BIODIVERSITAS
Volume 25, Number 11, November 2024
Pages: 4613-4626

ISSN: 1412-033X
E-ISSN: 2085-4722
DOI: 10.13057/biodiv/d251160

Exploring the microbial diversity and functional potential of Domas Crater of Mount Tangkuban Perahu, Indonesia, through shotgun metagenomics

HARYONO BUDI SANTOSO[1,2,♥], ANTONIUS SUWANTO[2,3], RAHADIAN PRATAMA[4]

[1]Saraswanti Genomics Institute. Jl. Rasamala No. 20, Bogor 16113, West Java, Indonesia. Tel. +62-251-7532348, ♥email: haryonobs@gmail.com
[2]Graduate Program of Biotechnology, School of Graduates, Institut Pertanian Bogor. Jl. Raya Dramaga, Kampus IPB, Dramaga, Bogor 16680, West Java, Indonesia
[3]Department of Biology, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor. Jl. Raya Dramaga, Kampus IPB, Dramaga, Bogor 16680, West Java, Indonesia
[4]Department of Biochemistry, Faculty of Mathematics and Natural Sciences, Institut Pertanian Bogor. Jl. Bungur No. 1, Dramaga, Bogor 16680, West Java, Indonesia

Manuscript received: 3 September 2024. Revision accepted: 30 November 2024.

Abstract. *Santoso HB, Suwanto A, Pratama R. 2024. Exploring the microbial diversity and functional potential of Domas Crater of Mount Tangkuban Perahu, Indonesia, through Shotgun Metagenomics. Biodiversitas 25: 4613-4626.* Domas Crater, an extreme environment in Indonesia, is known for high temperatures and acidic conditions, providing a unique habitat for specialized microbial communities. These extreme conditions increase the possibility of discovering thermophilic enzymes with valuable biotechnological applications. Therefore, this study aimed to explore the microbial diversity in Domas Crater using shotgun metagenomics to analyze both previously reported microbes and novel microorganisms comprehensively. Shotgun metagenomics is particularly advantageous in identifying microbial species that cannot be cultured using conventional methods, enabling the exploration of microorganisms with considerable potential. The application of next-generation sequencing technologies and bioinformatics tools allowed the successful reconstruction of eight high-quality Metagenome-Assembled Genomes (MAGs), a testament to the technical proficiency of the study. The genomes were further characterized based on the functional genes, including the enzymes in carbohydrate metabolism or Carbohydrate-active enzymes (CAZyme), biosynthetic gene clusters for secondary metabolite (BGCs), and genes associated with micronutrient metabolism. The results showed that the microbial community was dominated by Hydrogenobaculum and Sulfurisphaera, both known for adaptation to extreme environments. Moreover, the first Hydrogenobaculum and Thermocladium were recorded in Indonesia as the novel discoveries of the study. These findings highlight the significance of Domas Crater as a reservoir for novel microbial species, particularly in terms of thermophilic microorganisms with unique enzymatic properties.

Keywords: Bioprospecting, hot springs, metagenome-assembled genome, shotgun metagenomics, thermophiles

INTRODUCTION

Hot springs are extreme environments supporting unique microbial ecosystems and have been the focus of several studies for decades. Indonesia, with more than 130 active volcanoes and numerous craters spread across its archipelago, Indonesia is home to a diverse range of geothermal environments, including an abundance of hot springs, each offering unique ecological and biotechnological potential (Lischer et al. 2020). The diverse physicochemical conditions at each location lead to highly specific habitats for extremophilic microbes capable of surviving and thriving in different temperatures, acidity, and salinity levels. An important example of these locations is Domas Crater in Mount Tangkuban Perahu, West Java, which has an acidic pH ranging from one to four and high temperatures, showing the area as an ideal environment for extremophilic microbes (Safitri et al. 2020). Previous studies have extensively explored the microbial diversity in this crater and reported the presence of several microorganisms with potential characteristics, as presented in Table 1.

The studies on microbes in hot springs have led to the discovery of different enzymes and biological molecules with potential applications in biotechnology. Moreover, several extremophiles living in hot springs have produced heat-resistant enzymes considered usable in industrial processes requiring high temperatures. These enzymes have potential applications in biofuel production, the food industry, and other sectors. Several studies have also been widely conducted to explore thermally stable enzymes such as CAZymes (enzymes related to cellulose, hemicellulose, and oligosaccharides) (Sharma et al. 2020; Reichart et al. 2021; Saryono et al. 2022), lipolytic enzymes (López-López et al. 2015), lipases, proteases, xylanases, and others (Almando et al. 2019; Nguyen et al. 2020). A particularly interesting enzyme to explore from thermophilic microbes is DNA polymerase.

Exploring microbial diversity in extreme environments like Domas Crater requires various methods, including conventional culturing and advanced sequencing technologies such as Next Generation Sequencing (16S rRNA metabarcoding and shotgun metagenome). While conventional culturing techniques provide valuable insights,

they have limitations in identifying hard-to-cultivate microbes, which are often capable of producing highly stable and functional enzymes under extreme conditions (Mirete et al. 2016). The application of 16S rRNA metabarcoding has become a popular method for analyzing microbial diversity due to its ability to show microbial diversity without prior cultivation but is limited to only taxonomic analysis and does not provide information on the genetic functions of the identified microbes (Kamble et al. 2020). The method is limited to detecting bacteria and archaea, as it targets the 16S rRNA gene and cannot identify viruses or other microbes that lack this gene. Another disadvantage is the possibility of missing critical information on microbes with highly divergent or poorly conversed 16S rRNA gene sequences (Kamble et al. 2020; Sysoev et al. 2021).

Shotgun metagenomics is a more comprehensive alternative that has the ability to identify all genetic elements within an environmental sample, including bacteria, archaea, viruses, and even functional DNA fragments (Strazzulli et al. 2017; Quince et al. 2017). This method allows for higher-resolution microbial identification and enables functional analysis of the entire microbial community (New and Brito 2020). Shotgun metagenomics can assist in uncovering active metabolic pathways, detecting genes responsible for antibiotic resistance, and discovering new enzymes with significant industrial applications (Strazzulli et al. 2017). However, there is limited application of this method in geothermal microbial studies in Indonesia probably due to the technical challenges and higher costs compared to 16S rRNA metabarcoding (New and Brito 2020). Shotgun metagenomics has significant potential to provide deeper and broader insights into microbial diversity in Domas Crater and other geothermal locations.

This study aimed to use shotgun metagenomics to explore microbial diversity in Domas Crater and identify functional genes present within the microbial community. The method was expected to show the microbes inhabiting the acidic and extreme environment of Domas Crater as well as to understand the genetic functions contributing to microbial adaptation to harsh environmental conditions. The goal was to offer a more in-depth understanding of the potential of environmental microbes to produce thermostable enzymes and other biomolecules with significant biotechnological applications. Additionally, the study aims to promote the broader use of shotgun metagenomics in exploring Indonesia's valuable microbial resources in extreme environments, driven by the growing availability of advanced sequencing technologies and bioinformatics tools.

MATERIALS AND METHODS

Sample collection

Water samples were collected from Domas Crater located in Mount Tangkuban Perahu, Ciater, Subang, West Java, Indonesia at Lat = -6.76167 S 6' 45'42.021" and Long = 107.62608 E107'37'33.90312 as presented in Figure 1. The sampling procedure is referenced from the studies by (Mashzhan et al. 2021; Kato et al. 2022). The pH and temperature of the water were measured using a pH meter and a thermometer before the sampling process. Subsequently, 500 mL of water from a depth of 5-20 cm was collected aseptically into a sterile bottle, and some dissolved particles were also collected along with the water. The bottle containing the sample was placed in a styrofoam box with dry ice to keep the sample cool during transport (temperature around -20℃ to -5℃). The samples were brought to the laboratory within approximately five hours and stored in a freezer upon arrival for one day. Moreover, the water samples were later filtered using a 0.2 µm cellulose nitrate filter paper (Sartorius, Germany). The filtered sample paper was subsequently stored at -80°C before DNA extraction.

DNA extraction and shotgun metagenomic sequencing

Samples from three different locations in Candradimuka Crater with varying temperatures were combined to obtain a microbial representation from diverse environmental conditions within a single sample set. This combination aimed to provide a more holistic overview of the microbial community in the same area despite temperature differences. DNA was extracted from the filtered sample paper using the Zymobiomics DNA miniprep kit (Zymo Research, USA) in line with the guidelines provided by the manufacturer. The concentration and quality of the DNA extracted were evaluated using a Qubit 4 fluorometer (Thermo Fisher Scientific™, USA) and confirmed by agarose gel electrophoresis.

Table 1. Previous exploratory studies conducted at Domas Crater, Subang, West Java, Indonesia

Year	Exploration methods	The discovered microbes	Reference
1991	Culture on growth media, GC content, & DNA homology analysis	*Acidianus infernus, Thermoplasma* sp.	(Huber et al. 1991)
2012	Culture on Growth Media	*Sulfolobus, Thermoplasma*	(Handayani et al. 2012)
2015	Culture on growth media, 16S rRNA & 23S rRNA sequencing	*Sulfurisphaera javensis* sp.	(Tsuboi et al. 2018)
2020	Culture on growth media, 16S rRNA sequencing	*Bacillus, Geobacillus*	(Safitri et al. 2020)
2021	Culture on growth media, WGS (Whole Genome Sequencing)	*Metallosphaera* sp.	(Sakai et al. 2022)
2022	Culture on growth media, Random PCR	*Caldivirga, Metallosphaera*	(Sastroredjo et al. 2023)

Moreover, 100 ng of total DNA was used to construct libraries through the MGIEasy FS DNA Library Preparation Set (MGI Tech Co., Ltd, China), and the metagenomic sequencing was conducted on the DNBSEQ-G400 system (MGI Tech Co., Ltd, China) to produce 150 bp paired-end reads (PE150). Sequencing was conducted in three technical replicates (labeled KD1, KD2, KD3), each with an output of 5 Gb, to evaluate the consistency of microbial abundance detection. Post-sequencing adapter sequences were later removed from the data using SOAPnuke software (Chen et al. 2018).

