Novel Research in Microbiology Journal (2024), 8(6): 2694-2711

(Print) (ISSN 2537-0286)
(Online) (ISSN 2537-0294)
https://nrmj.journals.ekb.eg/

Research Article

DOI: 10.21608/NRMJ.2024.328298.1753

Applying metagenomics to structural and functional analyses of microbial communities from lagoon sediments in some provinces of central Vietnam

Tran Thi Hong[1]; Ngo Van Anh[1]; Nguyen Thi Ngoc Anh[2]; Nguyen Van Phuong[3*]

[1]Mientrung Institute for Scientific Research, Vietnam National Museum of Nature, Vietnam Academy of Science and Technology (VAST), Hanoi 10000, Vietnam; [2]Hanoi Open University, Hanoi, Vietnam; [3]University of Science and Technology of Hanoi (USTH), Vietnam Academy of Science and Technology (VAST), Hanoi 10000, Vietnam

[*]Corresponding author E-mail: nguyen-van.phuong@usth.edu.vn

Received: 14 October, 2024; Accepted: 11 November, 2024; Published online: 12 November, 2024

Abstract

Metagenomics is an important technique for discovering and screening significant bioactive compounds such as polyketide synthases, non-ribosomal peptide synthetases, antibiotics, and biocatalysts. The enormous potential of microbial diversity in coastal lagoons remains untapped because of the difficulty of culturing many microorganisms under laboratory conditions. Therefore, this study aimed to decipher the structural and functional diversity of microbial communities, including un-culturable species in lagoon sediments collected from Tam Giang, Nai, and Thi Nai lagoons, which were located in the central coast of Vietnam using metagenomics approach. The results showed that all three large lagoons exhibited a high diversity of fungal species among the eukaryotic species and a high diversity of bacterial and archaeal species. Among them, the most prominent phyla included Ascomycota (fungal phyla), Firmicutes and Proteobacteria (bacterial phyla), and Thaumarchaeota (archaeal phyla). Sequencing of the metagenome samples of microbiome in the three lagoons of Tam Giang, Nai, and Thi Nai using shotgun sequencing technology revealed the existence of 126 potential gene sequences from the Tam Giang lagoon, 346 sequences from the Nai lagoon, and 341 sequences from the Thi Nai lagoon. These sequences participated in various metabolic processes and belonged to more than 30 groups of biologically active compounds. Most of these genes were related to antibacterial, antifungal, antiviral, cancer cell inhibition, and antioxidant activities. This is the first study conducted on lagoon microorganisms in Vietnam, opening up prospects and potential in exploiting microorganisms and biologically active substances with important applications in life from the lagoon ecosystems of Vietnam in particular and the overall world in general.

Keywords: Metagenomics, Tam Giang, Thi Nai, Nai, Lagoons, Microbial diversity, Antimicrobial activity

1. Introduction

Nowadays, many metagenomic studies target marine ecosystems and coastal lagoons to explore the microbial diversity and develop biotechnological applications (Barone *et al*., 2014). Microorganisms contribute to about 98 % of primary productivity in the marine environments; either as independent microorganisms or by collaborating with other ones. This relationship encourages the formation of different bioactive complexes, which can arise from the microorganisms own metabolic processes or through partnerships with other microorganisms (Newman and Cragg, 2016). The microbiome was effectively investigating by using a metagenomic methodology, which provided insights into the different microbial communities inhabiting a certain environment (Chaouni *et al*., 2022). Metagenomics is a culture-independent approach that screens and predicts the microbial functions based on their sequences, resulting in the exploration and production of various biologically important compounds, including antibiotics, polyketide synthases, non-ribosomal peptide synthetases, and biocatalysts (Mahapatra *et al*., 2020). Consequently, metagenomic techniques represent a significant advancement in identifying the biosynthesis of previously unknown gene clusters in certain microorganisms, allowing for incorporation of synthetic genes into host microorganisms and their subsequent development *in vitro* (Kamble and Vavilala, 2018).

The lagoons of Vietnam are mainly found in the country's central region, which has abundant coastal and deposits, powerful wave dynamics, and low tides. There are 12 typical lagoons covering an area of around 458 km^2 and occupying approximately 21 % of Vietnam's coastline from Thua Thien Hue to Ninh Thuan. A representative example is the Tam Giang - Cau Hai lagoon system in Thua Thien Hue province with a length of 70 km and width of 216 km^2, which is the largest in the world and in Southeast Asia. A central lagoon is the Thi Nai lagoon, Binh Dinh, which has an area of 5000 ha, while the Nai lagoon represents the south-central coast of Ninh Thuan. All these three lagoons were evaluated as ecosystems with high biological productivity, which represent places to store aquatic species and diverse habitat types, including estuaries, swamps, water grasslands, and beaches. The swamp contains mangrove plants, muddy bottoms, tidal creeks, sandy tidal flats, and rocky tidal areas; therefore, its biodiversity is high. In particular, the Tam Giang lagoon is considered to be rich in biological resources and diverse aquatic flora, fauna, ecosystems, and microorganisms (Phan *et al*., 2016). The objective of this study was to use a DNA metagenomic approach to evaluate the structural and functional analyses of microbial communities from three lagoon sediments in some provinces of central Vietnam, which resulted in a new and diverse insight of microbial communities in these lagoons

2. Material and methods

2.1. Collection of samples

Bottom mud samples were collected from Tam Giang (16°62'02" N; 107°49'73" E), Nai (11°37'17" N; 109°1'41" E), and Thi Nai lagoons (13°49'44" N; 109°14'06" E) from April 8, 2022 to April 25, 2022. Five locations were chosen for sampling in each lagoon: the mouth, the middle of the lagoon, an estuary, a mangrove site, and a shallow area 100 m from the beach (Fig. 1). Using collecting pipes, four sediment samples were taken at a depth of 30 cm at each sampling location. For the purpose of extracting DNA, these samples were transferred to the lab in ice packs after being kept in sterile tubes. The PowerMax® Soil DNA Isolation Kit was used for

extraction of microbial DNA from the sediment samples following the instructions of the manufacturer (Mo Bio Laboratories, Carlsbad, CA, USA). A bulk DNA sample that represented each lagoon was prepared by equal combination of the DNA collected from each site.

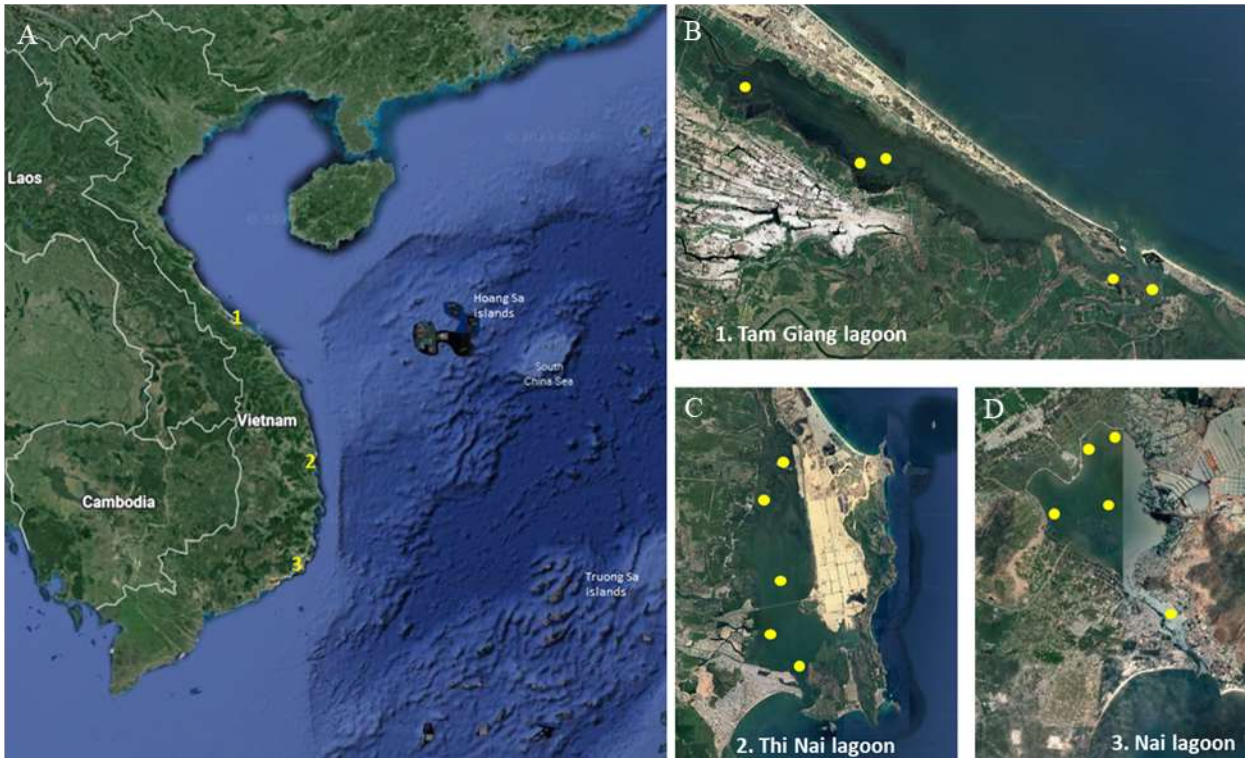

Fig. 1. Sampling sites in overview (A) with three lagoons of Tam Giang (1), Thi Nai (2), and Nai (3). Yellow dots in B, C, and D represent the positions of collected samples in the respective lagoon (Source: Google Earth, 2022)

