Vietnam Journal of Biotechnology **20**(1): 151-161, 2022

ANALYSIS OF MICROBIAL COMMUNITIES IN BINH CHAU HOT SPRING THROUGH METAGENOMIC DNA SEQUENCING

Tran Thanh Thuy[1], Nguyen Ngoc Lan[2,3], Nguyen Van Tung[2], Tran Dinh Man[1], Nguyen Huy Hoang[2,3], Nguyen Kim Thoa[1,3,]✉

[1]Institute of Biotechnology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam
[2]Institute of Genome Research, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam
[3]Graduate University of Science and Technology, Vietnam Academy of Science and Technology, 18 Hoang Quoc Viet Road, Cau Giay District, Hanoi, Vietnam

✉To whom correspondence should be addressed. E-mail: nkthoa@ibt.ac.vn

Received: 18.10.2021
Accepted: 02.12.2021

SUMMARY

Binh Chau hot spring locates in Bung Rieng commune, Xuyen Moc district, Ba Ria - Vung Tau province, and is the second hottest hot spring in Vietnam. Up to date, the microbial diversity of this hot spring has been mainly evaluated using the culture-dependent methods, thus, it could not be able to assess all of microorganisms there. In this report, the microbial community of Binh Chau hot spring was investigated by meta-analysis using Hiseq Illumina for sequencing. The results revealed that only 106,903 genes were annotated and 49,190 genes were unknown among 156,093 potential ORFs. The annotated genes consisted of 29,069 bacterial genes, 1,416 archaeal genes, 4 eukaryotic genes, 3 viral genes, and 76,411 unclassified genes. Among the annotated and classified genes, the microbial community of Binh Chau hot spring was characterized by the predominance of bacteria over archaea. At the phylum level, the most abundant phyla in Binh Chau hot spring were followed as Proteobacteria > Firmicutes > Cyanobacteria > Bacteroidetes > Planctomycetes > Thaumarchaeota > Actinobacteria > Euryarchaeota. These belonged mainly to the thermophilic group. A number of the phyla are reported to grow at high temperature, alkaline pH, and high salt concentrations. In addition to the culturable species, the metagenomic data is also indicated the presence of unculturable groups, for example phylum Acidobacteria and Chloroflexi of bacteria. The order Thermoplasmata, which is classified to class Thermoplasmatales of the phylum Euryarchaeota, as well as ultra-small cell and symbiotic archaea including phyla Parvarchaeota, Nanohaloarchaeota, and Nanoarchaeota were detected. The results in this study indicated an important insight into the microbial communities of Binh Chau hot spring, which revealed a promising thermophilic gene sources for their exploitation and conservation in Vietnam.

Keywords: Archaea, Binh Chau hot spring, DNA metagenomic analysis, microbial communities, thermophilic bacteria.

INTRODUCTION

Hot springs are the heated groundwater systems which occur in many countries. Each hot spring is characterized by its temperature, pH, the chemical composition (Hussein *et al*., 2017) as well as the dominant microbial group (Uribe-Lorio *et al*., 2019). Prokaryotic community in hot springs are mostly represented by bacteria and archaea. The biodiversity studies in alkaline hot springs (pH 7 - 9.8) with a range of temperature from 50 to 110 °C showed Proteobacteria and Firmicutes as the dominant phyla (Chan *et al*., 2015; Kambura *et al*., 2016), whereas Proteobacteria and Thermi are presented in slightly acidic hot springs (pH 5 - 7), with a lower temperature (45 – 90 ºC) (Paul *et al*., 2016). On the other hand, Thaumarchaeota, Euryarchaeota, Crenarchaeota are archaeal phyla that were found in most of hot springs (Chan *et al*., 2015; Kambura *et al*., 2016; Paul *et al*., 2016). Nowadays, there is an increasing interest in studying microbial from hot springs because of their ability to produce various novel thermostable enzymes which have potential to application in industry and biotechnology. Furthermore, studies on microbials of hot spring allow us to clearly know about the origin and evolution of earliest life as hot spring microorganisms are considered to be the most similar to the microorganisms living on the primitive Earth (Debnath *et al*., 2019).

Vietnam has a hot spring system stretching from north to south with approximately 300 geysers. The most hot springs have a range of temperature from 30 to 105°C. Binh Chau hot spring locates in Bung Rieng commune, Xuyen Moc district, Ba Ria - Vung Tau province, which is known as the second hottest hot spring in Vietnam with the temperature of hot spot is around 82°C. This is potential sources for exploiting of thermophiles producing novel thermostable enzymes as well as biocatalytic compounds (Nguyen, Tran, 2018; Nguyen *et al*., 2020). Similar to other hot springs, challenges in implementing microbial consortia in Binh Chau hot spring have been found due to the naturally low-cell density, and/ or very difficulty in sampling. On the other hand, almost thermophiles are known as unculturable species, culture - dependent approach can not reveal a whole picture of microbial diversity in geothermal area such as Binh Chau hot spring. Although culture – dependent methods have been used popularly for evaluation of microbial community here till now, the data was mainly obtained from the spore forming species such as *Geobacillus* genus (Tran *et al*., 2018). The development of sequencing techniques provides many advantages to access the microbial community in different escosystems. Next generation sequencing (NGS) such as Illumina-Hiseq system NGS has been used successully to evaluate microbial community and/ or functional genes in famous hot springs, for example the thermophile consortia in Lobios hot spring, Spain (Knapik *et al*., 2019), and basaltic hot spring of Unkeshwar, Maharashtra, India (Mehetre *et al*., 2015). From point of views, in this paper we have evaluated biodiversity of Binh Chau hot spring via DNA metagenomic analysis using Illumina HiSeq platform which resulted in a new and more diverse picture of microbial communitites than that of many previous studies. Moreover, these data could be served for further investigation on the potential thermophilic enzymes from microorganisms for the interest of the usage in industrial biotechnology.

