Journal of Marine Science and Technology

Volume 31 | Issue 4 Article 19

Microbial community and genome structure in coastal seawater from a commercial port and nearby offshore island in northern Taiwan

Chi-Yu Shih
Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan,
d97340002@email.ntou.edu.tw

Ching-Hsiang Chin
Departent of Bioscience and Biotechnology, National Taiwan Ocean University, Keelung, Taiwan

Shiow-Yi Chen
Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan

Chun-Ru Hsu
Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan

Lee-Kuo Kan
Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan

See next page for additional authors

Microbial community and genome structure in coastal seawater from a commercial port and nearby offshore island in northern Taiwan

Authors

Chi-Yu Shih, Ching-Hsiang Chin, Shiow-Yi Chen, Chun-Ru Hsu, Lee-Kuo Kan, and Wen-Shyong Tzou

RESEARCH ARTICLE

Microbial Community and Genome Structure in Coastal Seawater From a Commercial Port and Nearby Offshore Island in Northern Taiwan

Chi-Yu Shih [a,b,c,*], Ching-Hsiang Chin [d], Shiow-Yi Chen [a,d], Chun-Ru Hsu [a,b,d], Lee-Kuo Kang [a,b,c], Wen-Shyong Tzou [a,b,d,**]

[a] Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan
[b] Taiwan Ocean Genome Center, National Taiwan Ocean University, Keelung, Taiwan
[c] Bachelor Degree Program in Marine Biotechnology, National Taiwan Ocean University, Taiwan
[d] Department of Bioscience and Biotechnology, National Taiwan Ocean University, Keelung, Taiwan

Abstract

The pollution stemming from the use of petroleum by boats, the presence of deceased marine life, hazardous chemicals, and wastewater is a major issue in human-made commercial ports. This pollution poses a considerable challenge to the diverse organisms inhabiting the seawater. To comprehend the impact of this pollution on the microbiome, we conducted a study in which we collected surface water samples from a commercial port called Keelung Port, as well as a nearby offshore island known as Keelung Islet. These locations are situated in northern Taiwan, along the Northwestern Pacific Ocean.

Through the use of whole-metagenome shotgun sequencing, we made an intriguing discovery regarding the bacterial composition in the seawater of Keelung Port. Pseudomonadota emerged as the predominant phylum, with notable abundance observed in Oceanospirillaceae, Rhodobacteraceae, Halieaceae, and Arcobacteraceae families. In contrast, Keelung Islet exhibited the highest abundance of Flavobacteriaceae, alongside a noteworthy presence of Euryarchaeota and Uroviricota. These findings provide compelling evidence of the significant impact human activities have on the microbial community within the seawater of commercial ports. These human activities cause notable alterations in the original microbial composition observed in Keelung Islet. Functional analysis of Kyoto Encyclopedia of Genes and Genomes (KEGG) and Clusters of Orthologous Groups of proteins (COG) revealed that microbes in the seawater of Keelung Port have the ability to degrade oil pollution, adapt to their environment through gene products for sulfur and nitrogen metabolisms, and exhibit hallmarks such as chemotaxis, biofilm formation, and secretion systems.

Exploration of the microbial genome derived from the coastal seawater of the commercial port also led to the discovery of three distinct genomic islands, each harboring an array of genes. Within these genetic islands, we uncovered components such as tyrosine-type recombinase/integrase, DNA-invertase, immunity protein, helix-turn-helix domain, and glutathione-regulated potassium-efflux system protein. This significant finding indicates that genomic islands may act as pivotal elements in facilitating horizontal gene transfer. Such transfers enable microorganisms to adapt more effectively to the demanding and complex environment found within human-made port settings.

Keywords: Microbiome, Coastal seawater, Commercial port and offshore island, Genomic island, Environmental adaptation, Northwestern Pacific Ocean

Received 6 June 2023; revised 18 November 2023; accepted 21 November 2023.
Available online 15 December 2023

* Corresponding author at: Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan.

** Corresponding author at: Center of Excellence for the Oceans, National Taiwan Ocean University, Keelung, Taiwan.
E-mail addresses: d97340002@email.ntou.edu.tw (C.-Y. Shih), wstzou@ntou.edu.tw (W.-S. Tzou).

https://doi.org/10.51400/2709-6998.2726

1. Introduction

In recent years with the increasing awareness of environmental protection, there has been a rise in public concern for environmental issues. Studies of various industrial pollution [1], urban pollution, agricultural pollution [2], air pollution [3], water pollution [4], and soil pollution [5] have shown a rapid increase in research activities. Additionally, owing to the decreasing cost and widespread use of gene sequencing technologies in recent years, research on microbial species in the environment that are difficult to cultivate in the laboratory has also been growing. Next-generation sequencing techniques have been widely applied in studying environmental microbial communities, including 16S rRNA sequencing, 18S rRNA sequencing, and shotgun metagenomic sequencing [6]. Various software tools have also been developed to analyze sequencing data.

More and more metagenomics studies investigating the symbiotic bacteria in the intestines of animals as well as the unculturable microorganisms in various environments have emerged lately. For example, there are studies exploring the microbial composition of seawater, rivers, and harbors. One study conducted shotgun sequencing on samples collected from the Apatlaco River in Mexico to understand the water quality of the river and gain a deeper understanding of the relationship between the environment and microbial communities [7]. The research team further investigated the antibiotic resistance and heavy metal metabolism of microorganisms in the river. They discovered that anthropogenic pollution creates selective pressure, leading some bacteria to develop the ability to degrade polyethylene terephthalate (PET) and polystyrene [8]. In India, researchers conducted metagenomic studies on the Ganges River, one of the most heavily polluted rivers in the world, to explore whether the microbial communities have changed as a result of pollution. They collected samples from the Yamuna River, a tributary of the Ganges, before and after the monsoon season and found significant differences in microbial communities. The dominant bacteria in June were Acinetobacter, while in November, Aeromonas dominated. They also discovered numerous gene clusters related to nitrogen and sulfur metabolism, metal tolerance, xenobiotic degradation, and aromatic compound degradation (such as toluene, xylene, benzene, and phenol). Additionally, they found the presence of many antibiotic-resistance genes and observed that multidrug-resistant (MDR) genes were predominantly associated with efflux mechanisms [9]. Furthermore, in São Pedro stream in Brazil, researchers found that the upstream river water mainly consisted of common and widespread bacteria, whereas downstream river water, after sewage discharge, was contaminated with human pathogenic bacteria such as *Burkholderia*, *Shigella*, and *Salmonella*. These bacteria carried antibiotic resistance and metal tolerance genes, indicating the pollution and impact caused by human waste [10]. Chinese researchers discovered a decrease in nitrogen fixation ability in polluted mangroves sediment on Hainan Island, along with an increase in methane production and sulfate reduction, leading to an increase in greenhouse gas emissions. They also found metal tolerance genes and antibiotic resistance genes in the bacteria [11]. Another group of Chinese researchers conducted metagenomics analysis on the surface sediment layer (10 cm) of four sampling points in the Pearl River Delta of Greater Bay Area. They found that *Planctomycetes* and sulfate-reducing bacteria were correlated with the degree of contamination in the marine sediment. They also discovered a correlation between the presence of antibiotic resistance genes and drug efflux genes in bacteria and the degree of contamination in the marine sediment [12].

In the context of ports, most of the research has focused on commercial ports. The unique pollution from commercial ship fuel includes polycyclic aromatic hydrocarbons (PAHs), which exert selective pressure and further alter the microbial community composition [13]. For example, researchers studying commercial ports in the northwestern Mediterranean Sea found that bacteria in the sea surface microlayer (SML), which is the top 1 mm of the seabed surface, had the ability to degrade PAHs [14]. Furthermore, other researchers used Operational Taxonomic Units (OTUs) analysis techniques in the West Istria Sea in Croatia and identified a co-occurrence network among five different genera within the phylum Actinomycetota, which helps them adapt to heavy metal and PAH pollution [15]. Additionally, researchers collected microbial samples from sediments in the coastal areas of the Mediterranean Sea and the French Atlantic coast using the same techniques. They discovered a core co-occurrence network composed of bacteria from the phyla Gammaproteobacteria and Deltaproteobacteria, archaea from the phyla Thaumarchaeota and Bathyarchaeota, and eukaryotic organisms such as protozoa and dinoflagellates. PAHs had the greatest impact on the predator-prey relationships between dinoflagellates and Actinomycetota in this network [16]. In another instance, an oil spill caused

by a collision between two ships along the coast of the Bay of Bengal in January 28, 2017, led Indian scientists to collect seawater and coastal sediment samples near the accident site on February 1, 2017. They found a significant presence of Acinetobacter and Methylococcales within the γ-proteobacteria, and the metabolite analysis revealed a significant increase in the ability of microorganisms to degrde xenobiotic compounds [17].