2.2. 16S rRNA and Internal Transcript Spacer (ITS) amplicon sequencing

The primers *341-F* (5'-CTACGGGNGGCWGCAG-3') and *806-R* (5'-GGACTACNNGGGTATCT AAT-3') (NovogeneAIT, Singapore) were used to amplify the 16S rRNA gene (Lin *et al*., 2019). The ITS gene was amplified using primers *ITS5* (5'-GGAAGTAAAAGTCGTAACAAGG-3') and *ITS2* (5'-GCTGCGTTCTTCATCGATGC-3') (Bellemain *et al*., 2010). A reaction of 10 µl contained 0.5 µl of each primer (10 µM), 0.2 µl of DMSO, 4.5 µl of Phusion Master Mix (2x), and 4.3 µl of template DNA. The thermo-cycles for 16S rRNA gene amplification was set up as follows: pre-denaturation at 98 °C for 2 min., followed by 30 cycles of 30 sec each at 95 °C, 55 °C, and 72 °C, and finally 5 min. at 72 °C. Similarly, the ITS gene was with the following thermo-cycles: 1 min. of pre-denaturation at 98 °C, followed by 30 cycles of 10 sec at 98 °C, 30 sec at 55 °C, 30 sec at 72 °C, and finally 5 min. at 72 °C. The Polymerase Chain

Reactions (PCR) products were checked by electrophoresis on a 2 % agarose gel. Subsequently, the product bands were purified using a Qiagen Gel Extraction Kit (Qiagen, Germany).

2.3. Sequencing and data analysis

Index codes were inserted and a sequencing library was built using the NEBNext Ultra DNA Library Pre® Kit for Illumina (New England Biolabs, UK) in accordance with the manufacturer's instructions. An Agilent Bioanalyzer 2100 system and a Qubit @ 2.0 Fluorometer spectrometer (Thermo Scientific, USA) were used to evaluate the library's quality. Ultimately, 250 bp paired-end reads were used to sequence the library on an Illumina NovaSeq 6000 platform.

According to Schloss *et al*., (2009), the Mothur tool was used to analyze the community sequence data. The Silva database version 138 was used to align cleaned readings. The classify.seqs, cluster.split, and classify.otus functions were applied to classify the aligned reads. Sequences with 97 % similarity were assigned to Operational Taxonomic Units (OTUs); using the RPD version 18 training dataset (July 2020), and the 16S Silva version 138 full-length sequencing database. Sequences that were chimeric or non-bacterial (eukaryotes, mitochondria, chloroplasts, and unknown) were removed from the data.

2.4. Analysis of microbial diversity

The MetaPhlAn 3 toolkit was used for microbial diversity analysis (Beghini *et al*., 2021). MetaPhlAn 3 was designed to profile microbial diversity using species-level shotgun metagenomic sequence reads. MetaPhlAn applied as a comparative approach to the clade-specific marker gene database with approximately 1.1 million markers identified from more than 100,000 reference genomes, including more than 99,500 bacterial and archaeal genomes and more than 500 eukaryotic genomes. The MetaPhlAn3 database was used for this analysis.

2.5. Functional analysis

For functional analysis, a HUMAnN 3 software developed by Beghini *et al*., (2021) was used. HUMAnN 3 is a workflow that integrates various analytical tools in order to stratify the functional profiles of known and unknown microorganisms. The UniRef90 database was filtered to retain the sequences with EC codes. All genes appearing in the metagenomic samples were annotated using the BLAST tool (Camacho *et al*., 2009) in Diamond (Buchfink *et al*., 2014) and three major protein databases; mainly Kofam (Aramaki *et al*., 2020), Refseq (O'Leary *et al*., 2016), and Swissprot (Bairoch and Apweiler, 2000). Information about protein products, KO, and EC codes of each protein was extracted for functional searches of the KEGG pathway (Kanehisa *et al*., 2004).

3. Results

3.2. Characterization of the 16S rRNA amplicons

The V3V4 region of 16S rRNA gene and the ITS1 region of ITS gene were amplified by using the primers *341-F/ 806-R* and *ITS-2/ ITS-5*, respectively. Upon carrying out 2 % agarose gel electrophoresis, the PCR products were a single distinct band of about 450 bp for the *16S rRNA* gene and 280 bp for the *ITS1* gene.

Sequencing of the V3V4 region (16S rRNA) yielded over 16,000 sequence reads for each sample. Low-quality sequences were filtered using the make.contigs tool from Mothur. Along with chimeras and non-bacterial sequences, which were addressed using chimera.vsearch tool from Mothur. Ultimately, around 110,000 clean sequence reads per sample were assigned to OTUs, accounting for 70 % of the original data (Table 1).

Fig. 2A displays the rarefaction curves for the three samples containing OTUs and sequence reads. Among the three analyzed lagoons, Tam Giang lagoon had the most diverse species level, with 15,981 OTUs, followed by Thi Nai lagoon with approximately 8,418 OTUs, and then Nai lagoon with 7,118 OTUs (Table 1).

Table 1. Characteristics of 16S rRNA and ITS sequencing

Sample ID	Number of raw reads (16S rRNA/ ITS)	Number of cleaned reads (16S rRNA/ ITS)	Number of OTUs (16S rRNA/ ITS)
Tam Giang	160,954/ 142,085	113,450/ 101,285	15,981/ 296
Thi Nai	166,843/ 164,077	122,030/ 92,589	8,418/ 300
Nai	165,208/ 186,171	128,947/ 126,685	7,118/ 300

Where; OTUs: Operational Taxonomic Units

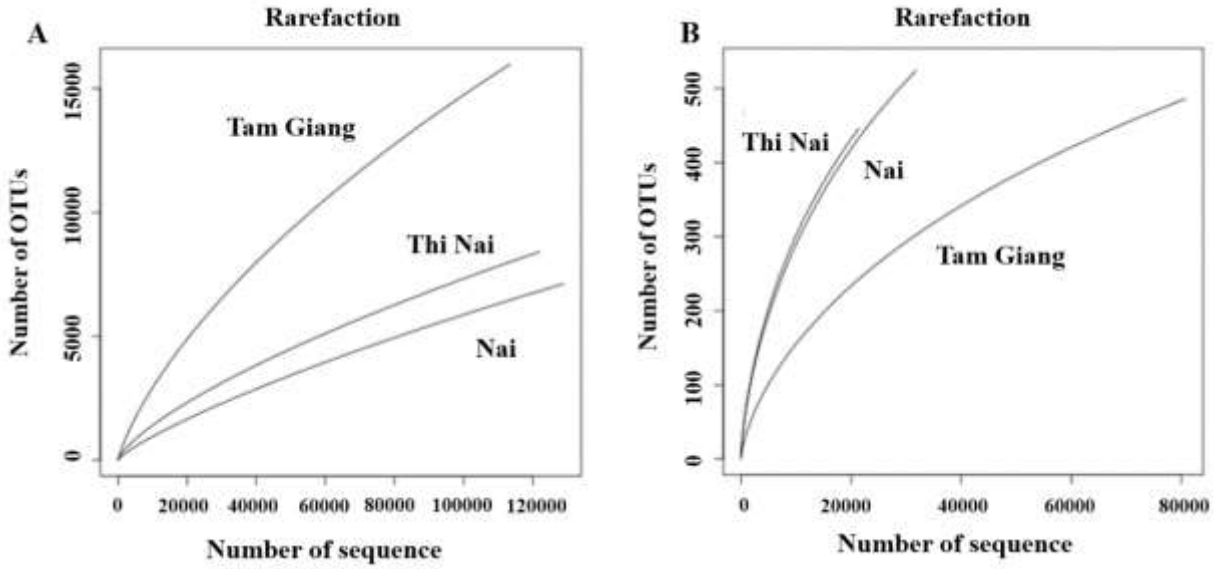

Fig. 2. The number of Operational Taxonomic Units (OTUs) and the corresponding number of reads in samples of three lagoons Tam Giang, Thi Nai, and Nai that were displayed by rarefaction curves. A and B represent sequencing of the V3V4 and ITS regions, respectively

Sequencing of ITS region of the fungal community showed that the Nai sample recorded the highest number of sequence reads (approximately 18,000), followed by the Thi Nai agoon (approximately 16,000), and the least number was observed from Tam Giang lagoon (approximately 14,000). After removing low-quality sequences, chimera sequences, and non-fungal sequences (UCHIME Algorithm, http://www.drive5.com/usearch/manual/uchime_algo.html), approximately 70 % of the clean sequences were recovered from the Nai and Tam Giang Lagoons and 56 % from the Thi Nai Lagoon. The rarefaction curves of the three samples with sequence reads and OTUs are shown in Fig. 2B.