MATERIALS AND METHODS

Samples

Water and mud samples were collected from Binh Chau hot spring, Vung Tau, Vietnam in May 2015. The water sample was pumped and filtered by an ultrafiltration with pore size of membrane around 0.01 µm. Muds were collected from downstream of greyser with temperature ranging from 45 to 65°C, pH 7.7 - 9.0.

Extraction of metagenome DNA

DNA Metagenome of water and mud samples were extracted by using Kit PowerWater® Bead Tube and Kit PowerMax® Soil DNA Isolation (MO BIO), respectively. The muds were mixed (1:1, w/w) together before weighting 10 grams to extract DNA metagenome. The final DNA metagenome of Binh Chau hot spring was prepared by mixing the DNA of water and mud with the ratio of 1:1 (v/v).

Metagenome sequencing and assembly

Metagenome DNA of Binh Chau hot spring was sequenced by using Illumina HiSeq^{TM} platform system in BGI Co., Ltd (Hong Kong). Quality checking and pre-processing raw sequence data were performed by using FastQC and Trimmomatic (Bolger *et al*., 2014). Preprocessed reads were assembled using SOAPdenovo2 (Luo *et al*., 2012) by default parameter. The Bowtie2 tool (Langmead, Salzberg, 2012) was used to align the high quality reads with assembled contigs, thus, single or incorrect paired-end reads were removed. The size of optimal kmer and assembled result were chosen depending on both contig N50 and mapping rate. After the assembly process, only contigs no less than 500 bp were kept for further analysis.

Gene prediction and microbial diversity analysis

MetaGeneMark (version 2.10) was used to predict open reading frames (ORFs) based on assembled results (Zhu *et al*., 2010). Genes from different samples were combined and clustered using CD-Hit with sequence identity threshold 95% and alignment coverage threshold 90% (Li, Godzik, 2006). Analysis of nr BLAST output files was performed using the MEGAN software (version 4.6) (Huson *et al*., 2007). This software reads the results of a BLAST comparison as input and attempts to place each read on a node in the NCBI taxonomy using lowest common ancestor (LCA) algorithm with default parameters (minScore = 50; topPercent = 10; minSupport = 50). The total relative abundance of all taxonomy in one sample was 100. The species having their abundance less than 0.5 in every sample were classified into “others” in other ranks. The total relative abundance of all genes in one sample is 1. The biodiversity was estimated by using Shannon index (Ortiz-Burgos, 2016). In the Shannon index, a is the proportion (n/N) of individuals of one particular species found (n) divided by the total number of individuals found (N), ln is the natural log, Σ is the sum of the calculations, and s is the number of species.

$$\mathbf{H'} = -\sum_{i=1}^{s} a_i ln a_i$$

RESULTS AND DISCUSSION

De novo metagenome assembly and gene predicting

DNA metagenome of Binh Chau hot spring was sequenced that resulted in 10.4 Gb of raw data. Using FastQC software identified the raw reads consisting of 0.04% ambiguous bases, 0.41% adapter sequences, 6.97% low quality bases (Q < 20), and 92.58% clean reads. Trimmomatic software was used to purify data, leading a total of 93,999,534 clean reads (9.4 Gb clean data) were obtained. The ratio of clean data to raw data is 89.98%. The Q20 and Q30 of clean data reaches 94.98 and 87.97%, respectively. SOAPdenovo2 assembled 93,999,534 clean reads to create 51,346 contigs with a total length of 172,103,446 bp, in which the N50 and N90 contig lengths are 9,791 bp and 1,059 bp, respectively. The longest contig is 1,767,609 bp, whearas the shortest contig is 500 bp. The average length of contigs is 3,351 bp and the mapping rate is 65.12%. MetaGeneMark v2.10 was used to predict 156,093 potential ORFs from 51,346 assembled contigs.

Composition of microbial communities in Binh Chau hot spring

The nr annotation and MEGAN were used to taxonomic assignment for 156,093 predicted genes. The result showed that there were 106,903 (68.5%) annotated genes and 49,190 (31.5%) unknown genes. Summary of taxonomic statistic are shown in Table1.