Metagenomic analysis

Quality control (QC) was determined for the sequencing data from each sample using fastp to eliminate low-quality bases (Chen 2023). For taxonomic classification, clean data were analyzed using Kaiju based on the nr_euk database (Menzel et al. 2016). The metagenomics analysis was initiated by merging the three sequencing datasets (KD1, KD2, KD3) into a single dataset using online tools available in KBase (Chivian et al. 2023). Genome assembly was also executed through the application of the tools in KBase, MEGAHIT (Li et al. 2016). Moreover, the metagenome-assembled genomes (MAGs) were obtained based on binning using MetaBAT2 (Kang et al. 2019), MaxBin2 (Wu et al. 2016), and CONCOCT (Alneberg et al. 2014), followed by refinement with DAS Tool (Sieber et al. 2018). The quality of each MAG was evaluated using CheckM (Parks et al. 2015). The initial sequencing data were later mapped back to each MAG using Bowtie2 (Langmead and Salzberg 2013) to determine relative abundance and genome coverage. Furthermore, the BAM files generated from the mapping process were subjected to quality checks using Qualimap2 (Okonechnikov et al. 2016). MAG prediction and annotation were conducted with Prokka (Seemann 2014), followed by the classification into CAZyme groups using dbCAN3 (Zheng et al. 2023), Biosynthetic Gene Clusters (BGCs) through antiSMASH v7 (Blin et al. 2023), and microTrait Bioelement (Karaoz and Brodie 2022). The last analysis was pangenome, which was achieved through Anvi'o v8 (Delmont and Eren 2018; Eren et al. 2021).

RESULTS AND DISCUSSION

Shotgun metagenomic sequencing data

The sequencing results using MGI technology (DNBSEQ-G400) yielded three datasets representing technical replicates. Each sequencing run produced approximately 40 million reads with a total of about 5 gigabases, as presented in Table 2. The average read length was 149 × 2 after quality control processing was conducted using fastp. Moreover, quality control was used to remove poor bases and filter out data with quality below Q20, leading to more than 98% of the reads above Q20, which were used for subsequent analysis. This sequencing data can be accessed in the NCBI repository under Bioproject number PRJNA1117210.

Table 2. Sequencing data statistics after quality control (QC)

Sample	Total reads	Total bases (Gbp)	Average of Read length	Bases >Q20 (%)	GC (%)
KD1	40,012,886	5.347197	149 x 2	98.39	39.43
KD2	40,273,240	5.120246	149 x 2	98.87	38.91
KD3	40,285,572	5.020576	149 x 2	98.85	40.86

Figure 1. A. Sampling locations at Domas crater, Subang, West Java, Indonesia: green circles show hot water emergence points while red circles represent sampling points; B. The sampling site at Domas crater at a water depth of 5-20 cm

Taxonomic abundance profile

The relative abundance of each sample replicate was examined using the Kaiju tool to classify each read into a taxon based on the similarity to hits in the nr_euk NCBI database used consisting of archaea, bacteria, viruses, fungi, and eukaryotic microbes. After taxonomic classification, the data was displayed at the genus level to provide an overview of the abundant microbes in each sample replicate. The percentages for KD1, KD2, and KD3 of the total reads classified into specific taxa were 42%, 31.16%, and 30.39%, respectively, as presented in Figure 2. On average, approximately 19% of reads were classified above the genus level and dominantly fell within the Thermoprotei and Aquificae classes (Figure 3). The relative abundance results for the average of the three replicates, listed in order of highest abundance, include *Hydrogenobaculum* (22.8%), *Pedobacter* (9.8%), *Sulfurisphaera* (7.8%), *Sulfolobus* (7%), *Vulcanisaeta* (5%), *Acidianus* (3.4%), *Achromobacter* (3.3%), *Metallosphaera* (3.2%), *Stygiolobus* (2.9%), *Acidilobus* (2.6%), and *Chromobacterium* (1.5%) (Figure 2). The remaining taxa had a relative abundance of less than 1%.

The abundant microbes detected are commonly found in acidic and high-temperature conditions which is the same environmental situation in Domas Crater. In the Archaea domain, the phylum Thermoproteota is dominant, with the majority of classified reads assigned to the orders of Sulfolobales and Thermoproteales. Moreover, the genera of the order Sulfolobales, such as *Sulfurisphaera*, *Sulfolobus*, *Acidianus*, *Metallosphaera*, and *Stygiolobus*, were thermoacidophilic and played roles in sulfur metabolism, allowing the growth in acidic and high-temperature conditions (Liu et al. 2021). The order Thermoproteales includes organisms living at high temperatures but having a broad pH range, such as *Vulcanisaeta*, *Caldivirga*, and *Thermocladium*, thriving in acidic pH. Meanwhile, *Thermoproteus* and *Pyrobaculum* are usually found in neutral temperatures (Jay et al. 2016). The relative abundance results for the order Thermoproteales in this study were similar to those reported in Yellowstone, USA, where *Vulcanisaeta*, *Caldivirga*, and *Thermocladium* were reported to be dominant (Jay et al. 2016). *Acidilobus*, also found in Domas Crater with an abundance of over 1%, is a hyperthermophilic organism identified in acidic pH and at 85°C in Kamchatka, Russia (Prokofeva et al. 2000).

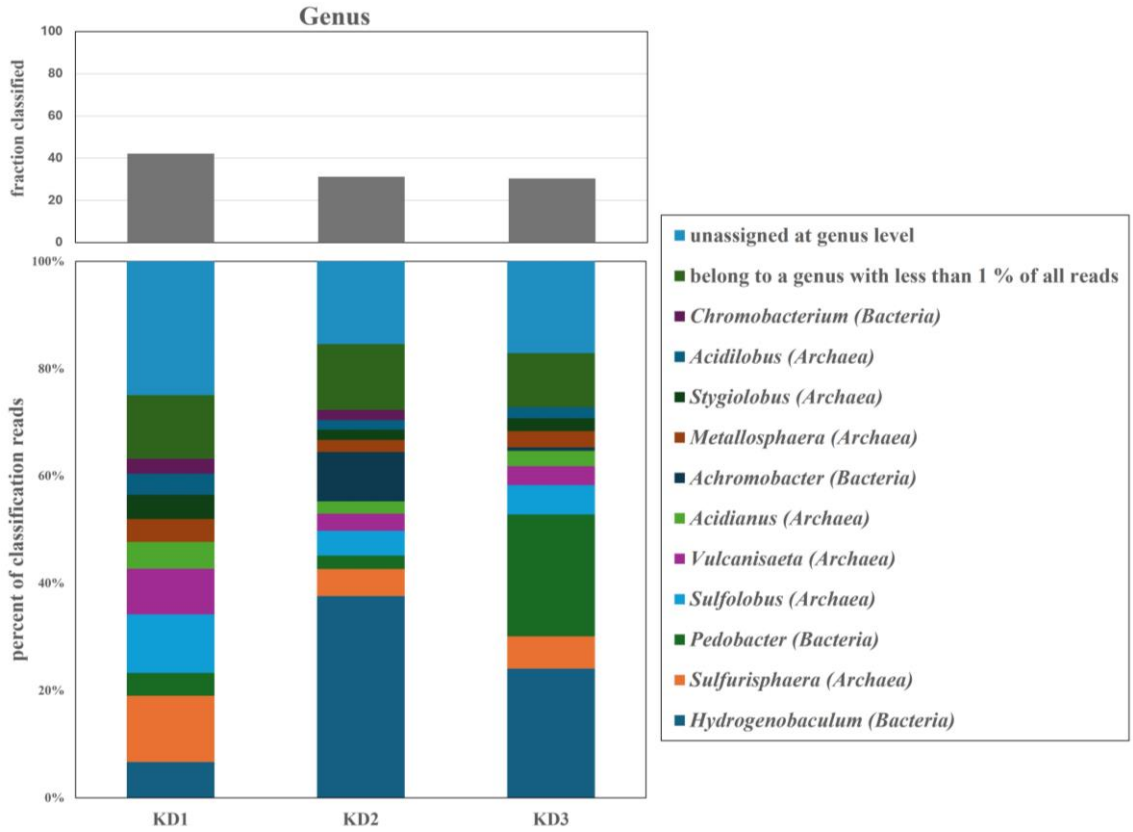

Figure 2. The bottom bar chart shows the Relative abundance of microbial diversity at the genus level. The upper bar chart shows the percentage of reads classified as specific taxa, with the remaining percentage representing unclassified reads