Horizontal Gene Transfer (HGT) has long been recognized as the primary mechanism for gene transfer, not only influencing an organism's selective advantage but also shaping the interconnected web of life as we know it [18]. In prokaryotes, HGT primarily occurs through three well-recognized mechanisms: conjugation, transformation, and transduction. Detecting HGT generally relies on identifying phylogenetic conflicts and the presence of foreign DNA, such as genomic islands. The detection of phylogenetic conflict involves the construction of reference species and gene trees to pinpoint discrepancies in branching patterns between them [19]. Another method of detecting HGT is by identifying the existence of genomic islands within a chromosome. Genomic islands are DNA segments that serve as vehicles for exchanging genetic material between microorganisms. Extensive evidence suggests that genomic islands play pivotal roles in environmental adaptation, pathogenesis, and evolution. By utilizing genomic islands as vehicles for genetic material, HGT becomes an evolutionary pattern that facilitates the evolution of microorganisms [20]. Several methods have been developed to discover genomic islands in microbial genomes, including the identification of biochemical features such as tRNA located on one side of the genomic island, the presence of direct repeats at the island's borders, analysis of sequence composition, GC content, and the presence of integrase/recombinase/transposase genes. Sequence comparisons between genome sequences are also used to identify genomic islands [21]. Genomic islands can serve various functions that confer advantages to their host organisms. These functions may include the integration of genes related to pathogenicity, virulence, symbiosis, antibiotic resistance, toxin production, and metabolic processes [20]. For example, Hackl et al. discovered a novel family of DNA transposons called "tycheposons" in the marine picocyanobacterium *Prochlorococcus* [22]. The presence of a nitrate assimilation cassette within the tycheposon of *Prochlorococcus* serves as clear evidence of environmental adaptation. Furthermore, an increasing body of evidence demonstrates a high frequency of horizontal gene transfer occurring among microorganisms in the oceans [23].

There is limited research on the microbial analysis of seawater, rivers, and ports in Taiwan. This led us to wonder if the microbial communities in Keelung Port, where a large number of cargo ships, cruise ships, and yachts navigate daily, may be different from those in other locations and what those differences might be. This study aimed to investigate the microbial species and communities present in the seawater of Keelung Port and nearby offshore Keelung Islet as a far less polluted control. We explored the dominant species in the microbial communities and predicted gene functions associated with microbes found in the seawater of two sampling sites.

2. Materials and methods

Seawater was collected from Keelung Port and Keelung Islet, with the latitude and longitude coordinates of the sampling locations being 25°7′53″N 121°44′26″E and 25°11′22″N 121°47′3″E, respectively (Fig. 1). The sampling times for the two locations were 8:30 am on July 19th and 8:30 am on July 20th, 2022, at Keelung Port and Keelung Islet in the sea, respectively. 10 L of surface seawater was collected and stored in two plastic bottles with a capacity of 6 L each. A total of 4 bottles containing approximately 20 L of seawater were transported to the laboratory at room temperature.

The seawater samples transported to the laboratory were quickly filtered through a 1.2 μm filter paper to remove sediment from the seawater, leaving a clean, sediment-free seawater sample in a sterile serum bottle. The clean, sediment-free seawater sample was then filtered through a 0.20 μm filter membrane, retaining the membrane and the microorganisms on it. The membrane was cut into strips of 1 cm width using sterile scissors and forceps, packed into a 50 mL centrifuge tube, labeled, and stored in a −20 °C freezer.

Total genomic DNA from samples was extracted using the column-based method (QIAamp Power-Fecal DNA Kit, Qiagen). All the samples were processed separately. Sequencing was performed on DNA extracted from seawater for whole genome sequencing to construct a library without PCR amplification or selection to avoid artificial errors caused by PCR amplification. After library construction, the sequencing samples went through the quality control steps and then were sequenced on the Illumina NovaSeq 6000 platform to obtain sequence information from the sample.

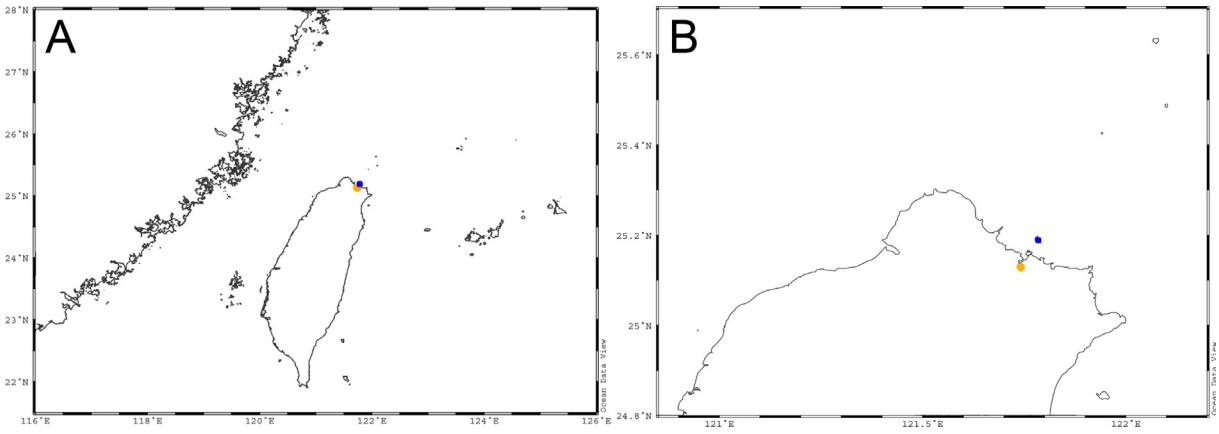


Fig. 1. Map of the sampling sites of seawater. (A) The sampling site is in northern Taiwan facing the Northwestern Pacific Ocean. (B) Sampling points. Sampling sites in a commercial port (Keelung Port) are denoted as orange circles and those in a nearby offshore island (Keelung Islet) as blue boxes.

SqueezeMeta software was used for sequence analysis [24,25]. SqueezeMeta-1.5.0 was used to conduct the analysis of raw data from the whole-metagenome shotgun sequencing. "Coassembly" mode in SqueezeMeta was adopted. In such a running mode, we used Trimmomatic [26]for filtering, MEGAHIT [27] for assembly, Bowtie2 [28] for mapping, diamond run for COG [29], Pfam [30] and KEGG [31] annotation, and MaxBin [32], MetaBAT2 [33], and DAS Tool [34] for binning. We used DIAMOND [35] runs for KEGG ID (orthologs) annotation against the latest publicly available version of KEGG database. KEGG pathway analysis was achieved by using SQMtools [25] using output of SqueezeMeta as input. TPM for KEGG and COG analysis was used as the measurement of abundance. Additionally, R and in-house Python scripts were used for data visualization to facilitate observation. We used IslandViewer 4 [36] to search for genomic islands in the assembled genomes. tRNAscan-SE [37] was used to search for tRNA sites. Pfam and hhsuite (https://github.com/soedinglab/hh-suite) [38], a remote homolog detection tool, was used to predict the molecular function of the protein. Only evalue (1e-10) was setup as the cutoff (for HMMER3 [39]). REPuter was used to reveal the direct repeat sequences on the borders of the genomic islands [40].

3. Results and discussion

3.1. Taxonomic analysis using whole-metagenome shotgun sequencing on the seawater from Keelung Port and Keelung Islet

We collected the seawater from two sampling sites, Keelung Port and Keelung Islet in July, 2022. Both sampling sites are located in the Northwestern Pacific Ocean and in the northeastern tip of Taiwan (Fig. 1). Keelung Port is the second biggest commercial port in Taiwan and Keelung Islet is an isolated island 7.8 km from Keelung port. Metagenome data were obtained from DNA extracted from the seawater samples (see Materials and Methods).