Moreover, MEGAN analysis also showed a total of 106,903 annotated genes consisting of 29,069 bacterial genes, 1,416 archaeal genes, 4 eukaryotic genes, 3 viral genes and 76,411 unclassified genes. The Shannon index was 1.75 which was lower than that of alkaline hot springs Yumthang (H '= 2.10) and Jakrem (H' = 1.96) from Indo-Burma region, North-east India (Panda *et al*., 2016). Among 106,903 annotated genes, the most eight dominant phyla were found in Binh Chau hot spring as following: Proteobacteria > Firmicutes > Cyanobacteria > Bacteroidetes > Planctomycetes > Thaumarchaeota > Actinobacteria > Euryarchaeota. Of which Thaumarchaeota and Euryarchaeota are belonged to archaeal kingdom and the remaining six phyla are classified to bacterial kingdom. The microbial diversity of Binh Chau hot spring is different as compared to that of the alkaline hot springs Unkeshwar, Maharashtra, India which owns the dominant phyla as following: Actinobacteria > Verrucomicrobia > Bacteriodes > Deinococcus – Thermus > Firmicutes > Viruses (Mehetre *et al.*, 2015). However, microbial communities in Binh Chau hot spring are similar to other hot springs where the temperature often exceeds the limit of eukaryotic life (~60°C). The dominance of Proteobacteria and Firmicutes phyla in Binh Chau hot spring is likely to microbial features of some alkaline hot springs such as Sungai Klah hot spring in Malaysia (pH 7 - 9, temperature 50 - 110°C) (Chan *et al.*, 2015) and hot spring in Kenya (pH 9.2-9.8) (Kambura *et al.*, 2016), whereas the dominant phyla in slightly acidic hot spring are often Proteobacteria and Thermi (Paul *et al.*, 2016). Together with temperature, pH factor plays an important role in microbial community features.

Table 1. Taxonomic statistic of annotated genes from DNA metabase of Binh Chau hot spring.

Total Gene number	Annotated Gene number	Annotated Kingdom number	Annotated Phylum number	Annotated Class number	Annotated Order number	Annotated Family number	Annotated Genus number	Annotated Species number
156,093	106,903	5	41	57	128	245	825	2,250

Bacterial diversity

Analysis of Binh Chau hot spring metabase showed existence of 29 bacterial phyla. Proteobacteria (68.68%) is the predominant phylum followed by Firmicutes (8.22%), Cyanobacteria (4.53%), Bacteroidetes (3.52%), Planctomycetes (2.81%), Actinobacteria (2.43), and other phyla (relative abundance <2%) (Table 2). The six dominant phyla are of culturable species, except for Acidobacteria, Chloroflexi. At class level of taxonomy, the biggest representatives are Gammaproteobacteria (24.87%), Alphaproteobacteria (19.61%), Betaproteobacteria (16.42%), and Deltaproteobacteria (6,91%) (Table 3). Members of the phylum Proteobacteria have been reported as major population in the Amazonian geothermal spring in Peru previously (Paul *et al.*, 2016) and in the hot springs in India with moderate high or higher temperature (Saxena *et al.*, 2016). The high prevalence of Proteobacteria in Binh Chau hot spring can be explained by their adaption to the high temperature and slightly basal pH (pH=7.5) which are similar to those of hot springs in the studies of Saxena *et al* (2016). Nevertheless, Firmicutes and Actinobacteria are reported as the dominating classes in the hot spring in Malaysia (Chan *et al.*, 2015) and basaltic hot spring (Mehetre *et al.*, 2015), respectively. It suggests that the dominant bacterial populations in hot spring ecology can be variable.

On the other hand, Aquificae, Thermodesulfobacteria, and Thermotogae are the well-known hyperthermophilic phyla, which represented a low abundance with 0.44, 0.29, and 0.06%, respectively, in Binh Chau metabase (Table 2). The composition of these phyla is presented in Figure 1. The Aquificae phylum is reported to consists of the most thermophilic bacterial species such as *Aquifex pyrophilus* (Huber *et al.*, 1992), *Aquifex aeolicus* (Deckert *et.al.*, 1998), *Thermocrinis ruber* (Huber *et al.*, 1998), *Sulfurihydrogenibium yellowstonense* (Nakagawa *et al.*, 2005), *Desulfurobacterium* sp., and *Hydrogenivirga* sp. Moreover, two major representatives of the phylum Thermodesulfobacteria are *Thermodesulfatator* *atlanticus* (Alain *et al.*, 2010) and *Thermodesulfobacterium thermophilum* (Rozanova, Khudiakova, 1974), which were decribed as a thermophilic, chemolithoautotrophic, sulfate-reducing bacteria species. *Thermotoga maritima* represented as the unique extremely thermophilic bacteria growing up to 90°C (Huber *et al.*, 1986), which was a major species in the phylum Thermotaga. This analysis implied that Binh Chau hot spring might be a suitable habitat of chemolitoautotrophic thermophiles and hyperthermophiles.

Table 2. Analysis of bacterial phyla in Binh Chau hot spring.