3.3. Bacterial community structure

The V3V4 region of the 16S rRNA gene was sequenced and the resulting data was analyzed using bioinformatics approaches to shed light on the bacterial diversity found in bottom mud samples from coastal lagoons in central Vietnam, specifically the Tam Giang, Nai, and Thi Nai lagoons. Using the Silva reference database, a 16S rRNA gene classification revealed that a significant amount of bacteria was present in the three lagoons. The OTUs were identified at the phylum, class, order, family, and genus levels. This study focused on three lagoons, where the maximum bacterial diversity was found in Tam Giang lagoon. The other two lagoons, Nai and Thi Nai, showed similar bacterial richness (Fig. 2A, Fig. 3). The number of species shared among Nai and Tam Giang, Nai and Thi Nai, and Tam Giang and Thi Nai were 169; 877; and 332, respectively (Fig. 3).

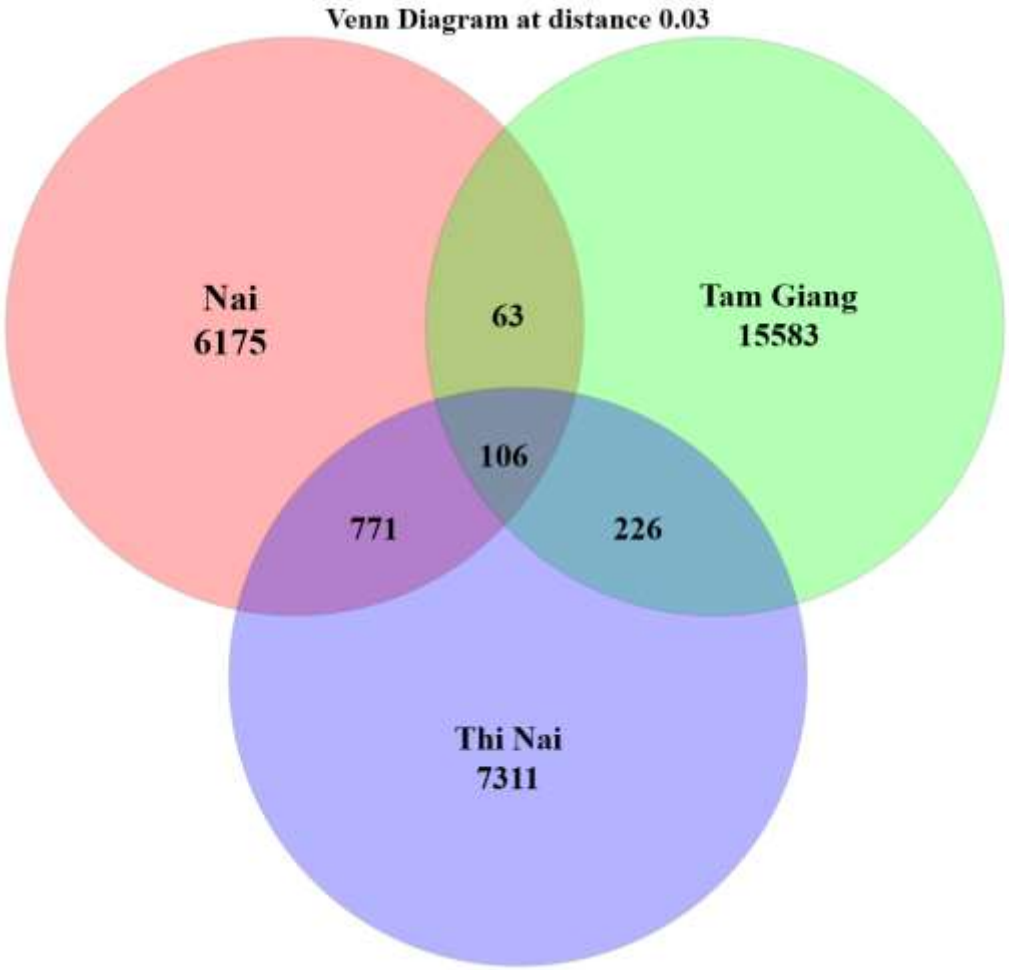

Fig. 3. Distribution of bacterial species composition (OTU) of the three lagoon samples

One aspect of biodiversity is the relative species richness, which indicates how frequent or rare a species is in relation to other species in a particular community. The percentage of a species of a particular type relative to all microorganisms in an area is known as relative abundance. The data was extracted from uncommon OTUs (appearing only once in a sample) and normalized to the same ratio in order to compare the relative abundances of the three lagoons (Tam Giang, Nai, and Thi Nai). The comparison results showed six phyla that were most abundant in all the three lagoons; respectively, which were Proteobacteria (25.35 %, 43.54 %, 55.60 %), Firmicutes (48.13 %, 32.68 %, 30.75 %), Bacteroidetes (3.56 %, 7.78 %, 1.26 %), Acidobacteria (2.93 %, 0.41 %, 0.11 %), Actinobacteria (0.82 %, 0.41 %, 0.31 %), and Chloroflexi (1.86 %, 0.24 %, 0.05 %) (Fig. 4). In the lagoons of Thi Nai and Nai, the genera *Lactobacillus* and *Vibrionaceae* contained the most groups of OTUs at the genus level. *Clostridium sensu stricto* was the genus with the highest proportion in the Tam Giang lagoon.

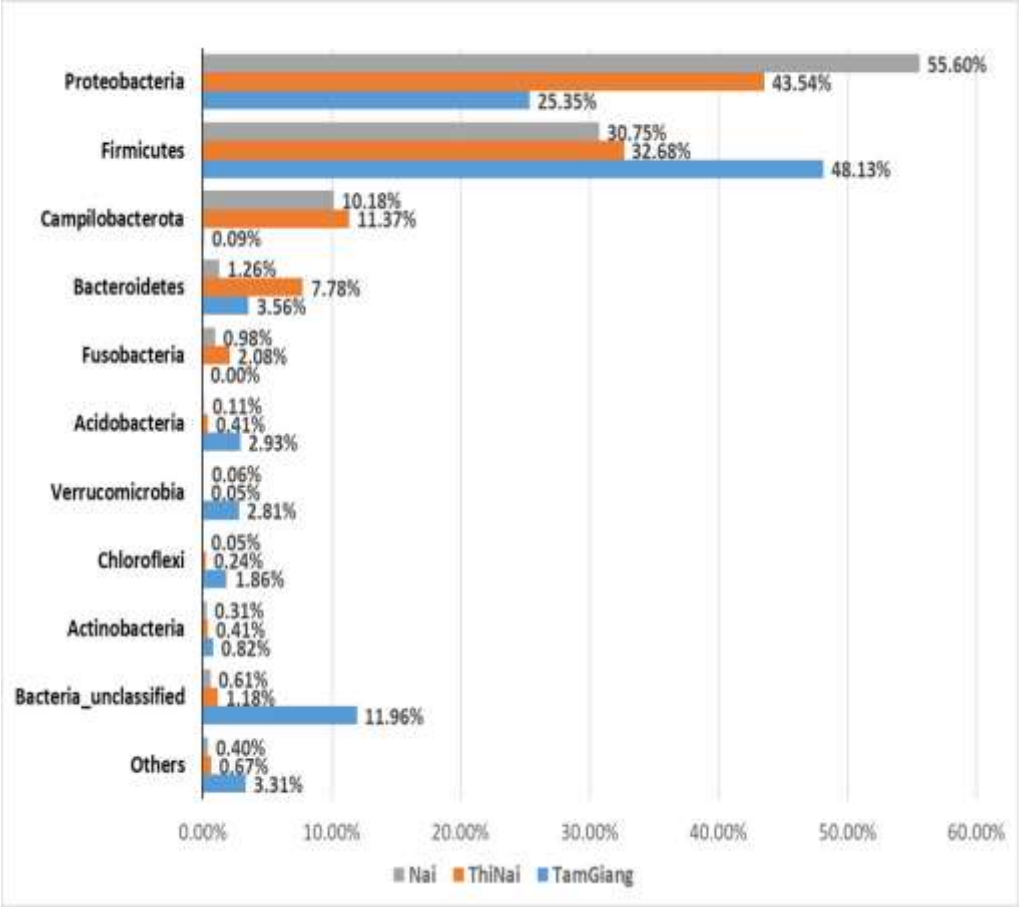

Fig. 4. Relative abundance of main bacterial phyla isolated in sediment samples from three lagoons of Tam Giang, Thi Nai, and Nai