We used SqueezMeta [24] to analyze the fastq data and found a total of 1,020,375 contigs with a total length of 799,913,040 base pairs (bps). The longest contig was 555,513 bp, while the shortest was 200 bp. N50 and N90 were 908 and 362 bps, respectively. In the seawater samples from Keelung Port and Keelung Islet, there were a total of 867,996 and 1,136,246 open reading frames (ORF) fragments, respectively. After comparing these fragments to KEGG database, 416,790 and 393,790 ORFs were identified and screened to obtain 389,685 and 362,192 high-quality ORFs. In addition to the KEGG database, comparisons were also made to COG database, resulting in 540,739 and 503,762 ORFs, which were subsequently screened to obtain 506,128 and 466,909 high-quality ORFs. Furthermore, comparisons to the Pfam protein family database yielded 338,401 and 298,638 ORFs for the two samples, respectively.

Subsequently, we investigated the taxonomy of contigs. At the phylum level, Pseudomonadota and Bacteroidota were the two most abundant phyla. In the seawater of Keelung Port, 53.1% and 16.2% of microbes detected belong to Pseudomonadota and Bacteroidota, respectively. 13.2% and 10.2% of microbes detected belonged to Unclassified and Unmapped. In the seawater of Keelung Islet, 25.5% and 15.9% of microbes detected belonged to Unclassified and Unmapped. 13.7% and 7.6% of microbes detected belong to Pseudomonadota and Bacteroidota, respectively (Fig. 2). Moreover,

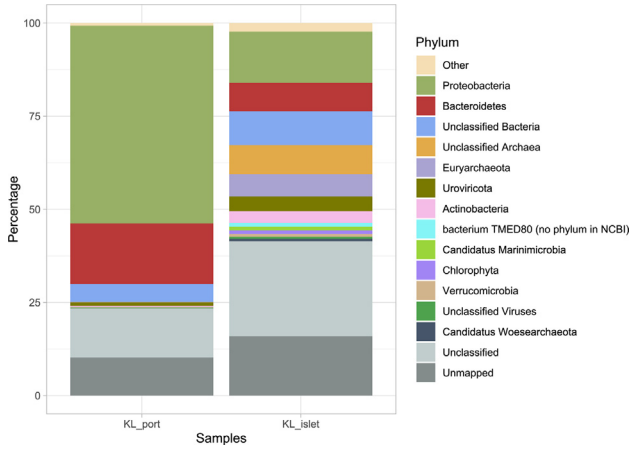


Fig. 2. Microbial composition (in percentage of phylum) in Keelung Port and Keelung Islet. The distribution of major contributing phyla based on the analysis of whole-metagenome shotgun sequencing.

Euryarchaeota and Uroviricota comprised 6.0% and 4.0%, respectively.

At the family level, the most abundant classified families in Keelung Port were Oceanospirillaceae (12.8%), followed by Rhodobacteraceae (9.3%), Flavobacteriaceae (7.2%), Halieaceae (3.4%) and Arcobacteraceae (3.3%). In Keelung Islet, the most abundant classified family was Flavobacteriaceae (3.9%) (Supplementary Fig. S1). At the species level, the classified species with the highest abundance in Keelung Port sample was Marinobacterium (5.8%).

Interestingly, we found a major phage Myoviridae (phylum Uroviricota) (2.4%) in Keelung Islet seawater. Myoviridae is a family of lytic phage with bacteria and archaea as the hosts. It had been found in the North Sea and the photic zone in the Pacific Ocean [41,42].

Numerous members within the Oceanospirillaceae family have the capability to break down petroleum compounds and secrete bactericidal compounds or melanin [43–47]. Rhodobacteraceae is a widely distributed species in seawater, especially in shallow water and surface sediments. Halieaceae is a gram-negative bacterium and is typically found in coastal, open, and deep-sea waters [48]. Arcobacteraceae has an unusually wide range of habitat distribution with the highest prevalence found in raw sewage and wastewater treatment plants, and some of its species may be pathogens of humans and animals [49].

In Keelung Islet, unlike Keelung Port, there was no clear dominance of specific species. Instead, the most abundant species were unknown, with only possible classification as bacteria, archaea, or eukaryotes. However, we could infer that the microbial community in Keelung Islet sample is very different from that in the Keelung Port sample, which was dominated by bacteria, because there was a significant proportion of archaea and eukaryotes in Keelung Islet.

3.2. Functional analysis using KEGG

We discovered 10 KEGG functions with a higher abundance in the KEGG functions in Keelung Islet than in Keelung Port (Fig. 3): ribonucleoside-diphosphate reductase alpha chain, ribonucleoside-diphosphate reductase beta chain, thymidylate synthase (FAD), T4 virus DNA polymerase, DNA primase/helicase, T7 virus DNA-directed DNA polymerase, twinkle protein, photosystem II P680 reaction center D1 protein, chaperonin GroEL, and trimeric autotransporter adhesin. These functions are involved with DNA synthesis, pyrimidine metabolism, carbon metabolism, DNA replication, protein folding, cell adhesion and infection of host cells, photosystems and viral DNA synthesis.

In contrast, KEGG functions with higher abundance in Keelung Port than in Keelung Islet were DNA polymerase I, iron complex outer membrane receptor protein, 3-oxoacyl-[acyl-carrier protein] reductase, putative transposase, and acyl-CoA dehydrogenase. These functions are involved in DNA synthesis, iron metabolism, fatty acid synthesis, transposition, and fat and nucleic acid metabolism.

KEGG function analysis also showed that the four most abundant functions associated with antibiotics resistance were more abundant in Keelung Port than in Keelung Islet (Fig. 4). These functions were K11004 (ATP-binding cassette, subfamily B, bacterial HlyB/CyaB, ABC transporters), K12541 (ATP-binding cassette, subfamily C, bacterial LapB, ABC transporters), K06147 (ATP-binding cassette, subfamily B), and K05595 (multiple antibiotic resistance protein).

These functions, associated with ATP-binding cassette transporters and multiple antibiotic resistance proteins, are well-known contributors to bacterial resistance against antibiotics [50,51]. The elevated abundance of these functions in Keelung Port indicates a potentially higher capacity for antibiotic resistance among microbial communities in this particular environment. The presence of higher levels of antibiotic resistance-associated functions in Keelung Port raises concerns about the potential for increased antibiotic resistance in the surrounding ecosystem. This finding underscores the importance of monitoring and understanding the factors that contribute to antibiotic resistance in order to develop effective strategies for mitigating the spread and impact of antibiotic resistance in aquatic environments like Keelung Port.

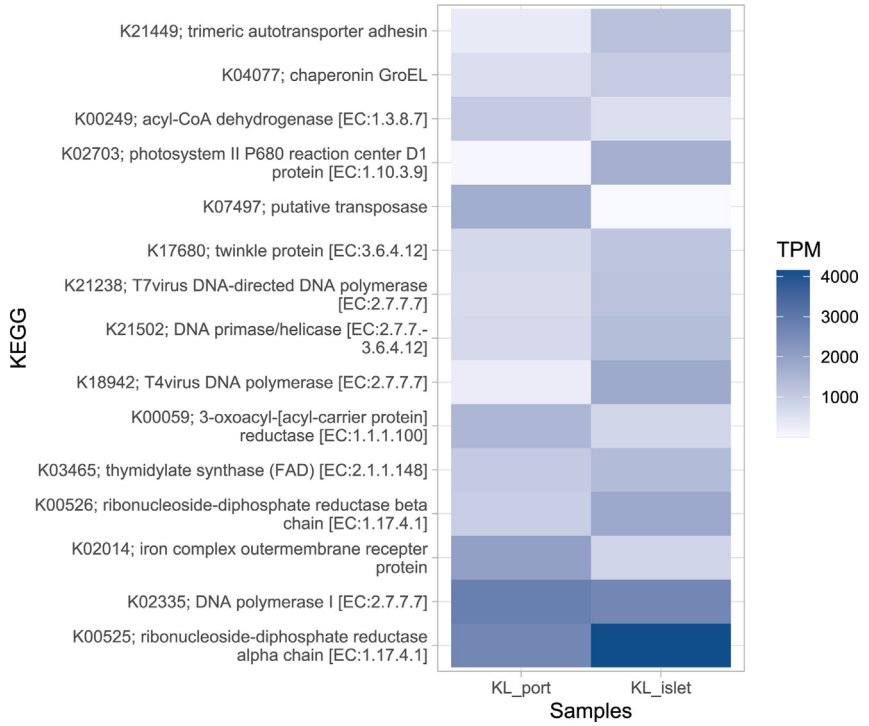


Fig. 3. Heatmaps showing the relative abundance (in tpm) of KEGG orthologs. Samples were collected from Keelung Port (KL_port) and Keelung Islet (KL_islet), respectively.