Phylum	No of read	Reative abundance (%)	Phylum	No of read	Reative rebundance (%)
Proteobacteria	19,966	68.68	Poribacteria	85	0.29
Firmicutes	2,389	8.22	Thermodesulfobacteria	85	0.29
Cyanobacteria	1,318	4.53	Acidobacteria	83	0.29
Bacteroidetes	1,023	3.52	Synergistetes	59	0.20
Planctomycetes	816	2.81	Fusobacteria	43	0.15
Actinobacteria	707	2.43	Dictyoglomi	36	0.12
Verrucomicrobia	481	1.65	Marinimicrobia	36	0.12
Spirochaetes	375	1.29	Lentisphaerae	35	0.12
Chloroflexi	367	1.26	Chlorobi	33	0.11
Nitrospirae	244	0.84	Candidatus Saccharibacteria	24	0.08
Nitrospinae	184	0.63	Chlamydiae	19	0.07
Gemmatimonadetes	134	0.46	Thermotogae	17	0.06
Aquificae	128	0.44	Deferribacteres	5	0.02
Deinococcus-Thermus	108	0.37	Tenericutes	5	0.02
Acetothermia	107	0.37	Unclassified	158	0.54

Table 3. Bacterial phyla and classes analysis from DNA metabase of Binh Chau hot spring.

Phylum	Class	No of read	Reative abundance (%)
Proteobacteria	Gammaproteobacteria	7,230	24.87
Proteobacteria	Alphaproteobacteria	5,702	19.61
Proteobacteria	Betaproteobacteria	4,773	16.42
Proteobacteria	Deltaproteobacteria	2,010	6.91
Cyanobacteria	NA	1,205	4.15
Firmicutes	Bacilli	1,151	3.96
Firmicutes	Clostridia	878	3.02
Planctomycetes	Planctomycetia	816	2.81
Actinobacteria	Actinobacteria	656	2.26
Spirochaetes	Spirochaetia	375	1.29
Bacteroidetes	Cytophagia	340	1.17
Bacteroidetes	Flavobacteriia	329	1.13
Firmicutes	Negativicutes	327	1.12
Verrucomicrobia	Verrucomicrobiales	274	0.94
Nitrospirae	Nitrospira	244	0.84
Nitrospinae	Nitrospinia	184	0.63

Chloroflexi	Chloroflexia	178	0.61
Unclassified	NA	158	0.54
Bacteroidetes	Bacteroidia	153	0.53
Chloroflexi	Ktedonobacteria	150	0.52
Bacteroidetes	Sphingobacteriia	149	0.51
Proteobacteria	Zetaproteobacteria	144	0.50
Gemmatimonadetes	Gemmatimonadetes	134	0.46
Aquificae	Aquificae	128	0.44
Cyanobacteria	Gloeobacteria	112	0.39
Deinococcus-Thermus	Deinococci	108	0.37
Acetothermia	NA	107	0.37
Poribacteria	NA	85	0.29
Thermodesulfobacteria	Thermodesulfobacteria	85	0.29
Verrucomicrobia	Spartobacteria	84	0.29
Verrucomicrobia	NA	82	0.28
Proteobacteria	Epsilonproteobacteria	76	0.26
Acidobacteria	NA	63	0.22
Synergistetes	Synergistia	59	0.20
Bacteroidetes	NA	52	0.18
Fusobacteria	Fusobacteriia	43	0.15
Verrucomicrobia	Opitutae	41	0.14
Chloroflexi	NA	39	0.13
Actinobacteria	Rubrobacteria	36	0.12
Dictyoglomi	Dictyoglomia	36	0.12
Marinimicrobia	NA	36	0.12
Lentisphaerae	Lentisphaeria	35	0.12
Chlorobi	Chlorobia	33	0.11
Proteobacteria	NA	31	0.11
Firmicutes	Erysipelotrichia	26	0.09
Candidatus Saccharibacteria	NA	24	0.08
Acidobacteria	Acidobacteriia	20	0.07
Chlamydiae	Chlamydiia	19	0.07
Thermotogae	Thermotogae	17	0.06
Actinobacteria	Coriobacteriia	9	0.03
Firmicutes	NA	7	0.02
Actinobacteria	Thermoleophilia	6	0.02
Deferribacteres	Deferribacteres	5	0.02
Tenericutes	Mollicutes	5	0.02
Chloroflexi	Dehalococcoidia	1	0.00

NA: Non-annotation.

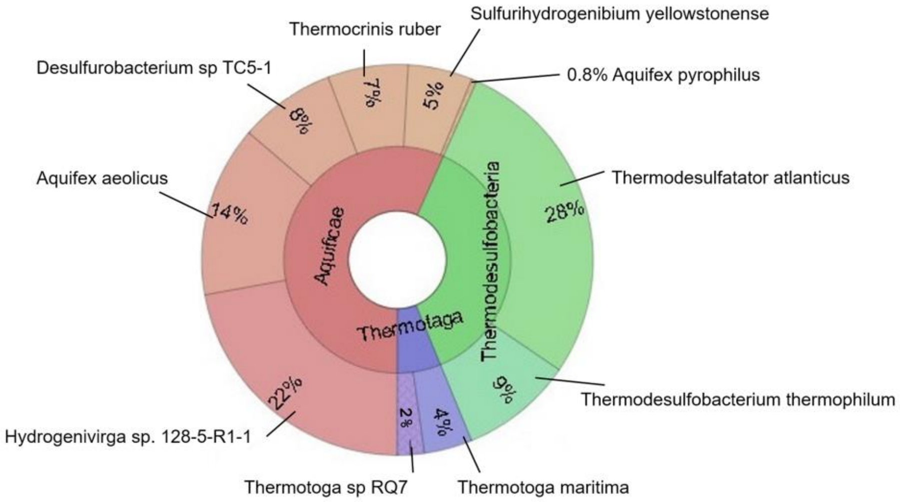

Figure 1. Krona chart representation of taxonomic classification of thermophilic bacteria in Binh Chau hot spring**.**

There were 106,903 genes from DNA metabase in Binh Chau hot spring which were annotated by using the Nr and MEGAN databases. However, a number of unclassified genes (78,487 genes), which was corresponded to 50.29% of annotated genes and 73,42% of predicted genes, prevents to evaluate the biodiversity at species taxonomy.

