UNIVERSIDADE TÉCNICA DO ATLÂNTICO
INSTITUTO DE ENGENHARIA E CIÊNCIAS DO MAR
WEST AFRICAN SCIENCE SERVICE CENTRE ON CLIMATE CHANGE AND ADAPTED LAND USE

Master Thesis

DIVERSITY OF MICROORGANISMS FOUND IN COASTAL WATERS OF SÃO VICENTE (CABO VERDE), WITH ASSESSMENT OF POTENTIALLY PATHOGENIC BACTERIA

OSVALDINA JULIÃO FERNANDES SOARES

Master Research Program on Climate Change and Marine Sciences

São Vicente
2023

UNIVERSIDADE TÉCNICA DO ATLÂNTICO
INSTITUTO DE ENGENHARIA E CIÊNCIAS DO MAR
WEST AFRICAN SCIENCE SERVICE CENTRE ON CLIMATE CHANGE AND ADAPTED LAND USE

Master Thesis

DIVERSITY OF MICROORGANISMS FOUND IN COASTAL WATERS OF SÃO VICENTE (CABO VERDE), WITH ASSESSMENT OF POTENTIALLY PATHOGENIC BACTERIA

Osvaldina Julião Fernandes Soares

Master Research Program on Climate Change and Marine Sciences

Supervisor | Dr. David Needham
Co-supervisor | Dr. António Pinto

São Vicente
2023

UNIVERSIDADE TÉCNICA DO ATLÂNTICO

INSTITUTO DE ENGENHARIA E CIÊNCIAS DO MAR

WEST AFRICAN SCIENCE SERVICE CENTRE ON CLIMATE CHANGE AND ADAPTED LAND USE

Diversity of Microorganisms Found in Coastal Waters of São Vicente (Cabo Verde), with Assessment of Potentially Pathogenic Bacteria

Osvaldina Julião Fernandes Soares

Master's thesis presented to obtain the master's degree in Climate Change and Marine Sciences, by the Institute of Engineering and Marine Sciences, Atlantic Technical University in the framework of the West African Science Service Centre on Climate Change and Adapted Land Use

Supervisor

Dr. David Needham
GEOMAR

Co-supervisor

Dr. António Pinto
Universidade Técnica do Atlântico

São Vicente
2023

UNIVERSIDADE TÉCNICA DO ATLÂNTICO

INSTITUTO DE ENGENHARIA E CIÊNCIAS DO MAR

WEST AFRICAN SCIENCE SERVICE CENTRE ON CLIMATE CHANGE AND ADAPTED LAND USE

Diversity of Microorganisms Found in Coastal Waters of São Vicente (Cabo Verde), With Assessment of Potentially Pathogenic Bacteria

Osvaldina Julião Fernandes Soares

Panel defense

President

Prof. Dr. Jailson Pereira

Examiner 1

MSc Mara Abu-Raya

Examiner 2

Dr. Rohan Sachdeva

Examiner 3

Dr. Yara Rodrigues

São Vicente
2023

Financial support

The German Federal Ministry of Education and Research (BMBF) in the framework of the West African Science Service Centre on Climate Change and Adapted Land Use (WASCAL) through WASCAL Graduate Studies Programme in Climate Change and Marine Sciences at the Institute for Engineering and Marine Sciences, Atlantic Technical University, Cabo Verde.

Dedication

In honor of my family.

For your sacrifice, encouragement and unwavering support.

Acknowledgements

First of all, I would like to acknowledge WASCAL who gave me this opportunity and the necessary means to make it all possible.

A very special thanks to my co-supervisor Dr. António Pinto and supervisor Dr. David Needham and his entire lab group at GEOMAR (Alyzza, Sukkie, Marjan, and Lara) especially Elisa, who helped me throughout the process of carrying out this work. Thank you for your patience, time, guidance, teaching, encouragement and good memories. And also, a special appreciation to Dr. Björn Fiedler for all his assistance during my stay in Germany to carry out the practical part of my thesis.

I am grateful for the help and encouragement given to me by all my friends and colleagues from the third batch throughout the journey we have taken together until here. Thanks also to Emmanuel Tora for all his patience and time in helping me.

My acknowledgment and thanks also go to the entire WASCAL Cabo Verde team who have tirelessly helped me and my colleagues so far.

Last but not least, I deeply thank God and my family who are my support and were with me the entire time and made all of this possible.

Resumo

Nas águas oceânicas de Cabo Verde, os microrganismos predominantes estão adaptados às condições do oceano aberto, prosperando em ambientes com nutrientes limitados, ao mesmo tempo que são influenciados por factores como salinidade, correntes oceânicas, contribuições da terra, condições climáticas, etc. Em qualquer caso, os microrganismos desempenham um papel crucial ao cumprirem uma variedade de propósitos ecológicos. Esses ambientes podem funcionar como reservatórios para uma ampla gama de microrganismos, abrangendo tipos benéficos e potencialmente prejudiciais, incluindo patógenos. Neste estudo foram usadas ferramentas da biologia molecular para extração do material genético das amostras da água do mar superficiais recolhidas em diferentes períodos e áreas da ilha de São Vicente. De seguida foram feitas quantificação e amplificação da região V4 e V5 usando primers universal Forward e Reverse (515F-926R) para sequenciamento para determinar a composição da comunidade geral bacteriana e possíveis patógenos associados à saúde humana. Os resultados demonstraram que proteobactérias, cianobactérias e bacteroidetes dominaram as amostras, o que vai de acordo com o achado de outros autores, sendo elas bastantes predominantes nas águas superficiais do Atlântico e ambientes oligotróficos. A presença de archaea foi notada, mas em quantidade não significante comparada com as bactérias. As bactérias potencialmente patogénicas foram maioritariamente identificadas em áreas de condições favoráveis e de intensa atividade antropogénica, como é o caso do mercado de peixe e a área próxima ao OSCM. Além disso, a maioria das bactérias potencialmente patogénicas foram detectadas durante a estação seca e fria em comparação com o período transicional para época de chuva que, no entanto, não chega a ser intensa na ilha. O reconhecimento de marcadores microbianos cruciais ligados a mudanças na qualidade da água pode fornecer informações para a criação de sistemas de alerta. Isto permite que as autoridades respondam prontamente, mitigando possíveis fontes de poluição e contaminação, reduzindo os riscos para a saúde e garantindo a segurança das áreas marinhas. Em última análise, isto salvaguarda a saúde e o bem-estar das comunidades e ecossistema.

Palavras-chave: Águas superficiais do mar, São Vicente, atividade antropogénica, comunidade bacteriana.

Abstract

In the oceanic waters of Cabo Verde, prevailing microorganisms are adapted to open ocean conditions, thriving in environments with limited nutrients, while being influenced by factors like salinity, ocean currents, inputs from the land, climatic conditions, etc. In any case, microorganisms play a crucial role by fulfilling a variety of ecological purposes. These environments can function as reservoirs for diverse microorganisms, encompassing both beneficial and potentially harmful types, including pathogens. This study used molecular biology tools to extract the genetic material from surface seawater samples collected in different periods and areas of São Vicente Island. Quantification and amplification of the V4 and V5 region were then performed using universal Forward and Reverse primers (515F-926R) for sequencing to determine the composition of the general bacterial community and possible pathogens associated with human health. The results showed that proteobacteria, cyanobacteria and bacteroidetes dominated the samples, which is in agreement with the findings of other authors, as they are quite predominant in surface waters of the Atlantic Ocean and oligotrophic environments. Presence of archaea was noted, but in a non-significant amount compared to bacteria. Potentially pathogenic bacteria were mostly identified in areas with favorable conditions and intense anthropogenic activity, such as the fish market and the area close to OSCM. Furthermore, most potentially pathogenic bacteria were detected during the dry and cold season compared to the transitional period for rainy seasons, which however is not intense on the island. Recognition of crucial microbial markers linked to changes in water quality can provide information for building early warning systems. This allows authorities to respond promptly, mitigating possible sources of pollution and contamination, reducing health risks and ensuring the safety of marine areas. Ultimately, this safeguards the health and well-being of communities and ecosystems.

Keywords: Surface seawater, São Vicente, anthropogenic activity, bacterial community.

Abbreviations and acronyms

WHO	World Health Organization
DNA	Deoxyribonucleic Acid
FIB	Fecal Indicator Bacteria
EPA	Environmental Protection Agency
qPCR	Quantitative polymerase chain reaction
ddPCR	Digital droplet polymerase chain reaction
rRNA	Ribosomal ribonucleic acid
RNA	Ribonucleic acid
NGS	Next Generation Sequencing
iSeq	Integrated Sequencer
MiniSeq	Mini Sequencer
MiSeq	Micro Sequencer
NextSeq	Next-generation Sequencer
GDT	General Directorate of Tourism
BP	Big Pipe
SP	Small Pipe
OSCM	Ocean Science Center Mindelo
CVAO	Cabo Verde Atmospheric Observatory
SF	Size fraction
HCl	Hydrochloric acid
PCR	Polymerase chain reaction
CCGA	Competence Center for Genomic Analysis
ASV	Amplicon sequence variants
WQDIV	Water quality and diversity

List of Contents

Figure index

Table index

1. Introduction

São Vicente Island has several beaches that are very often sought after for fishing and recreational activities both by locals and tourists, making it essential to provide conditions for safety and microbiological quality of beach water to ensure public health and avoid pollution. In order to do so, more information is needed about the baseline contextual conditions as well as conditions that can develop near population centers, and how they can change overtime due to natural variability, anthropogenic factors, and the combination of the two.

The water around Cabo Verde, in particular, is situated in an oligo-to-mesotrophic area (low nutrients) due to being surrounded by deep waters on all sides (Hill et al., 2012; Romero et al., 2016). Occasionally, nutrients are delivered from the propagation of eddies stemming from the complex hydrography of the region, especially the upwelling region at the mainland of Africa. Nevertheless, the waters are typically made up of microorganisms that dominate the open ocean and are adapted to low nutrient conditions (Martínez-García & Pinhassi, 2019).

Diverse physical phenomena, such as upwelling, turbulence driven by the wind, and the deposition of aeolian dust, have demonstrated their capacity to affect microbial communities near the surface water. These events may be connected not only to dust-derived nutrients but also to the associated wind. One of the major sources of fresh nutrients for surface water in Cabo Verde area arises from aeolian dust transported by trade winds from arid areas of Africa, primarily the Sahara Desert (Hill et al., 2012).

Microorganisms play a crucial role in coastal environments and provide a variety of ecological purposes. These functions include decomposition of organic matter, nutrient cycling and preservation of ecosystem balance. Marine microorganisms make up the base of the food chain, including phytoplankton which contributes about half of global primary productivity, heterotrophic bacteria and archaea, which consume about half of this material every day, as well as the microbial organisms that consume the plankton. It is important to emphasize that coastal waters can serve as reservoirs for both beneficial and potentially harmful microorganisms, including pathogens (Choi et al., 2015; Stec et al., 2022).

Waterbodies, especially the coastal oceans, serve as a source of sustenance, a place to work, play, and live, as well as the first line of protection against various natural and man-made risks and disasters. For long-term wellbeing of populations, it is crucial that these ecosystems remain functional and healthy. Due to changes in land use, coastal regions are harmed by pollution

inputs, and everyday inflows of wastes in enormous amounts resulting in change in water quality which can have significant impacts on their microbial communities (Massey, 2015).

According to the World Health Organization (WHO), beach water generally contains a mixture of microorganisms that may come from various sources, such as domestic animals, industrial waste, the population that uses the water (excretion), sewage effluents, etc. These microorganisms, when pathogenic and in sufficient numbers, can cause gastrointestinal diseases, respiratory tract infection, among others (WHO, 2013).

Pollution of coastal areas brought on by human activity is a significant environmental problem that requires immediate action. It brings a variety of risks to marine ecosystems and human health. Due to industrial discharge, improper waste management, and other human activities, coastal waters can become contaminated by pollutants such as chemicals, heavy metals, nutrients, microplastics, and other substances that have a terrible impact on marine habitats and biodiversity. It may lead to the degradation of habitats, the extinction of aquatic species, and the disturbance of delicate ecological balances, often signaled by the change in microbial diversity and composition, which acts as early indicators of environmental deterioration. Numerous factors make it crucial to examine microbial diversity in coastal waters as it provides insight into the state and functioning of marine ecosystems as a whole.

1.1. Objectives of the work

The main goal of this work is to research and characterize the diversity, distributions and abundance of microorganisms present in the coastal waters of São Vicente Island and to provide important insights into public health implications in order to develop an awareness of the population and to enable intervention from competent authorities to safeguard against the impacts of potential contamination.

- To examine the diversity, via DNA sequencing, of bacteria and archaea that are found in the waters surrounding São Vicente, coastal and further offshore;
- To compare the microbiome diversity between different periods of the year;
- To identify and quantify relevant bacteria which might be indicators of human activities.

2. Literature review

Numerous factors, such as salinity, ocean currents, inputs from the land, and climatic conditions, impact the diverse microbial organisms that live in the oceans. These forces influence the delicate balance of marine microorganisms, which directly affects both human health and the health of the ocean ecosystem in addition to driving crucial biogeochemical cycles necessary for life on Earth. The effects of these microorganisms can be profound and immediate, as they operate within fragile equilibria that can have significant consequences for both humans and the environment (American Academy of Microbiology, 2005).

Human activities have the potential to significantly impact the biodiversity of marine microorganisms at both global and local scales. Globally, all organisms in the oceans, including microorganisms, respond to the primary aspect of climate change: temperature fluctuations. Elevated temperatures have been associated with increased numbers of specific waterborne pathogens, while certain microbial genes, including virulence genes, can become activated, leading to disease in humans, corals, and other macroorganisms (Campbell et al., 2015). Locally, human activities such as releasing exotic ballast water can introduce invasive macroorganisms and potentially harmful microorganisms into coastal waters (Drake et al., 2007). Nutrient loading from sources like fish farming and runoff had substantially affected coastal environments, resulting in eutrophication and significant changes in local microbial communities (American Academy of Microbiology, *op cit*).

Human use of coastal waters dates back to antiquity and is used for fishing, recreational activities, sports, etc. Island countries have a high demand for these activities and Cabo Verde is no exception. In May 2015, Cabo Verde approved Decree-Law No. 30/2015, of May 18, which aims to establish the legal regime for the identification, management, monitoring and classification of sea bathing areas and the quality of beach water and provision of information to the public about them, to preserve, protect and to improve the quality of the environment and protecting human health. Also, according to the same Decree-Law, article 25, paragraph 3.b, the monitoring of bathing water must be carried out in areas where there is a greater risk of pollution, understood as the presence of microbiological contamination or other organisms that affect the quality of the water bathing areas and constitute a risk to the health of bathers and ecosystem (Imprensa Nacional de Cabo Verde, 2015). Unfortunately, on what level these rules are being followed needs to be clarified. Cabo Verde only started to have a microbiological analysis system for coastal water in 2015 and for some years these analyses

have been interrupted (Magalhães, 2018). When previously performed, it was only done in some regions of the country and other locations were totally neglected. Beaches with high demand by the local population and tourists must provide safe conditions and the microbiological quality of their water and sand must be a constant concern of public entities (Vieira et al., 2007). Coastal water is a natural type of recreational water where disinfection is impossible. It has a high risk of contamination from point sources such as sewage system leaks and industrial waste, and non-point sources such as streams, culverts, birds and bathers (Standard Methods Committee of the American Public Health Association, 2022a). The efficiency in the treatment of influents in sewage treatment plants, as well as the management and maintenance of these stations, is of great relevance in determining the destination and quality of the effluent, represented by the density of pollutants and contaminants present in it. Leakage from the sewage system due to technical failures in the structures or deliberate discharge of effluent into coastal areas is one of the most important sources of human fecal contamination in recreational aquatic environments (WHO, 2013). Water and wastewater can contain various microorganisms that can cause infections. Human beings are susceptible to contracting these infections through direct contact with these microorganisms that enter the body through ingestion, inhalation, skin wound, etc. (Standard Methods Committee of the American Public Health Association, 2022b). The acceptability of recreational water for its intended usage has been assessed using concentrations of particular microorganisms, also called recreational water quality indicators. They serve as markers whose concentrations in water can be quantitatively connected to the level of possible pathogen contamination and, consequently, dangers to human health from recreational usage. Fecal Indicator Bacteria (FIB), linked to gastrointestinal sickness in various epidemiological studies, are being used by modern methods to evaluate the quality of recreational water (Crain et al., 2021). For example, the Environmental Protection Agency (EPA) of the United State of America recommends using FIB (*Enterococcus spp*.) as markers of fecal pollution for marine water, in the 2012 Recreational Water Quality Criteria. The California Health and Safety Code 115881, article 2 (1995) mandates testing for microbiological pollutants, such as total coliforms, fecal coliforms, and Enterococci bacteria, in the water at all public beaches. In addition to the coastal water FIB, other bacteria are specific to sewage and vary in number according to geographic location and seasonality (WHO, *op cit*.). Fecal coliforms and *Escherichia coli* are still extensively utilized in environmental monitoring. However, Enterococci are becoming a more common target since they can be used to predict health concerns in recreational marine areas. According

to Ryu et al. (2014) there are more than twenty species in the genus *Enterococcus*, and research looking at the variety of Enterococci found in environmental waters have identified the majority of them as the strains: *E. faecalis, E. faecium, E. casseliflavus, E. hirae, E. durans,* and *E. mundtii.* Overall, the proportion of each species varies significantly. However, the first three have the highest reported environmental prevalence (Ryu et al., 2014).