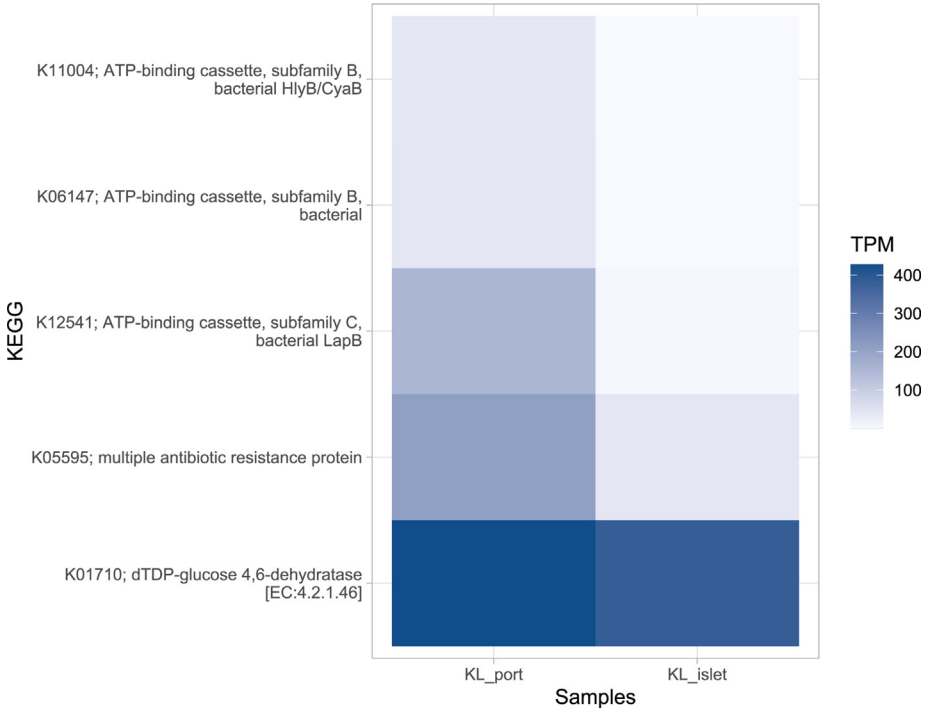


Fig. 4. Heatmaps showing the relative abundance (in tpm) of the antibiotics resistance. Samples were collected from Keelung Port (KL_port) and Keelung Islet (KL_islet), respectively.

3.3. Functional analysis using COG

The COG analysis demonstrated that the top 15 proteins with the highest abundance were identified as Histidine kinase, Ribonucleotide reductase (alpha subunit), DNA polymerase I -3′-5′ exonuclease and polymerase domains, DNA primase (bacterial type), NAD-dependent aldehyde dehydrogenase, Cation/multidrug efflux pump, Ribonucleotide reductase (beta subunit), Nucleoside-diphosphate-sugar epimerase, 5′-3′ exonuclease (including N-terminal domain and PolI), Dehydrogenase with different specificities (related to short-chain dehydrogenase), Thiol-disulfide isomerase and thioredoxins, predicted alternative thymidylate synthase, Acyl-CoA dehydrogenase, Enoyl-CoA hydratase/carnitine racemase, and Acyl-CoA synthetase (AMP-forming)/AMP-acid ligase II (Fig. 5).

Histidine Kinases, which are part of the two-component system, are conserved signaling elements present in bacteria and function as key regulators in regulating the virulence of pathogenic bacteria. Histidine kinase can not only sense the environmental cues but can also orchestrate the transcription [52].

The functions that showed higher abundance in the Keelung Port sample than in the Keelung Islet sample were related to signal and substance transmission, energy production, fatty acid metabolism and multidrug efflux. In Keelung Islet, the proteins related to DNA replication and repair as well as nucleotide transport and DNA excision showed higher abundance.

3.4. Biological pathway analysis using KEGG

We also discovered KEGG pathways highly abundant in the seawater of Keelung Port or Keelung Islet. Pathways abundant in Keelung Port include ko00010 Glycolysis/Gluconeogenesis, ko00624 Polycyclic aromatic hydrocarbon degradation, ko00910 Nitrogen metabolism, ko00920 sulfur metabolism, ko02030 Bacterial chemotaxis, ko05111 Biofilm formation, and ko03070 Bacterial secretion system. However, Pathways abundant in Keelung Islet include ko03030 DNA replication, ko03410 Base excision repair and ko03420 Nucleotide excision repair.

These results were consistent with the function analysis of KEGG and COG. Functional analysis of KEGG and COG revealed that microbes in the seawater of Keelung Port potentially have the ability to degrade oil pollution, adapt to their environment through gene products for sulfur and nitrogen metabolisms, and exhibit hallmarks such as chemotaxis, biofilm formation, and secretion systems.

We found a distinct microbiome and its associated functions in the seawater from Keelung Port, in comparison with those from Keelung Islet. This observation coincided with the water quality in Keelung Harbor, which has approximately 140 discharge outlets from Keelung city and collects the

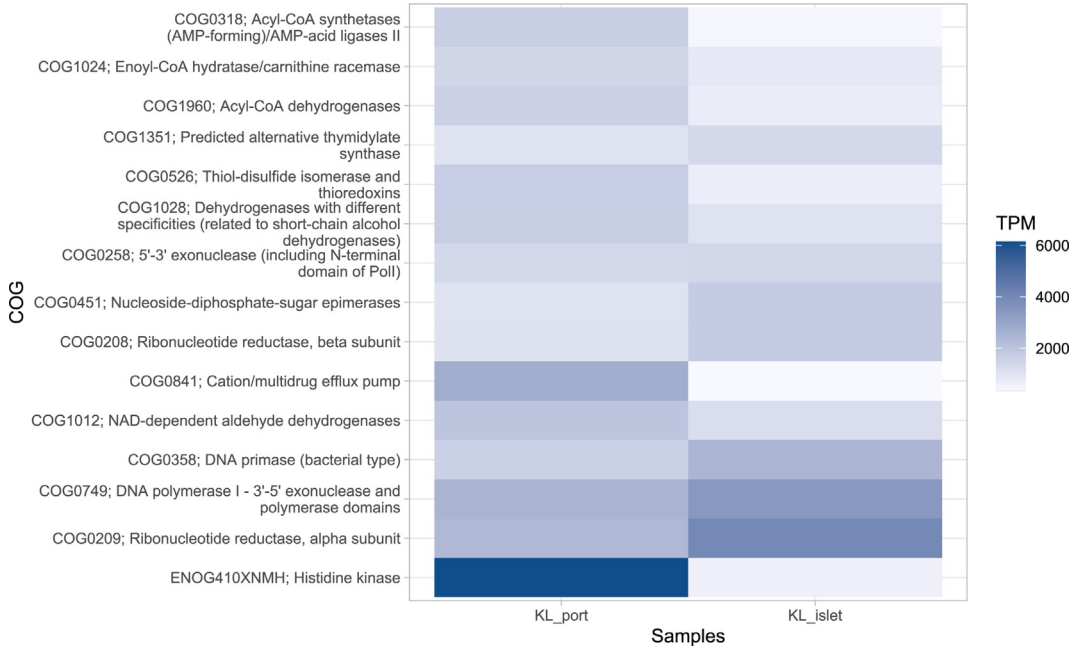


Fig. 5. Heatmaps showing the relative abundance (in tpm) of COG. Samples were collected from Keelung Port (KL_port) and Keelung Islet (KL_islet), respectively.

effluent of four drainage channels (Port of Keelung Environmental Report, https://www.twport.com.tw/Upload/A/RelFile/CustomPage/3834/5a2f2557-4f97-4bee-9a7f-73a23404f3a8.pdf). The polluted water provides nutrition and poses an environmental challenge for the microbes, despite the efforts of the runoff wastewater interceptors.

3.5. Genomic islands discovered from the binning results of metagenomes

We assembled the genomes using binning methods, resulting in 132 genomes. Subsequently, we chose two genomes with high completeness (>90%) and low contamination (<10%). This process of binning yielded two genomes meeting the criteria of high completeness and low contamination: Bin.01 (95.55% and 5.4%, completeness and contamination, respectively) and Bin.02 (92.02% and 7.7%). Bin.01 was classified to phylum Bacteroidota (family Flavobacteriaceae) and Bin.02 was classified to phylum Pseudomonadota (family Bacteriovoracaceae, genus Bacteriovorax). Both of them were more abundant in the seawater of Keelung Port than in Keelung Islet (202 fold more abundant for both Bins). Bin.01 is 2.14 Mbases (GC 37.12%) in length and Bin.02 3.67 Mbases in length (GC 37.24%).

