carbon fixation pathway in anaerobic conditions (Berg, 2011). However, no MAGs contained the genes encoding ACS/-CODH complex (CO-Dehydrogenase/Acetyl-CoA Synthase), a five-subunit enzyme complex which is responsible for the carbonyl branch of the Wood-Ljungdahl pathway. The reductive acetyl-CoA pathway was found to be one of the dominant carbon fixation pathways in deep biosphere (Magnabosco et al., 2016; Momper et al., 2017); however, our data suggested that this pathway does not work for the microorganisms in Bamucuo Lake.

The 3-hydroxypropionate/4-hydroxybutyrate cycle functions at high temperatures and also can function in either aerobic bacteria or anaerobic bacteria. Ruiz-Fernandez et al. concluded that the 3-hydroxypropionate/4-hydroxybutylate cycle is the dominant pathway in anoxic marine zones (Ruiz-Fernandez et al., 2020). However, in our investigation, the evidence for the existence of 3-hydroxypropionate/4-hydroxybutylate cycle was weak.

4.3. Methane metabolism

In methane metabolism, CO_2 is converted to CH_4, which is the second largest greenhouse gas, accounting for around 17% of global warming impact (Cai et al., 2016), with CO_2 accounting for approximately 60% (Nakicenovic et al., 2000). Our findings revealed that there was no clear evidence for the methanogenesis or the methane oxidation pathways (Fig. 6). In the methanogenesis pathway of microbial communities, the methane is generated through the key enzyme, MCR. However, in this investigation, the MCR genes were not identified in neither the MAGs' KEGG (Kanehisa and Goto, 2000) annotation results nor the metagenomic data, which was consistent with previous study reported the absence of MCR on Pangong Lake (Rathour et al., 2020). The presence of methanogens, such as Methanobacterium, Methanosarcina, and Methanocaldococcus in the contigs of the two samples, revealed that methanogens may survive in the hypoxic environment of Bamucuo Lake. However, no clear evidence for methane production was found in this environment.

Methane oxidation is an important process to slow down CH_4 emission (Canelhas et al., 2016), which is performed by CH_4 oxidizing bacteria (methanotrophs). Methanotrophs, such as Methylomonas, Methylomicrobium, and Methylococcus, were discovered in the two samples. However, in the KEGG annotation results, the functional genes for methane oxidation were not found, which might be due to insufficient sequencing depth.

4.4. Nitrogen cycle

Nitrogen and the nitrogen cycle play a crucial role in many microbial biological processes because it catalyzes several reactions. This study has found evidence for four essential nitrogen metabolism pathways,

including denitrification, nitrification, nitrogen fixation, and assimilatory nitrate reduction.

Denitrification is the primary source of nitrogen loss in QTP (Wang et al., 2019). The genes involved in the denitrification pathway were discovered in our investigation. Furthermore, denitrifying bacteria, such as *Pseudomonas* and *Bacillus*, were also identified in both samples.

Nitrification is a key process in biogeochemical nitrogen cycling, which oxidizes ammonia to nitrate via nitrite (Spasov et al., 2020). In our result, nitrosifyer (such as *Nitrosomonas* and *Nitrosococcus*) and nitrifying bacteria (such as *Nitrobacter* and *Nitrpspina*) were found in low abundance, and the evidence of the existence of nitrification pathway was weak, implying that the nitrification process in Bamucuo Lake may be hampered by the extreme environment.

It was previously reported that Proteobacteria plays an important role in the nitrogen cycle (Ren et al., 2018), and our findings also confirmed that the nitrogen fixation pathway has only been found in the Proteobacteria MAGs (genus *Hyphomonas* s05-03_bin.4, genus *Oceanicaulis* s05-03_bin.31 and genus *EhC02* s05-03_bin.35). As above described, we detected *nifH* in three Proteobacterial MAGs; however, the genes *nifD* and *nifK* encoding the essential subunits of nitrogenase were not identified, which at least indicated that the nitrogen fixation ability of the 75 bacteria we obtained was not strong. However, this conclusion cannot be applied to the whole bacterial community of Bamucuo Lake, because these 75 bacteria represented just a small portion of the bacterial community in Bamucuo Lake.