At the species taxonomy, *Marinobacter manganoxydans* (5.57%) was found to be the biggest representatives of Binh Chau hot spring followed *Roseobacter sp.* (4.81%), *Gamma proteobacterium* IMCC2047 (3.93%), *Thiobacillus thioparus* (1.11%), *Bdellovibrio bacteriovorus* (0.96%), *Thiobacillus denitrificans* (0.96%), *Caldimonas manganoxidans* (0.85%)) and other species (relative abundance < 0.85%). The most abundant species are known as the species isolated from marine. Almost species belonging to *Marinobacter* genus were found from various marine environments, for example, *Marinobacter manganoxydans* was isolated from a deep-sea hydrothermal vent in the Indian Ocean and capable of oxidizing manganese (Wang *et al*., 2012). Moreover, *Roseobacter* sp. CCS2 strain belonging to the *Roseobacter* genus was obtained from Pacific coastal waters and showed as an aerobic anoxygenic phototroph (AAnP) (Slightom, Buchan, 2009). *Gamma proteobacterium* IMCC2047 strain was isolated from a surface sea in the Yellow Sea, Korea (Kang *et al*., 2011). *Thiobacillus thioparus* species is a sulfur-oxidizing bacteria which can be isolated from marine microbial mat (Van den Ende, Gemerden, 1993). In fact Binh Chau hot spring locates very near to Ho Coc and Ho Tram coasts, therefore, it can develop the same evolution of microbial communities.

Archaeal diversity

Archaea is the second most abundant kingdom in Binh Chau hot spring. Analysis of 1,416 annotated genes, seven phyla including Thaumarchaeota (51.77%), Euryarchaeota (35.38%), Crenarchaeota (5.65%), Parvarchaeota (3.60%), Nanohaloarchaeota (1.77%), Korarchaeota (1.48%) and Nanoarchaeota (0.28%) were divided (Figure 2). Of these phyla, Thaumarchaeota - the most abundant archaeal phylum on earth – was also found in Binh Chau hot spring. Thaumarchaeota is known as chemoautotrophic, autotrophic ammonia-oxidizing (AOA) archaeal species and plays an important role in both carbon and nitrogen cycling (Pester *et al*., 2011). Moreover, the most abundant species was *Candidatus Nitrosoarchaeum limnia*, which belonged to Thaumarchaeota phylum. This is a thermophilic, cultured and capable ammonia-oxidizing archaeal species (Mosier *et al*., 2012). The second dominant archaeal phylum was Euryarchaeota. The members of Euryarchaeota were found as culturable archaea, belonging to several groups, such as methanogens (class: Methanobacteria, Methanococci, Methanomicrobia, and Methanopyri), halophilic (class: Halobacteria), acidic (class: Thermoplasmata), and thermophilic (class: Methanopyri and Thermococci) (Figure 2). Among the thermo archaeal groups of Binh Chau, *Methanopyrus kandleri* species belonged to Methanopyri class was able to grow at 122°C with the pressure of 20 MPa (Takai *et al*., 2008). Additionally, *Natronomonas pharaonis* species which classified in Euryarchaeota phylum, is known as an aerobic and haloalkaliphilic archaeon. This species grows optimally at 3.5 M NaCl and pH 8.5 (Falb *et al*., 2005). Besides the culturable species, the uncultured Euryarchaeota archaea were found and classified in Thermoplasmatales of Thermoplasmata order. In nature, Euryarchaeota involved in methane production, anaerobic methane oxidation and other biogeochemical cycles (Baker *et al*., 2020). Another dominant archaeal phylum in Binh Chau hot spring was Crenarchaeota, which composed of only Thermoprotei class (Figure 2). In Thermoprotei class, the orders Sulfolobales, Desulfurococcales, and Thermoproteales were identified. As chemolithoautotrophic archaea, the presence of Crenarchaeota plays an important role in geobiochemical cycles of hot springs because they are able to grow on H_2, S, or H_2S as substrates, and/or used O_2, NO_3^-, S, or Fe_3^+ as electron acceptors (Qi *et al*., 2010). Besides, Korarchaeota phylum was also found in Binh Chau hot spring metabase with only one species *Candidatus Korarchaeum cryptofilum*.