Interestingly, we found two genomic islands in Bin.01 and one Bin.02. For all genomic islands identified, characteristic low GC content, bordered direct repeat, and tRNA genes are all revealed (Figs. 6–8). In Bin.01, genomic islands 01 and 02 harbor six (Fig. 6) and seven genes (Fig. 7), respectively. In Bin.02, genomic island 01 contains 10 genes (Fig. 8). For all the genomic islands, the transcription orientation is either one direction or bi-directions divided in the middle of the island. While some of the genes encode protein with no functional annotation, we used remote homolog detection to discover the function.

Genomic island 01 of Bin.01 harbors Lipoprotein, Helix-turn-helix domain, Immunity protein 9, and tyrosine-type recombinase/integrase. Genomic island 02 of Bin.01 harbors Tyrosine recombinase XerC, 30S ribosomal protein S21, Acyl-CoA dehydrogenase, helix-hairpin-helix domain-containing protein, alanine:cation symporter family protein, Glutathione-regulated potassium-efflux system protein KefC and PspC domain-containing protein.

Genomic island 01 of Bin.02 harbors AsmA (assembly of outer membrane proteins), ATP-dependent RecD-like DNA helicase, AAA (ATPases Associated with diverse cellular Activities)-like domain, Peptidase C39 family, Replication initiation factor, helix-turn-helix domain-containing protein and tyrosine recombinase XerC.

Genomic islands [53] have been discovered in seawater and proved to confer tolerance of copper and oxidative stress on coastal marine *Synechococcus* [54,55] and have a wide distribution in the environment [56]. Even the diversification of a pelagic *Polynucleobacter* species was shown to be driven by genomic islands [57]. Our findings highlight the presence of genomic islands in the assembled

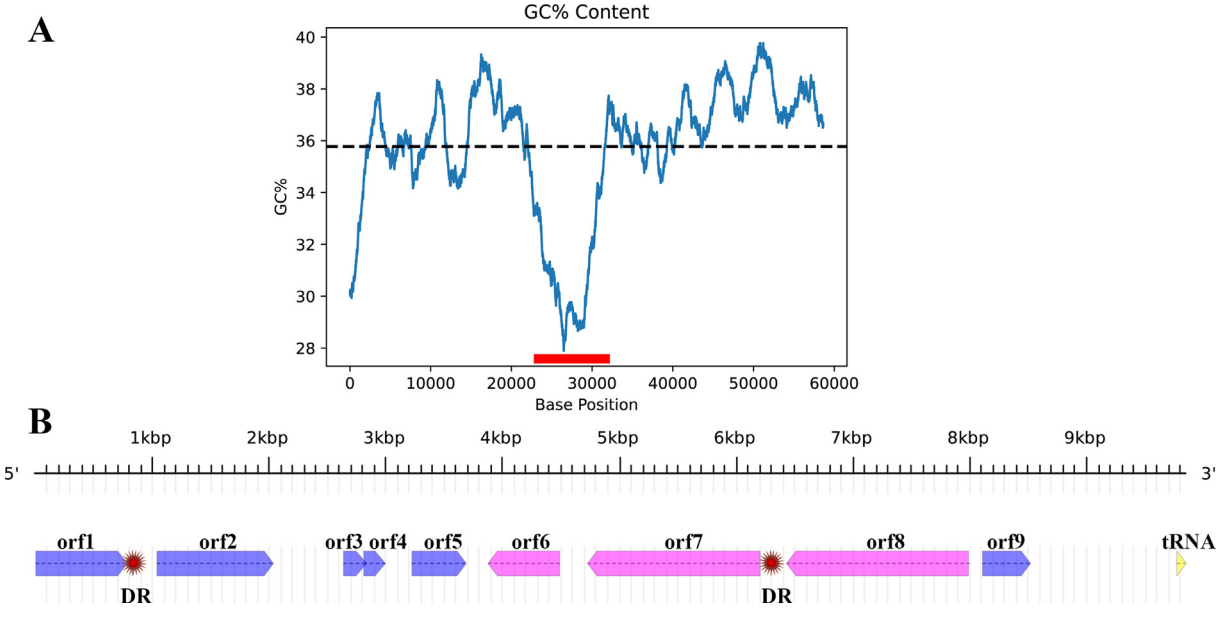


Fig. 6. GC content (A) and genomic island 01 (B) in the genome of Bin.01 (family Flavobacteriaceae) discovered in the costal seawater of Keelung Port. (A) GC content in percentage using the scanning window of 3000 bps. (B) Gene clusters in the genomic island are shown. tRNA genes are labeled with yellow triangles. "DR" (in a red circle) indicates a direct repeat on two flanking of the genomic islands.

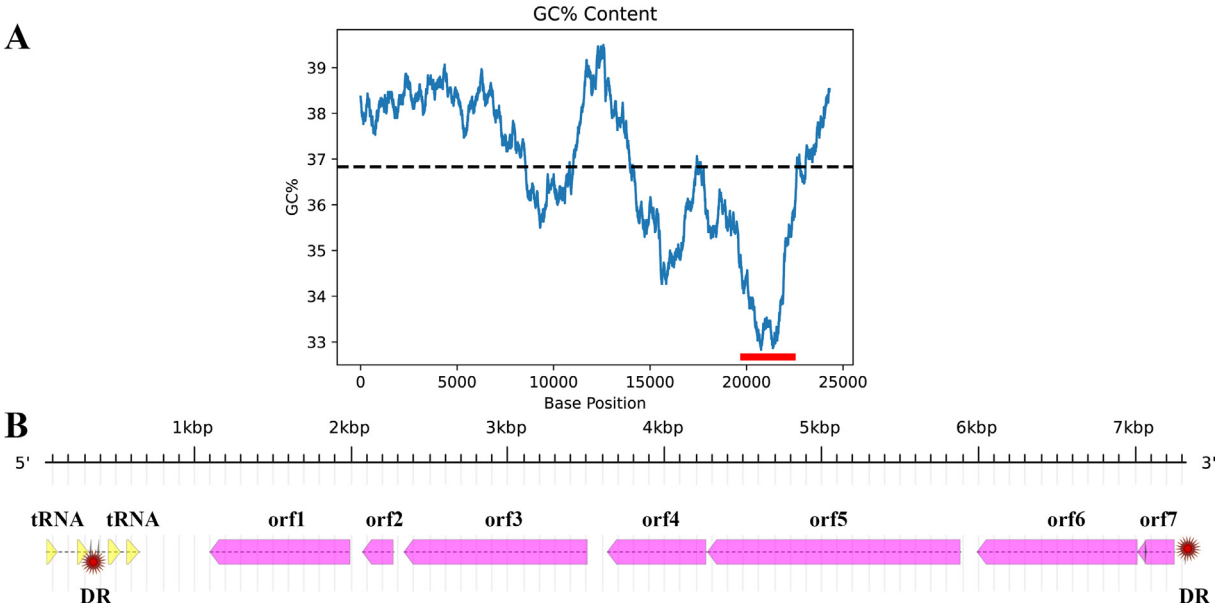


Fig. 7. GC content (A) and genomic island 02 (B) in the genome of Bin.01 (family Flavobacteriaceae) discovered in the costal seawater of Keelung Port. (A) GC content in percentage using the scanning window of 3000 bps. (B) Gene clusters in the genomic island are shown. tRNA genes are labeled with yellow triangles. "DR" (in a red circle) indicates a direct repeat on two flanking of the genomic islands.