Nitrate assimilation is the main source of nitrogen in soil. In this study, more genes related to the assimilatory nitrate reduction pathway were detected than other pathways in the nitrogen cycle, indicating that this pathway was primary nitrate reduction pathway in Bamucuo Lake, which is consistent with the report obtained from Pangong Lake, a unique brackish ecosystem (Rathour et al., 2020).

4.5. Sulfur cycle

The microbial sulfur cycle has been proposed as the driving force for bacterial survival (Cao et al., 2014). The sulfate-reducing bacteria (e.g., Desulfobacterales, Desulfovibrionales, Desulfotomaculum and Desulfosporosinus) in Bamucuo Lake, which can perform anaerobic respiration utilizing sulfate as the terminal electron acceptor for reducing sulfate to hydrogen sulfide, were found in the two samples; similar results in Pangong Lake were reported by Rathour et al. (2020). The assimilatory sulfate reduction pathway was first discovered in plants; however it has been shown to exist in certain bacteria, such as Allochromatium vinosum (Neumann et al., 2000). Our findings provided compelling evidence for the presence of an assimilatory sulfate reduction pathway in Bamucuo Lake (Fig. 6). However, the key functional gene encoding dissimilatory sulfite reductase for dissimilatory sulfate reduction had not been identified, indicating the lack of microorganisms that perform the functions of dissimilatory sulfate reduction.

Generally, the major source of sulfur is plutonic rocks containing sulfides, which are converted to sulfates during weathering, some of which dissolve, precipitate, and, depending on the conditions, reduce to elemental sulfur. The sulfur-oxidizing bacteria such as *Sulfuritalea* and *Thiobacillus* were present with a low abundance, and few genes related to the thiosulfate oxidation by SOX pathway were identified (Fig. 6). However, because the essential gene soxB, which is involved in thiosulfate oxidation via the SOX pathway, was detected in the majority of the MAGs, we concluded that sulfur oxidation is crucial in the energy metabolism of the microorganisms in Bamucuo Lake.

4.6. Metal resistance

The microbial community is the primary driving factor behind metal resistance gene distribution, and earlier research has shown that these genes are mostly found in heavy metal-contaminated settings (Song et al., 2019). However, in our study, abundant heavy metal resistance genes were identified (Fig. 7). A study supported that the pollution of heavy metals in QTP has been on the rise in recent years (Wu et al., 2018). In addition, rainwater and runoff may cause microorganisms to accumulate in the lake, being more conducive to the enrichment of heavy metal resistance genes in the water and nearshore soil microbial communities. Chen et al. reported that heavy metals can induce DNA damage and oxidative stress (Chen et al., 2018b). Our study showed that the DNA repair genes, such as *recB*, *recC*, *recD*, and *recF*, were found in majority of the MAGs. The explanation for this is because all eukaryotic and prokaryotic genomes require some DNA repair genes to repair DNA damage, and on the other hand, due to the ultraviolet rays in the environment at an altitude of 4500 m in Bamucuo Lake.

4.7. Biosynthetic gene clusters and antibiotics resistance genes

Antibiotics have long received special attention as secondary metabolites of bacteria (Atanasov et al., 2021). In recent years, some pathogens have evolved "drug resistance" to certain antibiotic drugs, and the number of new antibiotics has steadily decreased (Ventola, 2015). As a result, there is an urgent need to find new resources to obtain new antibiotic drugs. To discover novel antibiotics, it is necessary to search in the unexplored microbial ecological environments. Clearly, the QTP is a very valuable and novel microbial ecological environment. Several BGCs were found in many MAGs (Table S8), suggesting that Bamucuo Lake may contain many unique natural product biosynthetic gene clusters.