3.4. Fungal community structure

To analyze the fungal diversity present in the sediment samples from Tam Giang, Nai, and Thi Nai lagoons, the characteristic sequence fragment ITS1 was sequenced using the primer pairs *ITS-2* and *ITS-5*. Using the sequencing machine, Tam Giang lagoon obtained 142,085 original sequences, from Thi Nai lagoon 164,077 sequences, and from Nai lagoon 186,171 sequences. After filtering with bioinformatics tools (FLASH tool, GOLD database and UCHIME algorithm), the Tam Giang lagoon retrieved 101,285 clean sequences for diversity assessment (70 %), Thi Nai lagoon retrieved 92,589 sequences (56 %), and Nai lagoon retrieved 126,685 sequences (68 %). In the Tam Giang lagoon mud samples, out of a total of 101,285 ITS sequences, the data identified more than 90 thousand sequences belonging to 486 different OTUs, the majority of which were classified as belonging to the fungal kingdom; with more than 79 thousand sequences, accounting for nearly 99 % of all eukaryotic microbial sequences with 296 OTUs. For the Thi Nai lagoon bottom mud sample, 21,390 sequences with 446 OTUs were detected in eukaryotes. Of these, 6,327 sequences with 300 OTUs were identified in the fungal kingdom. For the Nai lagoon mud sample, 99,293 sequences belonging to 533 OTUs were classified, where 52,281 sequences belonged to the fungal kingdom with 300 OTUs.

Sequences were classified using the Mothur tool, the UNITE all Eukaryotes database, and the UNITE Fungi database. Classification of eukaryotic microorganisms in sediment samples of 3 lagoons showed that the proportion of fungi appearing in sediment sample of Tam Giang lagoon was the largest (≥ 98 %), followed by Thi Nai lagoon (29 %), the lowest was Nai lagoon (8.26 %). In addition, there were small proportions of zooplankton, including *Viridiplantae*, *Stramenopila*, *Alveolata*, *Heterolobosa*, *Metazoa*, and *Glaucocystoplantae* (Table 2).

The rarefaction curve (Fig. 5) demonstrated that the Thi Nai and Nai samples had high fungal diversity, whereas the Tam Giang sample had lower fungal diversity. The shared number of fungal species among Nai and Tam Giang, Nai and Thi Nai, and Tam Giang and Thi Nai were 125, 127, and 126, respectively (Fig. 6). Analysis of fungal species diversity using the UNITE Fungi database identified five common fungal phyla appearing in all three samples of Tam Giang, Thi Nai, and Nai, including Ascomycota, Mortierellomycota, Basidiomycota, Mucoromycota, and Chytridiomycota. Some other phyla appeared very little, which were only present in one lagoon such as Zoopagomycota, Entomophthoromycota, Olpidiomycota, and Glomeromycota phyla that only appeared in the Thi Nai lagoon, while the Blastocladiomycota phylum was only identified in the Nai lagoon. Thus, it can be noticed that Thi Nai lagoon had the highest fungal diversity, followed by Nai lagoon, and the lowest was recorded by Tam Giang lagoon (Fig. 7). The largest proportion was Ascomycota (98 %, 83 %, and 71 %, respectively), followed by Mortierellomycota and Basidiomycota. Analysis of species diversity in the phylum Ascomycota showed that, in Tam Giang lagoon, *Penicillium* (69 %) and *Candida* (28 %) were the two main genera. The main representative genus of *Penicillium* was *P. expansum* (66.79 %), while the species of genus *Candida* was *C. sake* (27.44 %). In the Thi Nai lagoon, *Penicillium* (31.15 %) and *Mortierella* (24.37 %) were the two main genera. The main representative of the genus *Penicillium* was the species of *P. expansum* (29.67 %), while for the genus *Mortierella*; the species was *M. rishikesha* (22.89 %). For the Nai lagoon, *Aspergillus* (20.10 %) and *Candida* (15.32 %) were the two main genera. The *Aspergillus* species was *A. penicillioides* (16.81 %), while the main representative species for the *Candida* genus was *C. sake* (14.83 %). Thus, two fungal species appeared abundantly in all three sediment samples of Tam Giang, Nai, and Thi Nai, which were *P. expansum* and *C. sake*.

Table 2. Diversity of eukaryotic groups in sediment samples at the bottom of Tam Giang, Thi Nai, and Nai lagoons

Microbial group	Proportion (%) in the sample		
	Tam Giang	Thi Nai	Nai
Fungi	98.52 %	29.40 %	8.26 %
Viridiplantae	1.24 %	0.36 %	0.79 %
Stramenopila	0.18 %	69.72 %	90.30 %
Alveolata	0.05 %	0.40 %	0.62 %
Heterolobosa	0.01 %	-	0.01 %
Metazoa	< 0.001 %	0.11 %	0.01 %
Glaucocystoplantae	< 0.001 %	< 0.001 %	-
Rhizaria	-	-	< 0.001 %

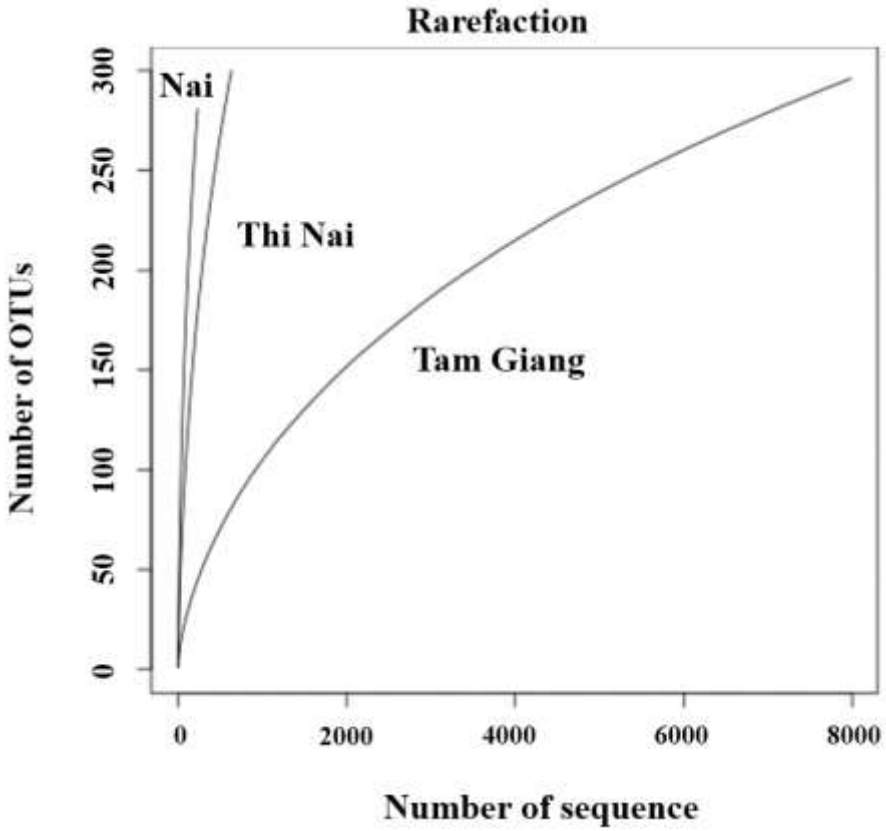

Fig. 5: The rarefaction curve of the three lagoon samples that were classified with the UNITE Fungi database

