Exploring planktonic microbiome diversity in the coastal ecosystems of the North of Portugal

Simão Matos Conceição Horta

Thesis to obtain the Master of Science Degree in

Microbiology

Supervisors: Prof. Catarina Maria Pinto Mora de Magalhães
and Prof. Rodrigo da Silva Costa

Examination Committee

Chairperson: Prof. Jorge Humberto Gomes Leitão
Supervisor: Prof. Catarina Maria Pinto Mora de Magalhães
Members of the Committee: Prof. Maria de Fátima Magalhães Carvalho

November 2022

Declaration

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Preface

The work presented in this thesis was performed at the EcoBiotec lab, of Interdisciplinary Center of Marine and Environmental Research, University of Porto, (Porto, Portugal), during the period September-August 2021/2022, under the supervision of Dr. Catarina Maria Pinto Mora de Magalhães and Dr Maria Paola Tomasino. The thesis was co-supervised at Instituto Superior Técnico by Prof. Rodrigo da Silva Costa.

Acknowledgments

The present thesis project was only possible thanks to the guidance, knowledge and experience from the senior researchers that supervised and guided this Master of Science thesis namely, Professor Catarina Maria Pinto Mora de Magalhães, and Dr. Maria Paola Tomasino, both from the Interdisciplinary Center of Marine and Environmental Research, University of Porto. I would like also to mention the important practical guidance and help from Eva Silva Lopes, and the vital bioinformatic analysis support provided by Francisco Daniel de Oliveira Pascoal, also from the Interdisciplinary Center of Marine and Environmental Research, University of Porto. Finally, I would like also to acknowledge Prof. Rodrigo da Silva Costa, from the Microbial Ecology and Evolution team of the Biological Sciences Research Group, Instituto Superior Técnico, University of Lisbon, for the supervision and revision of the present thesis.

I would like also to express thanks to the projects that financed this work, namely the project Ocean3R (NORTE-01-0145-FEDER-000064) and the project ATLANTIDA (ref. NORTE-01-0145-FEDER-000040) supported by the Norte Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement and through the European Regional Development Fund (ERDF). This study was also partially funded by the Strategic Funding UIDP/04423/2020 and UIDB/04565/2020 through national funds provided by FCT.

The personal acknowledgments go to all the professors from high school to the university, that instil in me the passion for true knowledge and science. A final personal acknowledgment to my family and friends that support me through this path of personal improvement.

Abstract

Microbial communities are essential in the marine ecosystems. With the advent of the industrial revolution, the environmental conditions are now changing at higher rates than before. In order to understand how the microbial communities are facing these changes, and to characterize the microplanktonic biodiversity, the present thesis aims to start evaluating and catalogue the biodiversity of the marine microorganisms on the north coast of Portugal. To do that, microbiome monitoring programs started in the Douro River estuary and in the NorthWest coast of Portugal. Periodic samples were collected and processed using standard methodologies including a 16S and 18S rRNA gene metabarcoding workflow.

Results from the Douro River estuary showed that the unicellular eukaryotic communities had an increased spatial dissimilarity, over the prokaryotic communities. It was also demonstrated that tides influence more strongly the downstream communities, over the upstream communities in the estuarine system studied (Douro Estuary). The surface communities were also compared with the water column bottom communities, and no significant influence was found registered in the estuarine stations. Temporally, the Autumn and Winter seasons significantly shaped both microbial communities, with a stronger influence in the prokaryotic microbiomes. The prokaryotic communities from the NW coast of Portugal showed an increased spatial dissimilarity over the unicellular eukaryotic communities. Additionally, the distribution of coastal prokaryotic communities was influenced by a greater number of environmental parameters than the unicellular eukaryotic communities. With this research our understanding of such important ecosystems will be enhanced, providing a theoretical foundation for the marine ecological health management.

Keywords: microbial ecology; marine ecology; metabarcoding; Douro estuary; Portugal north coast

Resumo

As comunidades microbianas são essenciais nos ecossistemas marinhos. Com o advento da revolução industrial, as condições ambientais estão a mudar com mais frequência do que antes. De forma a compreender como as comunidades microbianas estão a enfrentar estas alterações, e caracterizar a biodiversidade microplanctónica, a presente tese pretende iniciar a avaliação e catalogalização da biodiversidade dos microrganismos marinhos na costa norte de Portugal. Para tal, iniciaram-se programas de monitorização do microbioma no estuário do rio Douro e na costa noroeste de Portugal. Amostras periódicas foram coletadas usando metodologias padrão, incluindo um fluxo de trabalho de metabarcoding de genes 16S e 18S rRNA.

Os resultados do estuário do rio Douro mostraram que as comunidades eucarióticas unicelulares apresentaram maior dissimilaridade espacial, em relação às comunidades procarióticas. Foi também demonstrado que as marés influenciam mais fortemente as comunidades a jusante, do que as comunidades a montante (Estuário do Douro). As comunidades da superfície também foram comparadas com as comunidades do fundo da coluna da água, não sendo encontrada qualquer influência significativa. Temporalmente, as estações de outono e inverno moldaram significativamente ambas as comunidades microbianas, com uma influência mais forte nos microbiomas procarióticos. As comunidades procarióticas da costa NW de Portugal mostraram uma maior dissimilaridade espacial sobre as comunidades eucarióticas unicelulares. Além disso, a distribuição das comunidades procarióticas foi influenciada por um maior número de parâmetros ambientais do que as comunidades eucarióticas unicelulares. Com esta pesquisa, a compreensão destes ecossistemas será melhorada, fornecendo uma base teórica para a proteção da ecológica marinha.