Many terpenoids have important physiological activities and are important sources for obtaining natural bioactive products and new drugs, which are widely distributed in the genomes of many plants, fungi and bacteria (Yamada et al., 2015). We found 26 kinds of BGCs in 74 of the 75 MAGs (Table S8), and the three MAGs with the most BGCs were s05-02_ Bin.14 (21 types, soil), s05-02_ Bin. 2 (17 types, soil), and s05-03_ Bin.15 (14 types, water). Among all of the MAGs, terpenes appeared most frequently in our investigation (35.56%) (Table S8). Except for terpenes, NRPSs (14.81%) and PKSs (15.93%) were the most abundant types of BGCs in our study (Table S8), and NRPSs and PKSs are two families of modular mega-synthases (Cuadrat et al., 2018), which produce compounds such as antibiotics, antifungals, immunosuppressants, and iron-chelating molecules (Finking and Marahiel, 2004). Many NRPS and PKS gene clusters were found in Actinobacteria, Acidobacteria and Bacteroidetes MAGs, suggesting that Actinobacteria, Acidobacteria and Bacteroidetes are the potential natural product biosynthetic producers in Bamucuo Lake. In conclusion, QTP might be a useful resource for finding novel secondary metabolite clusters.

RpsL is an essential antibiotic resistance gene encoding bacterial 30S ribosomal protein S12, which is the streptomycin action site. When streptomycin binds to ribosomal protein S12, the strain's growth is hindered (Ochi and Hosaka, 2013). As a result, when a nonsense mutation arises in the rpsL gene, bacteria develop streptomycin resistance. The rpsL K88R mutation was only found in Actinobacteriata MAGs in our investigation (family Ilumatobacteraceae s05-03_bin.18, family S36–B12 s05-03_bin.21, family S36–B12 s05-03_bin.26, genus *CSBr16-110* s05-03_bin.29, and genus *UBA4592* s05-03_bin.48) (Fig. 8).

5. Conclusions

The microbial ecology of Bamucuo Lake in QTP was investigated in this study. Our study revealed the structure and metabolic processes of the soil and water microbial communities in Bamucuo Lake. The 3-hydroxypropionate bi-cycle pathway was shown to be the most typical carbon fixation pathway in Bamucuo Lake. Seventy-four novel MAGs in the water sample indicate a new contribution to global microbial diversity. The existence of *Pandoravirus salinus* in the soil provides significant information for further research into this unique microbe and the discovery of additional giant viruses. Because Actinobacteria was one of the dominant bacteria, QTP has the potential to become a

valuable resource pool where many Actinobacteria-related microorganisms can be employed to generate new bioactive compounds and new antibiotics. Many draft genomes and genome annotation information are provided in this paper for further in-depth study of the microbial ecology in QTP.

Credit author statement

Cai Wei: Investigation, Methodology, Software, Writing- Original draft preparation. **Dan Sun**: Resources, Data Curation. **Wenliang Yuan**: Software, Formal analysis. **Lei Li**: Resources, Data Curation. **Chaoxu Dai**: Data Curation, Resources. **Zuozhou Chen**: Visualization, Data Curation. **Xiaomin Zeng**: Formal analysis. **Shihang Wang**: Software, Data Curation. **Yuyang Zhang**: Formal analysis. **Shouwen Jiang**: Validation. **Zhichao Wu**: Validation. **Dong Liu**: Supervision: Writing - Review & Editing. **Linhua Jiang**: Project administration. **Sihua Peng**: Conceptualization, Supervision, Project administration, Writing - Review & Editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Sihua Peng reports financial support was provided by Natural Science Foundation of Shanghai. Linhua Jiang reports financial support was provided by National Science Foundation of China. Sihua Peng reports a relationship with Shanghai Ocean University that includes: employment.

Data availability

All the raw sequencing data were deposited in the NCBI Sequence Read Archive (SRA) under accession numbers SAMN17838283 (soil) and SAMN17838284 (water).

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.envres.2022.114847.

References

Alcock, B.P., et al., 2020. Card 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. Nucleic Acids Res. 48, D517–D525. https://doi.org/10.1093/nar/gkz935.

Alneberg, J., et al., 2014. Binning metagenomic contigs by coverage and composition. Nat. Methods 11, 1144–1146. https://doi.org/10.1038/nmeth.3103.

Alneberg, J., et al., 2018. Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. Microbiome 6, 173. https://doi.org/10.1186/s40168-018-0550-0.