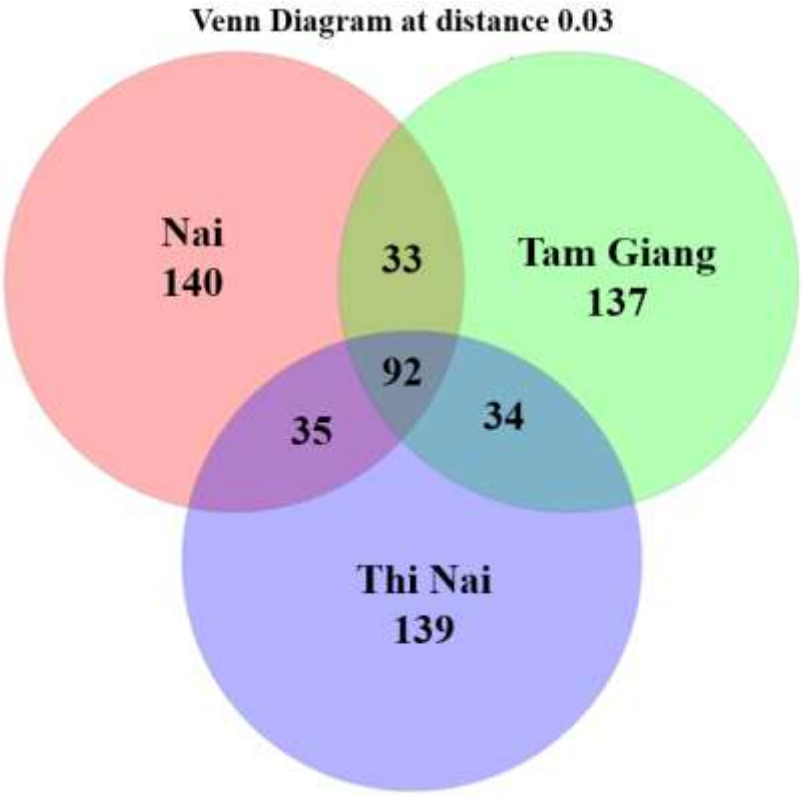

Fig. 6: Distribution of fungal species composition (OTU) of the three lagoon samples

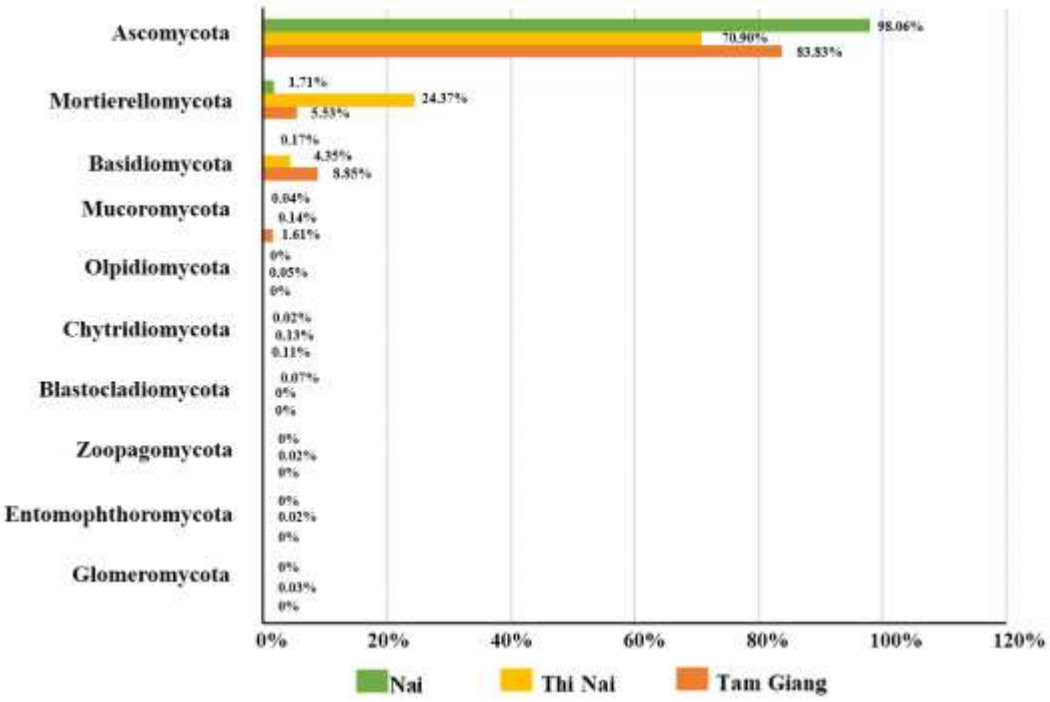

Fig. 7: Relative abundance of main fungal phyla detected in sediment samples from the three studied lagoons of Tam Giang, Thi Nai, and Nai

3.5. Structural analysis of archaea community

The data in Fig. (8) shows that the archaeal communities exhibited high diversity, comprising five common phyla in the Tam Giang, Nai, and Thi Nai lagoon sediment samples, including Thaumarchaeota, Woesearchaeota, Crenarchaeota, Euryarchaeota, and Pacearchaeota. However, the abundances of phyla differed among the lagoons. Sequences affiliated with Thaumarchaeota were dominant in Nai and Thi Nai (43.96 % and 26.35 %, respectively), whereas in Tam Giang, they were only 7.56 %. Meanwhile, Pacearchaeota accounted for the largest proportion in the Tam Giang lagoon (20.63 %) but not in a high proportion in the remaining two lagoonsof Nai and Thi Nai (12.84 % and 13.02 %, respectively). In contrast, Woesearchaeota was found to be a large proportion in Nai and Thi Nai lagoons (19.79 % and 14.60 %, respectively), but it was in a low proportion in Tam Giang lagoon (8.23 %). Euryarchaeota and Crenarchaeota were two phyla that appeared in both the Tam Giang and Thi Nai lagoons (16.31 %, 15.82 %, and 12.70 %, 13.65 %, respectively), but they were at low rates in the Nai lagoon (4.68 % and 2.27 %, respectively). In addition, other phyla such as Diapherotrites, Aenigmarchaeota, and Diapherotrites were also found in very low proportions in these lagoons.

3.5. Functional analysis of microbial communities

A shotgun metagenome approach was used to answer the question of what biological functions that microbes had in the Tam Giang, Nai, and Thi Nai lagoons?. In this study, the HUMAnN 3 software was used to analyze the gene function. To build a dataset with useful information, we annotated all of the genes appearing in the metagenome sample using the BLAST tool in Diamond and three major protein databases; mainly Kofam, Refseq, and Swissprot. Information about protein products, KO, and EC codes of each protein was extracted for functional searches of the KEGG pathway. For each lagoon, more than 100,000 proteins from the Kofam database, more than 300,000 proteins from the SwissProt database, and 380,000-500,000 proteins from the RefSeq database were annotated. There were three very large datasets. From these datasets, 126, 346, and 341 gene sequences related to the metabolic pathways of biologically active substances from the microbiota in mud samples obtained from the bottom of Tam Giang, Nai, and Thi Nai lagoons; respectively, were identified using the antiSMASH tool (Table 3). In order to screen genomic data from the metagenome data set, the antiSMASH tool used the gene sequences assembled by MegaHit as input, annotated the genes, and compared them with reported gene structures bearing the genes' biological activities, including antibacterial and anti-cancer properties. In total, 126-346 sequences belonging to different metabolic groups were identified in each lagoon (Fig. 9).

Data obtained from each lagoon identified a large number of coding sequences for short peptides involved in transcription and post-transcriptional protein modifications (Non-ribosomal peptide synthetase [NRPS], post-translationally modified peptide product-RiPP, and polyketide synthase (type I, type III Polyketide synthase - PKS). In addition, there were peptides belonging to the groups of cyclodipeptides, cyclic lactone auto-inducer peptides, lanthipeptides, lassopeptides, and Cysrich peptides. Sequences related to the protease inhibitor group included betalactone, butyrolactone, and other large groups of compounds, including aryl polyene, Ectoine, Furan, Ladderane, Phosphonate, Resorcinol, Siderophore and Terpene. These compounds exhibited antibacterial, antifungal, antiviral, and anti-cancer activities.