Palavras-chave: ecologia microbiana; ecologia marinha; metabarcoding; estuário do Douro; costa norte de Portugal

List of Contents

List of Figures

Figure 10: Alpha diversity (number of observed ASVs and Shannon index) of the Douro samples, collected in 2021/2022, for the different sampling depths (Bottom and Surface) and season (Autumn and Winter). The Figure 10a, was generated using the 16S rRNA gene dataset and the Figure 10b the 18S rRNA gene dataset.

Figure 11: Beta diversity analysis of the 16S rRNA gene Douro dataset, collected during 2021/2022; The Figure 11a PCoA ordination plot with the different depths (surface and bottom) distinguished by different shapes (circle and triangle), and the different seasons (Autumn and Winter) are notable with a different colour (red and clue); In the Figure 11b, hierarchical cluster analysis.

Figure 12: Beta diversity analysis of the 18S rRNA gene Douro dataset, collected during 2021/2022; The Figure 12a PCoA ordination plot with the different depths (surface and bottom) distinguished by different shapes (circle and triangle), and the different seasons (Autumn and Winter) are notable with a different colour (red and clue); In the Figure 12b, hierarchical cluster analysis.

Figure 13: Taxonomy diversity graph of the Douro 18S rRNA gene dataset, collected at 2021/2022, where is represented the taxonomic profiles in a bar plot. The Figure 13a, represents the 16S rRNA gene dataset and the Figure 13b the 18S rRNA gene dataset.

Figure 14: CCA ordination plot showing for the 16S rRNA gene (A.) and 18S rRNA gene (B.) data sets. The length of the environmental arrows is indicative of the degree of influence of the samples.

Figure 15: Environmental gradients of the Douro estuary where in the X axis is present each station sampled at different tides (H-High Tide; L-Low Tide); the Y axis represents the numeric scale where is measured the concentration of the temperature (C^{o}), Salinity (PPT), Nitrate (µM), Nitrite (µM), Phosphate (µM), Ammonia (µM), and pH.

Figure 16: Rarefaction curves of the Douro samples collected in 2016; where the Figure 16a displays the 16S rRNA gene dataset, while the 18S rRNA gene dataset is showed on the Figure 16b.

Figure 17: Alpha diversity (Observed and Shannon measure) of the Douro samples, collected in 2016, where the different sampling tides (High and Low) have different colour and shape; The Figure 17a, represents the 16S rRNA gene dataset and the Figure 17b the 18S rRNA gene dataset.

Figure 18: Beta diversity analysis of the Douro samples collected during 2016, for 16S rRNA gene rRNA (Figure 18a) and 18S rRNA gene rRNA (Figure 18b) datasets.; The analysis was graphically represented on the PCoA ordination plot, where the different stations have different colours, and the tides (High and low) are distinguished by different shapes (circle and triangle).

Figure 19: Heatmaps of the relative abundance of the different prokaryotic (A) and unicellular eukaryotic (B) phyla along the different Douro estuary stations in Winter (2016) at low and high tides.

Figure 20: Correlation of the environmental conditions with the 16S rRNA gene (Figure 20a) and 18S rRNA gene (Figure 20b) biological samples in a CCA ordination plot.

Figure 21: Map of the spatial distribution of the environmental gradients measured (Y), across predefined sampling stations (X) in the Portuguese NW coast. **a)** Plot of the temperature (C^{o}) parameter; **b)** Plot of the salinity (PPT) parameter; **c)** Plot of the pH parameter; **d)** Plot of the chlorophyll (µg/L) parameter; **e)** Plot of the ammonia (µM) parameter; f**)** Plot of the nitrites (µM) parameter; g**)** Plot of the silica (µM) parameter; **h)** Plot of the nitrates (µM) parameter; **i)** Plot of the phosphate (µM) parameter; **j)** Map where is represented the location of the different stations.

List of Tables/ Annexes

Annex 7- Table with the environmental data and the CTD parameters measured on the Douro estuary campaign in 2012 and 2022.

Annex 8- Table with the environmental data and the CTD parameters measured on the Douro estuary campaign in 2016.

Annex 9- Table with the environmental data and the CTD parameters measured NW coast of Portugal.

Annex 10- Table presenting the values of the PERMANOVA analysis for the Douro 16S rRNA gene (Annex 10a.) and 18S rRNA gene (Annex 10b.) dataset, collected in 2021 and 2022.

Annex 11- Table presenting the values of the PERMANOVA analysis for the Douro 16S rRNA gene (Annex 11a.) and 18S rRNA gene (Annex 11b.) dataset collected in 2016.

Annex 12- Table presenting the values of the PERMANOVA analysis for the costal samples of the 16S rRNA gene (Annex 12a.) and 18S rRNA gene (Annex 12b.) dataset.

List of Abbreviations

PCCC. Portugal coastal counter current

LUCA. *Last universal common ancestor*

DOM. Dissolved organic matter

DSi. Dissolved silicic acid

EMO BON. European Marine Omics Biodiversity Observation Network

NW. NorthWest

ASVs. Amplicon sequencing variants

OTUs. Molecular operational taxonomic units

HCL. Hydrochloric acid

e-DNA. Environmental DNA

PCA. Principal component analysis

NGS. New generation sequencing

1. Introduction

1.1 Historical perspective on the marine microbiome on Earth

Earth was formed 4.5 billion years ago[1]. At that time, the molten surface of this planet was too hot to bare any known form of life. After such a period, our planet started to gradually cool down, which allowed the emergence of a solid crust and liquid water, fundamental for life to thrive[1]. Life on Earth started 4.1 billion years ago, in the deep ocean hydrothermal systems where the abiotic conditions favoured early life[1]. Since then, natural selection allowed Bacteria and Archaea to diverge from the *last universal common ancestor* (LUCA). Over time, the diversity kept increasing allowing the beings to acquire a wide variety of metabolic arsenal.