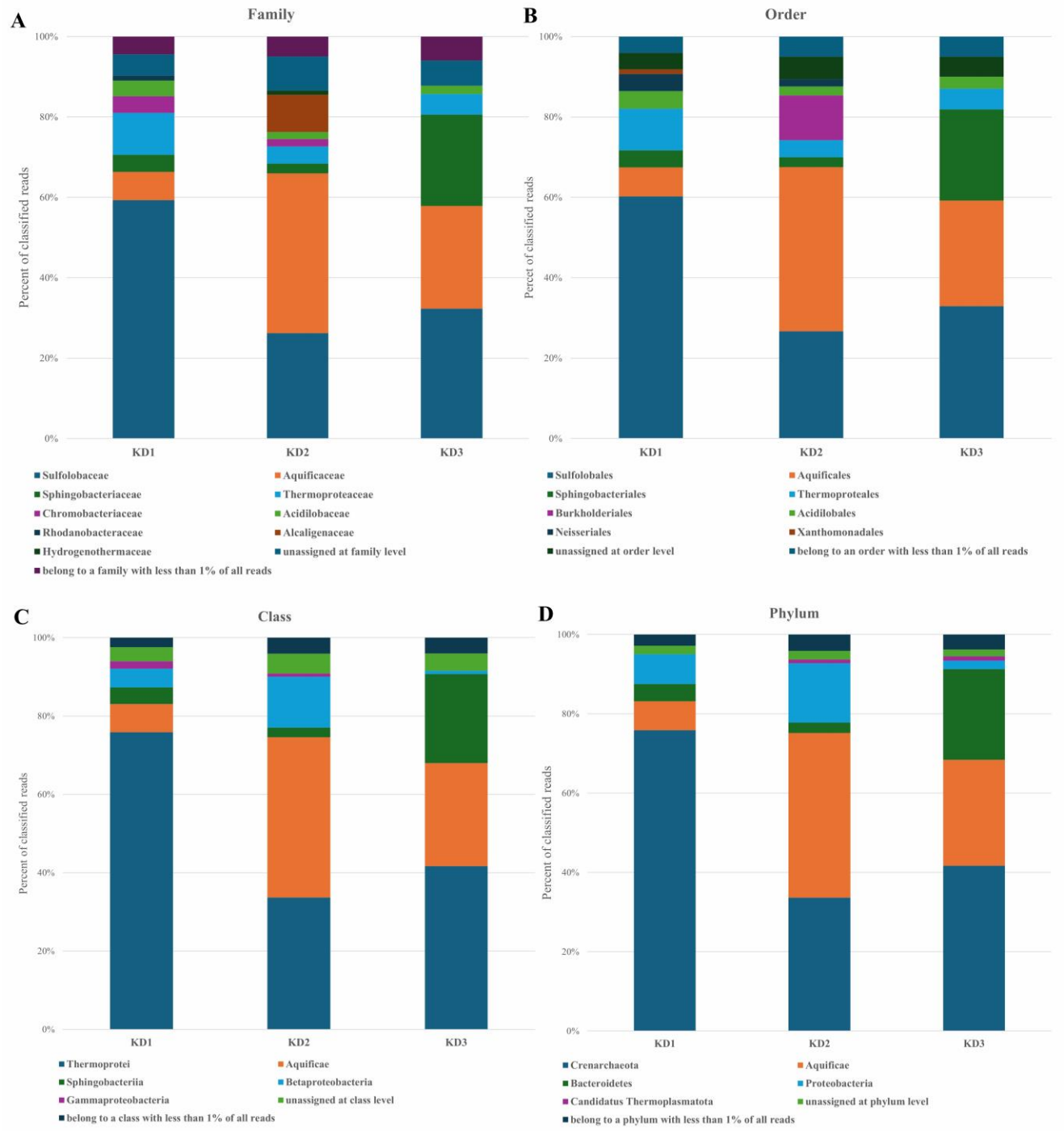

Figure 3. Relative abundance of microbial diversity at the level: A. Family; B. Order; C. Class; D. Phylum

In the Bacteria domain, *Hydrogenobaculum* was the most dominant (22.8%) genus found in Domas Crater and identified to be in the order Aquificales and the family Aquaficaceae. It was initially classified under the genus *Hydrogenobacter* but was reclassified due to the distant phylogenetic relationship and ability to survive in low pH conditions (Stöhr et al. 2001). *Pedobacter* was also found to be abundant (9.8%) and existed in different environmental conditions, including soil and water, and in association with other organisms such as planaria (Kangale et al. 2020). The microbe was reported to be in conditions as cold as 25-30°C (Tuyet and Kim 2024). However, no sources indicate that *Pedobacter* thrives in high-temperature and acidic pH environments. *Achromobacter* and *Chromobacterium* were also identified in this study, but no sources were found correlating both with hot and acidic environments. Both genera are typically found as infection sources and have been observed in aquatic environments (Gomes et al. 2022). The trend shows there is a need for further investigation on *Pedobacter*, *Achromobacter*, and *Chromobacterium*.

The identification results in this study captured approximately all the microbes previously obtained in Domas Crater, as reported in Table 1. *Acidianus*, *Sulfolobus*, *Sulfurisphaera*, and *Metallosphaera* were identified with an abundance of more than 1%, while *Thermoplasma*, *Bacillus*, and *Geobacillus* were also identified but below 0.5%. Meanwhile, the earlier studies included cultivation, which limited the microbes to those capable of growing in the media used.

Metagenome-Assembled Genomes (MAGs) reconstruction

The accuracy of the taxonomic classification results from the metagenomic data was assured by reconstructing the genomes of the microorganisms. The purpose was to ensure more confidence and conduct a deeper analysis of the functional genes found in the Domas Crater environment. In bioinformatics, there is the concept of MAGs, which is a strategy for reconstructing genomes from metagenomic datasets (Setubal 2021) through assembly and binning (Papudeshi et al. 2017). This study used bioinformatics tools, specifically MEGAHIT, for the assembly and a combination of Metabat2, Maxbin2, Concoct, and DAStools for binning. Both processes are crucial for constructing genomes from metagenomic data and subsequent grouping into genomes of the same organism (Ladoukakis et al. 2014).

The MAGs reconstruction from metagenomic data was conducted using combined samples KD1, KD2, and KD3, which were replicates obtained through DNA extraction from Domas Crater. A total of 25 MAGs were successfully reconstructed in Table 3 and categorized according to the "Minimum Information about Metagenome-Assembled Genomes (MIMAG) standards" based on completeness and contamination values (Bowers et al. 2017). The criteria for classification include high-quality MAGs at the completeness of >90% and contamination ≤5%, medium-quality is ≥50% and ≤10%, and low-quality is ≤50% and ≥10%, respectively. The classification led to the determination of eight high-quality, 15 medium-quality, and two low-quality MAGs in Table 3. However, this study only focused on discussing the eight high-quality MAGs, as presented in Table 4.

The eight High-quality MAGs are candidates for genomes in the contig form, ranging from 25 to 242, with sizes between 1.02 Mbp and 2.27 Mbp. The results showed that KD_010 was classified at the species level, five including KD_001, KD_009, KD_012, KD_018, and KD_024 at the genus level, and KD_005 and KD_023 at the family and order level of Conexivisphaerales. Moreover, the remapping sequencing data to each MAG showed that KD_012 and KD_018 were reconstructed with an average coverage of more than 1500x with relative abundances of 17.74% and 19.99%, respectively, as presented in Table 4. KD_012 was classified as *Sulfurisphaera* sp. while KD_018 was categorized as *Hydrogenobaculum* sp. These results strongly support the data presented in the previous taxonomic abundance profile in Figure 2, where the genera *Hydrogenobaculum* and *Sulfurisphaera* were identified to have a high abundance. Meanwhile, MAG KD_010 was classified at the species level as *Metallosphaera javensis*.

Table 3. All MAGs reconstructed from the Domas Crater

Bin name	Phylum	GTDB classification	Completeness	Contamination	MAGs quality
bin.001	Thermoproteota	*Caldivirga* sp.	99.26	1.16	High quality
bin.002	Aquificota	*Hydrogenobaculum* sp.	81.84	0.41	Medium quality
bin.003	Thermoplasmatota	*Thermoplasmatales archaeon*	81.29	2.58	Medium quality
bin.004	Nanoarchaeota	*Nanopusillaceae archaeon*	66.74	0	Medium quality
bin.005	Thermoproteota	*Conexivisphaerales archaeon*	90.29	0.97	High quality
bin.006	Nanoarchaeota	*Parvarchaeaceae archaeon*	57.63	0	Medium quality
bin.007	Thermoplasmatota	*Thermoplasmata archaeon*	84.81	2.02	Medium quality
bin.008	Thermoproteota	*Sulfurisphaera* sp.	98.81	16.07	Low quality
bin.009	Thermoproteota	*Caldisphaer*a sp.	94.42	1.27	High quality
bin.010	Thermoproteota	*Metallosphaera javensis*	98.81	0.6	High quality
bin.011	Thermoproteota	*Sulfurisphaera* sp.	83.43	5.65	Medium quality
bin.012	Thermoproteota	*Sulfurisphaera* sp.	92.86	1.49	High quality
bin.013	Thermoproteota	*Vulcanisaeta* sp.	72.79	5.88	Medium quality
bin.014	Thermoproteota	*Caldivirga* sp.	67.84	0.25	Medium quality
bin.015	Thermoproteota	*Conexivisphaerales archaeon*	48.22	0.97	Low quality
bin.016	Micrarchaeota	*Micrarchaeales archaeon*	76.64	0.93	Medium quality
bin.017	Nanoarchaeota	*Parvarchaeales archaeon*	77.1	0.93	Medium quality
bin.018	Aquificota	*Hydrogenobaculum* sp.	96.34	0.41	High quality
bin.019	Nanoarchaeota	*Parvarchaeaceae archaeon*	77.65	1.87	Medium quality
bin.020	Thermoproteota	*Vulcanisaeta* sp.	82.35	13.6	Medium quality
bin.021	Thermoproteota	*Thermoproteus* sp.	82.54	2.26	Medium quality
bin.022	Thermoproteota	*Vulcanisaeta* sp.	66.35	2.21	Medium quality
bin.023	Thermoproteota	*Conexivisphaeraceae archaeon*	95.15	3.16	High quality
bin.024	Thermoproteota	*Thermocladium* sp.	90.84	1.47	High quality
bin.025	Thermoproteota	*Marsarchaeales archaeon*	56.07	0.27	Medium quality

Functional analysis of HQ MAGs

Functional genes were analyzed on each high-quality MAG, and this was achieved by predicting the genes for each genome using Prokka with the count found to range from 1201 to 2517, as presented in Figure 4. The predicted genes were later identified based on functional groups, including carbohydrate metabolism-related enzymes (CAZyme), biosynthetic clusters for secondary metabolite formation, and those associated with bioelement metabolism, as presented through the heatmap in Figure 4.