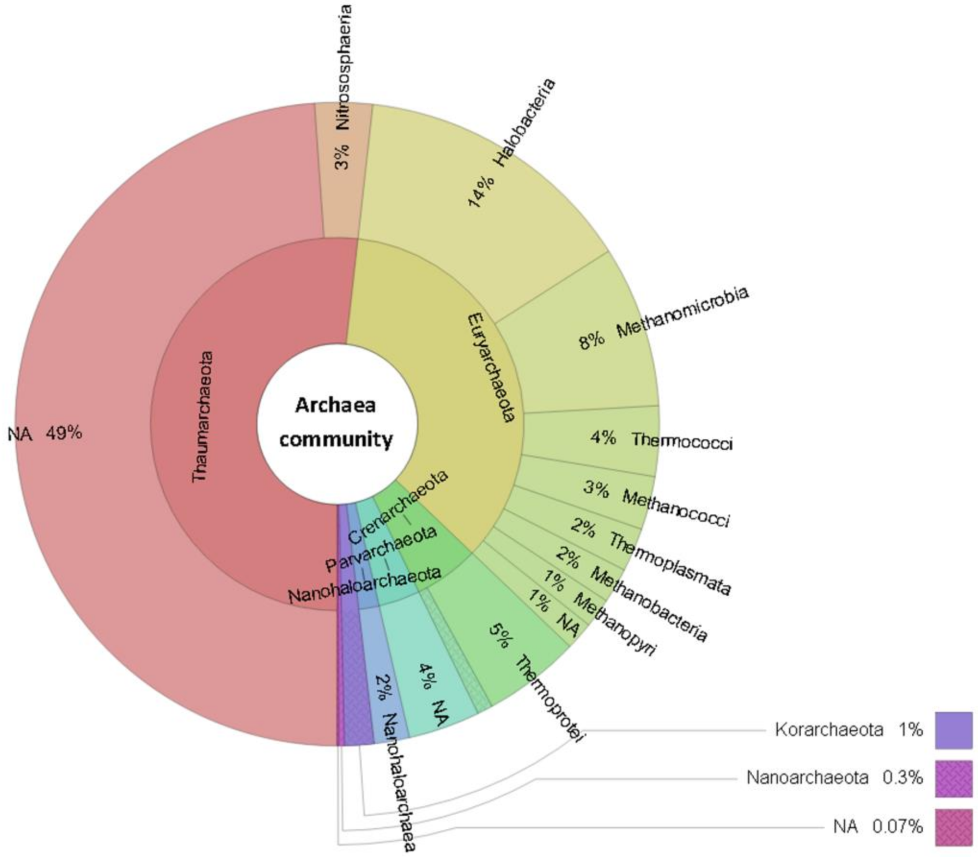

Figure 2. Krona chart representation of taxonomic classification of Archaea community in Binh Chau hot spring. NA, non-annotation.

Specially, several ultra-small cells were detected in the metabase of Binh Chau hot, including Parvarchaeota, Nanohaloarchaeota, and Nanoarchaeota phyla. Almost of them are unculturable anaerobic, and/or symbiotic relationships with other Euryarchaeota phylum (Castelle *et al*., 2015). Parvarchaeota, Nanohaloarchaeota and Nanoarchaeota phyla are able to adapt very well to extremely environments such as high temperature, acidic or alkali pH, and high salinity. These are involved in several metabolic pathways, for example, Parvarchaeota participates in iron oxidation (Chen *et al*., 2018), Nanohaloarchaeota usually presents in salinity environment and plays a role in Embden-Meyerhoff pathway (Narasingarao *et al*., 2012), whereas Nanoarchaeota is found in geothermal ecology (Cai *et al*., 2021). In Binh Chau hot spring metabase, *Nanoarchaeum equitans* was a unique representative of Nanoarchaeota, which owned the smallest genome (Waters *et al*., 2003). Taken together, a more diversed picture of microbial community in Binh Chau hot spring was revealed, but a number of unclassified genes are still existed. The analysis tools might be insufficient, therefore, it is necessary to develop for a deeper microbial assessment as well as for preservation of microbial genomes in special ecosystems such as hot springs in future.

CONCLUSION

A total of 156,093 genes were predicted from Binh Chau hot spring metabase. Among 156,093 genes, a ratio of 68.5% was annotated while remaining 31.5% was unknown. Analysis of annotated genes showed 29,069 bacterial, 1,416 archaeal, 4 eukaryotic, 3 viral, and 76,411 unclassified genes. For classification purpose, microbial communitites in Binh Chau hot spring were classified in 29 bacterial and 7 archaeal phyla. Several classes and species taxonomy analysis was discussed and revealed their roles in nature. Most of them were found firstly in this ecology, which contributed to hot spring genetic database for further exploitation and conservation in Vietnam.

Sequence data availability

The metagenome was deposited in the National Center for Biotechnology Information (NCBI) under accession number PRJNA415324.

Acknowledgements: *This research is funded partly by Ministry of Science and Technology (MOST), Vietnam under grant number: DTDLCN.15/14. We would like to thank Mr. Ngo Ba Huong in Saigon Binh Chau Corporation for sample processing.*

REFERENCES

Alain K, Postec A, Grinsard E, Lesongeur F, Prieur D, Godfroy A (2010) *Thermodesulfatator atlanticus* sp. nov., a thermophilic, chemolithoautotrophic, sulfate-reducing bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent. *Int J Syst Evol Microbiol* 60: 33–38.