Cabo Verde Decree-Law 30/2015 of May 18, article 27, paragraph 8.a,b regarding assessment of the quality of bathing water stated and I quote: 'Single samples are classified as follows: For coastal bathing waters, "water suitable for bathing" is considered when the value determined for the sample does not exceed 350 cfu/100 ml for fecal streptococci and intestinal enterococci or 1200 cfu/100 ml for *Escherichia coli*; and the water is considered "water unsuitable for bathing" when the values established in the previous paragraph are exceeded.'

Environmental waters can get contaminated by various fecal sources, each offering a unique set of health hazards. Identifying the primary sources of fecal pollution can be aided by knowledge of the prevalence of enterococcal species in the environment. The human fecal source is linked to more significant dangers, but non-human sources of pollution are getting more attention and have been a key factor in beach closures, which can have a considerable economic impact (Ryu et al., *op cit*). According to the same decreet-law, mentioned above, article 31 paragraph 1.f,h regarding management measures stated as follow: For the purposes of the provisions of this law, bathing water management measures must be adopted that include: Detection and assessment of the causes of pollution that may affect bathing water and harm the health of bathers. The development of actions to prevent bathers from being exposed to pollution and risks of natural origin.

Recently, molecular techniques have been strongly used to detect and quantify FIB. Although the culture method can predict health risks, when compared with molecular methods such as quantitative polymerase chain reaction (qPCR) and digital droplet PCR (ddPCR), the latter two proved to be more efficient as they can significantly reduce delays in public notification and facilitate same-day water quality warnings, as culture methods require 18 h to 96 h to produce results (Crain et al., 2021).

Despite the advantages of molecular methods, qPCR has some limitations, such as DNA inhibition due to the high molecular weight of compounds such as humic acid and other complex carbohydrates found in the water source that, when combined with metal ions, sequester the nucleic acid and prevent amplification and can lead to false negative results

(Crain, et al., 2021; Noble et al., 2010). The use of DNA extraction kits, dilution and addition of bovine serum albumin has been shown to be a great way to circumvent the problem of DNA inhibition during qPCR (Kreader, 1996; Noble et al., *op cit*). Targeting the 16S rRNA gene, which has higher sequence conservation and more sequencing data available - because the sequence database for function-specific genes of environmental enterococcal is much more limiting than for 16S rRNA genes - would be an excellent strategy to determine the presence of bacteria in the environmental (water) sample (Ryu et al., 2013).

Nowadays, tools to sequence genes more quickly and cheaply than traditional sequencing are available and are based on massively parallel processing of DNA or RNA fragments, offering ultra-high throughput, scalability and speed. They are called Next Generation Sequencing (NGS) or High-Performance Sequencing. Different platforms (equipment) are available on the market and they differ in technology and method used for mass sequencing. Among them, the most popular is Illumina, which can perform large sequencing with high quality reads for a more affordable price. The process simultaneously identifies nucleotides while incorporating them into a nucleic acid chain. Since its launch in 2007, Illumina has had several versions (iSeq, MiniSeq, MiSeq, NextSeq) differing from each other by the size of reads and number of reads per run. The primary distinction between qPCR and next-generation sequencing (NGS) technologies is discovery power. While qPCR can only detect known sequences, both enable sensitive and accurate variant detection. On the other hand, NGS is a method that does not rely on prior knowledge of sequence data and is hypothesis-free. NGS offers greater sensitivity to measure uncommon variants and transcripts and greater discovery capacity to identify novel genes (Illumina, Inc, 2023).

3. Materials and Methods

The study is based on qualitative and quantitative methods to determine the diversity of bacteria and those potentially pathogenic to humans and consequently determine the water quality in the coastal regions of São Vicente Island. For molecular characterizations of microbial communities via 16S rRNA gene sequencing, we used DNA extraction, DNA quantification, PCR amplification, gel electrophoresis, clean-up, genomic library construction, and DNA sequencing. All procedures were performed according to laboratory and manufacturers' established protocols.

Samples were collected across two distinct time periods: the end of June and July 2022, which are considered transitional months (delineating the shift from a dry to a rainy season) and March 2023, which falls within the dry and cold season on the island. Each time-point the sampling schemes were the same. Further details are found below in section 3.2 - Sample Collection and Filtration.

3.1. Study site

The study was carried out in the marine waters of the Cabo Verde archipelago, located about 500 km from the West African coast. Cabo Verde is an archipelago of volcanic origin located in the Atlantic Ocean, formed by ten islands, including São Vicente Island.

Due to its geographical location in the sub-Sahelian zone, with a subtropical climate, Cabo Verde has an arid and semi-arid, hot and dry climate, with an average annual temperature of around 24ºC and low rainfall. Two seasons can be identified in them, which define the climate of the islands: the breeze season (dry and cool season, from December to June) and the rainy season (hot and humid season, which normally takes place between August and October, July being the transition month) (Gomes, 2014; INE, 2022).

São Vicente is located between the latitudes of 16º55’19” to 16º46’21” N and the longitudes of 24º51’48” to 25º05’40” W (Figure 1) and Table 1 shows the geographic coordinates of the sampling sites. Geologically constituted by the products of a volcano, São Vicente Island has the highest population density in Cabo Verde (374.29 inhabitants/km^2), with 84.964 residents, of which more than 90% live in urban areas and more than 70% of the houses have a septic tank or sewage (INE, 2022). The island has the main port in the country, Porto Grande, and has the second highest GDP of Cabo Verde. Among the various economic activities, tourism has

emerged as one of the main axes of development in the country. Due to the lack of rain for agriculture on the island, fishing is a major primary activity and critical to the sustenance and economy of the island. Forming part of the Barlavento group of islands, São Vicente has a dimension of around 24 km, between *Ponta Machado* and *Ponta Calhau*. From a climatic perspective, the island is characterized by a warm and regular climate, with an average annual temperature of around 24º C, low rainfall compared to the other islands in the Barlavento group and being windy, especially from December to March. According to the General Directorate of Tourism (GDT), the island's coastline comprises black or white sand beaches, areas of dunes and stream valleys, pebble or gravel beaches and rocky, stony and sandy shallow areas. It constitutes one of the country's greatest potentials for economic development emphasizing tourism, maritime, aquaculture and fishing, port and industrial activities, civil construction of public and private works, etc. (GDT, n.d.).

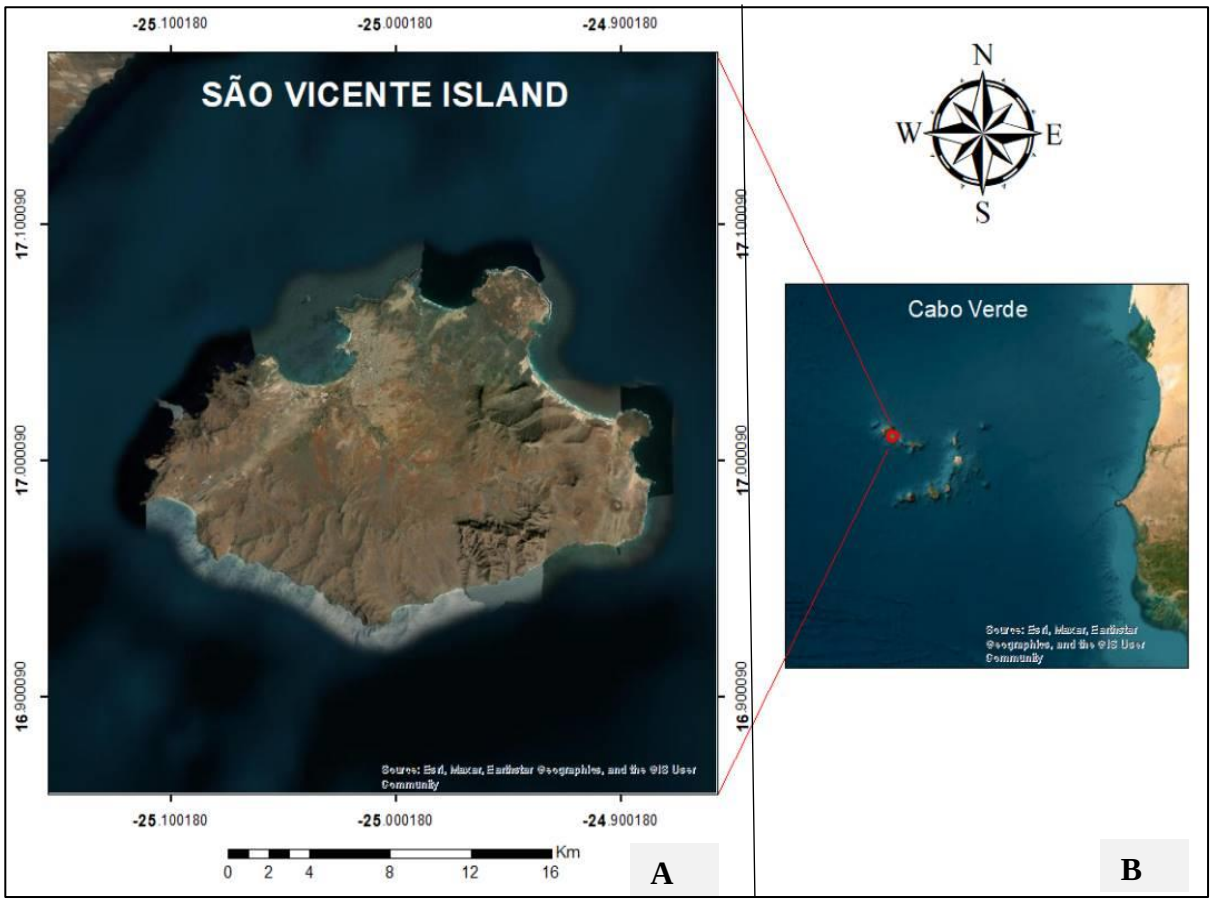


Figure 1. Collection site: Map of São Vicente Island (A) and the Cabo Verde archipelago located near the West African coast (B). Source: Esri, Maxar, Earthstar Geographics.

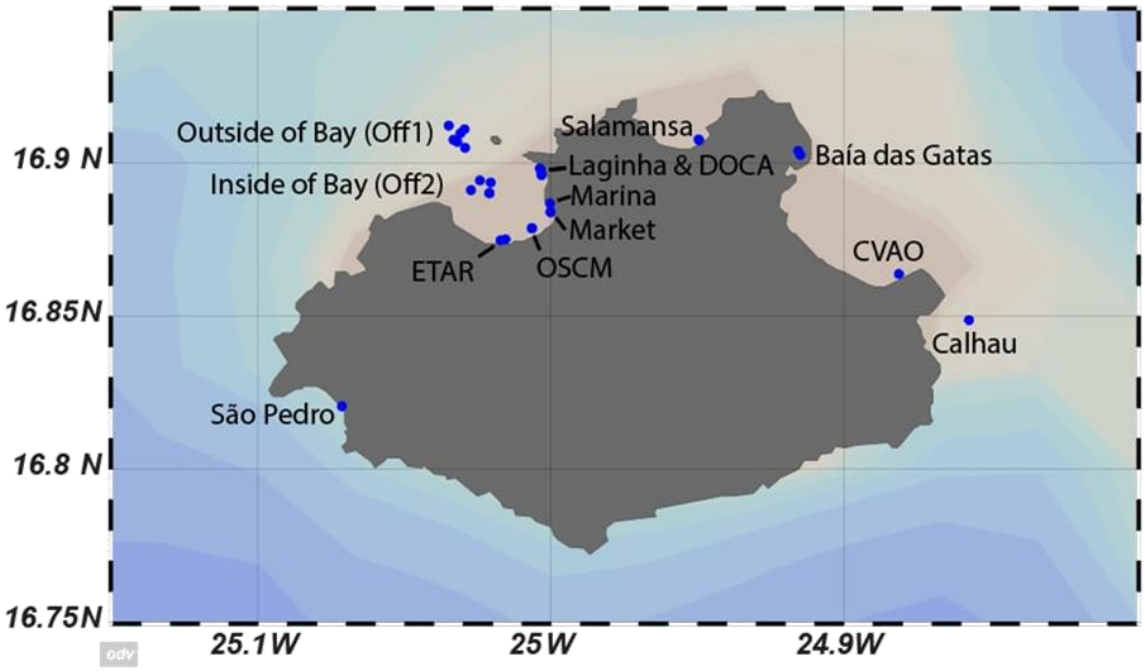


Figure 2. Study areas: Samples collection sites. Source: Ocean Data View.

Table 1. Geographic coordinates of the sampling sites on the island of São Vicente across both seasons.

Location	Latitude	Longitude	Dates
Marina	16°53'12.8"N	24°59'24.5"W	March 2023
Ribeira de Vinha (BP)	16°52'29.4"N	25°00'25.4"W	March 2023
Ribeira de Vinha (SP)	16°52'30.4"N	25°00'18.8"W	March 2023
Doca	16°53'52.5"N	24°59'36.1"W	March 2023
Laginha 1	16°53'46.1"N	24°59'34.9"W	March 2023
OSCM	16°52'43.6"N	24°59'46.6"W	March 2023
Baía	16°54'14.5"N	24°54'19.4"W	March 2023
Salamansa	16°54'27.8"N	24°56'21.6"W	March 2023
Middle of the bay	16°53'39.7"N	25°00'50.2"W	March 2023
Outside the bay	16°54'44.2"N	25°01'28.7"W	March 2023
Laginha 2	16°53'46.1"N	24°59'34.9"W	March 2023
Calhau	16°50'55.3"N	24°50'49.2"W	March 2023

Fish market	16°53'02"N	24°59'24"W	March 2023
Pipe (fish market)	16°53'03"N	24°59'24"W	March 2023
Offshore1 28jun	16°54'52.1"N	25°01'09.8"W	June 2022
Offshore2 28jun	16°53'19.1"N	25°00'44.8"W	June 2022
Offshore1 30jun	16°54'28.3"N	25°01'22.0"W	June 2022
Offshore2 30jun	16°54'25.9"N	25°01'21.8"W	June 2022
Offshore1 01jul	16°54'34.9"N	25°01'14.4"W	July 2022
Offshore2 01jul	16°53'24.7"N	25°00'39.0"W	July 2022
Offshore1 04jul	16°54'40.3"N	25°01'09.4"W	July 2022
Offshore2 04jul	16°53'26.8"N	25°00'31.8"W	July 2022
CVAO (Calhau)	16°51'50.0"N	24°52'15.4"W	July 2022
Salamansa	16°54'26.9"N	24°56'21.2"W	July 2022
Doca	16°53'53.8"N	24°59'37.1"W	July 2022
Baía das Gatas	16°54'14.5"N	24°54'19.4"W	July 2022
OSCM	16°52'43.6"N	24°59'46.6"W	July 2022

3.2 Sample collection and filtration

The coastal and offshore samples were collected from the top 1 m using a bucket previously rinsed with milli-Q water and dried. Special care was taken not to contaminate the collection container. Three washes were done with seawater, taking into account the sea current and the discarded sample was not re-introduced into the sampling location. The fourth time the sample itself was collected. This same water was previously used to rinse the bottle three times, where the sample was stored until the moment of filtration. Before containing the sample, the bottle was washed with milli-Q water and HCl 3%. Single samples were collected with the exception of Laginha where replicates were made. Environmental parameters such as temperature and salinity at the time of collection were obtained. For both measurements, a digital multiparameter tester 4-in-1 COM-401, model ACQ-16 was used. Samples were transported to the Ocean Science Center Mindelo (OSCM) within 1-2 hours and processed immediately. For filtration, a manifold (Figure 2A) that was connected by tubing to a vacuum trap which

was connected to a vacuum pump, with low pressure of less than 200 mbars, was used (Figure 2B). Special attention was paid to the vacuum pump by ensuring which channels are open or closed and by checking the pump gauge to ensure that all air is pumped from the system. The vacuum manifold was used to collect microorganisms on a filter membrane, a Supor (polyethyl sulfone) filter with 0.2 µm pore size (this pore size allows the collection of all microorganisms equal to or larger than bacteria size). Another manifold was used for gravity filtration (Figure 2C) where the microorganisms were gently collected on nucleopore filters with 3 µm pore size (this pore size collects larger microorganisms and the bacteria attached to them). In both cases 500 ml of the seawater sample was added to the filter and when the entire volume was filtered, the filter was removed from the device and folded in half (two to three times) with the help of two clean forceps. After that, it was placed in previously labeled cryotubes and preserved at -80^{o}C and then transported on the Polar Research Vessel POLARSTERN to the laboratory at GEOMAR Helmholtz Centre for Ocean Research in Kiel, Germany, where the next steps were carried out.