In the beginning, the Earth's atmosphere was mainly composed of carbon dioxide (CO_2). As a result, several microorganisms started to exploit this abundant resource. However, Cyanobacteria were unique in their approach. They also assimilated CO_2, but unlike the anoxygenic phototrophs, they were the first to release oxygen to the atmosphere, making them the first oxygenic phototrophs microorganisms that appeared on Earth[1]. Over time, the oxygen released by such microorganisms started to build up in the Earth´s atmosphere leading to the great oxidation event, around 2.4 billion years ago. From this event, a myriad of new life forms appeared due to the exploitation of oxidative respiration, whose advantage resides in a higher energy yield[1]. Additionally, the gradual conversion of oxygen (O_2) into ozone (O_3) starts to accumulate in the atmosphere, resulting in the ozone layer. This atmospheric layer still protects Earth from part of the UV solar radiation. Without this barrier, Earth´s surface would be a very challenging environment for life. With the advent of the ozone shield, organisms could start to explore terrestrial habitats resulting in an even-greater diversity[1].

Large multicellular organisms only appeared 0.6 billion years ago, meaning that microorganisms were the only inhabitants of the planet Earth for over 3,5 billion years. In that period, they profoundly shaped and regulate the Earth´s atmosphere, biosphere, and geosphere, which ultimately, allowed the development of favourable conditions for life to thrive on Earth[2]. Across the eras, microorganisms had to face a constantly changing environment, frequently characterized by climate fluctuations and geological upheavals that usually result in widespread mass extinctions[3–5]. Even so, microorganisms kept evolving and surviving showing their great resilience provided by their broad functional diversity[2].

1.1.1 *The advent of Mankind*

Homo sapiens established themselves as a species about 160 000 years ago[6]. They transitioned from nomadic to settlers with the advent of agriculture and the domestication of animals, 11 500 years ago. Two hundred years ago, the industrial revolution took place powered by the burning of fossil fuels like coal, petroleum, and natural gas. The burning of fossil fuels increases the accumulation of greenhouse gases in the atmosphere like carbon dioxide, nitrous oxide, and methane[7]. The greenhouse gases have the ability to retain heat from the sun´s radiation, maintaining the Earth´s temperature optimal for life[8].

However, the increase accumulation of such gases is currently triggering an increased greenhouse effect, which means, that the heat is trapped more effectively in the atmosphere, resulting in the Earth´s warm up[8]. Currently, the atmospheric concentration of carbon dioxide is the highest, in two million years [9]. Methane and nitrogen dioxide levels are at their peak concentration, since the last 800 000 years[9]. As result, between 2001-2020, the global surface temperature was one Celsius degree higher, than between 1850 to 1900 (Figure 1)[9].

Marine waters are also being disturbed by global warming. The ocean is continuing to warm in a multi-decadal trend that is well documented[10]. The increase in global warming waters will certainly have an impact on the biogeography patterns of organisms, ranging from phytoplankton to marine mammals[10]. Another problem caused by the increase in the global temperature is the increased ocean stratification[10]. The global warming is making the upper 200 meters ocean layer less dense and therefore inhibiting the water exchange between surface and deep waters. This event is currently disrupting the ocean nutrient cycles and negatively impacting the primary producers[10]. The increasing of carbon dioxide concentration in the atmosphere is also negatively affecting the marine waters[10]. For the past 50 million years, the open ocean pH was increasing. However, in the last decades, the pH is unusually low, when compared with the past 20-million-year period. This phenomenon can be mainly explained by the increased uptake rate of carbon dioxide by the oceans, which lead to the acidification of the waters[10]. There are also high confidence studies that demonstrate the drop of oxygen levels in many upper ocean regions since the mid-20th century [10]. This can be explained by the warming weather, associated with changes in the ocean´s physics and biogeochemistry [10]. Additionally, humans had also a huge impact on the surface melting of ice sheets and on the raising of the global sea levels[9].