Baker B, De Anda V, Seitz K, Dombrowski N, Santoro A, Lloyd K (2020) Diversity, ecology and evolution of Archaea. *Nat Microbiol* 5: 1-14.

Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15): 2114-2120.

Cai R, Zhang J, Liu R, Sun C (2021) Metagenomic Insights into the Metabolic and Ecological Functions of Abundant Deep-Sea Hydrothermal Vent DPANN Archaea. *Appl Environ Microbiol* 87(9): e03009-20.

Castelle CJ, Wrighton KC, Thomas BC, Hug LA, Brown CT, Wilkins MJ, Frischkorn KR, Tringe SG, Singh A, Markillie LM, Taylor RC, Williams KH, Banfield JF (2015) Genomic Expansion of Domain Archaea Highlights Roles for Organisms from New Phyla in Anaerobic Carbon Cycling. *Curr Biol* 25(6): 690-701.

Chan CS, Chan KG, Tay YL, Chua YH, Goh KM (2015) Diversity of thermophiles in a Malaysian hot spring determined using 16S rRNA and shotgun metagenome sequencing. *Front Microbiol* 6: 177.

Chen LX, Méndez-García C, Dombrowski N, Servín-Garcidueñas LE, Eloe-Fadrosh EA, Fang BZ, Luo ZH, Tan S, Zhi XY, Hua ZS, Martinez-Romero E, Woyke T, Huang LN, Sánchez J, Peláez AI, Ferrer M, Baker BJ, Shu WS (2018) Metabolic versatility of small archaea Micrarchaeota and Parvarchaeota. *ISME J* 12(3): 756-775.

Debnath T, Kujur RRA, Mitra R, Das SK (2019) Diversity of Microbes in Hot springs and Their Sustainable Use. In: Satyanarayana T, Johri BN, Das SK (eds) *Microbial Diversity in Ecosystem Sustainability and Biotechnological Applications*. *Springer, Singapore*, pp. 159-186.

Deckert G, Warren PV, Gaasterland T, Young WG, Lenox AL, Graham DE, Overbeek R, Snead MA, Keller M, Aujay M, Huber R, Feldman RA, Short JM, Olsen GJ, Swanson RV (1998) The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. Nature 392(6674): 353-358.

Falb M, Pfeiffer F, Palm P, Rodewald K, Hickmann V, Tittor J, Oesterhelt D (2005) Living with two extremes: conclusions from the genome sequence of *Natronomonas pharaonis*. *Genome Res* 15(10): 1336-1343.

Huber R, Eder W, Heldwein S, Wanner G, Huber H, Rachel R, Stetter KO (1998) *Thermocrinis ruber* gen. nov., sp. nov., A pink-filament-forming hyperthermophilic bacterium isolated from yellowstone national park. *Appl Environ Microbiol* 64: 3576–3583.

Huber R, Langworthy TA, König H, Thomm M, Woese CR, Sleytr UB, Stetter KO (1986) *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch Microbiol* 144: 324–333.

Huber R, Wilharm T, Huber D, Trincone A, Burggraf S, König H, Reinhard R, Rockinger I, Fricke H, Stetter KO (1992) *Aquifex pyrophilus* gen. nov. sp. nov., represents a novel group of marine hyperthermophilic hydrogen-oxidizing bacteria. *Syst Appl Microbiol* 15: 340–351.

Huson DH, Auch AF, Qi J, Schuster SC (2007) MEGAN analysis of metagenomic data. *Genome Res* 17(3): 377-386.

Hussein EI, Jacob JH, Shakhatreh MAK, Abd Al-Razaq MA, Juhmani ASF, Cornelison CT (2017) Exploring the microbial diversity in Jordanian hot springs by comparative metagenomic analysis. *MicrobiologyOpen* 6(6): e00521.

Kambura AK, Mwirichia RK, Kasili RW, Karanja EN, Makonde HM, Boga HI (2016) Bacteria and Archaea diversity within the hot springs of Lake Magadi and Little Magadi in Kenya. *BMC Microbiol* 16(1): 136.

Kang I, Kang D, Oh HM, Kim H, Kim HJ, Kang TW, Kim SY, Cho JC (2011) Genome sequence of strain IMCC2047, a novel marine member of the Gammaproteobacteria. *J Bacteriol* 193(14): 3688-3689.

Knapik K, Becerra M, González-Siso MI (2019) Microbial diversity analysis and screening for novel xylanase enzymes from the sediment of the Lobios Hot spring in Spain. *Sci Rep* 9(1): 11195.

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9(4): 357-359.

Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22(13): 1658-1659.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu SM, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam TW, Wang J (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* 1(1): 18-18.

Mehetre GT, Paranjpe AS, Dastager SG, Dharne MS (2015) Complete metagenome sequencing based bacterial diversity and functional insights from basaltic hot spring of Unkeshwar, Maharashtra, India. *Genom Data* 7: 140-143.

Mosier AC, Allen EE, Kim M, Ferriera S, Francis CA (2012) Genome sequence of "*Candidatus Nitrosoarchaeum limnia*" BG20, a low-salinity ammonia-oxidizing archaeon from the San Francisco Bay estuary. *J Bacteriol* 194(8): 2119-2120.