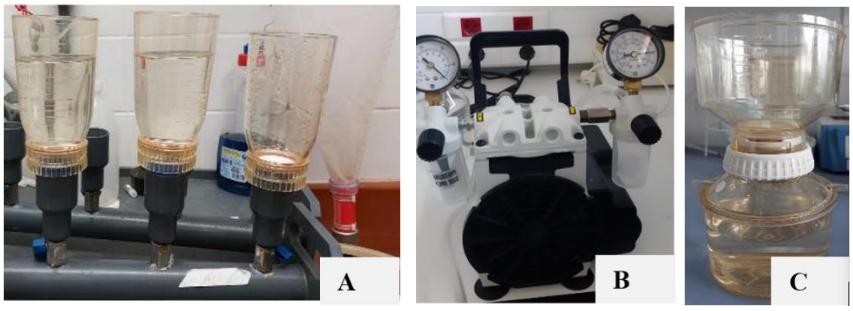


Figure 3. Filtration system: Manifold with filter size of >0.2 µm (A) that is connected to a vacuum pump (B). Manifold for filtration by gravity with a filter size of >3 µm (C).

3.3 DNA extraction and quantification

The filtered sample was used for DNA extraction, which was based on the Qiagen DNeasy Plant kit with minor modifications in incubation times and column steps plus the addition of a mechanical lysis step based on bead beating, freezing and thawing for samples filtered on the filter with size >0.2 µm.

To begin the DNA extraction process, glass beads (0.1 mm and 0.5 mm) were mixed in a 1:1 ratio, and roughly 100 µl was aliquoted into 2 mL screw cap microtubes. For >0.2 µm filters, clean were used to transfer the filter into 2 mL screw cap micro tubes containing glass

beads. The filters were then crushed using clean forceps to roughly doubles the DNA extracted since filters were cramped in the tubes. Then, spike-in ZymoBIOMICS™ Spike-in Control I (High Microbial Load) was added to each sample to ensure that the samples were accurately quantified and to enable cell number measurements. Next, 750 µl of buffer AP1 was added and the samples were exposed to physical lysis by freezing in liquid nitrogen, then transferred to a water bath for incubation. This freeze-thaw step was repeated three times. The tubes were transferred directly from the last thaw step to the Laboratory Mixer Mill, where the samples were bead-beaten for 2 minutes at 30 Hz and then pulse-centrifuge to reduce foam. 45 µl of Proteinase K was added to each sample and then incubated. After that, 4 µl of RNase A was added to each sample, followed by vigorous vortexing and another incubation process. To each lysate, 130 µl of Buffer P3 was added following the vortexing and incubation on ice for 10 minutes. After that, centrifugation was performed at 14,000 rpm to pellet precipitates and beads. The supernatant was transferred to a QIAshredder Mini Spin Column placed in a 2 mL collection tube and centrifuged. The flow-through was transferred, carefully, to a new 2 mL tube without disturbing any debris pellet. 750 µl of Buffer AW1 was added, then 650 µl of the mixture was transferred into a DNeasy Mini spin column. For this step, centrifugation was performed at 8,000 rpm for 1 minute and the flow-through was discarded. The DNeasy Mini spin column was placed into a new 2 mL collection tube, and 500 µl of Buffer AW2 was added and the same process above was followed. The collection tube was reused in the next step. To dry the membrane, another 500 µl of Buffer AW2 was added, and then centrifuged. The spin column was carefully removed from the collection tube and placed into a new and sterile 1.5 mL microcentrifuge tube. For samples from the filter size >0.2 µm, three elutions were performed using 30 µl of pre-heated TE Pipetted directly onto the DNeasy Mini Spin Column. Incubation was done at room temperature for 5 minutes, followed by centrifugation at 8,000 rpm for 1 minute for each elution. For samples from the filter size >3 µm, two elutions with 30 µl of pre-heated TE were performed using the same incubation and centrifugation steps. The final DNA extract was stored at -80°C.

To quantify the DNA extracted, PicoGreen assay was used. It is an ultra-sensitive fluorescent nucleic acid stain for quantitating double-stranded DNA (dsDNA). It was used to assess DNA concentration by staining the sample with a fluorescent marker, which is a fluorescent dye used to measure the intensity of dyes that bind to supply nucleic acids and fluoresce selectively when bound. Following manufacture's instructions, a standard curve was

produced for DNA masses ranging from 0.1 ng to 20 ng, and compared to the measurement of 1 µL of sample DNA extracts.

3.4 Polymerase Chain Reaction (PCR) amplification and library preparation

To amplify the rRNA gene from the DNA extractions, the V4 and V5 hypervariable was amplified by the polymerase chain reaction (PCR). 20 µL PCR reactions were prepared containing 5x KAPA (buffer), 10 µM dNTPs, HiFi HotStart (DNA polymerase), 5 µM of each primer, and 1 ng of sample DNA (except negative control for which 1 µL of molecular grade water was substituted). The final volume of each reagent, with the exception of the DNA template and the 5 µMF and 5 µMR primers (universal Forward and Reverse primers, 515F-926R as shown in Table 2), was calculated according to the number of samples. The 5′ ends of the primers were tagged with specific barcodes per sample (A-I-NNNN-barcode-515F and A-index-I-926R where A is the Illumina sequencing adapter and I is the Illumina primer) (Yeh et al., 2018). The cycling parameters used for the PCR were as follows, represented in Table 3. Molecular grade water was used as blank and a mock community, made up of 10 organisms (Appendix 1), was used as positive control. Successful amplification of samples and lack of amplification in the negative were assessed via electrophoresis (Appendix 2). PCR reactions were cleaned and concentrated with 0.8x ratio Beckman Coulter Agencourt Ampure XP beads following manufacturer's instructions. A purification and normalization using the Just-a-Plate^{TM}96 PCR Normalization and Purification kit from Charm Biotech, from all samples of 2022 and those filtered from >3 µm from 2023, was performed following manufacturer's instructions (Appendix 3).

Table 2. Universal bacterial primers used to target conserved regions in the gene during amplification.

Primers	Sequence	Amplified fragment size
515F	GTGYCAGCMGCCGCGGTAA	525 pb
926R	CCGYCAATTYMTTTRAGTTT	

Table 3. PCR cycling parameters used for the amplification of DNA target sequence.

	Parameters		
	Temperature (°C)	Time	
Step 1	95.0	5 min	Denaturation
Step 2	98.0	20 sec	
Step 3	55.0	12 sec	
Step 4	72.0	1 min	
Step 5	72.0	5 min	Extension
Step 6	16.0	∞	

30 cycles

3.5 *Illumina MiSeq* sequencing

After confirmation of the PCR products with agarose gel electrophoresis for the first batch and purification and normalization for the second batch of samples, the amplicons were now ready for sequencing. After cleaning and concentration, samples were normalized to equi-molar concentrations and submitted to the Competence Center for Genomic Analysis (CCGA), Kiel, for sequencing by *Illumina MiSeq* Nano 2x250 bp.

3.6 Data analysis

For the bioinformatics analysis, a metadata file was created according to the specifications of the company responsible for the sequencing. Each sample was identified correctly with i7 and i5 index, incorporating a unique barcode sequence for each sample.

Sequencing data was analyzed with Quantitative Insights into Microbial Ecology version 2 (QIIME 2 2023.5) software was extensible, accessible, and open source, allowing for quantitative and diversity assessments. The sequences were quality filtered by trimming the primers (using cutadapt function), separating 16S reads from 18S and formatting them (using bbsplit and bbmap function) by BLAST against SILVA database, as well as removing mitochondrial, chloroplast and unassigned samples and then 'denoised' to form taxonomic units called amplicon sequence variants (ASV), which could differ by as few as a single base pair.

The analysis was also done under Ubuntu, where packages like bbmap, commonly used for mapping and analyzing DNA/RNA sequencing data, were used. Inside QIIME 2, essential plugins were used: `qiime dada2 denoise-paired`, `qiime taxa barplot`, `qiime diversity core-metrics-phylogenetic`, `qiime diversity alpha-group-significance` and `qiime diversity beta-group-significance`. To perform diversity analysis (alpha and beta) qzv files were created and then visualized in Qiime 2 view. For that, `faith_pd_vector.qza` and ` evenness_vector.qza` metrics were used for alpha diversity. In the same way `unweighted_unifrac_dsitance_matrix.qza` was used for beta diversity analysis. Alpha diversity analysis helps assess the diversity and distribution of microorganisms within a specific area or sample, while beta diversity analysis assesses the differences in microbiome composition between multiple locations or communities. For whole community analyses, metrics like Bray-Curtis similarity were used to assess the similarity between the sites.

Heatmaps outside the QIIME 2 were created using Python packages/libraries like Seaborn, which helps explore and visualize the datasets.

4 Results

4.1 Physical characterization of the water sample

A careful physical characterization involving the examination of parameters like temperature and salinity was carried out to obtain information about the properties and composition of the water sample to characterize the conditions of the natural habitat of the microbial communities.

In 2023 collection sites, there was a temperature variation with an average of 24.1 ºC (Table 4), for 2022, the average was slightly lower at 23.7 ºC (Table 5). The highest temperature of 27.1 ºC was recorded in Calhau (2023 sampling) and 25.5 ºC was recorded in Baía (2022 sampling). Marina and Doca recorded the lowest temperatures of 21.8 and 23.1 respectively (2022 and 2023 samplings). Regarding salinity in 2020 sampling, an average of 35.2 ppt was recorded with highest value at offshore with 35.6 ppt and lowest at Baía with 33.6 ppt. While for 2023 it was an average of 34.3 ppt. The highest salinity for this case was measured at Doca and OSCM, both with 35.2 ppt and the lowest was at Calhau with 32.7 ppt.

Table 4. Physical parameters values measured at each collection site in March 2023.

	Parameters	
Sample sites	Temperature (ºC)	Salinity
Marina	23.1	34.6
Ribeira de Vinha (BP)	23.3	35.1
Ribeira de Vinha (SP)	23.5	34.5
Doca	23.8	35.2
Laginha 1	25.6	33.1
OSCM	23.6	35.2
Baía	25.8	34.7
Salamansa	23.9	34.7
Middle of the bay	26	33.6
Outside the bay	25.7	34.7
Laginha 2	25	33.6
Calhau	27.1	32.7

Table 5. Physical parameters values measured at each collection site in June and July 2022.

	Parameters	
Sample sites	Temperature (ºC)	Salinity
Offshore1 28jun	23.6	35.5
Offshore2 28jun	23.8	35.5
Offshore1 30jun	23.6	35.5
Offshore2 30jun	23.9	35.5
Offshore1 01jul	23.2	35.5
Offshore2 01jul	23.1	35.5
Offshore1 04jul	23.4	35.6
Offshore1 04jul	23.5	35.5
CVAO (Calhau)	24.7	35.5
Salamansa	25.2	35.5
Doca	21.8	35.2
Baía das Gatas	25.5	33.6
OSCM	24	34.8

4.2 Evaluation of the total DNA extraction PCR method of microbial communities

In regards to total DNA extracted from samples from 2022, the highest concentration of extracted DNA came from samples from Offshore filtered from both size filters (>3 μm => 222.72 ng, >0.2 μm => 753.93 ng), except for one offshore sample (off30jun) (Figure 3).

The samples from 2023 with the highest concentration of extracted DNA came from the Fish Market from both size filters (>3 μm = 2460.24 ng, >0.2 μm = 1355.13 ng) (Figure 4).

The concentrations presented here were multiplied by the number of elutions performed in each one.

Gel electrophoresis, based on the ladder, confirmed that the sizes of the fragments were approximately 525 pb, as expected (Figure 5).

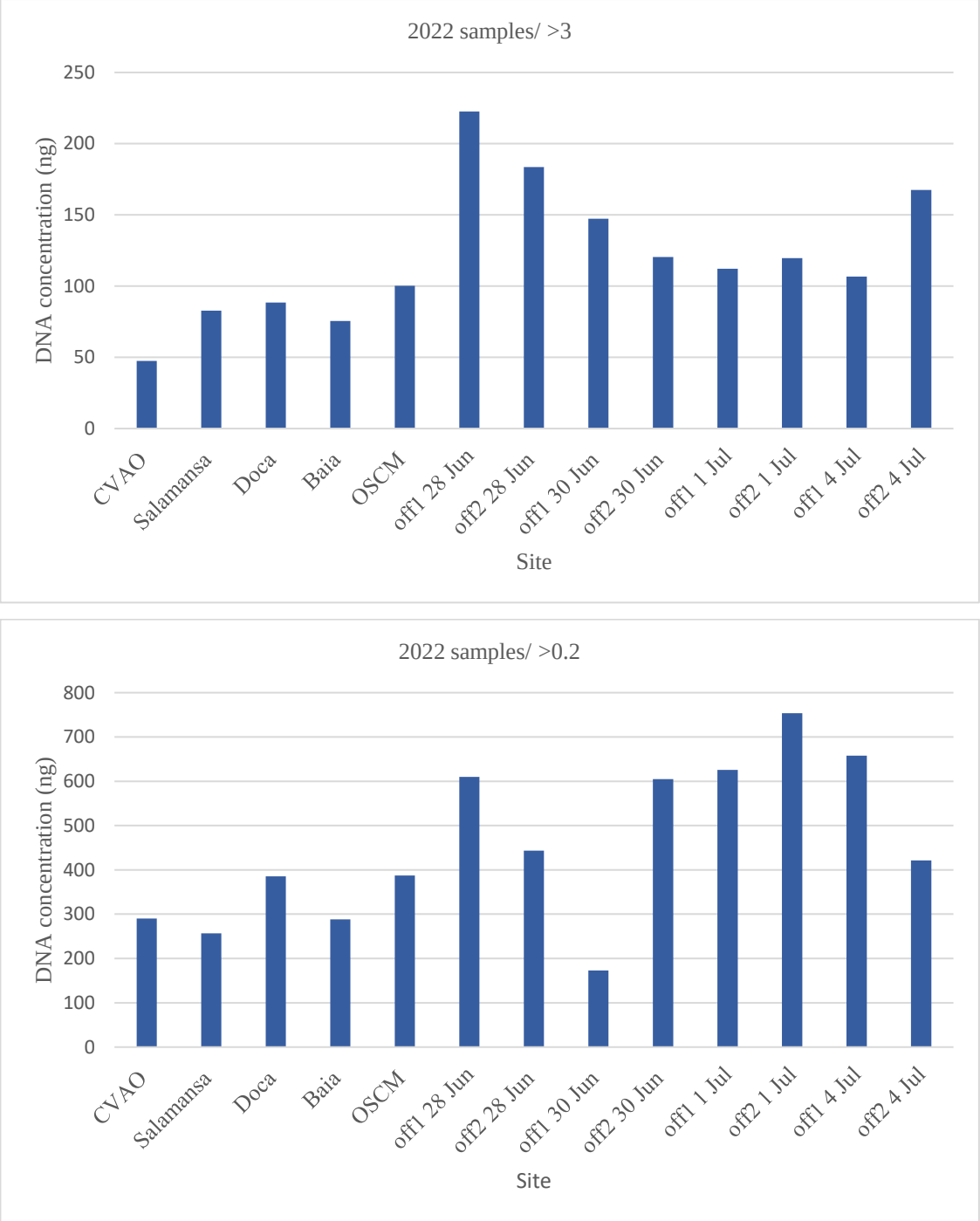
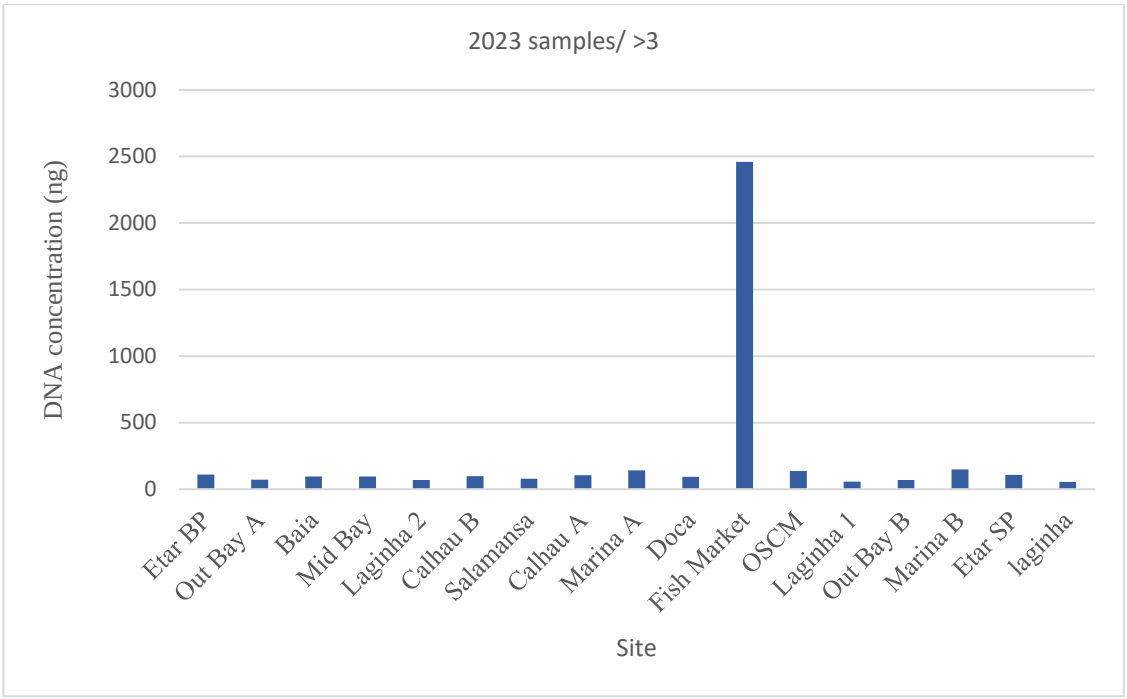