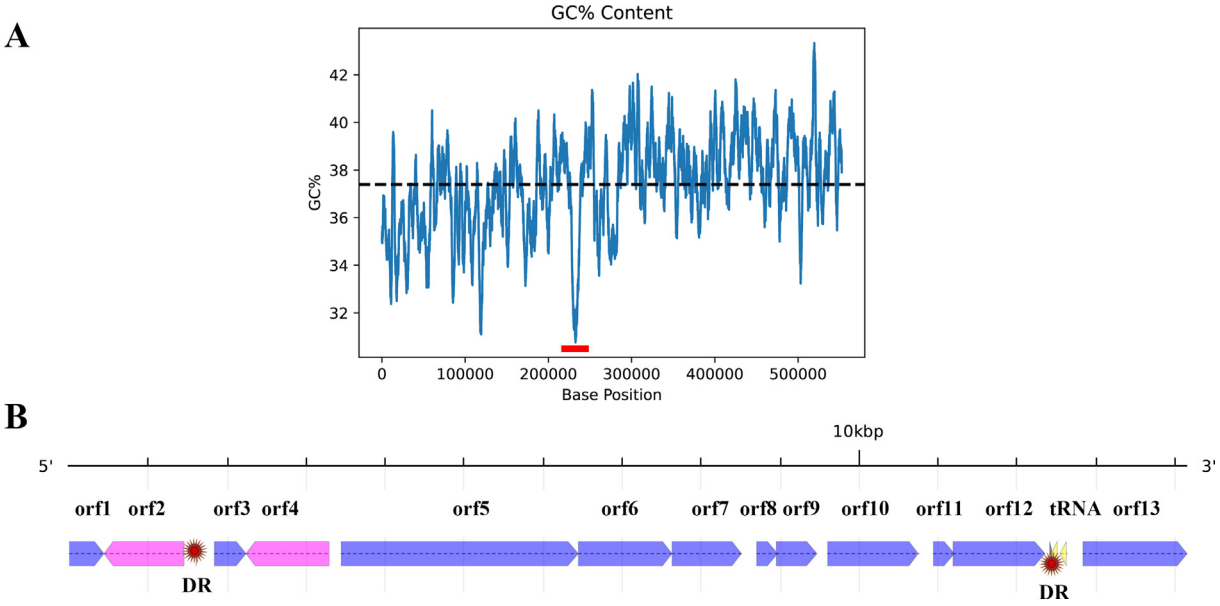


Fig. 8. GC content (A) and genomic island 01 (B) in the genome of Bin.02 (family Bacteriovoracaceae, genus Bacteriovorax) discovered in the costal seawater of Keelung Port. (A) GC content in percentage using the scanning window of 3000 bps. (B) Gene clusters in the genomic island are shown. tRNA genes are labeled with yellow triangles. "DR" (in a red circle) indicates a direct repeat on two flanking of the genomic islands.

genomes and their potential role in conferring unique functionalities to the microorganisms in Keelung Port. Acyl-CoA dehydrogenase (discovered in Genomic island 01 of Bin.01) is capable of catalyzing the initial step in fatty acid β-oxidation and associated with the regulation of redox status. Acyl-CoA dehydrogenase was found to be abundant in the seawater of Keelung Port in KEGG function analysis. Alanine:cation symporter family protein (discovered in Genomic island 02 of Bin.01) is an active amino acid transporter. AsmA, discovered in Genomic island 01 of Bin.02, is essential for the assembly of outer membrane proteins in *Escherichia coli*. It played a crucial role in the formation of pores that functioned in the passive diffusion of small molecules. The identification of these specific genes

within these islands provides valuable insights into the adaptive mechanisms and potential biological processes occurring in these microbial communities. Further research and functional characterization of these genomic islands and their associated genes would enhance our understanding of their contribution to the ecology and adaptation of microorganisms in Keelung Port.

4. Conclusion

The main bacterial phylum in the Keelung Port samples was Pseudomonadota, with high abundance of Oceanospirillaceae, Rhodobacteraceae, Halieaceae, and Arcobacteraceae. Oceanospirillaceae are bacteria associated with the degradation of petroleum compounds, which can be directly linked to the use of fossil fuels by ships in the port. Arcobacteraceae are associated with human and animal pathogens. In contrast, there were many unknown microbes in Keelung Islet seawater. The high abundance of genes associated with specific functions in Keelung Islet and Keelung Port seawater indicates the microbial adaptation to distinct environments. The identification of genomic islands offers insights into potential mechanisms of adaptation, potentially involving the horizontal transfer of genes. The identification of genomic islands within the commercial port indicates the possibility of horizontal gene transfer among microbes, which could potentially facilitate their adaptation to stressful environments.

Funding

The sources of financial support for this study is from Ministry of Science and Technology, Taiwan (Training Program for Interdisciplinary Talents of Precision Health). The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

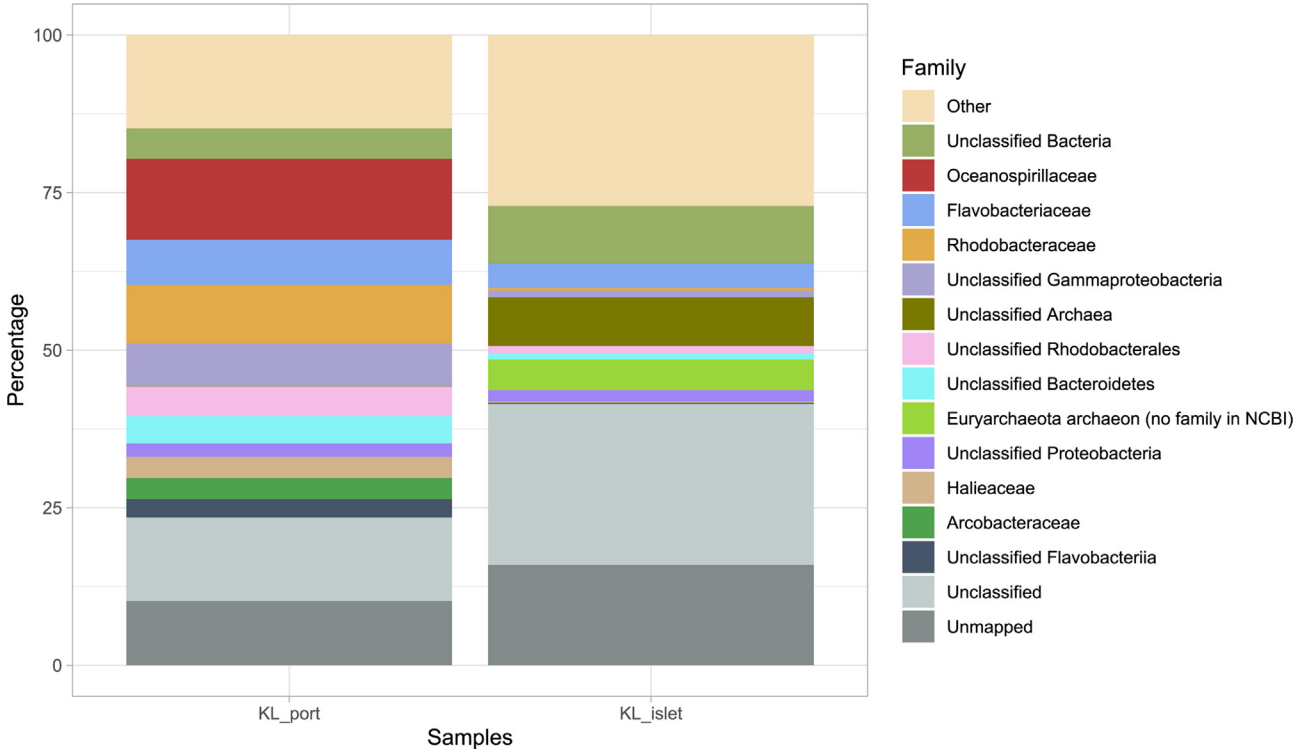
Conflict of interest

The authors have declared that no conflicting interests exist.

Acknowledgments

We would like to express our gratitude to the Keelung Port Authority for approving the sampling of seawater during the study period. We also thank Yu-An Wu for generating GC% plots for contig sequences.

Appendix A.



Supplementary Fig. S1. Microbial composition (in percentage of family) in Keelung Port (KL_port) and Keelung Islet (KL_islet). The distribution of major contributing families based on the analysis of whole-metagenome shotgun sequencing.

References

[1] Sun W, Chen Z, Wang D. Can Land Marketization Help Reduce Industrial Pollution? Int J Environ Res Publ Health 2019;16(12):2213. https://doi.org/10.3390/ijerph16122213.

[2] Sun B, Zhang L, Yang L, Zhang F, Norse D, Zhu Z. Agricultural non-point source pollution in China: causes and mitigation measures. Ambio 2012;41(4):370−9. https://doi.org/10.1007/s13280-012-0249-6. Epub 2012/02/09. PubMed PMID: 22311715; PubMed Central PMCID: PMCPMC3393061.