4. Discussion

Vietnam has many coastal lagoons. The Tam Giang-Cau Hai lagoon system is well known as one of the largest in the world and Southeast Asia. The lagoons in Vietnam are rich in biological resources and diverse ecosystems of aquatic plants, animals, and

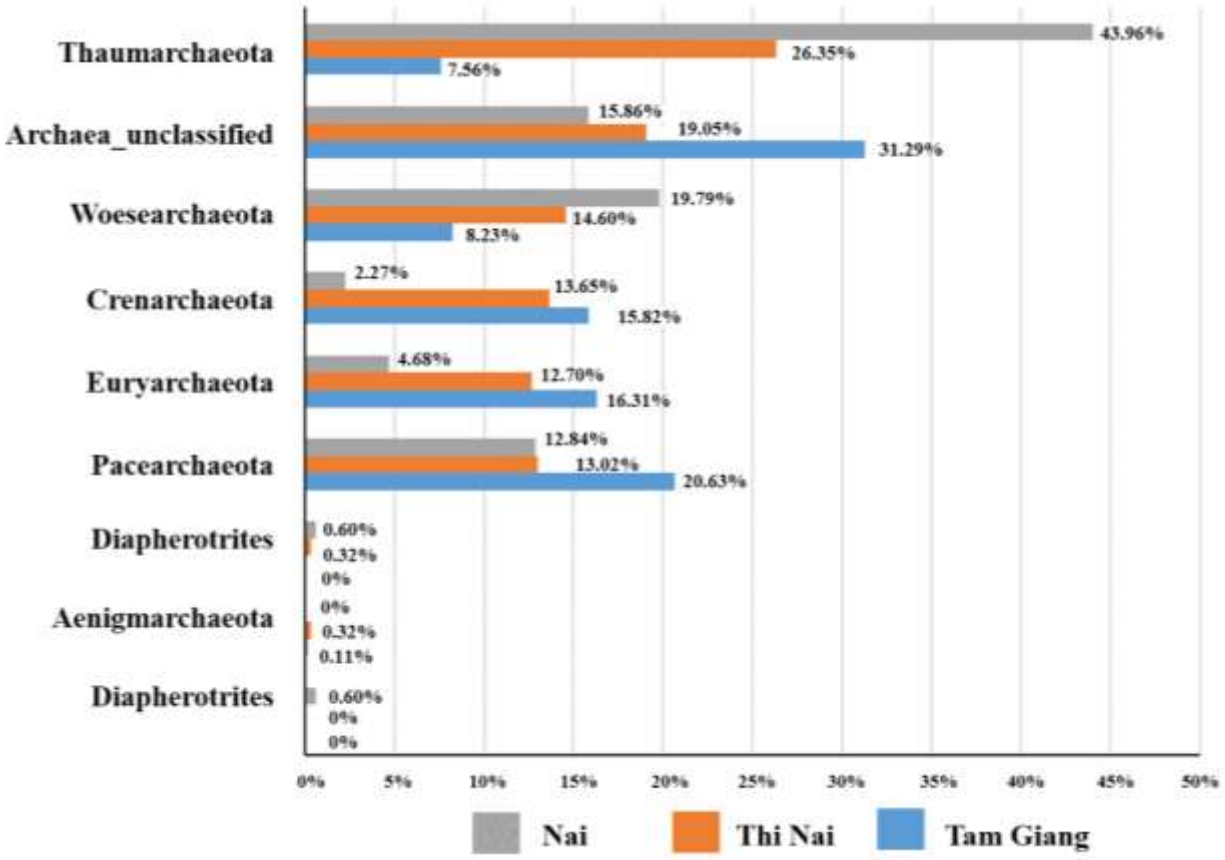

Fig. 8: Relative abundance of main archael phyla recorded in sediment samples collected from three lagoons of Tam Giang, Thi Nai, and Nai

Table 3: Statistics of annotated protein data obtained from metagenome data

Lagoons	Annotated protein numbers			
	Kofam	Swissprot	Refseq	antiSMASH
Tam Giang	100,240	306,141	380,419	126
Thi Nai	140,725	400,444	535,330	341
Nai	138,025	378,704	380,419	346

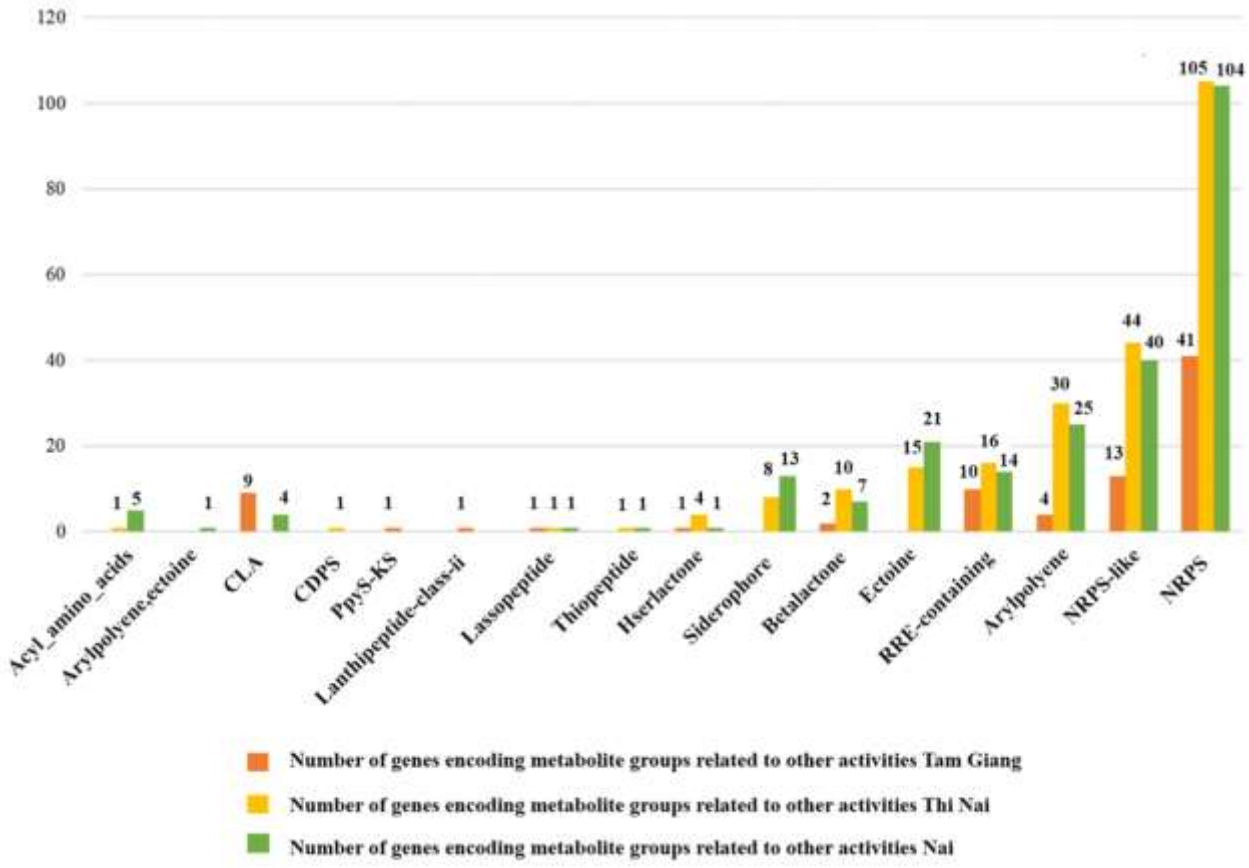

Fig. 9: Number of genes encoding metabolite groups related to other activities in the three studied lagoons of Tam Giang, Thi Nai, and Nai

microorganisms. However, researches on lagoons have mainly focused on water quality surveys (Cao *et al*., 2009), aquatic plant and animal resources (Luong and Nga, 2017), and assessing the pollution load entering lagoons (Le, 2021). This is the first study in Vietnam to use a high-throughput sequencing method to evaluate microbial diversity and determine their functions in the sediments of large lagoons in Vietnam, including the Tam Giang, Nai, and Thi Nai lagoons.