Figure 4. DNA concentration per site: Barplot shows the amount of DNA extracted in each sample of 2022 and with different filter size (>3 µm and >0.2 µm, respectively). Note that the samples >3 µm and >0.2 µm were eluted in approximately 60 µL and 90 µL of tris-EDTA (TE), respectively, which contributes to the concentrations shown here. Thus, the most straight-forward comparison is only within a graph.



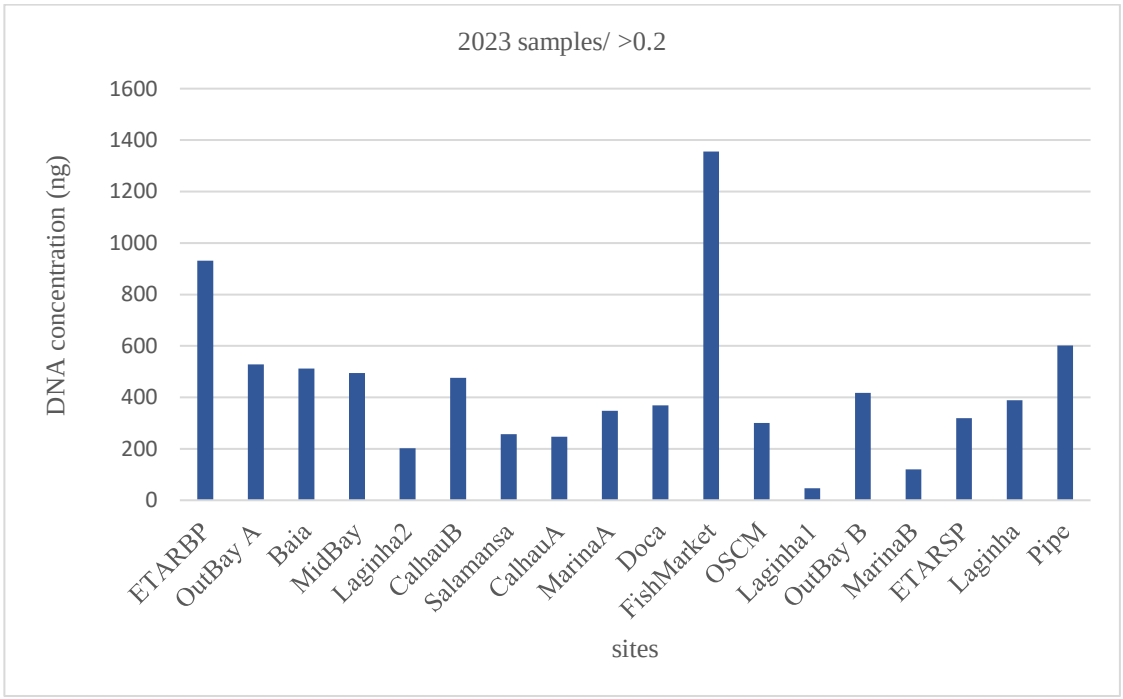


Figure 5. DNA concentration per site: Barplot shows the amount of DNA extracted in each sample of 2023 and with different filter size (>3 µm and >0.2 µm, respectively). In this case, both samples were eluted in approximately 90 µL of tris-EDTA (TE), which contributes to the concentrations shown here.

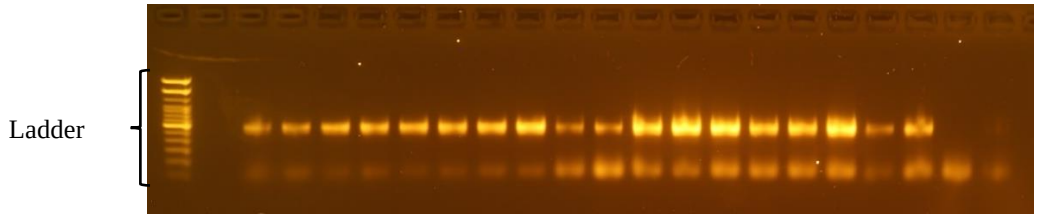


Figure 6. Gel electrophoresis: Representative gel shows the ladder on the left and bands of DNA fragments from >0.2 μm samples from 2023.

4.3 Overview of Bacterial Diversity and Community

A total of 212.091 rDNA sequence reads (individual sequences obtain from DNA fragment) were obtained from the 65 samples being the Fish Market samples with the highest number of reads. It should be noted that the samples filtered with pore filters >0.2 μm had higher numbers of reads than the others (Figure 6).

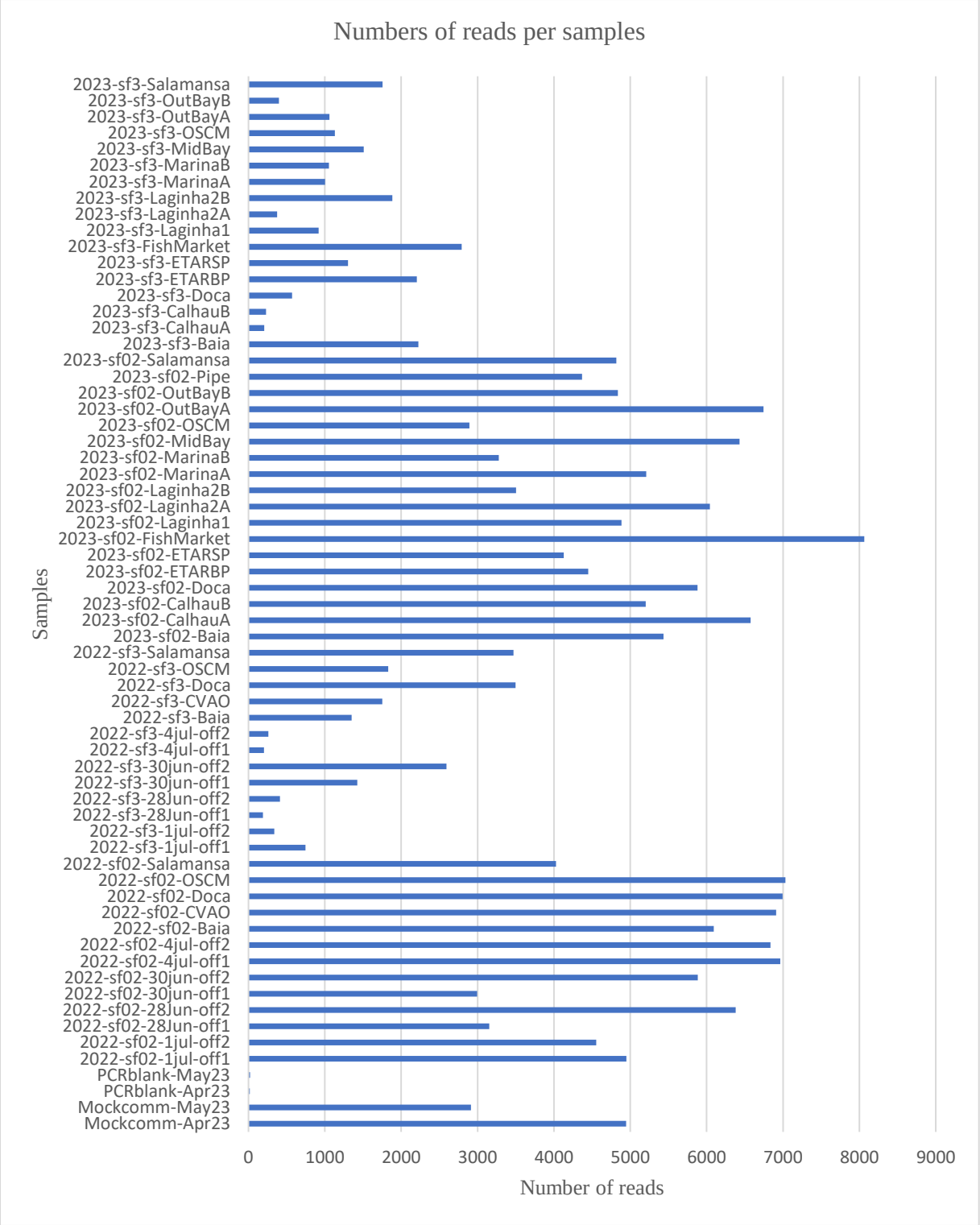


Figure 7. Number of reads per sample: Barplot shows the amount of reads each sample presented, with Fish Market having the most with 8065 reads and the fewest with the PCR Blanks with 18 reads.

The prokaryotic features (after removing mitochondrial, chloroplast and unassigned samples) were further assigned to different levels of taxonomic for further analysis. As shown on figure 7 bacterial communities constitute more than 90% of microorganisms found in the

samples. The remaining organisms belong to Archaea. The most abundant phyla found were Proteobacteria followed by Bacteroidota and Cyanobacteria (Figure 8).

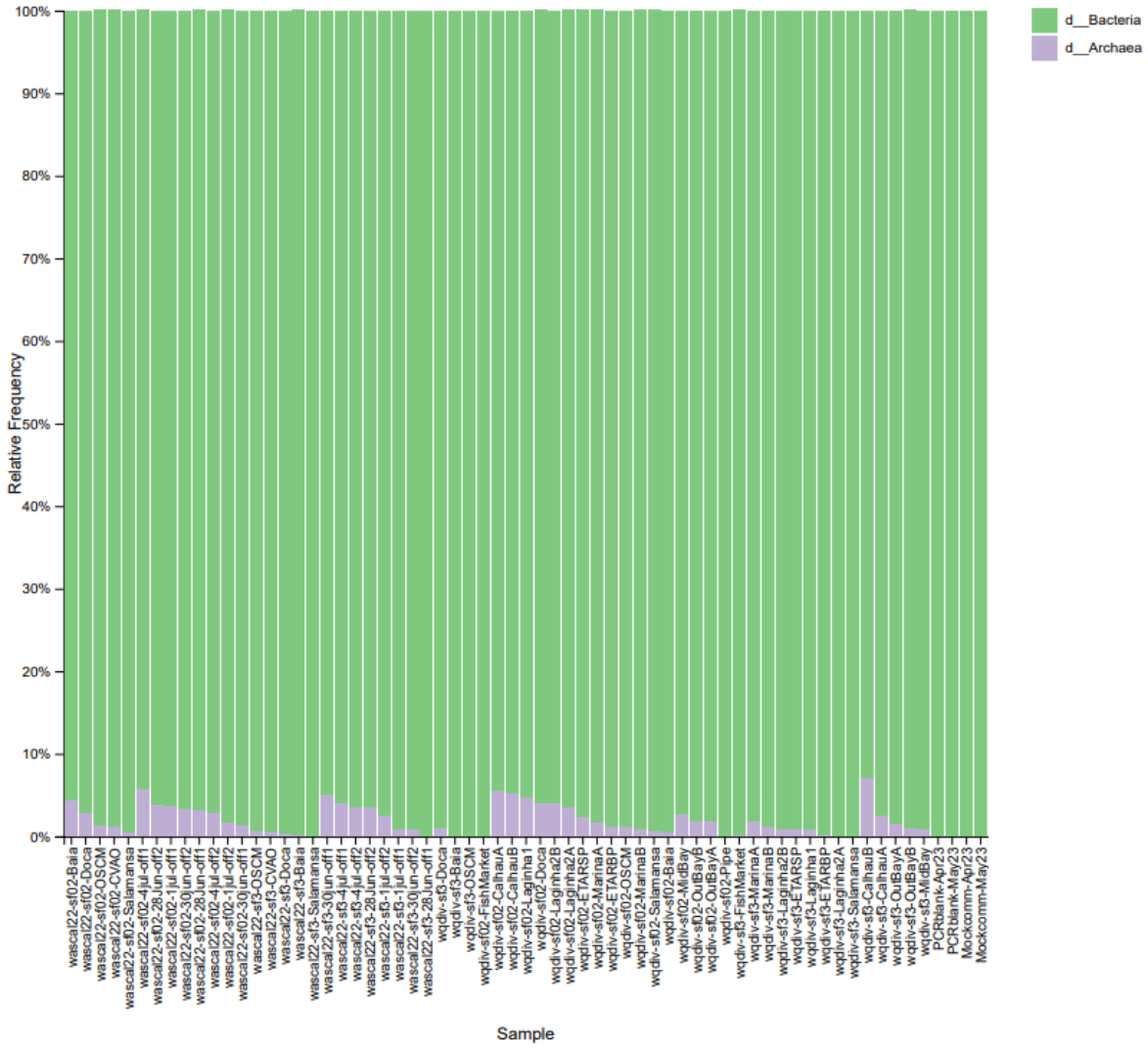


Figure 8. Abundance of two primary domains in samples: Relative abundance of bacteria and archaea found in the samples, presented in terms of percentage.