Ashburner, M., et al., 2000. Gene Ontology: tool for the unification of biology. Nat. Genet. 25, 25–29. https://doi.org/10.1038/75556.

Atanasov, A.G., et al., 2021. Natural products in drug discovery: advances and opportunities. Nat. Rev. Drug Discov. 20, 200–216. https://doi.org/10.1038/s41573-020-00114-z.

Berg, I.A., 2011. Ecological aspects of the distribution of different autotrophic CO2 fixation pathways. Appl. Environ. Microbiol. 77, 1925–1936. https://doi.org/10.1128/aem.02473-10.

Blin, K., et al., 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. 47, W81–W87. https://doi.org/10.1093/nar/gkz310.

Buchfink, B., et al., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. https://doi.org/10.1038/nmeth.3176.

Cai, Y.F., et al., 2016. Conventional methanotrophs are responsible for atmospheric methane oxidation in paddy soils. Nat. Commun. 7, 11728 https://doi.org/10.1038/ncomms11728.

Canelhas, M.R., et al., 2016. Methane oxidation at the water-ice interface of an ice-covered lake. Limnol. Oceanogr. 61, S78–S90. https://doi.org/10.1002/lno.10288.

Cao, H.L., et al., 2014. Microbial sulfur cycle in two hydrothermal chimneys on the southwest Indian ridge. mBio 5. https://doi.org/10.1128/mBio.00980-13 e00980-13.

Carbon, S., et al., 2017. Expansion of the gene Ontology knowledgebase and resources. Nucleic Acids Res. 45, D331–D338. https://doi.org/10.1093/nar/gkw1108.

Carbon, S., et al., 2019. The gene Ontology resource: 20 years and still GOing strong. Nucleic Acids Res. 47, D330–D338. https://doi.org/10.1093/nar/gky1055.

Cavicchioli, R., et al., 2019. Scientists' warning to humanity: microorganisms and climate change. Nat. Rev. Microbiol. 17, 569–586. https://doi.org/10.1038/s41579-019-0222-5.

Chaumeil, P.A., et al., 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36, 1925–1927. https://doi.org/10.1093/bioinformatics/btz848.

Chen, B.X., et al., 2014. The impact of climate change and anthropogenic activities on alpine grassland over the Qinghai-Tibet Plateau. Agric. For. Meteorol. 189–190, 11–18. https://doi.org/10.1016/j.agrformet.2014.01.002.

Chen, L.H., et al., 2016. Vfdb 2016: hierarchical and refined dataset for big data analysis-10 years on. Nucleic Acids Res. 44, D694–D697. https://doi.org/10.1093/nar/gkv1239.

Chen, S.F., et al., 2018a. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, 884–890. https://doi.org/10.1093/bioinformatics/bty560.

Chen, Y., et al., 2018b. Long-term and high-concentration heavy-metal contamination strongly influences the microbiome and functional genes in Yellow River sediments. Sci. Total Environ. 637, 1400–1412. https://doi.org/10.1016/j.scitotenv.2018.05.109.

Chen, Y.L., et al., 2017. Distinct microbial communities in the active and permafrost layers on the Tibetan Plateau. Mol. Ecol. 26, 6608–6620. https://doi.org/10.1111/mec.14396.

Cock, P.J.A., et al., 2010. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res. 38, 1767–1771. https://doi.org/10.1093/nar/gkp1137.

Creevey, C.J., et al., 2011. Universally distributed single-copy genes indicate a constant rate of horizontal transfer. PLoS One 6, e22099. https://doi.org/10.1371/journal.pone.0022099.

Cuadrat, R.R.C., et al., 2018. Recovering genomics clusters of secondary metabolites from lakes using genome-resolved metagenomics. Front. Microbiol. 9, 251. https://doi.org/10.3389/fmicb.2018.00251.

DelSontro, T., et al., 2018. Greenhouse gas emissions from lakes and impoundments: upscaling in the face of global change. Limnology and Oceanography Letters 3, 64–75. https://doi.org/10.1002/lol2.10073.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340.