The diversity of carbohydrate-active enzymes (CAZymes) in microbial communities, especially in extreme environments like hot springs, holds significant potential for various biotechnological applications. CAZymes are key players in the breakdown and modification of complex carbohydrates, which is crucial for processes such as biofuel production, waste management, and the development of industrial enzymes (Reichart et al. 2021). For example, thermophilic enzymes can be used in the production of biofuels from lignocellulosic biomass, where high-temperature processes can accelerate degradation (Rasool and Irfan 2024). Additionally, the identification of novel CAZymes may aid in the development of more efficient bioremediation strategies, where microbial enzymes are used to break down pollutants or complex organic compounds (Chettri et al. 2020). These enzymes also have applications in the food and pharmaceutical industries, where they can be used in the processing of fibers and the development of specialty products. In the CAZyme group, enzymes related to carbohydrate metabolism were classified into Glycoside Hydrolases (GH), Glycosyltransferases (GT), Polysaccharide Lyases (PL), Carbohydrate Esterases (CE), Carbohydrate-binding Modules (CBM), and Auxiliary Activities (AA) (Cantarel et al. 2009; Drula et al. 2022). Glycoside Hydrolases (GH) are required in hydrolyzing glycosidic bonds to break down compounds such as cellulose, hemicellulose, and other polysaccharides. Glycosyltransferases (GT) are responsible for synthesizing oligosaccharides, polysaccharides, and glyco-conjugates. At the same time, Carbohydrate Esterases (CE) facilitate the hydrolysis of carbohydrate esters and are considered important for degrading hemicellulose and pectin by removing ester groups. Moreover, Polysaccharide Lyases (PL) break down polysaccharides through non-hydrolytic processes, and Carbohydrate-binding Modules (CBM) are non-catalytic domains that enhance the efficiency of other catalytic enzymes by binding them to carbohydrates (Sidar et al. 2020; Shi et al. 2023). Auxiliary Activities (AA) are included in lignocellulose degradation through oxidative processes(Levasseur et al. 2013).

MAG KD_001 *Caldivirga* sp. was identified to have the highest number of CAZymes, totaling 71 genes, with GH being the most dominant group, as presented in Figure 4. This shows that KD_001 has a significant ability to hydrolyze glycosidic bonds in carbohydrates. *Caldivirga* was reported to be growing on different carbon sources at high temperatures and low pH and also had the ability to break down complex plant biomass due to the presence of CAZymes (Meng et al. 2015). Meanwhile, KD_018 *Hydrogenobaculum* sp. had the highest number of CAZymes in the GT group, and the microbes with more completion were observed to have a greater ability to metabolize carbohydrates. Other MAGs, such as those in the Sulfolobales group, *Metallosphaera*, and *Sulfurisphaera*, were identified to contain CAZymes but in smaller quantities. Furthermore, the ability of the species in the Sulfolobales order, such as *Sulfolobus solfataricus*, to produce novel glycoside hydrolases has been studied (Cobucci-Ponzano et al. 2010).

Another group of functional genes identified is the BGCs, which are regions capable of producing secondary metabolites in microorganisms or plants (Galal et al. 2023). BGCs are categorized based on the synthesis pathways of secondary metabolites such as NRPS (Non-Ribosomal Peptides), PKS (Polyketide Synthases), RiPP (Ribosomally Synthesized and Post-Translationally Modified Peptides), and Terpene.

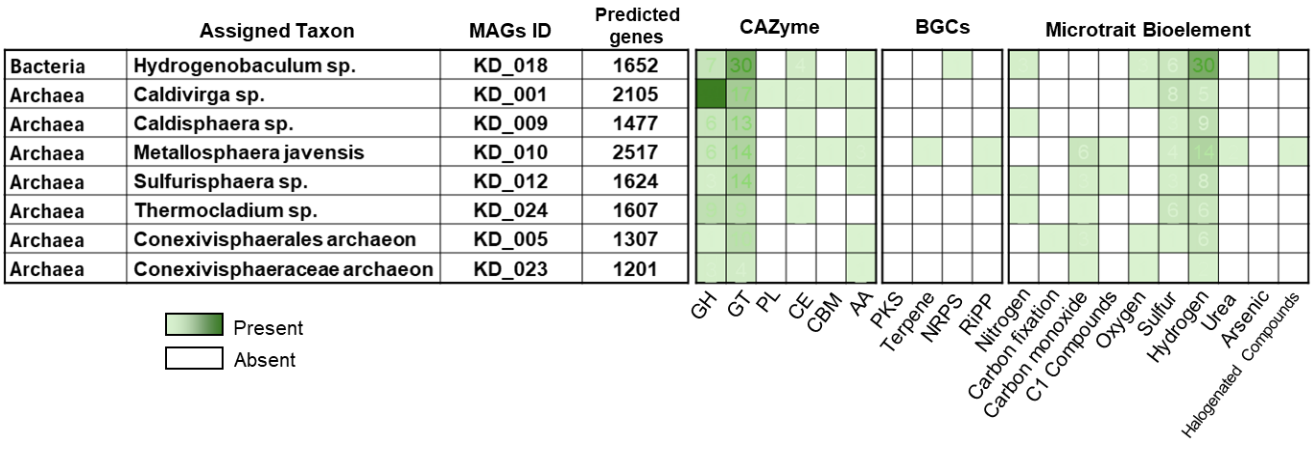

Figure 4. Functional genes classification of MAGs into CAZymes, BGCs, and MicroTrait

The analysis of MAGs from Domas Crater showed that only three were detected to have BGCs, including KD_018 *Hydrogenobaculum* sp. with one BGC in NRPS, KD_010 *Metallosphaera javensis* with two pathways in the form of Terpene and RiPP, and KD_012 *Sulfurisphaera* sp. with one BGC in RiPP as presented in Figure 4. Moreover, KD_010 and KD_012 were part of the Sulfolobales order. They identified to produce Pyrroloquinoline quinone (PQQ), which was considered important in bacterial dehydrogenase in addition to potential health benefits for mammals in the form of neuroprotection in neurodegenerative diseases (Canovai and Williams 2024). Sulphur-dependent archaebacteria have previously been known to produce quinone compounds (Elling et al. 2016). Furthermore, KD_018 *Hydrogenobaculum* sp. was detected to have BGCs associated with the synthesis of two-succinylbenzoate-CoA ligase which was an enzyme in vitamin K2 synthesis.

MicroTrait analysis was applied to predict functional genes from MAGs related to bioelement metabolism capabilities in the environment (Karaoz and Brodie 2022). The results estimated that all MAGs had genes associated with hydrogen and sulfur metabolism and were observed to be consistent with the high sulfur levels recorded in Domas Crater. MAG KD_018 *Hydrogenobaculum* had the highest number of genes correlated with hydrogen and was detected to have genes associated with arsenic metabolism. This was in line with previous studies that *Hydrogenobaculum* could oxidize arsenite and use hydrogen as the energy source (Donahoe-Christiansen et al. 2004). For nitrogen-related genes, four MAGs were identified, including KD_018, KD_009, KD_012, and KD_024. The analysis showed that MAG KD_010, in addition to sulfur, hydrogen, and carbon, also had genes correlated with urea and halogenated compounds. Furthermore, *Metallosphaera* has been reported to perform sulfide metal oxidation and contains genes related to metal and sulfur oxidation (Wang et al. 2020).

Pangenome analysis of HQ MAGs

Pangenome analysis was conducted to compare the MAGs successfully reconstructed from Domas Crater with several genomes in the NCBI RefSeq database. The list of genomes used in the analysis can be found in Table 5. Moreover, the genes present in all genomes were formed into orthologous clusters for subsequent categorization into different groups. The first is Core, which consists of gene clusters present in all compared genomes; the second is Accessory, consisting of those found in at least two or more compared genomes but not all; and the third is Singleton, including the gene clusters unique to a single genome and serve as differentiators between the genomes compared. Pangenomics proved valuable for understanding bacterial clades and developing strategies based on biological similarities and differences. The method contributed to the understanding of genetic diversity within species and populations (Golchha et al. 2024).

MAG KD_018, classified as *Hydrogenobaculum* sp., was compared with five genomes available in RefSeq NCBI and presented in Table 5. The results showed that none of the five genomes was classified at the species level and fell under *Hydrogenobaculum* sp. but all were strains obtained from Yellowstone National Park, USA. Moreover, it was observed from the six genomes, including MAG KD_018 and five references, in Figure 5.A that 1264 gene clusters (63.5%) were classified as Core, 359 (18%) as Accessory, and 368 (18.5%) as Singleton. The analysis did not show significant records of *Hydrogenobaculum* genomes available, but the five references had some similarities, with only GCF_000020785 observed to have a slight overlap of gene clusters with KD_018. A total of 280 gene clusters (14%) were observed to be unique to KD_018 showing the genome had the first record of *Hydrogenobaculum* in Indonesia and was considered a candidate for a new species.