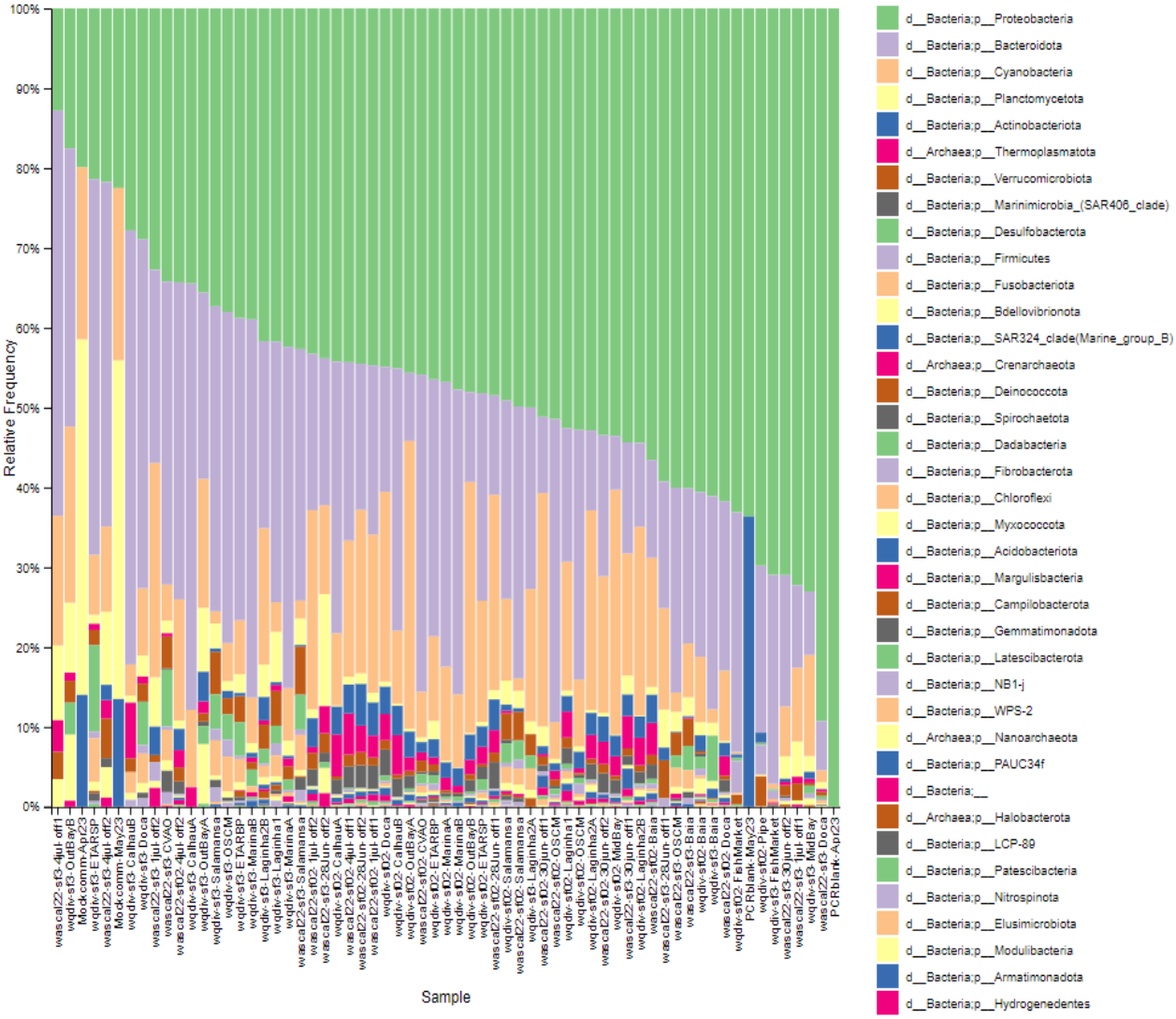


Figure 9. Abundance of main phyla in the samples: Relative abundance of phyla taxonomic level found in each sample, presented in terms of percentage. Sorted by proteobacteria taxonomic abundance, ascending from left to right. Being Proteobacteria represented by green color, Bacteroidota by purple and Cyanobacteria by orange.

The Faith Phylogenetic Diversity of composition exhibited a notable variation across the sampling areas (Figure 9), as indicated by the Kruskal-Wallis test with a significance level of p-value = 0.001. Specifically, when comparing the pairwise Baía and Marina (p-value = 0.02), Marina and OSCM (p-value = 0.02), Marina and Salamansa (p-value = 0.02) and Marina OSCM and Salamansa (p-value = 0.04), statistically significant distinctions were observed (with a significance threshold of $p<0.05$). Faith Phylogenetic Diversity showed high alpha diversity in Salamansa and a decline in alpha diversity in Offshore samples.

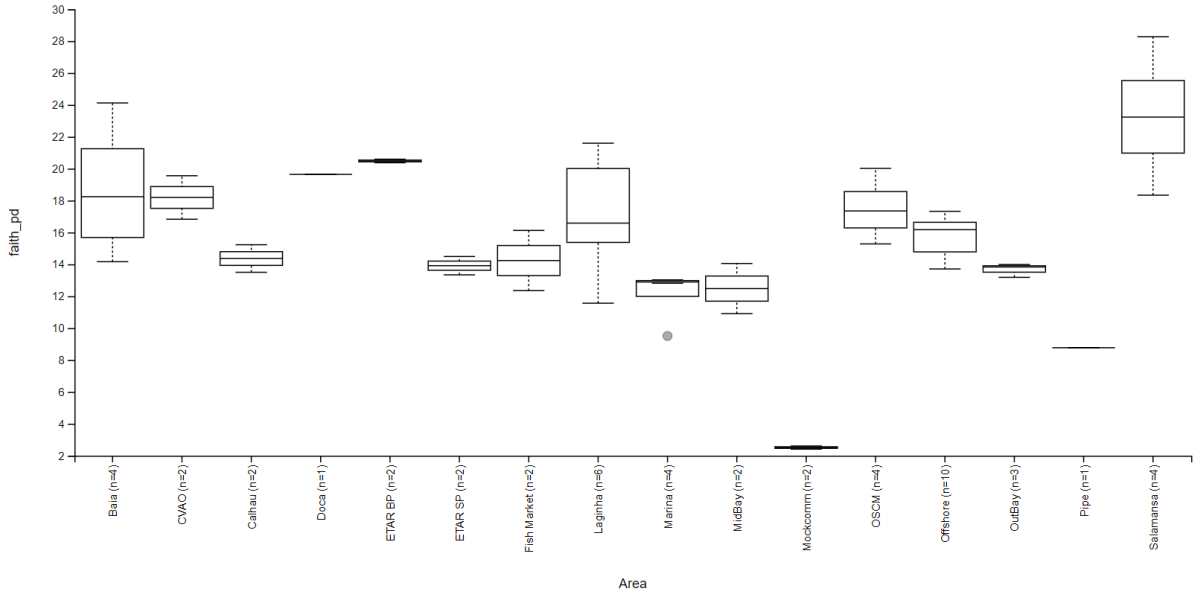


Figure 10. Faith Phylogenetic Diversity: Boxplot shows a qualitative measure of community richness per area plus mock community.

Regarding the different size fraction of the filters used, no statistically significant difference was observed based on the richness of the community (Kruskal-Wallis test with p-value = 0.24) (Figure 10).

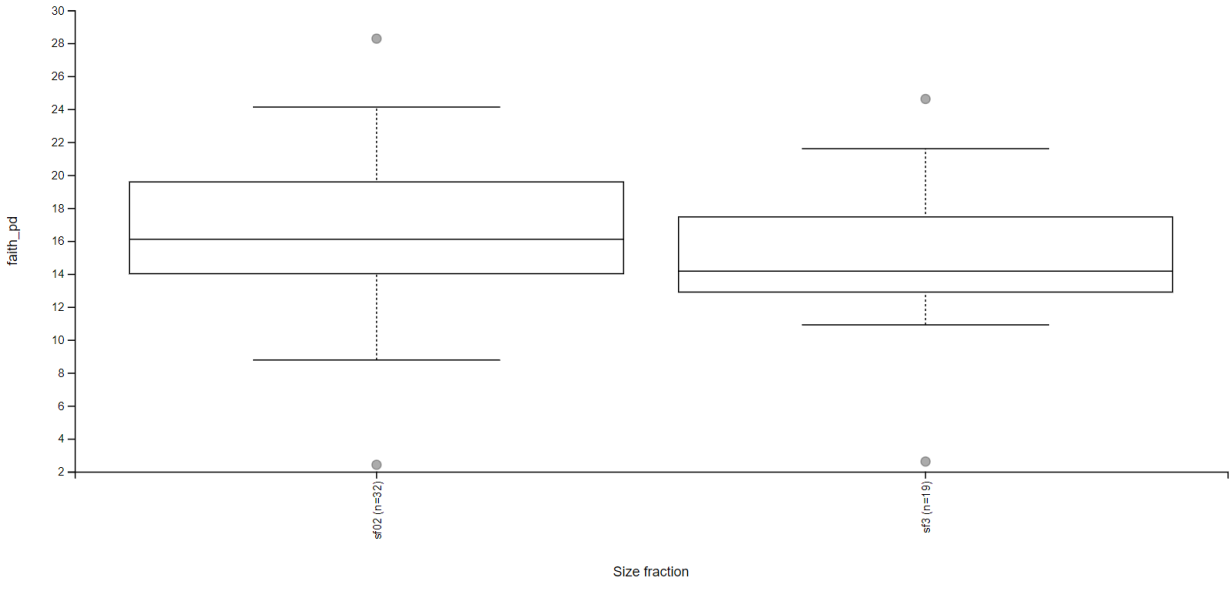


Figure 11. Faith Phylogenetic Diversity: Boxplot shows a qualitative measure of community richness per size fraction filters. Being sf02 the pore size >0.2 µm and sf3 the pore size 3 µm.

Regarding the monthly variation in diversity based on the richness of the community also no statistically significant difference was observed (Kruskal-Wallis test with p-value = 0.059). Among all groups the p-value was greater than 0.05. (Figure 11).

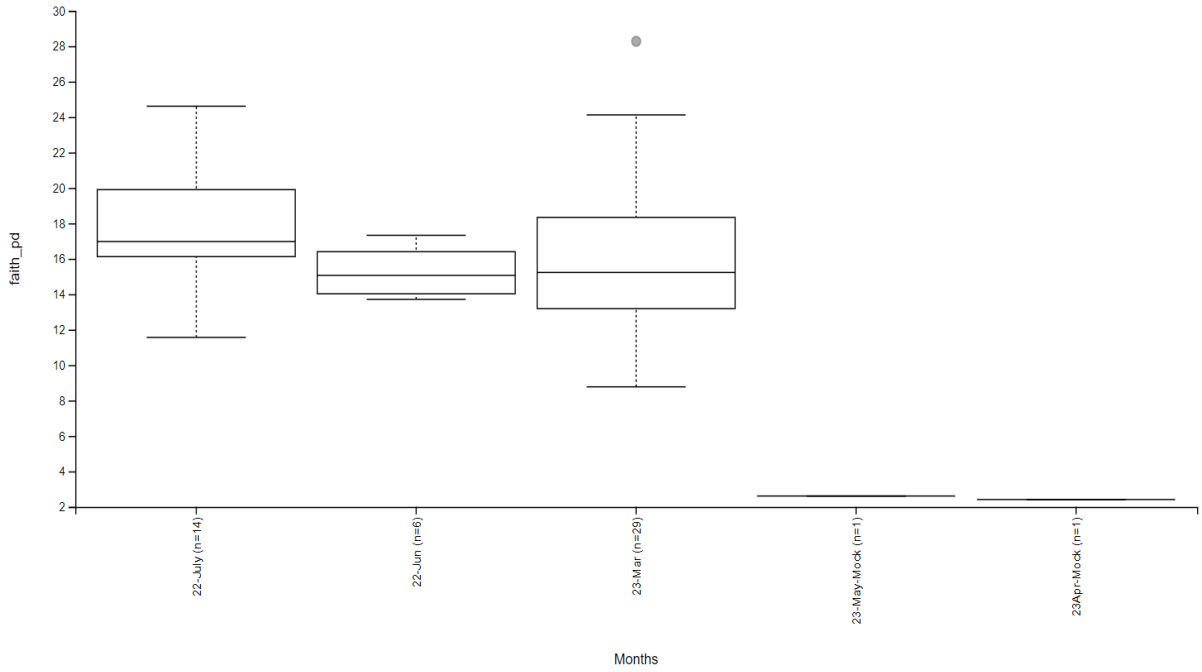


Figure 12. Faith Phylogenetic Diversity: Boxplot shows a qualitative measure of community richness per month plus mock community.

The composition evenness (Pielou's evenness) showed no significant difference among the areas having the Kruskal-Wallis test with p-value = 0.14 (Figure 12). The unique exception was observed between Marina and Salamansa with p-value = 0.02. In overall, mock community showed less evenness.

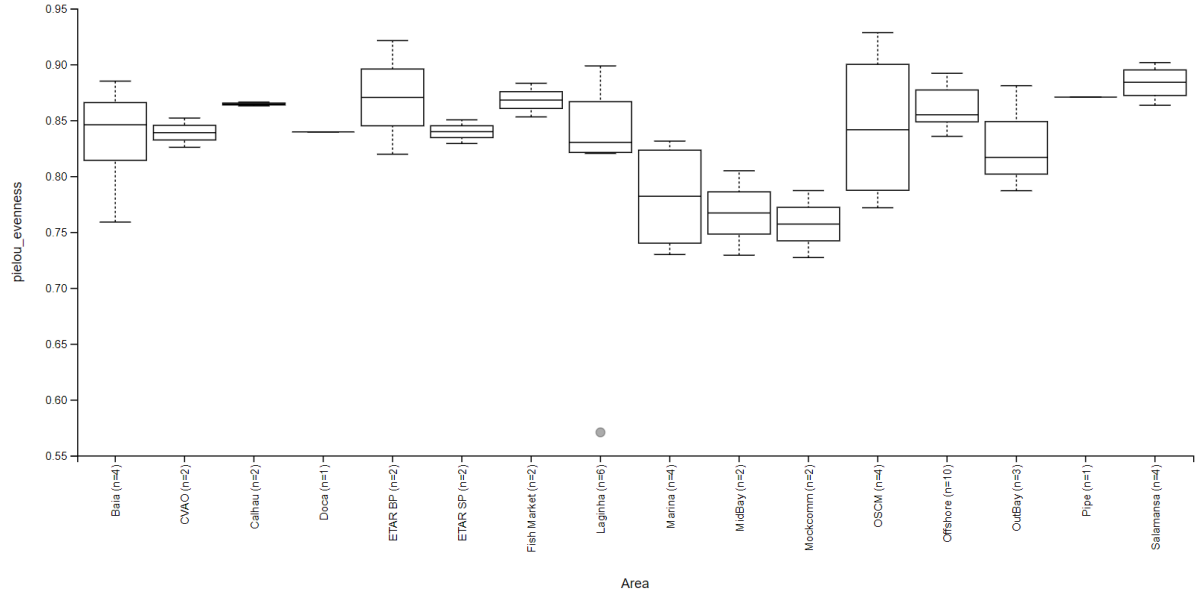


Figure 13. Pielou's Evenness diversity: Boxplot shows a distribution of the bacterial community per area plus mock community.

The same observation was made in composition evenness (Pielou's evenness) related to size fraction of filters and monthly variation. With the Kruskal-Wallis test with, p-value = 0.08 and p-value = 0.33, respectively (Figure 13 and Figure 14).

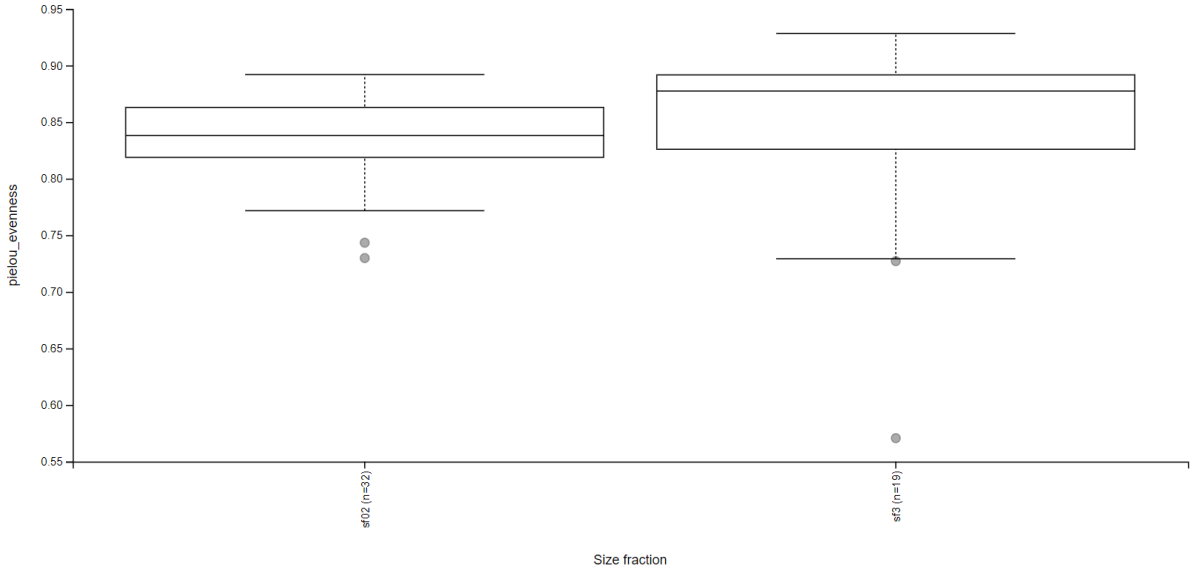


Figure 14. Pielou's Evenness diversity: Boxplot shows a distribution of the bacterial community per size fraction filters. Being sf02 the pore size >0.2 µm and sf3 the pore size 3 µm.