In this study, the microbial diversity and biological functions of sediment samples from the three lagoons Tam Giang, Nai, and Thi Nai in the central region of Vietnam were described. Amplicon sequencing (16S rRNA in bacteria and ITS in fungi) is a common approach to investigate microbial communities with different aspects such as composition, organization, and spatiotemporal patterns (Sinclair *et al*., 2015). Analysis of 16S rRNA genes helps to quantify microbial diversity and examines phylogeny in different natural environments using metagenomic approaches (Sharma *et al*., 2012). Microbialites worldwide maintain a similar composition at the phylum level, regardless of their geographical region. Actinobacteria, Bacteroidetes, Cyanobacteria, and Proteobacteria, are the common microbial components (Saurav *et al*., 2017). Our results are similar to those of previous studies on microbial diversity in lagoon environments. Most studies have indicated that

Proteobacteria are the most common bacterial phyla in lagoons. According to Yanez-Montalvo *et al*., (2020), Actinobacteria, Bacteroidetes, Cyanobacteria, and Proteobacteria are common components of microbialites in the Bacalar lagoon. The microbiota in the Bacalar lagoon was composed primarily of Proteobacteria (40-80 %), Cyanobacteria (1-11 %), Bacteroidetes (7-8 %), Chloroflexi (8-14 %), Firmicutes (1-23 %), Planctomycetes (1-8 %), and Verrucomicrobia (1-4 %) (Yanez-Montalvo *et al*., 2020). Okumura *et al*., (2023) results also indicated that Proteobacteria was the dominant phylum among the bacterial flora in the sediment of Nagatsura-Ura Lagoon in the northeastern Pacific coast of Japan. Proteobacteria is a bacterial phylum commonly identified not only in lagoon sediments but also in the lagoon surface water environment, and is also identified as the major phylum. Using 16S rRNA amplicon- and shotgun-based metagenomics techniques, Chaouni *et al*., (2022) studied surface water samples from the Marchica and Oualidia lagoons in Morocco during the summers of 2014 and 2015. The authors found that the majority of bacteria in the samples belonged to the phylum Proteobacteria (25-53 %, 29-29 %), followed by Cyanobacteria (34–12 %, 11-0.53 %), Bacteroidetes (24-16 %, 23-43 %), Actinobacteria (7-11 %, 13-7% %), and Verrucomicrobia (4-1 %, 15-14 %), in Marchica and Oualidia, respectively.

Less than 3 % of the 16S rRNA gene sequences found in the Tam Giang, Nai, and Thi Nai lagoons were found to be archaea; representing a negligible fraction of the metagenome. Our study's findings are consistent with those of Chaouni *et al*. (2022), who demonstrated that barely 1 % of the 16S rRNA gene sequences recovered from the metagenome of Moroccan water samples from the Marchica and Oualidia lagoons was attributable to archaea. These results corroborate other studies showing that archaea are a class of microorganisms that favor deep water, where they are found in notably large populations (Signori *et al*., 2014). In our study, Thaumarchaeota was the phylum that accounted for a large proportion in the Nai and Thi Nai lagoons, while Pacearchaeota was the largest proportion in the Tam Giang lagoon. Thaumarchaeota represents a unique phylum within the archaea domain that includes ammonia-oxidizing microorganisms found in soil, marine waters, and hot springs (Stieglmeier *et al*., 2014). An examination of the distribution and community structure of benthic bacteria and archaea was conducted in a stratified coastal lagoon located in the Southern Gulf of Mexico. The findings showed that whereas 35 phyla made up the bacterial community composition in the sediments from Celestún lagoon, only 22 lineages were represented in >1 % abundance. Proteobacteria, Chloroflexi, Bacteroidetes, and Planctomycetes were the principal taxa; in order of prevalence, followed by *Spirochaetae*, Acidobacteria, Actinobacteria, Cyanobacteria, and Verrucomicrobia. In each sediment sample that was studied, 26–34 % of the relative abundance was attributed to Proteobacteria (Gómez-Acata *et al*., 2023). The percentage of Chloroflexi varied from 7 to 28 %, with uncultivated Dehalococcoidia members (1-4 %) and *Anaerolineaceae* (3–10 %) being the most abundant. *Marinilabiaceae* (1–5 %), *Flavobacteriaceae* (1-4 %), and *Saprospiraceae* were identified within the Bacteroidetes (9-17 %). The *Planctomycetaceae* family accounted for 1–5 % of the primary components of the detected Planctomycetes (3–7 %). Twelve archaeal phyla were connected to the 16S rRNA gene sequences that were recovered from the Celestún sediments. With a range in their relative abundances of 15-54 %, 8-36 %, 7-30 %, 12-25 %, and 7-21 %; respectively, Euryarchaeota, Thaumarchaeota, Woesearchaeota (DHVEG–6), Bathyarchaeota, and Lokiarchaeota were the most represented groups in all the studied samples. Aenigmarchaeota, the ancient archaeal group (AAG), dipherotrites, and the WSA2 environmental groups were among the other lineages with modest abundances of ~2 % that were also found (Cadena *et al*., 2019). Thus, our research results are similar to those of many previous studies analyzing the structures of bacteria and archaea present in lagoons worldwide. Studies have shown that in most studied lagoons, Proteobacteria accounts for the largest

proportion of the bacterial community, while Thaumarchaeota accounts for the largest proportion of archaeal community present in the lagoon. No previous studies have assessed the diversity of fungi in lagoon environments. This is the first study that examines fungal diversity in a lagoon setting, as far as we are aware. Coastal lagoons are significant aquatic environments with notable physicochemical gradients and established microorganisms engagement in biogeochemical cycles. The diversity and distribution of microorganisms in these settings; however, has not received enough attention up to this point. As a result, this study offers helpful details regarding the microbial makeup of this natural setting. In this study, functional diversity of the microbial communities in each lagoon (Tam Giang, Thi Nai, and Nai) was determined using shotgun metagenomics. The results showed that several genes related to antibacterial, antifungal, and cancer cell inhibition activities accounted for the majority of the three studied lagoons. Amongst these genes, *NRPS* were the most numerous, accounting for 41-105 genes, followed by *RiPP*-like, *Aryl polyene*, and *RRE*-containing, accounting for 444 genes. However, *T1PKS*, *T3PKS*, *Ectoine*, *Terpene*, *HglE-KS*, *Betalactone*, *Siderophore*, *Resorcinol*, *Ranthipeptide*, *PUFA*, *Redox-cofactor*, and *Hserlactone* were less numerous, accounting for 1–21 genes. Additionally, a small number of genes related to lipolytic, antiviral, natural pigment, antioxidant activity, tuberculosis treatment, and cardiovascular and cancer diseases, were detected in the microflora of each lagoon. Using the shotgun metagenomics approach, analysis of the functional gene diversity of microbial communities in lagoon environments has also been conducted in many different lagoons around the world. Two Moroccan lagoons; mainly Marchica and Oualidia, possessed genes that have been implicated in protein, carbohydrate, and amino acid metabolism (Chaouni *et al*., 2022). Functional genes involved in methanogenesis (*mcrA*), methanotrophy (*pmoA*), and sulfate reduction (*dsrAB*) were revealed at Clipperton lagoon in the North Pacific Ocean; through combination of sequencing cloned bacterial and archaeal 16S rRNA genes and capillary electrophoresis single-strand polymorphism (CE-SSCP) fingerprinting (Galand *et al*., 2009). Metagenome analysis of the Venice lagoon; one of the largest Adriatic transitional systems, has highlighted significant differences among these sites with regard to biochemical processes (e.g., C, N, Fe, and S metabolism) and cell-cell interaction strategies (e.g., mobilome, regulation, and cell signaling) (Banchi *et al*., 2021). This study is the first to show that lagoon microorganisms possess a large number of functional genes related to biological activities that are valuable for diseases treatment. Thus, it can be observed that coastal lagoons have great potential to exploit microorganisms with useful biological activities.

Conclusion

This study is the first to evaluate the diversity of microbial communities in the sediments of three important lagoons in central Vietnam: The Tam Giang, Nai, and Thi Nai lagoons. Proteobacteria and Firmicutes were the two bacterial phyla with the largest proportions in the three lagoons; however, there were significant differences in the composition of each phylum. All three lagoons exhibited high fungal diversity among eukaryotic species, with the phylum Ascomycota accounting for the largest proportion in all three lagoons. Regarding archaea, Thaumarchaeota was the dominant phylum in the Thi Nai and Nai lagoons, whereas in Tam Giang Lagoon, the dominant archaeal phylum was Pacearchaeota. Overall, the Tam Giang lagoon had the highest microbial diversity, followed by the Thi Nai and Nai lagoons. This result is acceptable as the Tam Giang lagoon is considered to be the richest and most diverse in terms of biological resources in Vietnam's coastal lagoon system; with a lagoon area of up to 52 km^2, followed by the Thi Nai lagoon; with an area of 50 km^2, and the Nai lagoon having the smallest area of only 12 km^2. Microbial composition of sediments in these lagoons reflected the amount of organic matters and urban pollutants in the environment. This microbial composition has been found to be intimately related to the carbon, nitrogen, and sulfur cycles in the ecosystem. Metagenomics sequencing results demonstrated high abundance of

genes related to antibacterial, antifungal, antiviral, cancer cell inhibition, natural pigmentation, and antioxidant activities in the three tested lagoons. This indicated that the lagoon is a potential environment for exploiting biologically active microorganisms.

Acknowledgement

None.

Conflict of interests

The authors have no relevant conflicts of interest.

Funding source

This research was financially supported by Ministry of Science and Technology, Vietnam (Fund code no.: ĐTĐL.CN-114/21).

Ethical approval

Non-applicable.