In the pangenomic analysis of *Sulfurisphaera*, three reference genomes, including GCF_000011205.1: *Sulfurisphaera tokodaii*, GCF_009729055.1: *Sulfurisphaera ohwakuensis*, and GCF_041154675.1: *Sulfurisphaera javensis* were compared with MAG KD_012. The results showed 1376 gene clusters (37%) classified as Core across the four genomes and 1217 (33%) as Accessory, with most shared by the three reference genomes, while KD_012 shared similarities only with GCF_041154675.1 *Sulfurisphaera javensis*. For Singleton, there were 1135 gene clusters (30%), mostly unique to GCF_000011205.1, GCF_009729055.1, and GCF_041154675.1, with only a few in KD_012 as presented in Figure 5.B. These results showed that the entire MAG KD_012 genome was represented in the references but most similar to GCF_041154675.1 *Sulfurisphaera javensis* based on the ANI value.

MAG KD_010 was compared with seven species of *Metallosphaera* as the reference, and the results presented in Figure 5.C showed 1447 gene clusters (32.4%) as Core across all genomes, 1298 (29.1%) as Accessory, and 1722 (38.5%). It is important to state that a high number of singletons shows the existence of significant diversity in gene clusters among the genomes. It was further observed that KD_010 had similarity with GCF_022064045.1 *Metallosphaera javensis* based on ANI values, and both were closely related to clusters from GCF_005222525.1 *Metallosphaera prunae* and GCF_000016605.1 *Metallosphaera sedula* but most distantly related to GCF_000243315.1 *Metallosphaera yellowstonensis*.

MAG KD_009 *Caldisphaera* was compared with two genomes in RefSeq NCBI, including GCF_000317795.1 *Caldisphaera lagunensis* from the Philippines and GCF_023256325.1 *Caldisphaera* sp. from Russia. The results in Figure 6.A showed that 1189 gene clusters (62.2%) were classified as Core, 212 (11.1%) as Accessory, and 510 (26.7%) as Singleton. Most of the Accessory gene clusters were shared by GCF_000317795.1 *Caldisphaera lagunensis* and GCF_023256325.1 *Caldisphaera* sp., ensuring the two reference genomes were more closely related compared to MAG KD_009. Moreover, the distribution of the 510 Singleton gene clusters was even across the genomes, suggesting the three compared had unique gene clusters distributed evenly.

Table 4. Reconstructed High-Quality MAGs from Domas crater samples

MAGs ID	Taxonomic classification	Completeness (%)	Contamination (%)	Num of contig	Genome size (Mbp)	GC (%)	Mean coverage (x)	Relative abundance (%)	Accession number
KD_001	*Caldivirga* sp.	99.26	1.16	78	1.94	49.95	82	1.03	SAMN43335211
KD_005	*Conexivisphaerales archaeon*	90.29	0.97	25	1.18	55.84	232	1.89	SAMN43335212
KD_009	*Caldisphaera* sp.	94.42	1.27	187	1.42	45.54	82	0.72	SAMN43335213
KD_010	*Metallosphaera javensis*	98.81	0.6	160	2.27	49.04	73	1.09	SAMN43335214
KD_012	*Sulfurisphaera* sp.	92.86	1.49	136	1.43	35.21	1875	17.74	SAMN43335215
KD_018	*Hydrogenobaculum* sp.	96.34	0.41	74	1.56	34.04	1867	19.99	SAMN43335216
KD_023	*Conexivisphaeraceae archaeon*	95.15	3.16	140	1.02	55.49	81	0.52	SAMN43335217
KD_024	*Thermocladium* sp.	90.84	1.47	242	1.44	52.47	59	0.54	SAMN43335218

Table 5. The reference genomes from NCBI RefSeq used for Pangenome analysis

MAGs	Accession number	Name	Location	pH	Temp	submitter
	GCF_000018305.1	*Caldivirga maquilingensis* IC-167	USA			Joint Genome Institute, U.S. Department of Energy
Caldivirga	GCF_001663375.1	*Caldivirga* sp. MU80	USA: Yellowstone National Park, WY			University of Cincinnati
	GCF_023256255.1	*Caldivirga* sp. KMA_Bin28	Russia: Kamchatka, Mutnovsky	3.5-4	70℃	University of Goettingen
	GCF_002506515.1	*Caldivirga* sp. UBA161	Taiwan: Shi-Huang-Ping hot spring			University of Queensland
Caldisphaera	GCF_000317795.1	*Caldisphaera lagunensis* DSM 15908	Phillipines		75℃	Joint Genome Institute
	GCF_023256325.1	*Caldisphaera* sp. KMA_Bin23	Russia: Kamchatka, Mutnovsky	3.5-4	70℃	University of Goettingen
	GCF_000016605.1	*Metallosphaera sedula* DSM 5348	USA			Joint Genome Institute
	GCF_000204925.1	*Metallosphaera cuprina* Ar-4	China: Yunnan province			Environmental Microbiology Research Center (EMRC)
	GCF_003201675.2	*Metallosphaera hakonensis* DSM 7519	Japan: Hakone National Park, Kanagawa			North Carolina State University
Metallosphaera	GCF_005222525.1	*Metallosphaera prunae* Ron 12/II	Germany:Ronneburg			North Carolina State University
	GCF_013343295.1	*Metallosphaera tengchongensis* Ric-A	China: Yunnan, Tengchong hot spring	2.5	65℃	Institute of microbiology
	GCF_000243315.1	*Metallosphaera yellowstonensis* MK1	Yellowstone National Park, USA	2-3"	65℃	Joint Genome Institute
	GCF_022064045.1	*Metallosphaera javensis* J1 (ex Hofmann	Indonesia			TU Bergakademie Freiberg
	GCF_000011205.1	*Sulfurisphaera tokodaii* str. 7	Japan:Oita, Beppu		75℃	Biological Resource Center, NITE (NBRC)
Sulfurisphaera	GCF_009729055.1	*Sulfurisphaera ohwakuensis* TA-1	Japan: Ohwaku Valley, Hakone			North Carolina State University
	GCF_041154675.1	*Sulfurisphaera javensis* KD-1	Indonesia			Soka University
	GCF_000020785.1	*Hydrogenobaculum* sp. Y04AAS1	Yellowstone National Park			Joint Genome Institute
	GCF_000341855.1	*Hydrogenobaculum* sp. HO	Yellowstone National Park		58℃	Joint Genome Institute
Hydrogenobaculum	GCF_000348765.2	*Hydrogenobaculum* sp. SN	Yellowstone National Park		58℃	Joint Genome Institute
	GCF_000213785.1	*Hydrogenobaculum* sp. 3684	USA: Yellowstone National Park			Joint Genome Institute
	GCF_000215065.1	*Hydrogenobaculum* sp. SHO	Yellowstone National Park			Joint Genome Institute
Thermocladium	GCF_014646535.1	*Thermocladium modestius* JCM 10088	Japan			The University of Tokyo
Conexivisphaerales	GCF_013340765.1	*Conexivisphaera calida* NAS-02	Japan:Tochigi, Oku-Shiobara, Arayu			Japan Collection of Microorganisms