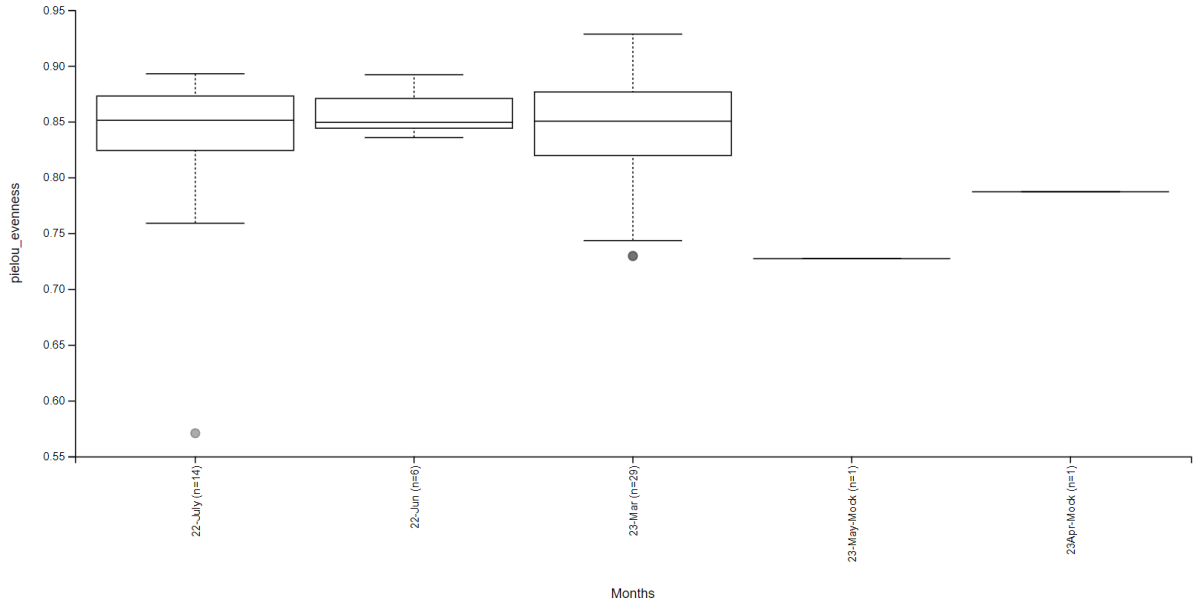


Figure 15. Pielou's Evenness diversity: Boxplot shows a distribution of the bacterial community per month.

A slight dissimilarity between coastal and offshore samples was observed. As expected, the Fish Market and Pipe samples showed similarities with the coastal samples, which were dispersed along the axis, but at the same time clustered between them. From the same coastal

samples, five outliers were observed showing similarity with the offshores samples, which were also grouped together (Figure 15, A). The samples filtered with the same filter size, regardless of the year they were collected, were clustered together. However, three outlier values were observed: samples from Fish Market, Pipe and Salamansa filtered with a filter size >0.2 µm showed greater similarity with samples filtered with a filter size >3 µm then with the rest of the same filter size of >0.2 µm (Figure 15, B). In both cases the mock communities were dispersed and were not classified into any of the groups.

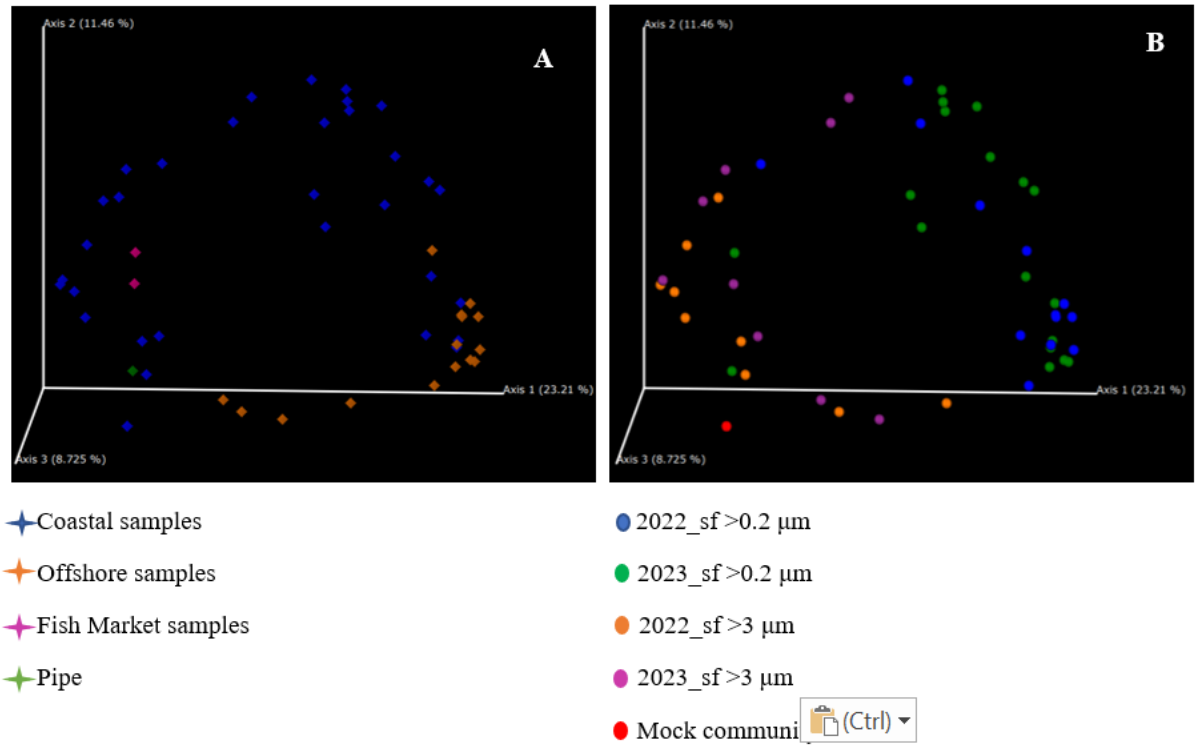


Figure 16. Emperor plots based on Bray-Curtis similarity: Shown the similarity and dissimilarity between regions (A) and between samples collected in different years with specific size filter (B). Principal Coordinates Analysis (PCoA) was used.

4.4 Overview of taxonomic diversity of bacterial families containing pathogenic microorganisms with human health implications

In analyzing the microbial composition in the samples, a diverse range of bacterial families with microorganisms considered pathogenic was identified (Table 6). Notably, the presence of the Neisseriaceae family was observed within the phylum Proteobacteria. Additionally, the families Rickettsiaceae, Enterobacteriaceae, Yersiniaceae, Pseudomonadaceae, Legionellaceae, Aeromonadaceae and Vibrionaceae also belong to the Proteobacteria phylum, were detected, evidencing the prevalence of this phylum and its potential importance in the

studied ecosystem. The pathogenic bacteria from the families above were mostly found in the Pipe and Fish Market, all belonging to March 2023 samples. Going beyond the Proteobacteria, the phylum Bacteroidota exhibited representation by the Flavobacteriaceae and Weeksellaceae families, demonstrating prevalence also in the Pipe and Fish Market. In contrast, the phylum Spirochaetota was represented by the Spirochaetaceae family, whose highest detection occurred in OSCM samples from March 2023. The Firmicutes phylum, on the other hand, featured the Bacillaceae and Enterococcaceae families which had the highest occurrences in OSCM and Laginha, respectively, both samples collected in March 2023 (Figure 16). It should be noted that these taxa all or often have naturally (non-anthropogenically derived) occurring counterparts.

Table 6. Human associated diseases caused by pathogenic bacteria: Taxonomic Classification, distribution, and sources.

Phylum	**Family**	**Genus/species**	**Sites**	**Common sources**	**Common disease**
Proteobacteria	Vibrionaceae	*Vibrio vulnificus*	Baía (2022)	Environmental waters, seafood	Wound infection, fulminant septicemia (Sampaio et al., 2022)
	Rickettsiaceae	*Orientia tsutsugamushi*	OutBay (2023)	Animals	Scrub typhus (Nicholson & Paddock, 2023)
	Neisseriaceae	_____	Pipe (2023)	Human	Meningococcal disease (CDC, 2022a; Morse, 1996)
	Enterobacteriaceae	*Klebsiella* *Shigella*	OSCM (2023) Fish Market (2023)	Human	Pneumonia, Gastroenteritis (CDC, 2010, 2023)
	Yersiniaceae	*Serratia spp*	OSCM (2022)	Environment	Urinary and respiratory tract infections, wound infections (Khanna et al., 2013)

	Pseudomonadaceae	*Pseudomonas stutzeri* *Pseudomonas putida*	MidBay (offshore-2023) Laginha (2023)	Environment	Ocular infection, Skin infection (Alhalimi et al., 2022; Thomas et al., 2013)
	Legionellaceae	*Legionella*	Salamansa (2023)	Water system, animals	Legionnaires' disease (CDC, 2021; Rodrigues & Cunha, 2017)
	Aeromonadaceae	*Aeromonas*	Fish Market (2023)	Human	Gastroenteritis, septicemia (Rodrigues & Cunha, 2017)
Bacteroidota	Flavobacteriaceae	*Myroide*	Fish Market (2023)	Environment	Urinary tract infections, pneumonia, meningitis (Mammeri et al., 2002)
	Weeksellaceae	*Chryseobacterium*	Pipe (2023)	Environment, human, animals	Bacteremia, pneumonia, meningitis (Izaguirre-Anariba & Sivapalan, 2020; Mwanza et al., 2022)
Spirochaetota	Spirochaetaceae	*Treponema*	OSCM (2023)	Human	Yaws, Pinta (Morse, 1996)
Firmicutes	Bacillaceae	*Bacillus*	OSCM (2023)	Animals	Cutaneous anthrax, intestinal anthrax (CDC, 2022b; Morse, 1996)
	Enterococcaceae	*Enterococcus*	Laginha (2023)	Environment, human	Intra-abdominal infections, meningitis (Said et al., 2022)

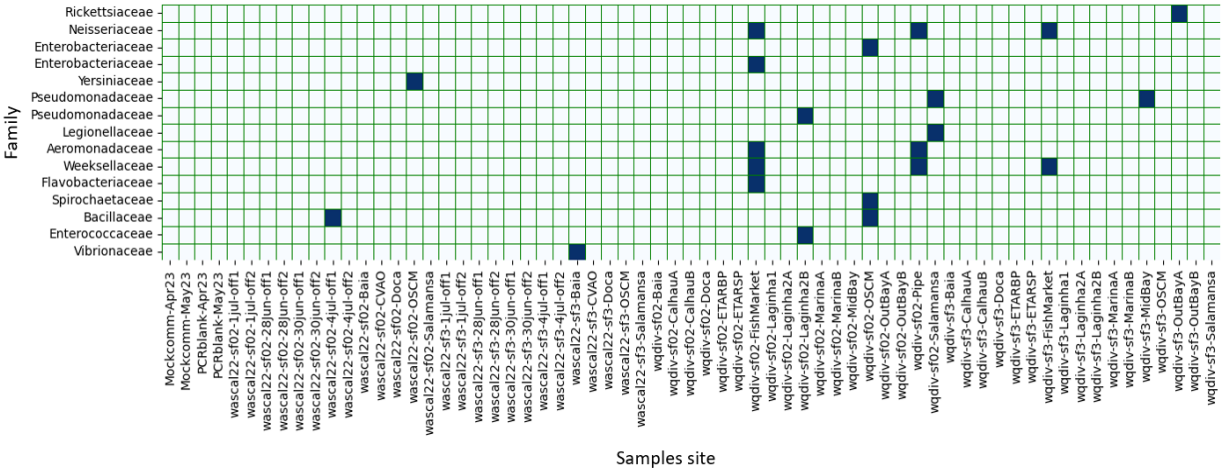


Figure 17. Presence and absence of relevant taxonomic families in each sample: Heatmap show the presence of bacterial families that have organisms potentially pathogenic.

By categorizing the samples into broader regions, it became evident that the samples collected from the areas around the city of Mindelo exhibited the greatest occurrence and diversity of bacterial families containing microorganisms known for their pathogenic characteristics, with the Weeksellaceae family being found in greater quantity (51%), particularly in Mindelo (Figure 17).

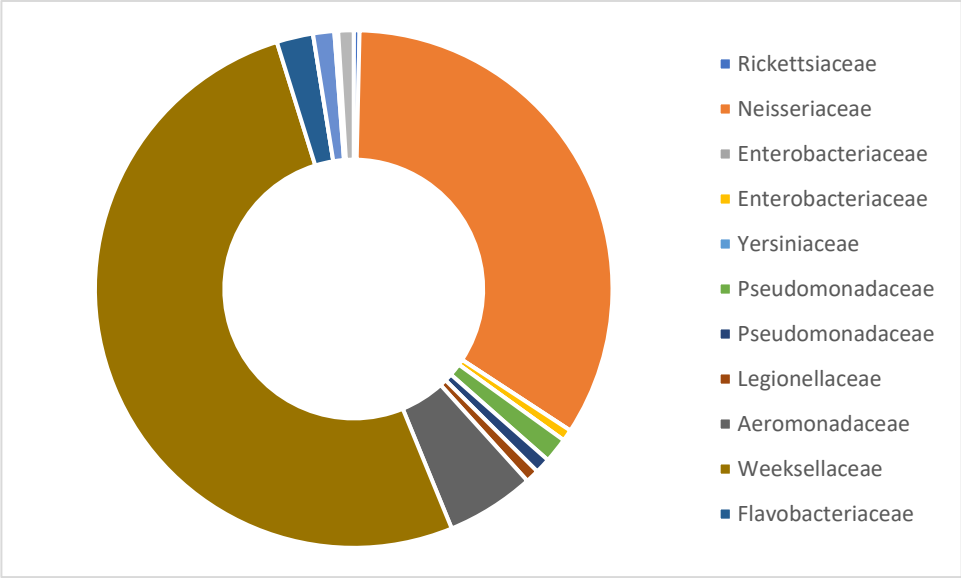


Figure 18. Quantity of taxonomic family with microorganisms potentially pathogenic: Pie chart showing the proportion of each family, in percentage.

Among the families of bacteria that have organisms considered pathogenic to humans, 86.6% came from samples collected in March 2023. The exception was only for samples from Baia and OSCM from July 2022 where *Vibrio vulnificus* of the Vibrionaceae family and *Serratia* spp. of the Yersiniaceae family, respectively, were found. And 80% of the samples from these relevant families were filtered using a filter size >0.2 μm.

5 Discussion

5.1 Diversity of the bacterial community

It is known that populations of microorganisms vary across the ocean surface. These variations are observed between areas close to the coast and areas of open ocean, in addition to being influenced by latitude and by places where waters from the mesopelagic zone emerge to the surface.

According to Gomes (2014), in Cabo Verde there are two layers of water on the surface: the first, up to 150 to 200 meters deep, comes from the north of the Atlantic Ocean and is saltier (36 ppt) compared to the second, which comes from the south of the Atlantic Ocean and is less salty. Although it is still mentioned that during the rainy season, torrential flows do not modify the salinity of the sea in the archipelago, we obtained a slight decrease in salinity.

Gomes (2014) also mentions that, in Cabo Verde the temperature of the sea surface presents a variation with an average of 24°C throughout the year, which coincides with the measurements of the physical parameters carried out in this study. According to the same author, the temperature distribution along the water column in Cabo Verde is also notable in the first 100 meters of depth with a maximum temperature stratification. In periods of stratification, the upper euphotic zone in tropical and mid-latitude regions transforms into an extreme environment, with intense light but limited nutrient availability (Gomes, 2014). This is the case of Cabo Verde, which is considered a low nutrient (oligo-mesotrophic) location since the wind regime in the archipelago does not cause significant "upwelling", as the configuration of the coast is not favorable, with the exception of cases where certain events, such as cyclonic circulations and vertical water movements, can cause local enrichment similar to "upwelling" having a highly beneficial impact on biological activity and primary production by bringing nutrient-rich waters from the depths to the brighter zone of the ocean surface

(Direção Geral dos Recursos Marinhos, 2019). Cabo Verde ocean waters generally consist of microorganisms that prevail in the open ocean and are suitable to thrive in environments with limited nutrients (Martínez-García & Pinhassi, 2019). In this study, highlighting the method of sample collection up to 1 m deep, most of the phyla of bacteria found were proteobacteria (Alphaproteobacteria class), Bacteriodetes (Bacteroidia class) and Cyanobacteria (genus *Prochlorococcus*), coinciding with studies carried out by Giovannoni & Vergin (2012) and Schattenholfer et al. (2009) in which these organisms dominated the water surface in the coastal and offshore regions of the Atlantic.