Finking, R., Marahiel, M.A., 2004. Biosynthesis of nonribosomal peptides. Annu. Rev. Microbiol. 58, 453–488. https://doi.org/10.1146/annurev.micro.58.030603.123615.

Han, R., et al., 2017. Microbial community structure and diversity within hypersaline Keke Salt Lake environments. Can. J. Microbiol. 63, 895–908. https://doi.org/10.1139/cjm-2016-0773.

Hu, W.G., et al., 2014. Diversity and community structure of fungi through a permafrost core profile from the Qinghai-Tibet Plateau of China. J. Basic Microbiol. 54, 1331–1341. https://doi.org/10.1002/jobm.201400232.

Hu, W.G., et al., 2016. Characterization of the prokaryotic diversity through a stratigraphic permafrost core profile from the Qinghai-Tibet Plateau. Extremophiles 20, 337–349. https://doi.org/10.1007/s00792-016-0825-y.

Huerta-Cepas, J., et al., 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. Mol. Biol. Evol. 34, 2115–2122. https://doi.org/10.1093/molbev/msx148.

Huerta-Cepas, J., et al., 2019. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 47, D309–D314. https://doi.org/10.1093/nar/gky1085.

Hukic, M., et al., 2018. The effect of lysozyme on reducing biofilms by Staphylococcus aureus, Pseudomonas aeruginosa, and gardnerella vaginalis: an in vitro examination. Microb. Drug Resist. 24, 353–358. https://doi.org/10.1089/mdr.2016.0303.

Jain, C., et al., 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat. Commun. 9, 5114. https://doi.org/10.1038/s41467-018-07641-9.

Jousset, A., et al., 2017. Where less may be more: how the rare biosphere pulls ecosystems strings. ISME J. 11, 853–862. https://doi.org/10.1038/ismej.2016.174.

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto Encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30. https://doi.org/10.1093/nar/28.1.27.

Kang, D.W.D., et al., 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7, e7359. https://doi.org/10.7717/peerj.7359.

Kuang, X.X., Jiao, J.J., 2016. Review on climate change on the Tibetan Plateau during the last half century. J. Geophys. Res. Atmos. 121, 3979–4007. https://doi.org/10.1002/2015JD024728.

La Cono, V., et al., 2018. Contribution of bicarbonate assimilation to carbon pool dynamics in the deep mediterranean sea and cultivation of actively nitrifying and CO2-fixing bathypelagic prokaryotic consortia. Front. Microbiol. 9, 3. https://doi.org/10.3389/fmicb.2018.00003.

Lannes, R., et al., 2019. Carbon fixation by marine ultrasmall prokaryotes. Genome Biology and Evolution 11, 1166–1177. https://doi.org/10.1093/gbe/evz050.

Letunic, I., Bork, P., 2019. Interactive Tree of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47, W256–W259. https://doi.org/10.1093/nar/gkz239.

Li, D.H., et al., 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102, 3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.

Loder, A.J., et al., 2016. Reaction kinetic analysis of the 3-hydroxypropionate/4-hydroxybutyrate CO2 fixation cycle in extremely thermoacidophilic archaea. Metab. Eng. 38, 446–463. https://doi.org/10.1016/j.ymben.2016.10.009.

Lombard, V., et al., 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495. https://doi.org/10.1093/nar/gkt1178.

Lu, J., et al., 2017. Bracken: estimating species abundance in metagenomics data. Peerj Computer Science 3, e104. https://doi.org/10.7717/peerj-cs.104.

Ma, A.A., et al., 2020. Phylogenetic and physiological diversity of cultivable actinomycetes isolated from alpine habitats on the Qinghai-Tibetan plateau. Front. Microbiol. 11, 555351 https://doi.org/10.3389/fmicb.2020.555351.

Magnabosco, C., et al., 2016. A metagenomic window into carbon metabolism at 3 km depth in Precambrian continental crust. ISME J. 10, 730–741. https://doi.org/10.1038/ismej.2015.150.

McKee, L.S., et al., 2021. Polysaccharide degradation by the Bacteroidetes: mechanisms and nomenclature. Environmental Microbiology Reports 13, 559–581. https://doi.org/10.1111/1758-2229.12980.