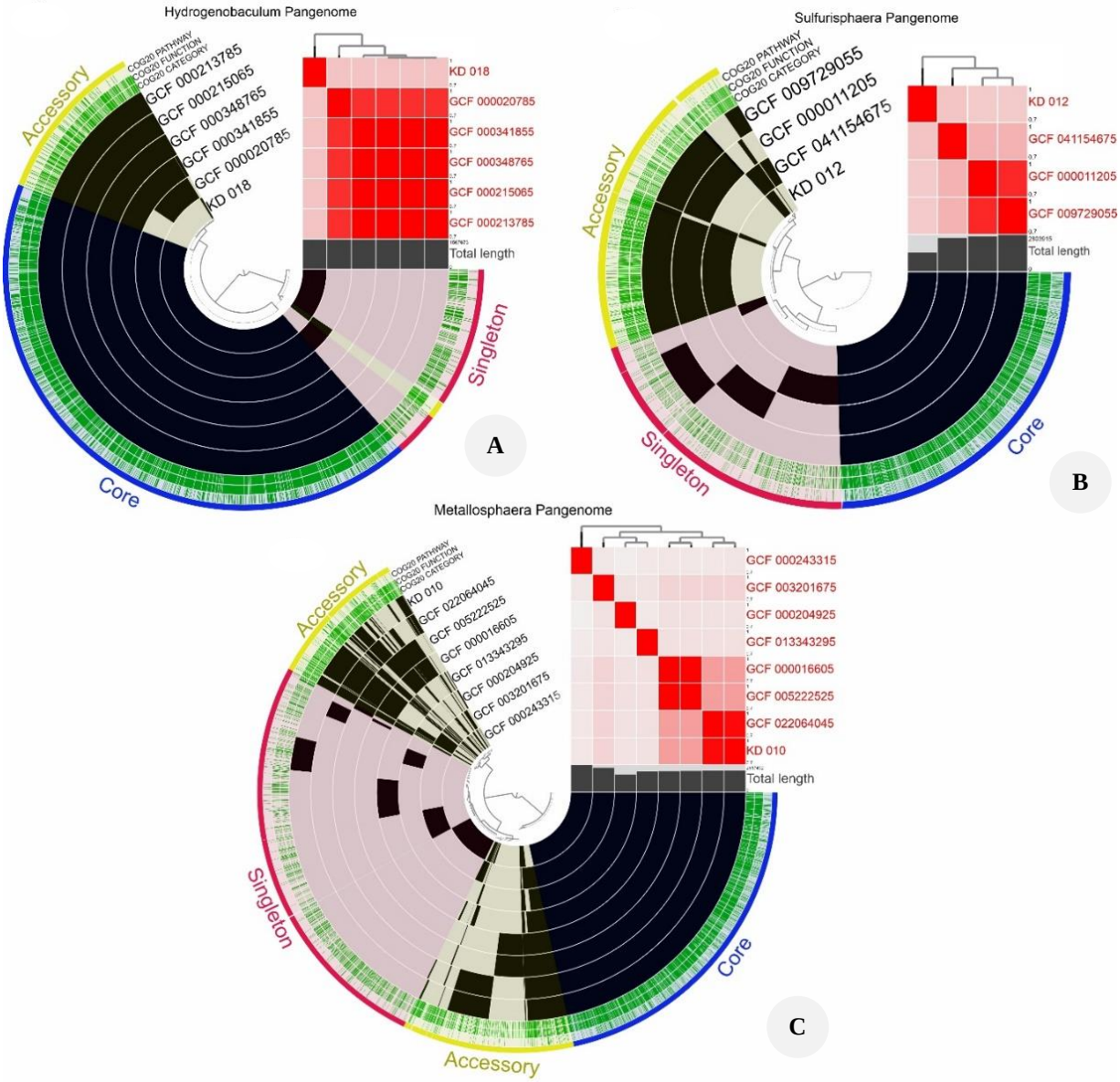

Figure 5. Pangenome analysis of A. MAGs KD_018; B. MAGs KD_012; and C. MAGs KD_010 with relevant Refseq public genomes constructed using Anvi'o. The figure shows the clustering of the genomes based on the presence and absence of gene clusters, categorized into Core, Accessory, and Singleton. The circle phylogram presents information arranged from the inner to outer regions with a focus on the presence or absence of gene clusters in each genome, COG classification based on function and pathway, as well as the dendrogram and red heatmap representing the Average Nucleotide Identity (ANI) percentage identity among the genomes, with red shading showing higher ANI value

For the *Caldivirga*, pangenomic analysis was conducted with four reference genomes, and the results in Figure 6.B showed that 1208 gene clusters (34.44%) were classified as Core, 790 (22.52%) as Accessory, and 1510 (43.04%) as Singleton. MAG KD_001 had the fewest cores compared to the references, showing the genome was the least similar but observed to be most closely related to GCF_001663375.1 *Caldivirga* sp. followed by GCF_000018305.1 *Caldivirga maquilingensis*, both obtained from the USA. Meanwhile, GCF_000018305.1 *Caldivirga maquilingensis* showed closer similarity to GCF_002506515.1 and GCF_023256255.1 from Taiwan and Russia, respectively.

The pangenomic analysis of MAG KD_024 *Thermocladium* was compared with one genome, GCF_014646535.1 *Thermocladium modestius* which was the only reference found in RefSeq NCBI. The results in Figure 6.C showed that 1350 gene clusters (62.13%) were classified as Core and shared between both genomes, while 823 (37.87%) were Singleton, with 610 identified in GCF_014646535.1 *Thermocladium modestius* and 213 in MAG KD_024. *Thermocladium* had few records and was only found in Australia, the USA, the Philippines, and Japan, while MAG KD_024 was the first identified in Indonesia.

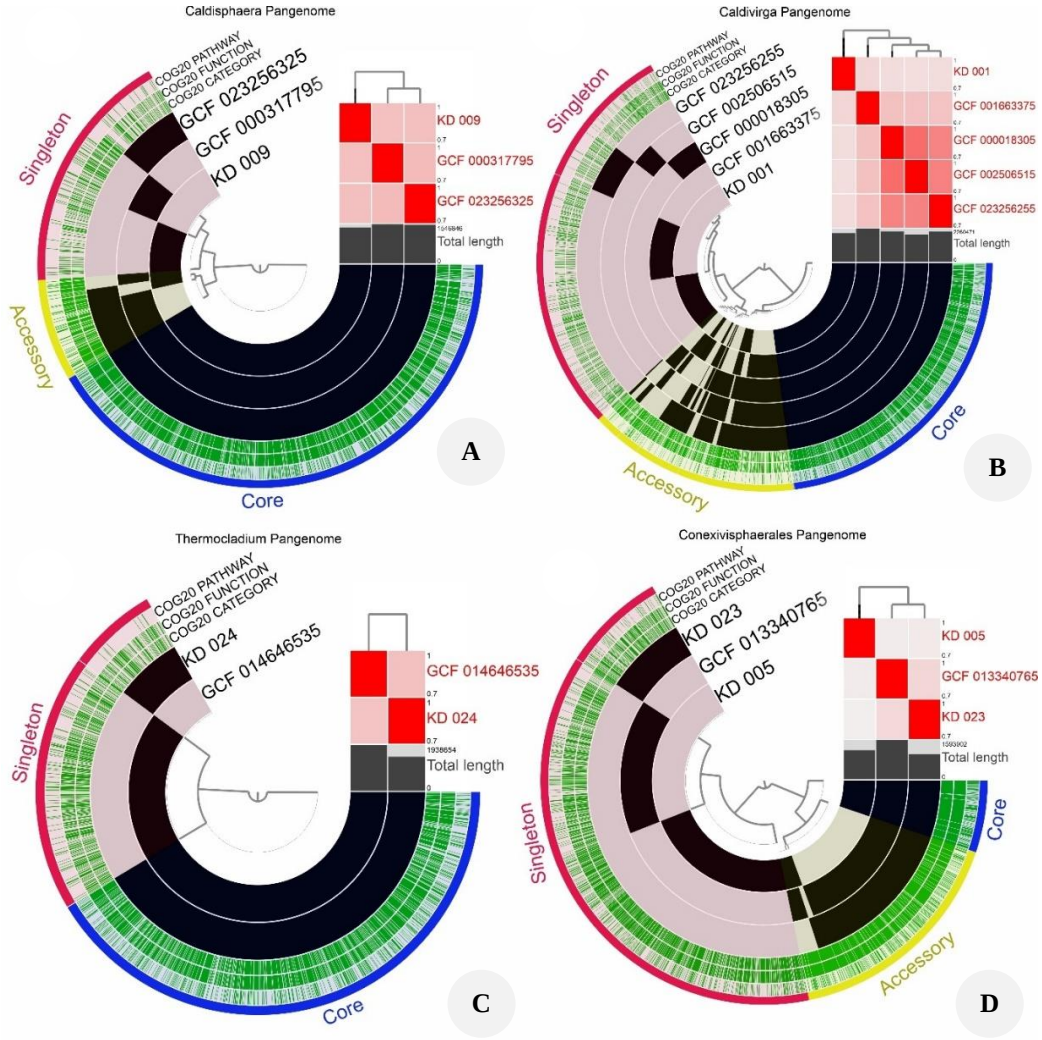

Figure 6. Pangenome analysis of A. MAGs KD_009; B. MAGs KD_001; C. MAGs KD_024; D. MAGs KD_005 & KD_023 with relevant Refseq public genomes, constructed using Anvi'o. The figure shows the clustering of the genomes based on the presence and absence of gene clusters, categorized into Core, Accessory, and Singleton. The circle phylogram presents information arranged from the inner to outer regions with a focus on the presence or absence of gene clusters in each genome, COG classification based on function and pathway, as well as the dendrogram and red heatmap representing the ANI percentage identity among the genomes, with red shading showing higher ANI value

MAGs KD_005 and KD_023 were classified as Conexivisphaerales archaeon (order level) and Conexivisphaeraceae archaeon (family level), respectively. The analysis showed that MAG KD_023 had more similarity with GCF_013340765.1 categorized in Accessory at approximately ±24.8%. Meanwhile, only 217 gene clusters (7.7%) were classified as Core for all three genomes in Figure 6.D. This result was expected because the comparison was only within the same order group, Conexivisphaerales, which caused the uniqueness or Singleton of most gene clusters to each genome. The trend was observed from the classification of 1,877 clusters (67.3%) as Singleton, which were distributed evenly among the genomes.

The pangenome analysis of the microbial communities from Domas Crater revealed significant differences among the individual MAGs, shedding light on the evolutionary adaptations that enable survival in this extreme, acidic environment. For instance, core gene clusters shared among the MAGs highlight essential metabolic pathways for sulfur and nitrogen cycling, critical for thriving in a sulfur-rich, high-temperature ecosystem. However, accessory and singleton genes show substantial variation across the MAGs, suggesting that specific adaptations have occurred independently in different microbial lineages, such as the ability to tolerate high acidity or utilize unique carbon sources. For example, MAGs from the genus *Sulfurisphaera* exhibit a higher number of genes associated with sulfur oxidation. At the same time, Hydrogenobaculum MAGs are enriched in genes related to hydrogen metabolism, reflecting their distinct ecological roles in the crater (Figure 4). These differences not only highlight the metabolic flexibility and niche specialization of microbes in Domas Crater but also suggest evolutionary pathways that might contribute to their survival under extreme environmental pressures.