Compared with the other samples, phylum proteobacteria dominated the Doca sample (89.27%), specifically, the genus *Ruegeria* (50.01%), which is widespread and some species was isolated from marine sediment from northwest Africa (Kim et al., 2019). *Ruegeria* can also be isolated from upper ocean waters and exhibits a positive correlation with pheopigments and chlorophyll, suggesting an association with algae (Sonnenschein et al., 2017).

The alphaproteobacteria, especially the SAR11 clade, found occupying greater abundance among the bacteria found in Laginha (28.24%), is one of the most prevalent in aquatic environments, making up approximately a quarter of the planktonic cells found in the upper sections of the ocean photic zone (Haro-Moreno et al., 2020). Because they are photoheterotrophs with simplified genomes which, together with the ability to oxidize a wide variety of carbon compounds and the use of light by proteorhodopsin they are particularly suitable in the stratified oligotrophic gyros of most aquatic environments, as is the case in Cabo Verde. It also plays a key role in the carbon cycle in low-nutrient seas (Giovannoni, 2017). Looking at the case of the high abundance of proteobacteria phylum found in the offshores samples we can have an understanding of the relationship between the survival of these organisms in low nutrient conditions.

According to Schattenholfer et al. (2009), Bacteroidetes are abundant in coastal regimes, cope well with the limiting conditions of the oligotrophic gyrus and are involved in the degradation of polymers such as polysaccharides and proteins that are compounds easily found in the environment where there is an exoskeleton of animals, algae present or other runoff products, etc., as in the case of samples collected at CVAO (Cabo Verde Atmospheric Observatory). As mentioned before, ocean waters within the Cabo Verde generally consist of microorganisms that prevail in the open ocean, such as small phytoplankton, *Prochlorococcus* (cyanobacteria). So as expected it was in our offshore samples that we found the most *Prochlorococcus*. The fundamental role of phytoplankton in the stability of aquatic ecosystems

involves the absorption of carbon dioxide and nutrients, in addition to the production of oxygen. It is anticipated that the continued increase in temperature will result in decreased amounts of larger phytoplankton, such as diatoms, compared to smaller forms giving cyanobacteria a selective advantage over other groups of phytoplankton. The composition of this community is also linked to other environmental conditions, mainly the presence of nutrients and light availability.

According to Schattenholfer et al. (2009), the increase in the most abundant groups of bacteria found in this study may be linked to the fact that the increase in phytoplankton (cyanobacteria) increases the sudden release of nutrients, consequently causing the proliferation of proteobacteria and bacteroidetes.

5.2 Alpha diversity analysis

To analyze the bacterial diversity and richness index as well as its distribution within the samples, they were grouped by areas for a better visualization and interpretation of the results. For the analysis of ASV in different months of collections, they were separated by the 3 main months of collection, which were June 2022, July 2022 and March 2023. Regarding the size filters, the samples were separated according to the filters pore size used, which was >0.2 µm and >3 µm.

As shown in the results, only the analysis of Faith Phylogenetics Diversity in different areas was significant, indicating that the change of location plays an important role in the change of bacterial community diversity. Factors such as physical-chemical conditions, types of marine ecosystems, biological interaction, nutrient availability, pollution, etc., strongly influence the difference in diversity from one location to another.

According to Trujillo & Thurman (2010), the variability of the environment is the main cause of the emergence of different species, "the more variable the environment, the more species are usually present." On the other hand, there are fewer species in the open ocean due to its generally consistent and uniform conditions, which do not force organisms to adapt. This relationship can be observed in offshore samples, which are more similar to the open ocean, when compared to Salamansa, which presents greater diversity. However, future analyzes will be necessary to know which of the factors mentioned above confers this particular characteristic to Salamansa. The same author also mentions that in the case of rocky coasts,

such as Salamansa, the diversity of species increases as you approach the line that is reached at low tide. This is because the conditions in these areas near the low tide line are more favorable for marine life, providing a more suitable environment for a wider range of species to establish and thrive (Trujillo & Thurman, 2010).

No significant differences in bacterial diversity between years or different filter sizes could be a result of sample size. Sample size can influence the power of the test to detect differences. The Kruskal-Wallis test, despite being suitable for our type of data (non-parametric), is sensitive to small changes. If the change is small, it will not show statistical significance, especially when seasonal variations are expected.

5.3 Beta diversity analysis

Using Bray-Curtis metrics to do a quantitative analysis of similarities and differences between the samples it was noticed that the samples collected at the coastal beaches showed similarity between them as well as with the samples collected at the Fish Market and Pipe. These two sampling sites are also located in the coastal region of the island, giving characteristics and conditions similar to coastal samples compared to offshore samples, as they can be equally influenced by environmental and anthropogenic factors. The exception occurred in some coastal samples, as is the case of Laginha (Mindelo) and Baía, which was grouped with offshore samples, giving it smaller amounts of ASV. In the first case (Laginha) it can be explained by the fact that a quantity of the sample was lost during processing. Samples from the Bay can be explained by the fact that in the area there is a large deposition and sedimentation of materials in the water, causing some nutrients to be deposited at the bottom of the water, reducing the amount of nutrients in the water column available for the growth of organisms.

Different pore sizes in filters, such as >3 μm and >0.2 μm, will have different impacts on the composition and type of microorganisms captured in the samples. Using a filter with pores larger than 3 μm will allow the capture of larger particles and microorganisms, such as some species of phytoplankton. Using a filter with pores larger than 0.2 μm will allow the capture of a wider range of microorganisms. This filter will capture a much larger portion of the total microbial community, allowing analysis of a wider range of microorganisms present in the sample. From the results obtained, the agglomeration/similarity between the filtered groups with the same size was noted, regardless of the season in which they were collected. With the

exception of some outlier samples from the >0.2 μm group that appeared clusters at >3 μm. One explanation would be the fact that smaller microorganisms have the ability to adhere to larger particles (Dudek et al., 2020).

5.4 Diversity of pathogenic bacteria with human health implications

More than 80% of the bacteria considered pathogenic found in the samples belong to those collected in coastal regions during the dry and cool season (March, 2023) (table 6). According to WHO several pathogenic microorganisms live freely in water or, once introduced, are able to colonize the environment, as is the case of *Vibrio* species found in the Baía sample. People who do water activities are unlikely to be ingesting large enough amounts of vibrio to cause gastrointestinal disease. But there is a risk for infections outside the digestive system, such as skin or ear infections. Even though it is not known exactly how much it takes for these infections to occur (WHO, 2013). However, there are also different sources of pollution of water habitat which include resuspension of sand, animal and human feces, rainwater runoff, human skin microorganisms, etc. (Stec et al., 2022).

Another case of free-living bacteria was found at the Fish Market. *Aeromonas spp*. are ubiquitous in fresh and marine surface waters, and sewage can also contain high numbers of *Aeromonas* (WHO, 2013). The fish market provides an environment conducive to the proliferation of this particular strain of bacteria. This is attributed not only to the ubiquitous presence of seawater but also to the utilization of freshwater for diverse operational processes and the disposal of processed fish viscera. The combined factors of available aquatic mediums and the introduction of freshwater from various sources create an environment that is particularly favorable for the growth and survival of *Aeromonas*.

The initial classification of places for recreational water activities is made based on combined evidence of the degree of influence of fecal material covering at least the three most important sources of fecal contamination: sewage, wastewater discharge near water recreation areas, and bathers' contamination, including excrement. Where human inputs are minimal, investigation of animal fecal inputs should be explored (WHO, *op cit*.). In samples from Laginha, one of the most popular beaches on the island, *Enterococcus* was detected, indicating possible fecal contamination. In the same place, the detection of *Pseudomonas putida* was also evident. Another fecal contamination was detected at Fish Market samples by indicating the presence of *Shigella*.

Many opportunistic pathogens have the ability to survive and proliferate in aquatic environments containing small amounts of nutrients, as is the case in our study area (Fang et al., 2011). These conditions can create episodes of microbial exposure and contamination associated with recreational and bathing activities. For example, it is assumed that gastrointestinal disorders and respiratory pathogens are transmitted from bather to bather through water and air and the population those most likely to be affected by disease include the young, the elderly and those with weakened immune systems (Fewtrell & Kay, 2015; WHO, 2013).

Among all collection sites, the Fish Market, Pipe and OSCM samples from the dry season were the ones with the greatest diversity of pathogenic bacteria, as shown in table 6. In these two environments, despite the lack of rainfall by that period of time to carry sediment and other pollutants to the water, strong anthropogenic and animal activities were observed, as well as deposition of animal remains, depletion of organic remains, etc.

Overall, samples collected in urban areas (Mindelo, corresponding to Laginha, Doca and Marina samples), characterized by a significant population density and, consequently, intense anthropogenic activity, revealed a greater diversity of pathogenic organisms.

However, more research must be done to identify the main source of pollution in these environments, that is, to better discern whether fecal contamination, for example, occurred through animals, defecation of bathers or the discharge of contaminating effluents directly into the water, etc. It would also be valuable to more closely examine the exact taxa that are present through more detailed analysis of sequences recovered, specific primers, or using more finely resolving methods than universal 16S approaches. In the end, it would also be relevant to carry out future research to assess and better understand the epidemiological situation of the island with its relation to the diseases that can be acquired through pathogens present in marine environments.

6 Conclusions

As a typical Atlantic environment, the waters of São Vicente present significant quantities of bacteria such as Proteobacteria, Bacteroidota and Cyanobacteria in both coastal and offshore areas, however, the composition of its bacterial community is not limited to just these three phyla. Contrary to what one might imagine, there was no significant difference in the diversity of the bacterial community between samples collected in different time period, but interesting findings were made, such as the identification of bacteria indicative of human activities, consequently indicating pollution, and also the identification of potential pathogenic organisms.

The quality of life of populations residing in coastal areas and island communities reveals an intricate interdependence with the oceans, in a relationship that operates in a bidirectional way. Oceans, notably coastal regions, also depend on human actions to maintain a healthy marine ecosystem. The findings of this research on microbial diversity hold significant implications for public health. The sequences from lineages that have pathogenic representatives are not very abundant compared to the general bacterial community. This is understandable since non-pathogenic strains can have a greater variety of functions, leading to more diverse sequences. Due to some evolutionary constraints, pathogens generally have simplified genomes that are adapted to their specific mode of parasitism or infection, which can lead to a reduction in global genetic diversity within pathogenic lineages. Another factor may be sampling bias. Our current dataset is constrained by its temporal scope, encompassing a mere two distinct weeks within an approximate eight-month timeframe. Consequently, in order to strengthen the reliability and representativeness of our analysis, there is a need to expand the sample size, thereby capturing a more comprehensive and temporally diverse range of data points.

Understanding the intricate makeup of microbial communities within various environments provides valuable insights into potential health risks. As diverse species coexist, the dynamic interactions between them can also lead to the emergence of pathogenic organisms that might threaten human well-being.

The application of management practices aimed at identifying and mitigating anthropogenic activities and other organisms that may cause damage and create favorable conditions for the proliferation of pathogenic microorganisms assume a crucial role in this context. In advance, according to the guidelines of the World Health Organization (WHO), it

is imperative to conduct annual sanitary inspections and microbial assessments of water quality in a combined way to determine possible changes in the sources of pollution. Even so, it is fundamental to recognize that the monitoring of water quality must be continuous and constant, based on the perception that the health of marine ecosystems. Therefore, human health is inextricably linked to maintaining the balance between anthropic activity and the vitality of the oceans.

7 Recommendations

It should be noted that the present study focuses on the characterization of the bacterial community in a comprehensive way, including an analysis of potentially harmful pathogenic species for humans. It is noteworthy that the approach employed does not aim to quantitatively quantify the organisms present, in order to determine whether their concentrations reach levels that may induce specific diseases. Instead, the focus is on the identification and qualitative assessment of bacterial species, considering both commensal and pathogenic ones to understand microbial diversity and its potential impact on public health.

By elucidating patterns and variations in microbial diversity, this research contributes to the identification of areas with the presence of harmful organisms, potential sources of contamination and the assessment of their impact on water quality. Such knowledge is fundamental for the development of effective strategies to manage and mitigate the proliferation of harmful microorganisms.

Research findings relating to microbial diversity have profound implications for public health considerations. The diverse composition of microbial communities, revealed by this study, supports the complex interaction between organisms that inhabit aquatic environments. This intricate web of interactions has the potential to influence the presence and prevalence of pathogenic microorganisms. Such knowledge allows proactive measures to be taken to prevent or manage potential waterborne disease outbreaks. Furthermore, identification of key microbial indicators associated with changes in water quality can inform the development of early warning systems, allowing authorities to promptly respond and mitigate health risks to ensure the safety of recreational environments, safeguarding and ultimately the health and well-being of communities.

8 References

Alhalimi, A. A., Alshammari, L. T., Al-Qurayn, A. K., & Al Rashed, A. S. (2022). Infective Endocarditis Caused by Pseudomonas stutzeri in a Pediatric Patient: A Case Report and Literature Review. *American Journal of Case Reports*, *23*(1), 105–109. https://doi.org/10.12659/AJCR.935743

American Academy of Microbiology. (2005). *Marine Microbial Diversity_ The key to Earth Habitability*.

Campbell, A. H., Marzinelli, E. M., Gelber, J., & Steinberg, P. D. (2015). Spatial variability of microbial assemblages associated with a dominant habitat-forming seaweed. *Frontiers in Microbiology*, *6*(MAR), 1–10. https://doi.org/10.3389/fmicb.2015.00230

CDC. (2010). *Klebsiella pneumoniae in Healthcare Settings | HAI | CDC*. https://www.cdc.gov/hai/organisms/klebsiella/klebsiella.html

CDC. (2021). *Legionnaires Disease Cause and Spread | CDC*. https://www.cdc.gov/legionella/about/causes-transmission.html

CDC. (2022a). *Meningococcal Disease | CDC*. https://www.cdc.gov/meningococcal/index.html

CDC. (2022b). *What is Anthrax? | CDC*. https://www.cdc.gov/anthrax/basics/index.html

CDC. (2023). *Sources of Infection and Risk Factors | Shigella – Shigellosis | CDC*. https://www.cdc.gov/shigella/infection-sources.html

Choi, J. K., Noh, J. H., Orlova, T., Park, M. O., Lee, S. H., Park, Y. J., Son, S., Stonik, I., & Choi, D. H. (2015). Phytoplankton and primary production. *Oceanography of the East Sea (Japan Sea)*, 217–245. https://doi.org/10.1007/978-3-319-22720-7_10

Crain, C., Kezer, K., Steele, S., Owiti, J., Rao, S., Victorio, M., Austin, B., Volner, A., Draper, W., Griffith, J., Steele, J., & Seifert, M. (2021). Application of ddPCR for detection of Enterococcus spp. in coastal water quality monitoring. *Journal of Microbiological Methods*, *184*(June 2020), 106206. https://doi.org/10.1016/j.mimet.2021.106206

Direção Geral dos Recursos Marinhos. (2019). *Plano de Gestão de Pesca 2019-2023*.

Drake, L. A., Doblin, M. A., & Dobbs, F. C. (2007). Potential microbial bioinvasions via ships' ballast water, sediment, and biofilm. *Marine Pollution Bulletin*, *55*(7–9), 333–341. https://doi.org/10.1016/J.MARPOLBUL.2006.11.007

Dudek, K. L., Cruz, B. N., Polidoro, B., & Neuer, S. (2020). Microbial colonization of microplastics in the Caribbean Sea. *Limnology And Oceanography Letters*, *5*(1), 5–17. https://doi.org/10.1002/lol2.10141

Fang, Y., Xie, G. lin, Lou, M. miao, Li, B., & Muhammad, I. (2011). Diversity analysis of Burkholderia cepacia complex in the water bodies of West Lake, Hangzhou, China. *Journal of Microbiology*, *49*(2), 309–314. https://doi.org/10.1007/s12275-011-0267-2

Fewtrell, L., & Kay, D. (2015). Recreational Water and Infection: A Review of Recent Findings. *Current Environmental Health Reports*, *2*(1), 85–94. https://doi.org/10.1007/s40572-014-0036-6

GDT. (n.d.). *INVENTÁRIO DOS RECURSOS TURÍSTICOS DO*.