Momper, L., et al., 2017. Energy and carbon metabolisms in a deep terrestrial subsurface fluid microbial community. ISME J. 11, 2319–2333. https://doi.org/10.1038/ismej.2017.94.

Moriya, Y., et al., 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35, W182–W185. https://doi.org/10.1093/nar/gkm321.

Nahar, S., et al., 2018. Advances and future prospects of enzyme-based biofilm prevention approaches in the food industry. Compr. Rev. Food Sci. Food Saf. 17, 1484–1502. https://doi.org/10.1111/1541-4337.12382.

Nakicenovic, N., et al., 2000. Special Report on Emissions Scenarios, Working Group III, Intergovernmental Panel on Climate Change. Cambridge University Press.

Nayfach, S., et al., 2021. A genomic catalog of Earth's microbiomes. Nat. Biotechnol. 39, 499–509. https://doi.org/10.1038/s41587-020-0718-6.

Neumann, S., et al., 2000. Characterization of the cys gene locus from Allochromatium vinosum indicates an unusual sulfate assimilation pathway. Mol. Biol. Rep. 27, 27–33. https://doi.org/10.1023/a:1007058421714.

Ochi, K., Hosaka, T., 2013. New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. Appl. Microbiol. Biotechnol. 97, 87–98. https://doi.org/10.1007/s00253-012-4551-9.

Ondov, B.D., et al., 2011. Interactive metagenomic visualization in a Web browser. BMC Bioinf. 12, 385. https://doi.org/10.1186/1471-2105-12-385.

Pal, C., et al., 2014. BacMet: antibacterial biocide and metal resistance genes database. Nucleic Acids Res. 42, D737–D743. https://doi.org/10.1093/nar/gkt1252.

Parks, D.H., et al., 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055. https://doi.org/10.1101/gr.186072.114.

Parks, D.H., et al., 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat. Biotechnol. 36, 996–1004. https://doi.org/10.1038/nbt.4229.

Parks, D.H., et al., 2020. A complete domain-to-species taxonomy for Bacteria and Archaea. Nat. Biotechnol. 38, 1079–1086. https://doi.org/10.1038/s41587-020-0501-8.

Patro, R., et al., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nat. Methods 14, 417–419. https://doi.org/10.1038/nmeth.4197.

Philippe, N., et al., 2013. Pandoraviruses: amoeba viruses with genomes up to 2.5 Mb reaching that of parasitic eukaryotes. Science 341, 281–286. https://doi.org/10.1126/science.1239181.

Price, M.N., et al., 2010. FastTree 2-approximately maximum-likelihood trees for large alignments. PLoS One 5, e9490. https://doi.org/10.1371/journal.pone.0009490.

Ranjan, R., 2016. Analysis of the microbiome: advantages of whole genome shotgun versus 16S amplicon sequencing. Biochem. Biophys. Res. Commun. 469, 967–977. https://doi.org/10.1016/j.bbrc.2015.12.083.

Rathour, R., et al., 2020. A comparative metagenomic study reveals microbial diversity and their role in the biogeochemical cycling of Pangong lake. Sci. Total Environ. 731, 139074 https://doi.org/10.1016/j.scitotenv.2020.139074.

Raymond, P.A., et al., 2013. Global carbon dioxide emissions from inland waters. Nature 503, 355–359. https://doi.org/10.1038/nature12760.

Ren, M., et al., 2018. Diversity and contributions to nitrogen cycling and carbon fixation of soil salinity shaped microbial communitiesin tarim basin. Front. Microbiol. 9, 431. https://doi.org/10.3389/fmicb.2018.00431.

Rui, J.P., et al., 2015. Responses of bacterial communities to simulated climate changes in alpine meadow soil of the Qinghai-Tibet Plateau. Appl. Environ. Microbiol. 81, 6070–6077. https://doi.org/10.1128/aem.00557-15.

Ruiz-Fernandez, P., et al., 2020. Autotrophic carbon fixation pathways along the redox gradient in oxygen-depleted oceanic waters. Environmental Microbiology Reports 12, 334–341. https://doi.org/10.1111/1758-2229.12837.