While our study primarily focused on the unique microbial diversity within the Domas Crater, it also highlights several non-novel yet crucial insights into microbial ecology and functional potential in extreme environments. For instance, the discovery of several taxa aligns with findings from other geothermal habitats, reinforcing the adaptability and evolutionary convergence of microbial communities in high-temperature, low-nutrient ecosystems. These observations contribute to our understanding of bioprospecting and biotechnological applications, such as enzyme discovery for industrial processes that require stability under extreme conditions. Additionally, the methodological approach we employed, combining metagenomics and pangenomics, could serve as a framework for future studies in similar settings, providing a replicable strategy to unveil both common and unique microbial dynamics. In the future, continued exploration of these environments may expand our knowledge of microbial diversity and foster the development of novel biotechnologies rooted in extremophilic capabilities.

In conclusion, shotgun metagenomics was an effective method to show the microbial diversity and the potential of functional genes in biotechnology. This was based on the method's ability to successfully identify both previously reported and newly discovered microbes. Moreover, eight MAGs were successfully reconstructed and published in the NCBI repository to ensure availability for further functional analysis. *Hydrogenobaculum* sp. and *Sulfurisphaera* sp. were found to be the most abundant microbes according to both Kaiju analysis and the remapping of reads from the MAGs obtained. The functional analysis of the eight MAGs showed that *Caldivirga* sp. had a high abundance of genes participating in CAZyme. Meanwhile, *Hydrogenobaculum* sp., *M. javensis*, and *Sulfurisphaera* sp. contained BGCs, and approximately all MAGs had genes in sulfur and hydrogen metabolism. Pangenome analysis also showed that each MAG had unique gene clusters serving as differentiators from other previously recorded species.

ACKNOWLEDGEMENTS

All shotgun metagenomics and high-quality MAGs have been deposited in NCBI under BioProject accession number: PRJNA1117210.

REFERENCES

Almando AG, Wanguyun AP, Harianto S. 2019. Bioprospecting thermostable enzymes-producing thermophiles from Indonesia. Ecol Environ Conserv 25 (2019): S75-S79.

Alneberg J, Bjarnason BS, De Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. Nat Methods 11 (11): 1144-1146. DOI: 10.1038/nmeth.3103.

Blin K, Shaw S, Augustijn HE, Reitz ZL, Biermann F, Alanjary M, Fetter A, Terlouw BR, Metcalf WW, Helfrich EJN, Van Wezel GP, Medema MH, Weber T. 2023. AntiSMASH 7.0: New and improved predictions for detection, regulation, chemical structures and visualisation. Nucleic Acids Res 51 (W1): 46-50. DOI: 10.1093/nar/gkad344.

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35 (8): 725-731. DOI: 10.1038/nbt.3893.

Canovai A, Williams PA. 2024. Pyrroloquinoline quinone: A potential neuroprotective compound for neurodegenerative diseases targeting metabolism. Neural Regen Res 20 (1): 41-53. DOI: 10.4103/nrr.nrr-d-23-01921.

Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The carbohydrate-active enzymes database (CAZy): An expert resource for glycogenomics. Nucleic Acids Res 37 (Suppl 1): 233-238. DOI: 10.1093/nar/gkn663.

Chen S. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta 2 (2): 107. DOI: 10.1002/imt2.107.

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, Zhang X, Wang J, Yang H, Fang L, Chen Q. 2018. SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience 7 (1): 1-6. DOI: 10.1093/gigascience/gix120.

Chettri D, Verma AK, Verma AK. 2020. Innovations in CAZyme gene diversity and its modification for biorefinery applications. Biotechnol Rep (Amst) 28: e00525. DOI: 10.1016/j.btre.2020.e00525.

Chivian D, Jungbluth SP, Dehal PS, Wood-Charlson EM, Canon RS, Allen BH, Clark MM, Gu T, Land ML, Price GA, Riehl WJ, Sneddon MW, Sutormin R, Zhang Q, Cottingham RW, Henry CS, Arkin AP. 2023. Metagenome-assembled genome extraction and analysis from microbiomes using KBase. Nat Protoc 18 (1): 208-238. DOI: 10.1038/s41596-022-00747-x.

Cobucci-Ponzano B, Aurilia V, Riccio G, Henrissat B, Coutinho PM, Strazzulli A, Padula A, Corsaro MM, Pieretti G, Pocsfalvi G, Fiume I, Cannio R, Rossi M, Moracci M. 2010. A new archaeal β-glycosidase from *Sulfolobus solfataricus*. J Biol Chem 285 (27): 691-703. DOI: 10.1074/jbc.m109.086470.

Delmont TO, Eren EM. 2018. Linking pangenomes and metagenomes: The Prochlorococcus metapangenome. PeerJ 2018: 4320. DOI: 10.7717/peerj.4320.

Donahoe-Christiansen J, D'Imperio S, Jackson CR, Inskeep WP, McDermott TR. 2004. Arsenite-oxidizing hydrogenobaculum strain isolated from an acid-sulfate-chloride geothermal spring in Yellowstone National Park. Appl Environ Microbiol 70 (3): 1865-1868. DOI: 10.1128/AEM.70.3.1865-1868.2004.

Drula E, Garron ML, Dogan S, Lombard V, Henrissat B, Terrapon N. 2022. The carbohydrate-active enzyme database: Functions and literature. Nucleic Acids Res 50: 571-577. DOI: 10.1093/nar/gkab1045.

Elling FJ, Becker KW, Könneke M, Schröder JM, Kellermann MY, Thomm M, Hinrichs KU. 2016. Respiratory quinones in Archaea: Phylogenetic distribution and application as biomarkers in the marine environment. Environ Microbiol 18 (2): 692-707. DOI: 10.1111/1462-2920.13086.

Eren AM, Kiefl E, Shaiber A et al. 2021. Community-led, integrated, reproducible multi-omics with anvi'o. Nat Microbiol 6 (1): 3-6. DOI: 10.1038/s41564-020-00834-3.

Galal A, Abou ES, Saleh AH, Ahmed AI, Abdelrahman MM, Kamal MM, Khalel RS, Ziko L. 2023. A survey of the biosynthetic potential and specialized metabolites of archaea and understudied bacteria. Curr Res Biotechnol 5: 2590-2628. DOI: 10.1016/j.crbiot.2022.11.004.

Golchha NC, Nighojkar A, Nighojkar S. 2024. Bacterial pangenome: A review on the current strategies, tools and applications. Medinformatics. DOI: 10.47852/bonviewMEDIN42022496.

Gomes RP, Oliveira TR, Gama AR, Alves KR, Santos RK, Vieira JDG, Pires DJ, Carneiro LC. 2022. Phenotypic plasticity and characterization of *Chromobacterium* isolates from aquatic environment. Res Soc Dev 11: 15. DOI: 10.33448/rsd-v11i15.36821.

Handayani S, Santoso I, Freisleben HJ, Huber H, Andi, Ardiansyah F, Mulyanto C, Luthfa Z, Saleh R, Freisleben SKU, Wanandi SI, Thomm, M. 2012. Archaeal life on Tangkuban Perahu-sampling and culture growth in Indonesian laboratories. Hayati J Biosci 19: 150-154. DOI: 10.4308/hjb.19.3.150. DOI: 10.4308/hjb.19.3.150.

Huber G, Huber R, Jones BE, Lauerer G, Neuner A, Segerer A, Stetter KO, Degens ET. 1991. Hyperthermophilic archaea and bacteria occurring within indonesian hydrothermal areas. Syst Appl Microbiol 14: 397-404. DOI: 10.1016/S0723-2020(11)80317-6.

Jay ZJ, Beam JP, Kozubal MA, Jennings R de M, Rusch DB, Inskeep WP. 2016. The distribution, diversity and function of predominant Thermoproteales in high-temperature environments of Yellowstone National Park. Environ Microbiol 18: 4755-4769. DOI: 10.1111/1462-2920.13366.

Kamble A, Sawant S, Singh H. 2020. 16S ribosomal RNA gene-based metagenomics: A review. Biomed Res J 7: 5. DOI: 10.4103/bmrj.bmrj_4_20.

Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 7: e7359. DOI: 10.7717/peerj.7359.

Kangale LJ, Raoult D, Ghigo E, Fournier PE. 2020. *Pedobacter schmidteae* sp. nov., a new bacterium isolated from the microbiota of the planarian *Schmidtea mediterranea*. Sci Rep 10 (1): 6113. DOI: 10.1038/s41598-020-62985-x.

Karaoz U, Brodie EL. 2022. microTrait: A toolset for a trait-based representation of microbial genomes. Front Bioinforma 2: 918853. DOI: 10.3389/fbinf.2022.918853.

Kato S, Masuda S, Shibata A, Shirasu K, Ohkuma M. 2022. Insights into ecological roles of uncultivated bacteria in Katase hot spring sediment from long-read metagenomics. Front Microbiol 13: 1045931. DOI: 10.3389/fmicb.2022.1045931.

Ladoukakis E, Kolisis FN, Chatziioannou AA. 2014. Integrative workflows for metagenomic analysis. Front Cell Dev Biol 2: 11. DOI: 10.3389/fcell.2014.00070.

Langmead B, Salzberg S. 2013. Fast gapped-read alignment with Bowtie 2. Nat Method 9 (4): 357-359. DOI: 10.1038/nmeth.1923.

Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. 2013. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnol Biofuels 6: 41. DOI: 10.1186/1754-6834-6-41.

Li D, Luo R, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. 2016. MEGAHIT v1.0: A fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102: 3-11. DOI: 10.1016/j.ymeth.2016.02.020.