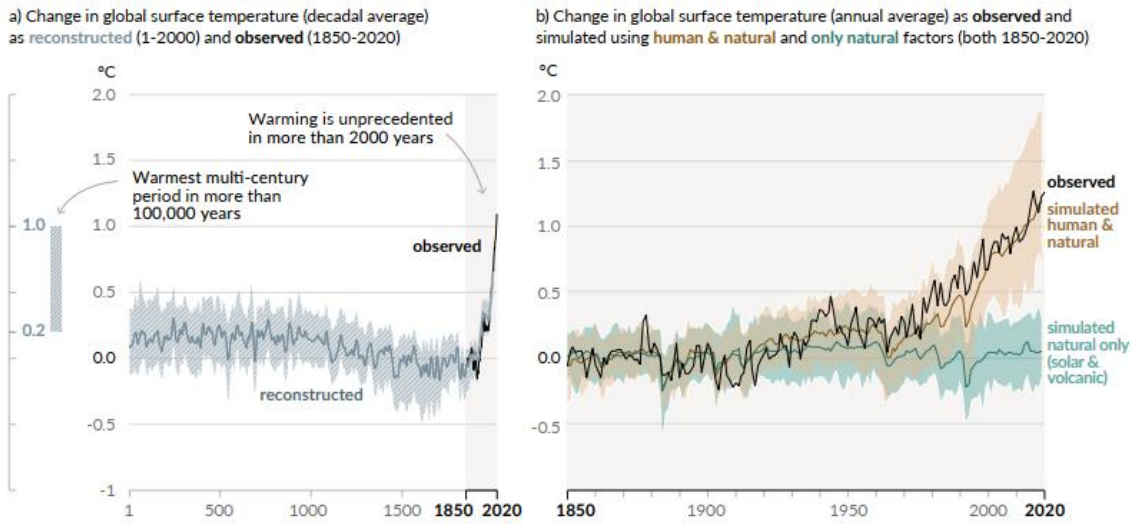


Figure 1: Changes in global surface temperature; a) Reconstruction of global surface temperature (decadal average) from 1 to 2000 and observed data from 1850 to 2020; The x-axis is temporal scale, and the y-axis is the variation of the temperature. b) Changing in global surface temperature (annual average) from 1850 to 2020, where is compared the temperature observed (black) with the simulated temperature resulting from of human and natural factors (brown) and with the simulated temperature resulting only from natural factors (green); The x-axis is temporal scale, and the y-axis is the variation of the temperature. Image taken from IPCC_AR6_WGI_2021[9].

Unfortunately, the growing human impact on planet Earth goes far beyond the burning of fossil fuels. Population growth and food production can seriously impact the costal water bodies [11]. Large amounts of domestic, industrial, cattle raising, and agricultural waste are produced every day. If these residues are poorly treated and discharged into the waters, they can cause the pollution of that environment [12]. Pollution is the contamination that results in biological adverse effects for individuals or communities[13]. Domestic sewage and industrial effluents are another major source of pollutants that can transform a healthy ecosystem in a toxic environment[14]. The main pollutants found in these discharges are heavy metals that can bioaccumulate in the food chains causing several nefarious effects on the species present in the environment [15]. Chemical fertilizers are another major source of pollution. The main sources are the intensive animal industries, which produce large quantities of nutrient-rich wastewater, and the intensive agriculture, which rely heavily on the use of nutrients like nitrogen and phosphorus[16]. These substances can reach a river or infiltrate the soil, that by its turn, will flow until the sea[17]. The increased supply of organic matter to the coastal ecosystems will enable eutrophication to occur. Eutrophication is the increased growth of phytoplankton biomass which can lead to hypoxic and later, to anoxic conditions in the environment[18]. Plastic debris is recently one of the main sources of pollution in the marine environment [14]. Since they are non-degradable, their presence in the environment can be extended for a long time. When this debris reaches the water bodies, they start to break into small pieces due to exposure to the UV lights[14]. These small debris are now accessible to marine life that will interact with it, disrupting the marine food webs and their natural equilibrium[14].

In conclusion, after the industrial revolution, climate changes were no longer just driven by natural factors. The release of greenhouse gases and pollution change the climate periodicity to an unstable global climate where more extreme hot periods will happen [19]. In the coastal areas, these events promoted the loss of the water quality and the loss of biodiversity. This phenomenon is currently threatening the balanced marine ecosystem with direct implications for humans.

1.2 The relevance of marine microorganism

In the marine ecosystem, the first organisms that are going to respond to stressors, like global warming and pollution, are the marine microbial communities [10].Microorganisms are the foundations of every ecosystem in the world (including the marine trophic chains), representing the major fraction of Earth's biomass. They set the biotic conditions, upon which more complex life forms arose[20]. These organisms can appear mostly as undifferentiated single cells, but some can form complex structures [1]. Since they typically live in a microbial community, their activities are regulated by interactions with each other, with the environment and interactions with more complex organisms [1].

In the marine ecosystem, microorganisms are present everywhere[21]. They have been evolving for nearly 4 billion years, which has resulted in an enormous biodiversity and metabolic versatility[21]. This diversity allowed them to be essential for biogeochemical cycling processes and therefore, crucial for the marine ecosystems and Earth´s climate[21]. Additionally, they are involved in more than half of the Earth´s primary production, and 80% of the world's oxygen is produced by photosynthetic marine microorganisms[22].

Specially in the marine realm, they are essential to provide organic carbon to the environment, sustaining the very foundation of the marine food webs[23]. Microorganism have also the ability restore a contaminated environment [24]. However, when the contamination takes higher proportions, the microbial communities will not be able to remediate the situation, leading often to the disrupting of the nutrient cycle. As consequence, harmful algae, and pathogen organisms start to appear[24].

1.2.1 *Plankton Definition*

Plankton has its origins in the ancient Greek word "planktos", which means drifter. Plankton is all the organisms commonly found in the water column, that are drifting free-living bodies, being passively transported by the ocean currents, where they only can actively change their buoyancy[25]. In the plankton community, we can distinguish a variety of nutrition styles. The autotrophic organisms are the foundations of the marine food web, directly and indirectly feeding virtually all kinds of marine creatures[26]. They are powered by the sun radiation, nutrients, and CO_2, that will drive photosynthesis. These planktonic forms are called the phytoplankton[27]. Zooplankton organisms are the primary consumers and phagotrophic predators in plankton. They graze smaller organisms, connecting the primary producers to the more complex multicellular life forms[28]. They vary widely in range size going from microscopic unicellular protists to massive and complex organisms[28].