Authors' Contributions

T.T.H.: Conceptualization, Methodology, Formal analysis, Software, Writing – Review and Edit-ing, Supervision; N.V.A.: Conceptualization, Methodology, Formal analysis, Software, Writing – Review and Editing; N.T.N.A.: Conceptualization, Methodology, Formal analysis, Software, Writing - Original Draft; N.V.P.: Conceptualization, Methodology, Formal analysis, Software, Writing – Review and Editing.

5. References

Aramaki, T.; Blanc-Mathieu, R.; Endo, H.; Ohkubo, K.; Kanehisa, M.; Goto, S. et al. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics. 36: 2251-2252. https://doi.org/10.1093/bioinformatics/btz859.

Bairoch, A. and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. Nucleic Acids Research. 28: 45-48. https://doi.org/10.1093/nar/28.1.45.

Banchi, E.; Del Negro, P.; Celussi, M. and Malfatti, F. (2021). Sediment Features and Human Activities Structure the Surface Microbial Communities of the Venice Lagoon. Frontiers in Marine Science. 8: 1-20. https://doi.org/10.3389/fmars.2021.762292.

Barone, R.; De Santi, C.; Esposito, F.P.; Tedesco, P.; Galati, F.; Visone, M. et al. (2014). Marine metagenomics, a valuable tool for enzymes and bioactive compounds discovery. Frontiers in Marine Science. 1: 38. https://doi.org/10.3389/fmars.2014.00038.

Beghini, F.; McIver, L.J.; Blanco-Míguez, A.; Dubois, L.; Asnicar, F.; Maharjan, S. et al. (2021). Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with biobakery 3. Elife. 10: 1-42. https://doi.org/10.7554/eLife.65088.

Bellemain, E.; Carlsen, T.; Brochmann, C.; Coissac, E.; Taberlet, P. and Kauserud, H. (2010). ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases. BMC Microbiology. 10: 189. https://doi.org/10.1186/1471-2180-10-189.

Buchfink, B.; Xie, C. and Huson, D.H. (2014). Fast and sensitive protein alignment using DIAMOND. Nature Methods. 12: 59-60. https://doi.org/10.1038/nmeth.3176.

Cadena, S.; Aguirre-Macedo, M.L.; Cerqueda-García, D.; Cervantes, F.J.; Herrera-Silveira, J.A. and García-Maldonado, J.Q. (2019). Community structure and distribution of benthic Bacteria and Archaea in a stratified coastal lagoon in the Southern Gulf of Mexico. Estuarine, Coastal and Shelf Science. 230: 106433. https://doi.org/10.1016/j.ecss.2019.106433.

Camacho, C.; Coulouris, G.; Avagyan, V.; Ma, N.; Papadopoulos, J.; Bealer, K. et al. (2009). BLAST^{+}:

Architecture and applications. BMC Bioinformatics. 10: 421. https://doi.org/10.1186/1471-2105-10-421.

Cao, T.T.T.; Vu, T.L. and Nguyen, H.C. (2009). Water quality in Tam Giang - Cau Hai lagoon. In: Proceedings of 2009 Vietnam Conference of Marine Science, Hue, Vietnam. pp. 123-132.

Chaouni, B.; Azami, A.I.; Essayeh, S.; Arrafiqui, E.H.; Bailal, A.; Raoui, S. et al. A (2022). Moroccan Lagoon Microbiomes. Water. 14: 1715. https://doi.org/10.3390/w14111715.

Galand, P.E.; Casamayor, E.O.; Kirchman, D.L. and Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. Proceedings of the National Academy of Sciences of the United States of America. 106: 22427–22432. https://doi.org/10.1073/pnas.0908284106.

Gómez-Acata, E.O.; Teutli, C.; Falcon, L.I.; García-Maldonado, J.Q.; Prieto-Davó, A.; Yanez-Montalvo, A. et al. (2023). Sediment microbial community structure associated to different ecological types of mangroves in Celestún, a coastal lagoon in the Yucatan Peninsula, Mexico. PeerJ. 11: e14587. https://doi.org/10.7717/peerj.14587.

Kamble, P. and Vavilala, S.L. (2018). Discovering novel enzymes from marine ecosystems: A metagenomic approach. Botanica Marina. 61(2): 161-175. https://doi.org/10.1515/bot-2017-0075.

Kanehisa, M.; Goto, S.; Kawashima, S.; Okuno, Y. and Hattori, M. (2004). The KEGG resource for deciphering the genome. Nucleic Acids Research. 32: D277–D280. https://doi.org/10.1093/nar/gkh063.

Le, T.T.T. (2021). Salinity intrusion and livelihoods of communities living along Thi Nai lagoon, Binh Dinh province. Journal of Scientific and Environmental Resources. 36: 82-96.

Lin, C.; Wan, J.; Lu, Y.; Zhang, H.; Chen, X.; Su, Y. et al. (2019). Active bacterial communities of pig fecal microbiota transplantation suspension prepared and preserved under different conditions. AMB Express. 9: 63. https://doi.org/10.1186/s13568-019-0787-4.

Luong, C.V. and Nga, N.T. (2017). Research on carbon storage capability of sea grass through biomass (in the case of Thi Nai lagoon, Binh Dinh province). Vietnam Journal of Marine Science and Technology. 17: 63-71. https://doi.org/10.15625/1859-3097/17/1/7900.

Mahapatra, G.P.; Raman, S.; Nayak, S.; Gouda, S.; Das, G. and Patra, J.K. (2020). Metagenomics Approaches in Discovery and Development of New Bioactive Compounds from Marine Actinomycetes. Current Microbiology.77:645-656. https://doi.org/10.1007/s00284-019-01698-5.

Newman, D.J. and Cragg, G.M. (2016). Drugs and Drug Candidates from Marine Sources: An Assessment of the Current "state of Play". Planta Medica. 82: 775-789. https://doi.org/10.1055/s-0042-101353.

Okumura, Y.; Furutono, T.; Ito, K. and Hara, M. (2023). Influence of submarine topography and sediment environment on microbial assemblages in a coastal lagoon in northeastern Japan. Marine Pollution Bulletin. 194: 115404. https://doi.org/10.1016/j.marpolbul.2023.115404.

O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R. et al. (2016). Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. Nucleic Acids Research. 44: D733-D745. https://doi.org/10.1093/nar/gkv1189.

Phan, D.N.; Vo, S.T.; Nguyen, V.L.; Hua, T.T. and Nguyen, A.K. (2016). Characteristics of benthic resources at central Vietnam Lagoons. Journal of Marine Science and Technology. 16(1): 80-88. https://doi.org/10.15625/1859-3097/16/1/6491.

Saurav, K.; Costantino, V.; Venturi, V. and Steindler, L. (2017). Quorum sensing inhibitors from the sea discovered using bacterial N-acyl-homoserine

lactone-based biosensors. Marine Drugs. 15: 53. https://doi.org/10.3390/md15030053.

Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B. et al. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. Applied and Environmental Microbiology. 75: 7537-7541. https://doi.org/10.1128/AEM.01541-09.

Sharma, P.K.; Sharma, M. and Kaur, J. (2012). Metagenomic Approach to Explore Microbial Diversity and Genetic Potential of Uncultured Microorganisms from Different Environment Niches. Research Journal of Pharmaceutical, Biological and Chemical Sciences. 3(2): 947–956.

Signori, C.N.; Thomas, F.; Enrich-Prast, A.; Pollery, R.C.G. and Sievert, S.M. (2014). Microbial diversity and community structure across environmental gradients in Bransfield Strait, Western Antarctic Peninsula. Frontiers in Microbiology. 5: 647. https://doi.org/10.3389/fmicb.2014.00647.

Sinclair, L.; Osman, O.A.; Bertilsson, S. and Eiler, A. (2015). Microbial community composition and diversity *via* 16S rRNA gene amplicons: Evaluating the illumina platform. PLoS One. 10: 1-18. https://doi.org/10.1371/journal.pone.0116955.

Stieglmeier, M.; Alves, R.J.E. and Schleper, C. (2014). The Phylum Thaumarchaeota. In: The Prokaryotes - Other Major Lineages of Bacteria and The Archaea. Rosenberg, E. (Editor). Springer-Verlag, Berlin, Heidelberg. pp. 347-362. https://doi.org/10.1007/978-3-642-38954-2.

Yanez-Montalvo, A.; Gómez-Acata, S.; Águila, B.; HernándezArana, H. and Falcón, L.I. (2020). The microbiome of modern microbialites in Bacalar Lagoon, Mexico. PLoS One. 15 (3): e0230071. https://doi.org/10.1371/journal.pone.0230071.