[3] Tiotiu AI, Novakova P, Nedeva D, Chong-Neto HJ, Novakova S, Steiropoulos P, Kowal K. Impact of Air Pollution on Asthma Outcomes. Int J Environ Res Publ Health 2020;17(17):6212. https://doi.org/10.3390/ijerph17176212.

[4] Reddy B, Dubey SK. Exploring the allochthonous pollution influence on bacterial community and co-occurrence dynamics of River Ganga water through 16S rRNA-tagged amplicon metagenome. Environ Sci Pollut Control Ser 2021;28(21):26990−7005. https://doi.org/10.1007/s11356-021-12342-w.

[5] Lillini R, Tittarelli A, Bertoldi M, Ritchie D, Katalinic A, Pritzkuleit R, et al. Water and Soil Pollution: Ecological Environmental Study Methodologies Useful for Public Health Projects. In: A literature review. Springer International Publishing; 2020. p. 179−214.

[6] Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. Metagenomic biomarker discovery and explanation. Genome Biol 2011;12(6):R60. https://doi.org/10.1186/gb-2011-12-6-r60.

[7] Breton-Deval L, Juárez K, Sánchez-Flores A, Vera-Estrella R. Water quality assessments and metagenomic analysis of the polluted river Apatlaco. Mexico. bioRxiv 2019:536128. https://doi.org/10.1101/536128.

[8] Breton-Deval L, Sanchez-Reyes A, Sanchez-Flores A, Juarez K, Salinas-Peralta I, Mussali-Galante P. Functional Analysis of a Polluted River Microbiome Reveals a Metabolic Potential for Bioremediation. Microorganisms 2020;8(4). https://doi.org/10.3390/microorganisms8040554. Epub 2020/04/16. PubMed PMID: 32290598; PubMed Central PMCID: PMCPMC7232204.

[9] Mittal P, Prasoodanan Pk V, Dhakan DB, Kumar S, Sharma VK. Metagenome of a polluted river reveals a reservoir of metabolic and antibiotic resistance genes. Environmental Microbiome 2019;14(1):5. https://doi.org/10.1186/s40793-019-0345-3.

[10] Medeiros JD, Cantao ME, Cesar DE, Nicolas MF, Diniz CG, Silva VL, et al. Comparative metagenome of a stream impacted by the urbanization phenomenon. Braz J Microbiol 2016;47(4):835−45. https://doi.org/10.1016/j.bjm.2016.06.011. Epub 2016/08/16. PubMed PMID: 27522532; PubMed Central PMCID: PMCPMC5052392.

[11] Li Y, Zheng L, Zhang Y, Liu H, Jing H. Comparative metagenomics study reveals pollution induced changes of microbial genes in mangrove sediments. Sci Rep 2019;9(1):5739. https://doi.org/10.1038/s41598-019-42260-4. Epub 2019/04/07. PubMed PMID: 30952929; PubMed Central PMCID: PMCPMC6450915.

[12] Chen J, McIlroy SE, Archana A, Baker DM, Panagiotou G. A pollution gradient contributes to the taxonomic, functional, and resistome diversity of microbial communities in marine sediments. Microbiome 2019;7(1):104. https://doi.org/10.1186/s40168-019-0714-6. Epub 2019/07/17. PubMed PMID: 31307536; PubMed Central PMCID: PMCPMC6632204.

[13] Miettinen H, Bomberg M, Nyyssonen M, Reunamo A, Jorgensen KS, Vikman M. Oil degradation potential of microbial communities in water and sediment of Baltic Sea coastal area. PLoS One 2019;14(7):e0218834. https://doi.org/10.1371/journal.pone.0218834. Epub 2019/07/03. PubMed PMID: 31265451; PubMed Central PMCID: PMCPMC6605675.

[14] Sauret C, Tedetti M, Guigue C, Dumas C, Lami R, Pujo-Pay M, et al. Influence of PAHs among other coastal environmental variables on total and PAH-degrading bacterial communities. Environ Sci Pollut Res Int 2016;23(5):4242−56. https://doi.org/10.1007/s11356-015-4768-0. Epub 2015/07/01. PubMed PMID: 26122564.

[15] Duran R, Bielen A, Paradzik T, Gassie C, Pustijanac E, Cagnon C, et al. Exploring Actinobacteria assemblages in coastal marine sediments under contrasted Human influences in the West Istria Sea, Croatia. Environ Sci Pollut Res Int 2015;22(20):15215−29. https://doi.org/10.1007/s11356-015-4240-1. Epub 2015/02/26. PubMed PMID: 25712885.

[16] Jeanbille M, Gury J, Duran R, Tronczynski J, Agogue H, Ben Said O, et al. Response of Core Microbial Consortia to Chronic Hydrocarbon Contaminations in Coastal Sediment Habitats. Front Microbiol 2016;7:1637. https://doi.org/10.3389/fmicb.2016.01637. Epub 2016/10/30. PubMed PMID: 27790213; PubMed Central PMCID: PMCPMC5061854.

[17] Neethu CS, Saravanakumar C, Purvaja R, Robin RS, Ramesh R. Oil-Spill Triggered Shift in Indigenous Microbial Structure and Functional Dynamics in Different Marine Environmental Matrices. Sci Rep 2019;9(1):1354. https://doi.org/10.1038/s41598-018-37903-x.

[18] Soucy SM, Huang J, Gogarten JP. Horizontal gene transfer: building the web of life. Nat Rev Genet 2015;16(8):472−82. https://doi.org/10.1038/nrg3962. Epub 2015/07/18. PubMed PMID: 26184597.

[19] Williams D, Fournier GP, Lapierre P, Swithers KS, Green AG, Andam CP, Gogarten JP. A rooted net of life. Biol Direct 2011;6:45. https://doi.org/10.1186/1745-6150-6-45. Epub 20110921. PubMed PMID: 21936906; PubMed Central PMCID: PMCPMC3189188.

[20] da Silva Filho AC, Raittz RT, Guizelini D, De Pierri CR, Augusto DW, dos Santos-Weiss ICR, Marchaukoski JN. Comparative Analysis of Genomic Island Prediction Tools. Front Genet 2018;9. https://doi.org/10.3389/fgene.2018.00619.

[21] Langille MGI, Hsiao WWL, Brinkman FSL. Detecting genomic islands using bioinformatics approaches. Nat Rev Microbiol 2010;8(5):373−82. https://doi.org/10.1038/nrmicro2350.

[22] Hackl T, Laurenceau R, Ankenbrand MJ, Bliem C, Cariani Z, Thomas E, et al. Novel integrative elements and genomic plasticity in ocean ecosystems. Cell 2023;186(1):47−62. https://doi.org/10.1016/j.cell.2022.12.006. e16. Epub 2023/01/08. PubMed PMID: 36608657.

[23] McDaniel LD, Young E, Delaney J, Ruhnau F, Ritchie KB, Paul JH. High frequency of horizontal gene transfer in the oceans. Science 2010;330(6000):50. https://doi.org/10.1126/science.1192243. Epub 2010/10/12. PubMed PMID: 20929803.

[24] Tamames J, SqueezeMeta Puente-Sánchez F. A Highly Portable, Fully Automatic Metagenomic Analysis Pipeline. Front Microbiol 2019;9. https://doi.org/10.3389/fmicb.2018.03349.

[25] Puente-Sánchez F, García-García N, Tamames J. SQMtools: automated processing and visual analysis of 'omics data with R and anvi'o. BMC Bioinf 2020;21(1):358. https://doi.org/10.1186/s12859-020-03703-2.

[26] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30(15):2114−20. https://doi.org/10.1093/bioinformatics/btu170.

[27] Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 2015;31(10):1674−6. https://doi.org/10.1093/bioinformatics/btv033.

[28] Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9(4):357−9. https://doi.org/10.1038/nmeth.1923.

[29] Huerta-Cepas J, Szklarczyk D, Forslund K, Cook H, Heller D, Walter MC, et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res 2015;44(D1):D286−93. https://doi.org/10.1093/nar/gkv1248.