Lischer K, Putra ABRD, Guslianto BW, Avilla F, Sitorus SG, Nugraha Y, Sarmoko. 2020. Short communication: The emergence and rise of indigenous thermophilic bacteria exploration from hot springs in Indonesia. Biodiversitas 21: 5474-5481. DOI: 10.13057/biodiv/d211156.

Liu LJ, Jiang Z, Wang P, Qin YL, Xu W, Wang Y, Liu SJ, Jiang CY. 2021. Physiology, taxonomy, and sulfur metabolism of the sulfolobales, an order of thermoacidophilic archaea. Front Microbiol 12: 768283. DOI: 10.3389/fmicb.2021.768283.

López-López O, Knapik K, Cerdán ME, González-Siso MI. 2015. Metagenomics of an alkaline hot spring in Galicia (Spain): Microbial diversity analysis and screening for novel lipolytic enzymes. Front Microbiol 6: 1291. DOI: 10.3389/fmicb.2015.01291.

Mashzhan A, Javier-López R, Kistaubayeva A, Savitskaya I, Birkeland NK. 2021. Metagenomics and culture-based diversity analysis of the bacterial community in the zharkent geothermal spring in Kazakhstan. Curr Microbiol 78 (8): 2926-2934. DOI: 10.1007/s00284-021-02545-2. 10.1007/s00284-021-02545-2.

Meng DD, Ying Y, Zhang K Di, Lu M, Li FL. 2015. Depiction of carbohydrate-active enzyme diversity in *Caldicellulosiruptor* sp. F32 at the genome level reveals insights into distinct polysaccharide degradation features. Mol Biosyst 11 (11): 64-73. DOI: 10.1039/c5mb00409h.

Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. Nat Commun 7 (1): 11257. DOI: 10.1038/ncomms11257.

Mirete S, Morgante V, González-Pastor JE. 2016. Functional metagenomics of extreme environments. Curr Opin Biotechnol 38: 143-149. DOI: 10.1016/j.copbio.2016.01.017.

New FN, Brito IL. 2020. What is metagenomics teaching us, and what is missed?. Annu Rev Microbiol 74: 117-135. DOI: 10.1146/annurev-micro-012520-072314.

Nguyen KT, Nguyen NL, Tung N Van, Nguyen HH, Milhim M, Le TTX, Lai THN, Phan TTM, Bernhardt R. 2020. A novel thermostable cytochrome p450 from sequence-based metagenomics of binh chau hot spring as a promising catalyst for testosterone conversion. Catalysts 10 (9): 1083. DOI: 10.3390/catal10091083.

Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: Advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32: 292-294. DOI: 10.1093/bioinformatics/btv566.

Papudeshi B, Haggerty JM, Doane M, Morris MM, Walsh K, Beattie DT, Pande D, Zaeri P, Silva GGZ, Thompson F, Edward RA, Dinsdale EA. 2017. Optimizing and evaluating the reconstruction of metagenome-assembled microbial genomes. BMC Genom 18: 915. DOI: 10.1186/s12864-017-4294-1.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25 (7): 1043-1055. DOI: 10.1101/gr.186072.114

Prokofeva MI, Miroshnichenko ML, Kostrikina NA, Chernyh NA, Kuznetsov BB, Tourova TP, Bonch-Osmolovskaya EA. 2000. *Acidilobus aceticus* gen. nov., s.p. nov., a novel anaerobic thermoacidophilic archaeon from continental hot vents in Kamchatka. Intl J Syst Evol Microbiol 50 (6): 2001-2008. DOI: 10.1099/00207713-50-6-2001.

Quince C, Walker AW, Simpson JT, Loman NJ, Segata N. 2017. Shotgun metagenomics, from sampling to sequencing and analysis. Nat Biotechnol 35 (9): 833-844. DOI: 10.1038/nbt.3935.

Rasool G, Irfan M. 2024. The role of microbial diversity in lignocellulosic biomass degradation: A biotechnological perspective. ChemBioEng Rev 11: 613-635. DOI: 10.1002/cben.202300073.

Reichart NJ, Bowers RM, Woyke T, Hatzenpichler R. 2021. High potential for biomass-degrading enzymes revealed by hot spring metagenomics. Front Microbiol 12: 668238. DOI: 10.3389/fmicb.2021.668238.

Safitri R, Kusumawardhani DP, Annisa, Partasasmita R, Asharina S, Maskoen AM. 2020. Characterization and identification of three thermophilic *Bacillus* strain isolated from domas crater, Mt. Tangkuban Perahu, Indonesia. Biodiversitas 21 (8): 3444-3453. DOI: 10.13057/biodiv/d210805.

Sakai HD, Nur N, Kato S, Yuki M, Shimizu M, Itoh T, Ohkuma M, Suwanto A, Kurosawa N. 2022. Insight into the symbiotic lifestyle of DPANN archaea revealed by cultivation and genome analyses. Proc Natl Acad Sci 119: e2115449119. DOI: 10.1073/pnas.2115449119.

Saryono, Novianty R, Suraya N, Piska F, Devi S, Pratiwi NW, Ardhi A. 2022. Molecular identification of cellulase-producing thermophilic fungi isolated from Sungai Pinang hot spring, Riau Province, Indonesia. Biodiversitas 23: 1457-1465. DOI: 10.13057/biodiv/d230333.

Sastroredjo S, Almira S, Akhmaloka A. 2023. Exploration of gene encoded thermostable enzymes by using random pcr from natural sample of Domas Crater. J Sci Informatics Soc 1 (1): 19-26. DOI: 10.57102/jsis.v1i1.23.

Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30: 2068-2069. DOI: 10.1093/bioinformatics/btu153.

Setubal JC. 2021. Metagenome-assembled genomes: Concepts, analogies, and challenges. Biophys Rev 13 (6): 905-909. DOI: 10.1007/s12551-021-00865-y.

Sharma N, Kumar J, Abedin MM, Sahoo D, Pandey A, Rai AK, Singh SP. 2020. Metagenomics revealing molecular profiling of community structure and metabolic pathways in natural hot springs of the Sikkim Himalaya. BMC Microbiol 20: 246. DOI: 10.1186/s12866-020-01923-3.

Shi Q, Abdel-Hamid AM, Sun Z, Cheng Y, Tu T, Cann I, Yao B, Zhu W. 2023. Carbohydrate-binding modules facilitate the enzymatic hydrolysis of lignocellulosic biomass: Releasing reducing sugars and dissociative lignin available for producing biofuels and chemicals. Biotechnol Adv 65: 108-126. DOI: 10.1016/j.biotechadv.2023.108126.

Sidar A, Albuquerque ED, Voshol GP, Ram AFJ, Vijgenboom E, Punt PJ. 2020. Carbohydrate binding modules: Diversity of domain architecture in amylases and cellulases from filamentous microorganisms. Front Bioeng Biotechnol 8: 871. DOI: 10.3389/fbioe.2020.00871.

Sieber CMK., Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nat Microbiol 3 (7): 836-843. DOI: 10.1038/s41564-018-0171-1.

Stöhr R, Waberski A, Völker H, Tindall BJ, Thomm M. 2001. *Hydrogenothermus marinus* gen. nov., sp. nov., a novel thermophilic hydrogen-oxidizing bacterium, recognition of *Calderobacterium hydrogenophilum* as a member of the genus *Hydrogenobacter* and proposal of the reclassification of *Hydrogenobacter acidophilus* as *Hydrogenobaculum acidophilum* gen. nov., comb. nov., in the phylum "*Hydrogenobacter*/*Aquifex*.". Intl J Syst Evol Microbiol 51 (5): 1853-1862. DOI: 10.1099/00207713-51-5-1853.

Strazzulli A, Fusco S, Cobucci-Ponzano B, Moracci M, Contursi P. 2017. Metagenomics of microbial and viral life in terrestrial geothermal environments. Rev Environ Sci Biotechnol 16: 425-454. DOI: 10.1007/s11157-017-9435-0.

Sysoev M, Grötzinger SW, Renn D, Eppinger J, Rueping M, Karan R. 2021. Bioprospecting of novel extremozymes from prokaryotes-the advent of culture-independent methods. Front Microbiol 12: 630013. DOI: 10.3389/fmicb.2021.630013.

Tsuboi K, Sakai HD, Nur N, Stedman KM, Kurosawa N, Suwanto A. 2018. *Sulfurisphaera javensis* sp. nov., a hyperthermophilic and acidophilic archaeon isolated from indonesian hot spring, and reclassification of *Sulfolobus tokodaii* Suzuki et al. 2002 as *Sulfurisphaera tokodaii* comb. nov. Intl J Syst Evol Microbiol 68 (6): 1907-1913. DOI: 10.1099/ijsem.0.002765.

Tuyet NLT, Kim J. 2024. Genome-based classification of *Pedobacter albus* sp. nov. and *Pedobacter flavus* sp. nov. isolated from soil. Diversity 16 (5): 292. DOI: 10.3390/d16050292.

Wang P, Li LZ, Qin YL, Liang ZL, Li XT, Yin HQ, Liu LJ, Liu SJ, Jiang CY. 2020. Comparative genomic analysis reveals the metabolism and evolution of the thermophilic archaeal genus *Metallosphaera*. Front Microbiol 11: 1192. DOI: 10.3389/fmicb.2020.01192.

Wu YW, Simmons BA, Singer SW. 2016. MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 32: 605-607. DOI: 10.1093/bioinformatics/btv638.

Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. 2023. DbCAN3: Automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res 51 (W1): 115-121. DOI: 10.1093/nar/gkad328.