Giovannoni, S. J. (2017). SAR11 Bacteria: The Most Abundant Plankton in the Oceans. *Annual Review of Marine Science*, *9*(1), 231–255. https://doi.org/10.1146/annurev-marine-010814-015934

Giovannoni, S. J., & Vergin, K. L. (2012). Seasonality in ocean microbial communities. *Science*, *335*(6069), 671–676. https://doi.org/10.1126/science.1198078

Gomes, N. (2014). *Modelação da circulação oceânica no Arquipélago de Cabo Verde Engenharia do Ambiente*. Instituto Superior Técnico. Ulisboa. 67

Haro-Moreno, J. M., Rodriguez-Valera, F., Rosselli, R., Martinez-Hernandez, F., Roda-Garcia, J. J., Gomez, M. L., Fornas, O., Martinez-Garcia, M., & López-Pérez, M. (2020). Ecogenomics of the SAR11 clade. *Environmental Microbiology*, *22*(5), 1748–1763. https://doi.org/10.1111/1462-2920.14896

Hill, P. G., Heywood, J. L., Holland, R. J., Purdie, D. A., Fuchs, B. M., & Zubkov, M. V. (2012). Internal and External Influences on Near-Surface Microbial Community Structure in the Vicinity of the Cape Verde Islands. *Microbial Ecology*, *63*(1), 139–148. https://doi.org/10.1007/s00248-011-9952-2

Imprensa Nacional de Cabo Verde. (2015). Boletim Oficial. *World Journal of Critical Care Medicine*, *5*(1), 27.

INE. (2022). *Anuário Estatístico 2020 - Cabo Verde*. *1*, 1–14.

Izaguirre-Anariba, D. E., & Sivapalan, V. (2020). Chryseobacterium indologenes, an Emerging Bacteria: A Case Report and Review of Literature. *Cureus*, *12*(1). https://doi.org/10.7759/cureus.6720

Khanna, A., Khanna, M., & Aggarwal, A. (2013). Serratia marcescens- a rare opportunistic nosocomial pathogen and measures to limit its spread in hospitalized patients. *Journal of Clinical and Diagnostic Research*, *7*(2), 243–246. https://doi.org/10.7860/JCDR/2013/5010.2737

Kim, J., Kim, D. Y., Yang, K. H., Kim, S., & Lee, S. S. (2019). Ruegeria lutea sp. Nov., isolated from marine sediment, Masan Bay, South Korea. *International Journal of Systematic and Evolutionary Microbiology*, *69*(9), 2854–2861. https://doi.org/10.1099/ijsem.0.003568

Kreader, C. A. (1996). Kreader Relief of Amplification Inhibition PCR with T4 Gene 32 Protein. *Applied and Enviromental Microbiology*, *62*(3), 1102–1106.

Magalhães, C. (2018). *Praias de Cabo Verde sem análises de qualidade*. https://expressodasilhas.cv/pais/2018/08/19/praias-de-cabo-verde-sem-analises-de-qualidade/59618

Mammeri, H., Bellais, S., & Nordmann, P. (2002). Chromosome-encoded β-lactamases TUS-1 and MUS-1 from Myroides odoratus and Myroides odoratimimus (formerly Flavobacterium odoratum), new members of the lineage of molecular subclass B1 metalloenzymes. *Antimicrobial Agents and Chemotherapy*, *46*(11), 3561–3567. https://doi.org/10.1128/AAC.46.11.3561-3567.2002

Martínez-García, S., & Pinhassi, J. (2019). Adaptations of microorganisms to low nutrient environments: Managing life in the oligotrophic ocean. *Encyclopedia of Microbiology*, 9–21. https://doi.org/10.1016/B978-0-12-809633-8.90696-4

Massey, J. (2015). *Decision Making and Coastal Risks : A Good Practice Guide , Decision Making and Coastal Risks : A Good Practice Guide Atlantic Network for Coastal Risks Management*. *January 2013*.

Morse, S. A. (1996). Neisseria, Moraxella, Kingella and Eikenella. *Medical Microbiology*. https://www.ncbi.nlm.nih.gov/books/NBK7650/

Mwanza, E. P., Hugo, A., Charimba, G., & Hugo, C. J. (2022). Pathogenic Potential and Control of Chryseobacterium Species from Clinical, Fish, Food and Environmental Sources. *Microorganisms*, *10*(5), 1–16. https://doi.org/10.3390/microorganisms10050895

Nicholson, W., & Paddock, C. (2023). *Rickettsial Diseases | CDC Yellow Book 2024*. https://wwwnc.cdc.gov/travel/yellowbook/2024/infections-diseases/rickettsial-diseases#table504

Noble, R. T., Blackwood, A. D., Griffith, J. F., McGee, C. D., & Weisberg, S. B. (2010). Comparison of rapid quantitative PCR-Based and conventional culture-based methods for enumeration of enterococcus spp. and escherichia coli in recreational waters. *Applied and Environmental Microbiology*, *76*(22), 7437–7443. https://doi.org/10.1128/AEM.00651-10

Rodrigues, C., & Cunha, M. Â. (2017). Assessment of the microbiological quality of recreational waters: indicators and methods. *Euro-Mediterranean Journal for Environmental Integration*, *2*(1). https://doi.org/10.1007/s41207-017-0035-8

Romero, O. E., Fischer, G., Karstensen, J., & Cermeño, P. (2016). Eddies as trigger for diatom productivity in the open-ocean Northeast Atlantic. *Progress in Oceanography*, *147*, 38–48. https://doi.org/10.1016/j.pocean.2016.07.011

Ryu, H., Grond, K., Verheijen, B., Elk, M., Buehler, D. M., & Santo Domingo, J. W. (2014). Intestinal microbiota and species diversity of Campylobacter and Helicobacter spp. in migrating shorebirds in Delaware Bay. *Applied and Environmental Microbiology*, *80*(6), 1838–1847. https://doi.org/10.1128/AEM.03793-13

Ryu, H., Henson, M., Elk, M., Toledo-Hernandez, C., Griffith, J., Blackwood, D., Noble, R., Gourmelon, M., Glassmeyer, S., & Santo Domingo, J. W. (2013). Development of quantitative PCR assays targeting the 16s rRNA genes of enterococcus spp. and their application to the identification of enterococcus species in environmental samples. *Applied and Environmental Microbiology*, *79*(1), 196–204. https://doi.org/10.1128/AEM.02802-12

Said, M., Tirthani, E., & Lesho, E. (2022). *Enterococcus Infections - PubMed*. https://pubmed.ncbi.nlm.nih.gov/33620836/

Sampaio, A., Silva, V., Poeta, P., & Aonofriesei, F. (2022). Vibrio spp.: Life Strategies, Ecology, and Risks in a Changing Environment. *Diversity*, *14*(2), 1–26. https://doi.org/10.3390/d14020097

Schattenhofer, M., Fuchs, B. M., Amann, R., Zubkov, M. V., Tarran, G. A., & Pernthaler, J. (2009). Latitudinal distribution of prokaryotic picoplankton populations in the Atlantic Ocean. *Environmental Microbiology*, *11*(8), 2078–2093. https://doi.org/10.1111/j.1462-2920.2009.01929.x

Sonnenschein, E. C., Nielsen, K. F., D'Alvise, P., Porsby, C. H., Melchiorsen, J., Heilmann, J., Kalatzis, P. G., López-Pérez, M., Bunk, B., Spröer, C., Middelboe, M., & Gram, L.

(2017). Global occurrence and heterogeneity of the Roseobacter-clade species Ruegeria mobilis. *ISME Journal*, *11*(2), 569–583. https://doi.org/10.1038/ismej.2016.111

Standard Methods Committee of the American Public Health Association. (2022a). *9213 RECREATIONAL WATERS - Standard Methods For the Examination of Water and Wastewater*. https://www.standardmethods.org/doi/abs/10.2105/SMWW.2882.187?gclid=Cj0KCQiAutyfBhCMARIsAMgcRJQbfRaTKwdeRja5XvZg4tQMTyTMAddyXlGBTro_NCk24WAHC5fjGRsaAjAfEALw_wcB

Standard Methods Committee of the American Public Health Association. (2022b). *9260 INTRODUCTION TO DETECTING PATHOGENIC BACTERIA - Standard Methods For the Examination of Water and Wastewater*. https://www.standardmethods.org/doi/full/10.2105/SMWW.2882.201?role=tab

Stec, J., Kosikowska, U., Mendrycka, M., Stępień-Pyśniak, D., Niedźwiedzka-Rystwej, P., Bębnowska, D., Hrynkiewicz, R., Ziętara-Wysocka, J., & Grywalska, E. (2022). Opportunistic Pathogens of Recreational Waters with Emphasis on Antimicrobial Resistance—A Possible Subject of Human Health Concern. *International Journal of Environmental Research and Public Health*, *19*(12). https://doi.org/10.3390/ijerph19127308

Thomas, B. S., Okamoto, K., Bankowski, M. J., & Seto, T. B. (2013). A lethal case of pseudomonas putida bacteremia due to soft tissue infection. *Infectious Diseases in Clinical Practice*, *21*(3). https://doi.org/10.1097/IPC.0b013e318276956b

Trujillo, A., & Thurman, H. (2010). *Essentials of Oceanography*. https://doi.org/10.1177/014662168400800314

Vieira, R. H. S. F., Oliveira, A. C. N., & Sousa, O. V. (2007). Monitoramento Microbiológico Das Águas E Areias Das Praias Do Meireles E Do Futuro (Fortaleza – Ceará). *Boletim Técnico Científico Do CEPNOR*, *7*(1), 17–26. https://doi.org/10.17080/1676-5664/btcc.v7n1p17-26

WHO. (2013). *WHO | Guidelines for safe recreational water environments*. *1*, 118–127. http://www.who.int/water_sanitation_health/bathing/srwe1/en/

Yeh, Y.-C., Needham, D. M., Sieradzki, E. T., & Fuhrman, J. A. (2018). Taxon Disappearance from Microbiome Analysis Reinforces the Value of Mock Communities as a Standard in Every Sequencing Run. *MSystems*, *3*(3). https://doi.org/10.1128/msystems.00023-18

Appendix

1. Mock community organisms

SAR11_DQ009197.1
OCS155_DQ009124.1
Proc_DQ009356.1
SAR86_DQ009149.1
Plancto_AF029077
Chrysochromulina_KF621005.1
DinoGroupII_KC582918.1
Guinardia_delicatula_KT861015.1
Lingulo_polyedr_AB693196.1
Cryothecomonas_sp_AB252744.1

2. Gel electrophoresis protocol

For the procedure, first, the base (plate) and the electrophoresis comb were chosen according to the number of samples. 2.25g of agarose + 150 ml of 1xTAE was mixed and heated slowly in the microwave until completely dissolved without showing bubbles. Then, 6 drops of ethidium bromide (dye) were added and homogenized. The mixture was poured onto the plate and allowed to cool and solidify. It is important not to let bubbles form near the comb. Then the gel was carefully placed in the electrophoresis machine previously filled with 1xTAE. A parafilm was prepared and then 2 µl of blue juice (loading dye) was added to it in the form of small drops (number of drops depends on the number of samples). 5 µl of the sample (pcr product) was also added to each drop of loading dye and mixed. Then, a total of 7 µl of each mixture was Pipetted into the well of the gel. For the molecular weight markers, 1 µl of Invitrogen ladder (100 bp), 1 µl of loading dye and 8 µl of molecular H2O were mixed in the parafilm and placed on the first tooth of the comb. Finally, the electrophoresis power source was turned on with the negative and positive poles properly connected, at 100 volts and observed every 20 min. It was left for a total of 40 min.

Gel electrophoresis (1.5%) was used to measure the size of DNA fragments by comparison with solutions with DNA fragments of known size, called molecular weight markers (ladder) to assess the purity and integrity of the sample. Charged nucleic acid molecules migrated through the gel matrix at different speeds, depending on their size and charge. Larger molecules

traveled more slowly through the gel, while smaller molecules migrated through it more quickly. A variety of staining techniques, such as the ethidium bromide stain that was used in this study, can be used to visualize the unique bands or patterns that are produced as a result of this gel separation.

3. PCR products clean-up/ purification and normalization protocol

PCR reaction tubes with the samples were centrifuged using a microcentrifuge in a short spin. Once the beads were warm, they were added to the PCR tubes at a 0.8x ratio (12 µl of beads for 15 µl of PCR product) to achieve a lower concentration of beads, which was preferred to remove primer dimer while retaining enough DNA for sequencing. The strip tubes containing the PCR product were vortexed for several seconds, followed by a 5-minute waiting period for the beads and DNA to bind. After the incubation, the tubes were placed in a magnetic separator and left for 3-5 minutes to enable complete separation. The clear buffer was removed and discarded as the DNA adhered to the magnetic beads. Next, 100 µl of freshly made 80% ethanol was added to the tubes, and the mixture was vortexed for 5-6 seconds. Once again, the PCR tubes were placed on the magnet for separation, and the ethanol was removed while keeping the tubes on the magnet. This addition of 100 µl of ethanol was repeated and removed. The tubes were spun for several seconds to collect any remaining ethanol in the bottom and then placed back on the magnetic plate. The remainder of the ethanol was removed, and the tubes were allowed to dry on the magnet with the tubes open for approximately 5 minutes, ensuring that the beads did not crack to prevent DNA loss. After drying, 10 µl of TE Buffer was added to the beads, and the mixture was Pipetted several times to break up the bead pellet and elute the DNA. The tubes were incubated for about 5 minutes and then separated on the magnet. Approximately 9.5 µl of the TE buffer containing the DNA was Pipetted off, taking care not to remove any beads (although a small amount of beads was acceptable but not ideal). The DNA was then quantified using PicoGreen with a 1 to 3 µl dilution of molecular grade water. Finally, PCR products were stored at -20 degrees to potentially reduce index crosstalk.

After completing the PCR and equilibration at room temperature, ten (10) µl of each reaction mixture from the PCR plate was transferred to the Binding Plate (BP5) using a multi-channel Pipettor. The Binding Buffer (NB7) was thoroughly mixed by swirling the bottle, and then 10 µl of Binding Buffer was added to each sample well. The PCR products were mixed well with the Binding Buffer by Pipetting the liquid up and down 5-6 times, carefully avoiding

introducing bubbles. Next, the Binding Plate was sealed with an adhesive plate cover and incubated at room temperature for 30 minutes. After the incubation period, the adhesive plate cover was removed. The liquid from the Binding Plate was then removed using the following method: the solution was decanted by quickly flipping the plate over a waste container, shaking it briskly, and then placing the inverted plate on absorbent paper to remove excess liquid. To wash the plate, 25 µl of Wash Buffer (WB2) containing ethanol was added to each well, and the solution was mixed by Pipetting up and down 1-2 times. The plate was incubated for 30 seconds at room temperature, and then the Wash Buffer was removed using the methods described in the previous step. Steps 1 and 2 (transfer of PCR products and addition of Binding Buffer) were repeated once for two washes. After the final wash, the plate was blotted dry on absorbent paper and air-dried in a lab hood for 5-10 minutes to remove residual liquid. Next, 20 µl of Elution Buffer (EB1) was added to each well of the Binding Plate. The plate was sealed with a new adhesive cover and vortexed for 30 seconds. It was then briefly centrifuged at maximum speed to collect all the liquid at the bottom of the wells. The purified and normalized products are substantially free of contaminants such as salts, dNTP and primers and can be used for next-generation sequencing analysis.

4. Library preparation

For this, two pools of indexed samples were made. Samples with DNA concentration less than 0.7 ng/µl were pooled 1 µl and greater than 0.7 ng/µl were pooled 2 µl. A clean-up was performed for both pools with a ratio of 0.8x of bead clean-up with 15 µl of final elution for each pool. After quantification (PicoGreen), a dilution was necessary to adjust the volume for the final pool with both batches together, totalizing the final volume of 20 µl.

Data availability

The data obtained in this study can be obtained by contacting my supervisor Dr. David Needham or through WASCAL Cabo Verde.