Saunois, M., et al., 2016. The global methane budget 2000-2012. Earth Syst. Sci. Data 8, 697–751. https://doi.org/10.5194/essd-8-697-2016.

Seemann, T., 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Sharpton, T.J., 2014. An introduction to the analysis of shotgun metagenomic data. Front. Plant Sci. 5, 209. https://doi.org/10.3389/fpls.2014.00209.

Song, W.J., et al., 2019. Bacterial community rather than metals shaping metal resistance genes in water, sediment and biofilm in lakes from arid northwestern China. Environ. Pollut. 254, 113041 https://doi.org/10.1016/j.envpol.2019.113041.

Spasov, E., et al., 2020. High functional diversity among Nitrospira populations that dominate rotating biological contactor microbial communities in a municipal wastewater treatment plant. ISME J. 14, 1857–1872. https://doi.org/10.1038/s41396-020-0650-2.

Sunagawa, S., et al., 2013. Metagenomic species profiling using universal phylogenetic marker genes. Nat. Methods 10, 1196–1199. https://doi.org/10.1038/nmeth.2693.

Tatusov, R.L., et al., 1997. A genomic perspective on protein families. Science 278, 631–637. https://doi.org/10.1126/science.278.5338.631.

Uritskiy, G.V., et al., 2018. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6, 158. https://doi.org/10.1186/s40168-018-0541-1.

Ventola, C.L., 2015. The antibiotic resistance crisis: part 1: causes and threats. P T : a peer-reviewed journal for formulary management 40, 277–283.

Wang, S.Y., et al., 2019. Denitrification is the main microbial N loss pathway on the Qinghai-Tibet Plateau above an elevation of 5000 m. Sci. Total Environ. 696, 133852 https://doi.org/10.1016/j.scitotenv.2019.133852.

Wei, S.P., et al., 2014. Diversity and distribution of archaea community along a stratigraphic permafrost profile from Qinghai-Tibetan plateau, China. Archaea-an International Microbiological Journal 12, 240817. https://doi.org/10.1155/2014/240817.

Winnenburg, R., et al., 2006. PHI-base: a new database for pathogen host interactions. Nucleic Acids Res. 34, D459–D464. https://doi.org/10.1093/nar/gkj047.

Wiseschart, A., et al., 2019. Shotgun metagenomic sequencing from Manao-Pee cave, Thailand, reveals insight into the microbial community structure and its metabolic potential. BMC Microbiol. 19, 144. https://doi.org/10.1186/s12866-019-1521-8.

Wood, D.E., et al., 2019. Improved metagenomic analysis with Kraken 2. Genome Biol. 20, 257. https://doi.org/10.1186/s13059-019-1891-0.

Wu, J., et al., 2018. Pollution, ecological-health risks, and sources of heavy metals in soil of the northeastern Qinghai-Tibet Plateau. Chemosphere 201, 234–242. https://doi.org/10.1016/j.chemosphere.2018.02.122.

Wu, Y.W., et al., 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32, 605–607. https://doi.org/10.1093/bioinformatics/btv638.

Xing, R., et al., 2019. Large-scale distribution of bacterial communities in the Qaidam Basin of the Qinghai-Tibet Plateau. Microbiologyopen 8, e909. https://doi.org/10.1002/mbo3.909.

Yamada, Y., et al., 2015. Terpene synthases are widely distributed in bacteria. Proc. Natl. Acad. Sci. U.S.A. 112, 857–862. https://doi.org/10.1073/pnas.1422108112.

Yin, Y.B., et al., 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 40, W445–W451. https://doi.org/10.1093/nar/gks479.

Zhang, X.X., et al., 2018. Metagenomics reveals microbial diversity and metabolic potentials of seawater and surface sediment from a hadal biosphere at the Yap Trench. Front. Microbiol. 9, 2402. https://doi.org/10.3389/fmicb.2018.02402.

Zhou, J.Z., et al., 2015. High-Throughput metagenomic technologies for complex microbial community analysis: open and closed formats. mBio 6. https://doi.org/10.1128/mBio.02288-14 e02288-14.