The smallest living organisms on the seawater are the picoplankton that ranges from 0.2 to 2 µm[29]. It´s been proven that they appear ubiquitously and at high densities. On plate count, their numbers can range from 10^5 cells/ml in the oligotrophic realms, to about 10^6 cells/ml in eutrophic coastal areas [29]. Picoplankton is constituted mainly by prokaryotes, namely the Cyanobacteria, that are ones of the responsible for the photosynthesis that happens in the oligotrophic waters[29]. The nanoplankton can range from 2 to 20 µm. The pigmented flagellates, the chlorophytes and some tiny diatoms are examples of the photosynthetic diversity [29]. At this level is also possible to identify phagotrophic organisms like non pigmented flagellates[29]. Usually, they are less abundant than picoplankton, presenting numbers on surface water, that can range from 10^3 to 10^4 cells/ml [29]. Microplankton organisms are sized from 20 to 200 µm, which will include the photosynthetic diatoms and larger dinoflagellates [29]. This group also possess phagotrophic beings that include the ciliates, dinoflagellates, radiolaria and achantharia. Their abundance level is between 1 to 10 organisms per milliliter [29].

1.2.2 *The marine environment*

The coastal areas are the marine environment that host the highest diversity of microorganisms. This phenomenon can be explained by the gradients of light, temperature, nutrients, and salinity that are important to promote diversity shifts in microbial communities' taxonomy. In addition, microbial communities can also vary due to anthropogenic impacts like the contamination from domestic and industrial sewage or agricultural run-off that [24]. Stressors that change their community composition will ultimately affect interactions between a huge range of marine organisms, with unforeseen

consequences[10]. The coastal areas have always attracted human settlements being the shelter for 61% of the world's population [12]. From that population, 71% of people live within 50 km of an estuary and because of that the world's most important urban areas are often located around estuaries[12]. Coastal areas are occurring when the land meets the ocean. In these areas several marine habitats can be present like the estuarine ecosystem. Estuaries are a coastal body of water that are partially enclosed by land and partially open to the sea. This water body receives discharge from a river(s), and thus its salinity is often at lower levels than the littoral zone. The salinity can vary temporally and along its length[30]. The coastal areas offer several services to the communities that live nearby since these areas are rich in food and water sources. They offer protection from extreme climate events and can be also used for transportation and recreational opportunities[11].Cycling of nutrients and the carbon sequestering is also a relevant process that happens here because it allows the water purification from these elements. In addition, estuaries also provide essential nursery habitats for many different species, especially for human food sources like fishes[11].

1.2.3 *The microbial loop*

In 1983, Azam et al. (1983)[31] described for the first time the microbial loop as the continuous delivery of dissolved organic matter (DOM), from the lower trophic levels, at the microbial realm, to the increasingly higher trophic levels [31]. For the microbial loop to happen, primary production is supplied by the picoplankton and nanoplankton phototrophic prokaryotes and eukaryotes [29]. Here, at the lower trophic chains, size is important because the competition for dissolved nutrients favors the small organisms, that are typically ranging from 0.3 to 1 µm[32], and have a high surface-to-volume ratio. These features give them the competitive advantage to uptake nutrients even at low concentrations[31]. In the euphotic zone, the phytoplankton thrives due to the sunlight and the nutrient abundance, enabling optimal conditions for inorganic carbon to turn into organic structural cell material[21].

The phytoplankton will grow abundant, but their numbers are going to be controlled by the grazing activity of small protozoa and a large variety of flagellates[32]. As a result of the grazing activity, fecal pellets will be produced[21]. Viruses are another important grazing agent that can cause similar rates of mortality, where phytoplankton cells will be lysed by viral attack[32]. As consequence, a large part of the primary production is released to the environment into the form of DOM[31]. It is estimated that 5 to 50% of the carbon fixated is released as DOM[31]. This organic matter is then utilized by the heterotrophic bacteria and archaea for their own growth[31]. Here, size is still important because organisms tend to feed on particles one order of magnitude smaller than themselves[31]. In this case, the organisms sized between 0.3 to 1 µm feed on DOM, sized between <1 nm to 0.45 µm.

In its turn, heterotrophic flagellates, sized between 3 to 10 µm, feed on those heterotrophic bacteria and archaea (0.3 to 1 µm), who fed on DOM[31]. The flagellates also feed on autotrophic Cyanobacteria in the same size range (0.3 to 1 µm)[31]. Heterotrophic flagellates are lysed by viruses and preyed by phagotrophic microplankton, that will also release DOM into the environment[21]. Once again, the heterotrophic microplankton is subsequently grazed by the mesozooplankton like giant protozoa,

ciliates, or copepods, which in their turn are often a direct food source for more complex organisms, like fish (Figure 2) [33,34].

In summary, the phytoplankton provides the primary source of carbon throughout photosynthesis. These photoautotrophic agents are grazed by the small zooplankton that, by its turn, are grazed by bigger heterotrophic organisms. This cycle repeats itself in different size magnitudes, ultimately allowing the feeding of the macroscopic organisms like fishes. In addition, this loop is also responsible for the recycling of nutrients, fundamental to sustain the marine food webs[31].