Nakagawa S, Shtaih Z, Banta A, Beveridge TJ, Sako Y, Reysenbach AL (2005) *Sulfurihydrogenibium yellowstonense* sp. nov., an extremely thermophilic, facultatively heterotrophic, sulfur-oxidizing bacterium from Yellowstone National Park, and emended descriptions of the genus *Sulfurihydrogenibium*, *Sulfurihydrogenibium subterraneum* and *Sulfurihydrogenibium azorense.* *Int J Syst Evol Microbiol* 55: 2263–2268.

Narasingarao P, Podell S, Ugalde JA, Brochier-Armanet C, Emerson JB, Brocks JJ, Heidelberg KB, Banfield JF, Allen EE (2012) *De novo* metagenomic assembly reveals abundant novel major lineage of Archaea in hypersaline microbial communities. *ISME J* 6(1): 81-93.

Nguyen KT, Nguyen NL, Nguyen VT, 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.

Nguyen Kim Thoa, Tran Thanh Thuy (2018) Characteristics of an alkaline Carboxymethyl cellulase produced by a thermophilic bacterial strain isolated from Binh Chau hot spring. *National Biotechnology Conference*: 1230-1235.

Ortiz-Burgos S (2016) Shannon-Weaver Diversity Index. In: Kennish MJ (ed). Encyclopedia of Estuaries. *Dordrecht: Springer Netherlands***:** 572-573.

Panda AK, Bisht SS, De Mandal S, Kumar NS (2016) Bacterial and archeal community composition in hot springs from Indo-Burma region, North-east India. *AMB Express* 6(1): 111-111.

Paul S, Cortez Y, Vera N, Villena G, Gutiérrez-Correa M (2016) Metagenomic analysis of microbial community of an Amazonian geothermal spring in Peru. *Genom Data* 9: 63-66.

Pester M, Schleper C, Wagner M (2011) The Thaumarchaeota: an emerging view of their phylogeny and ecophysiology. *Curr Opin Microbiol* 14(3): 300-306.

Qi SZ, Chen JQ, Jiang H, Zhou EM, Tang SK, Zhi XY, Zhang LX, Zhang L, Li WJ (2010) Diversity of Crenarchaeota in Terrestrial Hot springs in Tengchong, China. *Extremophiles: life under extreme conditions* 14(3): 287-296.

Rozanova EP, Khudiakova AI (1974) A new non-spore forming thermophilic organism, reducing sulfates, *Desulfovibrio thermophilus* nov. sp. *Mikrobiologiia* 43(6): 1069–1075.

Saxena R, Dhakan DB, Mittal P, Waiker P, Chowdhury A, Ghatak A, Sharma VK (2016) Metagenomic analysis of hot springs in central India reveals hydrocarbon degrading thermophiles and pathways essential for survival in extreme environments. *Front Microbiol* 7: 2123.

Slightom R, Buchan A (2009) Surface Colonization by Marine Roseobacters: Integrating Genotype and Phenotype. *Appl Environ Microbiol* 75: 6027-6037.

Takai K, Nakamura K, Toki T, Tsunogai U, Miyazaki M, Miyazaki J, Hirayama H, Nakagawa S, Nunoura T, Horikoshi K (2008) Cell proliferation at 122°C and isotopically heavy CH₄ production by a hyperthermophilic methanogen under high-pressure cultivation. *Proc Natl Acad Sci USA* 105(31): 10949-10954.

Tran K, Le VTT, Ngo DD, Hoang K, Nguyen PV, Nguyen TH (2018) Isolation and optimization of the growth conditions of thermophilic microorganism from hot springs. *The Journal of Agriculture and Development* 17(3): 55-60.

Uribe-Lorío L, Brenes-Guillén L, Hernández-Ascencio W, Mora-Amador R, González G, Ramírez-Umaña CJ, Díez B, Pedrós-Alió C (2019) The influence of temperature and pH on bacterial community composition of microbial mats in hot springs from Costa Rica. *MicrobiologyOpen* 8(10): e893.

Van den Ende FP, Gemerden HV (1993) Sulfide oxidation under oxygen limitation by a *Thiobacillus thioparus* isolated from a marine microbial mat. *FEMS Microbiol Ecol* 13(1): 69-77.

Wang H, Li H, Shao Z, Liao S, Johnstone L, Rensing C, Wang G (2012) Genome sequence of deep-sea manganese-oxidizing bacterium *Marinobacter manganoxydans* MnI7-9. *J Bacteriol* 194(4): 899-900.

Waters E, Hohn MJ, Ahel I, Graham DE, Adams MD, Barnstead M, Beeson KY, Bibbs L, Bolanos R, Keller M, Kretz K, Lin X, Mathur E, Ni J, Podar M, Richardson T, Sutton GG, Simon M, Soll D, Stetter KO, Short JM, Noordewier M (2003) The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism. *Proc Natl Acad Sci U S A* 100(22):12984-12988.

Zhu W, Lomsadze A, Borodovsky M (2010) Ab initio gene identification in metagenomic sequences. *Nucl Acids Res* 38(12): e132.