[30] Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar Gustavo A, Sonnhammer ELL, et al. Pfam: The protein

families database in 2021. Nucleic Acids Res 2020;49(D1): D412−9. https://doi.org/10.1093/nar/gkaa913.
[31] Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27−30. https://doi.org/10.1093/nar/28.1.27. Epub 1999/12/11. PubMed PMID: 10592173; PubMed Central PMCID: PMCPMC102409.
[32] Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics 2015;32(4):605−7. https://doi.org/10.1093/bioinformatics/btv638.
[33] Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. PeerJ 2019;7:e7359. https://doi.org/10.7717/peerj.7359. Epub 2019/08/08. PubMed PMID: 31388474; PubMed Central PMCID: PMCPMC6662567.
[34] Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. Nature Microbiology 2018;3(7):836−43. https://doi.org/10.1038/s41564-018-0171-1.
[35] Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. Nat Methods 2015;12(1):59−60. https://doi.org/10.1038/nmeth.3176.
[36] Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing G, Lau BY, Hoad G, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. Nucleic Acids Res 2017;45(W1):W30. https://doi.org/10.1093/nar/gkx343. W5. Epub 2017/05/05. PubMed PMID: 28472413; PubMed Central PMCID: PMCPMC5570257.
[37] Chan PP, Lowe TM. tRNAscan-SE: Searching for tRNA Genes in Genomic Sequences. Methods Mol Biol 2019;1962: 1−14. https://doi.org/10.1007/978-1-4939-9173-0_1. Epub 2019/04/26. PubMed PMID: 31020551; PubMed Central PMCID: PMCPMC6768409.
[38] Steinegger M, Meier M, Mirdita M, Vöhringer H, Haunsberger SJ, Söding J. HH-suite3 for fast remote homology detection and deep protein annotation. BMC Bioinf 2019;20(1):473. https://doi.org/10.1186/s12859-019-3019-7.
[39] Eddy SR. A new generation of homology search tools based on probabilistic inference. Genome Inform 2009;23(1): 205−11. PubMed PMID: 20180275.
[40] Kurtz S, Choudhuri JV, Ohlebusch E, Schleiermacher C, Stoye J, Giegerich R. REPuter: the manifold applications of repeat analysis on a genomic scale. Nucleic Acids Res 2001; 29(22):4633−42. https://doi.org/10.1093/nar/29.22.4633.
[41] Hurwitz BL, Sullivan MB. The Pacific Ocean virome (POV): a marine viral metagenomic dataset and associated protein clusters for quantitative viral ecology. PLoS One 2013;8(2): e57355. https://doi.org/10.1371/journal.pone.0057355. Epub 2013/03/08. PubMed PMID: 23468974; PubMed Central PMCID: PMCPMC3585363.
[42] Wichels A, Biel SS, Gelderblom HR, Brinkhoff T, Muyzer G, Schütt C. Bacteriophage diversity in the North Sea. Appl Environ Microbiol 1998;64(11):4128−33. https://doi.org/10.1128/aem.64.11.4128-4133.1998. Epub 1998/10/31. PubMed PMID: 9797256; PubMed Central PMCID: PMCPMC106618.
[43] Hedlund BP, Geiselbrecht AD, Bair TJ, Staley JT. Polycyclic aromatic hydrocarbon degradation by a new marine bacterium, Neptunomonas naphthovorans gen. nov., sp. nov. Appl Environ Microbiol 1999;65(1):251−9. https://doi.org/10.1128/AEM.65.1.251-259.1999. PubMed PMID: 9872786; PubMed Central PMCID: PMCPMC91009.
[44] Yakimov MM, Giuliano L, Gentile G, Crisafi E, Chernikova TN, Abraham WR, et al. Oleispira antarctica gen. nov., sp. nov., a novel hydrocarbonoclastic marine bacterium isolated from Antarctic coastal sea water. Int J Syst Evol Microbiol 2003;53(Pt 3):779−85. https://doi.org/10.1099/ijs.0.02366-0. PubMed PMID: 12807200.
[45] Yakimov MM, Giuliano L, Denaro R, Crisafi E, Chernikova TN, Abraham WR, et al. Thalassolituus oleivorans gen. nov., sp. nov., a novel marine bacterium that obligately utilizes hydrocarbons. Int J Syst Evol Microbiol 2004;54(Pt 1):141−8. https://doi.org/10.1099/ijs.0.02424-0. PubMed PMID: 14742471.
[46] Teramoto M, Suzuki M, Okazaki F, Hatmanti A, Harayama S. Oceanobacter-related bacteria are important for the degradation of petroleum aliphatic hydrocarbons in the tropical marine environment. Microbiology (Read) 2009;155(Pt 10): 3362−70. https://doi.org/10.1099/mic.0.030411-0. Epub 20090618. PubMed PMID: 19541999.
[47] Teramoto M, Ohuchi M, Hatmanti A, Darmayati Y, Widyastuti Y, Harayama S, Fukunaga Y. Oleibacter marinus gen. nov., sp. nov., a bacterium that degrades petroleum aliphatic hydrocarbons in a tropical marine environment. Int J Syst Evol Microbiol 2011;61(Pt 2):375−80. https://doi.org/10.1099/ijs.0.018671-0. Epub 2010/03/23. PubMed PMID: 20305063.
[48] Suzuki T, Yazawa T, Morishita N, Maruyama A, Fuse H. Genetic and Physiological Characteristics of a Novel Marine Propylene-Assimilating Halieaceae Bacterium Isolated from Seawater and the Diversity of Its Alkene and Epoxide Metabolism Genes. Microb Environ 2019;34(1):33−42. https://doi.org/10.1264/jsme2.ME18053. Epub 20190116. PubMed PMID: 30651420; PubMed Central PMCID: PMC PMC6440738.
[49] Venâncio I, Â Luís, Domingues F, Oleastro M, Pereira L, Ferreira S. The Prevalence of Arcobacteraceae in Aquatic Environments: A Systematic Review and Meta-Analysis. Pathogens 2022;11(2):244. https://doi.org/10.3390/pathogens11020244. PubMed PMID:.
[50] Akhtar AA, Turner DP. The role of bacterial ATP-binding cassette (ABC) transporters in pathogenesis and virulence: Therapeutic and vaccine potential. Microb Pathog 2022;171: 105734. https://doi.org/10.1016/j.micpath.2022.105734. Epub 2022/08/26. PubMed PMID: 36007845.
[51] Garmory HS, Titball RW. ATP-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. Infect Immun 2004;72(12):6757−63. https://doi.org/10.1128/iai.72.12.6757-6763.2004. Epub 2004/11/24. PubMed PMID: 15557595; PubMed Central PMCID: PMCPMC529116.
[52] Stock AM, Robinson VL, Goudreau PN. Two-component signal transduction. Annu Rev Biochem 2000;69:183−215. https://doi.org/10.1146/annurev.biochem.69.1.183. Epub 2000/08/31. PubMed PMID: 10966457.
[53] Juhas M, van der Meer JR, Gaillard M, Harding RM, Hood DW, Crook DW. Genomic islands: tools of bacterial horizontal gene transfer and evolution. FEMS Microbiol Rev 2009;33(2):376−93. https://doi.org/10.1111/j.1574-6976.2008.00136.x. Epub 2009/01/31. PubMed PMID: 19178566; PubMed Central PMCID: PMCPMC2704930.
[54] Stuart RK, Brahamsha B, Busby K, Palenik B. Genomic island genes in a coastal marine Synechococcus strain confer enhanced tolerance to copper and oxidative stress. ISME J 2013;7(6):1139−49. https://doi.org/10.1038/ismej.2012.175. Epub 2013/01/25. PubMed PMID: 23344240; PubMed Central PMCID: PMCPMC3660668.
[55] Sanfilippo JE, Nguyen AA, Karty JA, Shukla A, Schluchter WM, Garczarek L, et al. Self-regulating genomic island encoding tandem regulators confers chromatic acclimation to marine Synechococcus. Proc Natl Acad Sci U S A 2016;113(21):6077−82. https://doi.org/10.1073/pnas.1600625113. Epub 2016/05/07. PubMed PMID: 27152022; PubMed Central PMCID: PMCPMC4889380.
[56] Siebor E, Neuwirth C. Overview of Salmonella Genomic Island 1-Related Elements Among Gamma-Proteobacteria Reveals Their Wide Distribution Among Environmental Species. Front Microbiol 2022;13. https://doi.org/10.3389/fmicb.2022.857492.
[57] Hoetzinger M, Schmidt J, Jezberova J, Koll U, Hahn MW. Microdiversification of a Pelagic Polynucleobacter Species Is Mainly Driven by Acquisition of Genomic Islands from a Partially Interspecific Gene Pool. Appl Environ Microbiol 2017;83(3). https://doi.org/10.1128/AEM.02266-16. Epub 2016/11/12. PubMed PMID: 27836842; PubMed Central PMCID: PMCPMC5244307.