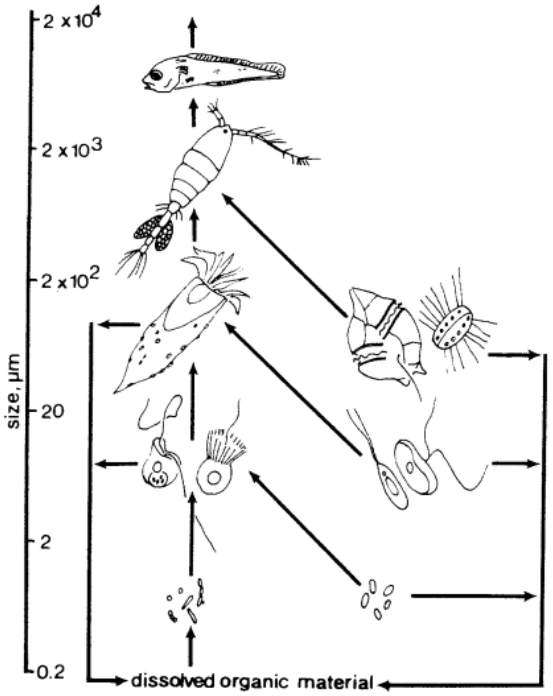


Figure 2: Schematic representation of the microbial loop where the trophic levels are related to the size of the representative beings. The represented organisms on the left are heterotrophs and the ones on the right are photoautotrophic. Image from the paper "Marine Plankton Food Chains"[29].

1.2.3 *Biogeochemical Cycling*

1.2.3.1 Carbon

The carbon cycle is a vital biochemical cycle that supports life on Earth[21]. Carbon has the ability to perform stable bounds with another four atoms creating a myriad of molecular forms. This characteristic gave rise to the organic molecular complexity found in each cell. Due to that, carbon is commonly referred to as the nuclear element necessary for life because it is found in every organic molecule, ranging from proteins, lipids, carbohydrates, and nucleic acids.

The marine carbon cycle is primarily characterized by the conversion of inorganic carbon to organic carbon, and vice-versa. For instances, in photosynthesis, carbon dioxide or bicarbonate (inorganic forms) are converted into cell biomass (organic forms)[21]. In seawater, the carbon cycle takes place

mainly in the euphotic zone, which has the best conditions of light and nutrients for photosynthesis to occur[21]. Carbon dioxide is the other key compound for photosynthesis, that becomes available to marine organisms due to its high solubility and reactiveness with water. From the interaction with water, CO_2 is transformed into carbonic acid (CO_3^{-2}), which is quickly converted into bicarbonate (HCO_3^-)[21]. Photosynthetic microorganisms will uptake the inorganic carbon (HCO_3^-) and convert it into organic matter in the form of their own biomass, which will serve as the basis of the marine food web[21]. For the organic carbon to become available to the higher trophic levels, grazing protists and viral lysis are essential. Most of the organic matter is decomposed by the chemotrophic community in the euphotic zone. However, the organic matter that is hard to degrade will sink in the water column to be utilized at deeper ocean levels [21]. Additionally, the exportation of particulate organic carbon from watersheds to the coastal zone is another important source of organic carbon that reaches the oceans[35].

As discussed above, organic matter will be available to the community through the release of the faecal pallets by the grazing protists, or in the form of particulate organic matter, released by viral lyses[36]. The organic matter that is not up-taken and recycled will aggregate to suspended material or to other organisms, and will sink into deeper ocean levels until, eventually, reach the seafloor [36]. The falling organic matter is called marine snow. During the journey, the organic matter will be degraded even further, where most of the nutrients are going to be recycled. At the bottom of the sea, the organic matter is extremely poor in nutrients and recalcitrant[36]. Here, the organic matter is buried and suffer diagenetic processes that take place at geological timescale[37]. Another contribution to the rates of carbon depositing at the seafloor is the active migration of zooplankton[38]. This transportation of fixed carbon from the euphotic zone to the deep sea is called the carbon pump. This process is currently counteracting the anthropogenic increase of carbon dioxide in the atmosphere, by burying it in the ocean sediments[37]. Because of that, the ocean floor is the largest carbon reservoir on the Earth, which has currently absorbed more than a quarter of the carbon dioxide anthropogenically released [38].

Climate changes are bringing new phenomena that could irreversibly unbalance the carbon cycle [38]. Several models had been put forward to understand how climate change will affect the carbon cycle. Most of them agreed that, in the future, the capacity of oceans to absorb atmospheric CO_2 will decrease[39,40]. The predictions foresee an induction of a positive feedback loop, where the increase in carbon dioxide will increasingly warm up the Earth. The carbon cycle will continually lose strength to counter react the increase of carbon dioxide in the atmosphere because it is expected that the CO_2 uptake rates will decrease as temperatures rises [39,40]. In the future, as the Earth warm-up, the efficiency of the carbon pump will continually decrease, exacerbating even further the CO_2 atmospheric accumulation and consequentially intensifying global warming[41].

1.2.3.2 Nitrogen

Nitrogen is an important nutrient for all life on Earth since it is one of the fundamental building blocks that are present in proteins, cell walls and nucleic acids [21]. Although, dinitrogen gas (N_2) makes 80% of the earth´s atmosphere, this form of nitrogen is not available for most of the organisms [42]. The nitrogen-

fixing bacteria have a crucial role in the transformation of N_2 into biologically available nitrogen. This is mainly explained because the triple bond, that connects the two nitrogen atoms in the dinitrogen gas makes this molecule almost inert, meaning that the cells need to invest a significant amount of energy for the nitrogen fixation to occur [21]. The fixation of atmospheric nitrogen (N_2) is commonly associated with different classes of Cyanobacteria [21,42].

For nitrogen fixation to occur, N_2 needs to be solubilized in the water in order to diffuse into the cell, where the nitrogenase enzyme complex catalyses the reaction. Initially, it occurs the fixation reaction where N_2 is fixed to ammonia (NH_3) (Figure 4)[21,42]. NH_3 can then be subjected to ammonification, which is the reaction that turns ammonia (NH_3) into ammonium (NH_4^+). Ammonium can now be directly immobilized and assimilated for the different proposes of the cell [21]. At this level, nitrogen can now become available to the ecosystem by processes of grazing and viral lysis. Organisms that cannot produce their own organic nitrogen will use the free ammonium (NH_4^+) as their primary source of free reduced nitrogen[21,42]. The free organic nitrogen (NH_4^+) that was not up-taken by these organisms will sink below the euphotic zone, turning into particulate nitrogen[42]. Here, it will suffer remineralization, turning into ammonium again and then processes of nitrification will take place (Figure 4)[42]. Nitrification reactions are mainly performed by chemolithotrophic microorganisms and are characterized by the oxidation of the reduced organic nitrogen (NH_4^+) into nitrites (NO_2^-) and subsequently into nitrates (NO_3^-) (Figure 4) [21]. Nitrates, and nitrites are also highly bioavailable nitrogen forms commonly selected by the marine plankton[42]. The first reaction (NH_4^+ to NO_2^-) is mainly catalysed by the archaea ammonia oxidizers like Crenarchaea or Thaumarchaeota[21]. The second reaction (NO_2^- to NO_3^-) is mainly performed by nitrite oxidizers bacteria like *Nitrospina*[42]. In both reactions, the oxidation of the reduced forms is involved in the ATP synthesis because they participate in the production of a proton motive force due to the electrons given[43]. At this stage, nitrates (NO_3^-) tend to accumulate in the realm between the euphotic zone and the oxygen minimum zone. Eventually, they will be transported back to the euphotic zone to be recycled again or will sink even deeper in the ocean. The fate of the nitrates will depend on the physical processes that are occurring in the area, like local stratification, temperature differences, upwelling or the hydrography in general[21].

Nitrates are the most oxidated forms of nitrogen in the cycle, meaning that from now on the reactions that will occur are mainly reductions. Microorganisms can now benefit from the reduction of nitrites (NO_2^-) and nitrates (NO_3^-) because their oxidation-reduction potential is very close to the oxygen, meaning that can be used as an efficient terminal electron acceptor[42]. All the consecutive reductive reactions that will occur are called the denitrification reactions, which will ultimately culminate in the loss of nitrogen to the atmosphere, either in the form of dinitrogen (N_2), nitrous oxide (N_2O) or nitric oxide (NO) (Figure 4)[44]. For the denitrification reactions to happen low oxygen concentrations need to be verified[42]. The first denitrification reaction is the reduction of nitrate (NO_3^-) to nitrite (NO_2^-), by the enzyme nitrate reductase [44]. The second reaction is catalysed by the nitric oxide reductase, where nitrites will be reduced to nitric oxide (NO), which can be released to the atmosphere. Thirdly, nitric oxide is reduced to nitrous oxide (N_2O), which can also escape to the atmosphere[44]. And finally, nitrous oxide can suffer a fourth modification catalysed by the nitrous oxide reductase. The product of this reaction is the

gaseous dinitrogen (N_2) that will close the nitrogen cycle (Figure 4)[44]. Nevertheless, ammonia or nitrites can also produce dinitrogen as a direct product by the anammox reactions[42].

As discussed above, ocean acidification is a clear consequence of the increase of the anthropogenically released CO_2 into the atmosphere. Such changes have a huge potential to affect the nitrogen cycle because, changes in the acid/base balance of the seawater may affect the sensitive oxidation-reduction reactions that the nitrogen needs to go through in the cycle [45]. This event raises the concern how key organisms will respond to pH changes in the environment [45].

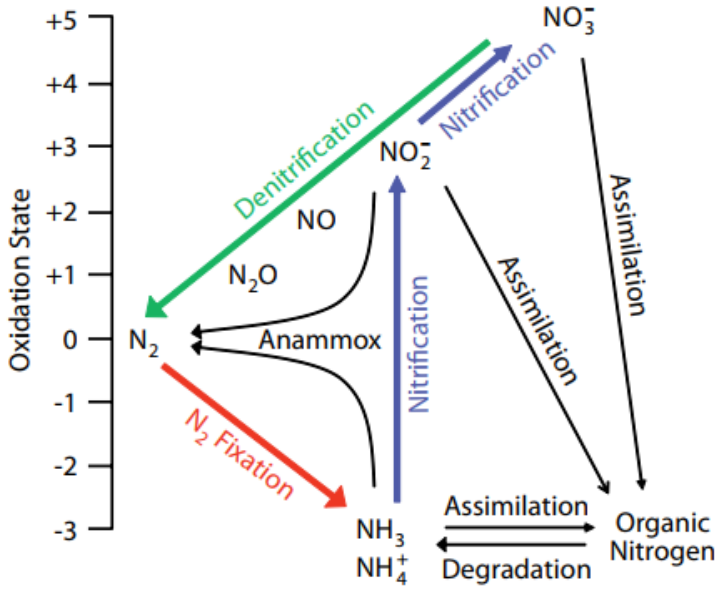


Figure 3: Most important chemical reactions happening in the nitrogen cycle. The different nitrogen forms are plotted according to their oxidative state, present in the x-axis. Image from "Nutrient cycles and Marine Microbes in CO_2 -enriched Ocean"[45].

1.2.3.3 Phosphorus

Phosphorus (P) is another important element that is present in many essential metabolic paths in the cell[46]. It is present in huge proportions in the cell membranes, mainly related to phosphoproteins and phospholipids[46]. P is also a crucial building block of the phosphate-ester backbone of DNA and RNA and is vital for the transmission of chemical energy through the ATP molecule[46]. Although phosphorus is an important and abundant element, is not easily accessible to marine organisms because this nutrient it is largely present in the earth´s crust[46]. There, P is mainly present in the rocks and soil (Apatite) as inorganic phosphate or organic phosphate derivatives[47]. These compounds will be dissolved by processes of weathering[47], and subsequently, riverine fluxes and groundwater will transport the dissolved and particulate phosphorus to the estuaries and coastal environments[46]. Another important source of P can be provided by the recycling of organic matter[42], the transportation of dust particles by the wind, from the continents to the ocean, and the volcanic activity[46]. The phosphorus that reached the

seawater can oscillate between two states, either is dissolved in the water column or is present as a particulate form[46]. The particulate state will eventually sink and deposit in the marine seafloor where the oxygen concentrations are low enough to prevent the degradation of the particulate forms[47]. When particulate phosphorus is exposed to oxygen for a long time it will be converted into dissolved phosphorus [47]. The dissolved inorganic phosphorus is up-taken by phytoplankton that turns it into the organic form. Phosphorus at this state can become available for the ecosystem by processes of grazing and viral lyses [46]. Phosphorus and nitrogen are often considered the limiting nutrients in marine ecosystems, meaning that their levels can influence the rates of the primary production in the carbon cycle[46].

As understood, the role of microorganisms in the P cycle is extremely passive[46]. However, they are the ones to provide the organic forms to the marine ecosystem. Nowadays, intensive agriculture, which uses indiscriminately phosphorus as a fertilizer, will also provide large quantities of phosphorus in the marine ecosystems[46]. The phosphorus poured in the fields will be washed away into rivers and groundwater, that eventually will reach the coastal areas and induce eutrophication of the environment [46].

1.2.3.4 Silica

Silica (Si) is another key element for the balance of the marine microbial communities because it is essential for the growth of several photosynthetic microorganisms such as diatoms, silicoflagellates, radiolarians and choanoflagellates[48]. This means that the silica cycle has the ability to regulate the abundance and composition of the phytoplankton, that ultimately will influence the carbon cycle [49]. Silica is mainly delivered to the marine microorganisms as dissolved silicic acid (DSi)[49], that is mainly chemically weathering of the Earth's crust. The silicic acid will flow into the rivers and groundwater and ultimately, will be delivered to the coastal areas. Some silicate can also be delivered from deep-sea hydrothermal systems and aeolian dust[48]. At the coastal area the diatoms are major players in the silica cycle since the Si present in their skeleton can be recycled in the euphotic zone. Some material that escapes dissolution will sink as particles, to be buried in the oceanic floor[48].

Recently, the anthropogenic disturbance is also withdrawing significant quantities of silica from the environment. The conversion of forests into farmland is altering the rock weathering rates[48], and the constructing of dams that retains sediments (among which particulate silica), are diminishing the silica stocks in the costal areas[48]. As consequence, the rates of diatoms are decreasing which will favour the non-silica organism to overgrown, resulting in the disruption of the microbial communities. In addition, when the low rates diatoms scenario is present, the large fish populations are often directly affected[49].

1.3 Monitoring and Evaluation

As noticed previously, the importance of marine microbial communities is notorious since they are key players in the marine ecosystems. Before industrial revolution, these microbial communities adapted to

the natural climate changes that happened in the marine environment. With the advent of the industrial revolution, the environmental conditions are now changing at higher rates than before, turning the global climate even more unpredictable and unstable. These new stresses and challenges are specially noticed by the marine microorganisms since they can rapidly adapt to the new conditions. As consequence, changes in the microbial communities´ composition will affect the higher trophic levels since they are the foundations of many marine processes. Their sensitive and rapid responses to perturbations, combined with their effects on higher tropic levels, make the microbial communities a suitable bioindicators of stress, allowing the access to the overall marine environment health[24]. In order to understand if the microbial communities are actually changing, and to understand how they are changing, the monitoring programs will serve to catalogue and evaluate the biodiversity of marine microorganisms. This will allow us to be aware of the whole environmental picture, over time and space, and understand the possible consequences of a certain factor.

To keep track of the alterations in the microbial composition, constant and significant data must be available[50]. With such meaningful information available, researchers, alongside with politicians, can establish a solid maritime spatial planning, to successfully meet ecosystem preservation goals [50]. After the project being implemented, monitoring must continue in order to understand if the goals proposed are being achieved or not [50]. Evaluation should allow the identification of gaps in the information that are necessary to adjust. The initial conditions of the study might also change throughout the research and therefore is essential to incorporate the new information and adjust the plan to the current circumstances[50]. Monitoring will also allow to make projections for the anticipation of outcomes in order to better adapt the current plan[50]. Within the marine monitoring, there are three main approaches. The state-of-the-system will focus on assessing just one parameter like the status of biodiversity, or the quality of water. State-of-the environment will assess several variables in numerous locations, either at the local or at the international scale [50].Performance monitoring is another form of monitoring system where direct measures of the area are taken to determine if a specific change was caused by a specific measured factor [50]. After choosing the approach, the indicators must be selected in order to understand how a specific condition is affecting the microbial communities[50]. In order to select the most significant indicators, several studies have been done, that can shed a light in why and how the microbial communities can change.

Marine microbial communities can change over space, time and over several environmental conditions. The following topics will describe in more detail how each condition, can shape the marine microbial communities.

1.3.1 Biogeography

1.3.1.1 Distribution over space

Biogeography is the field of study that aims to describe the distribution of biodiversity over time and space [51,52]. Understanding biogeography is important to delineate the key biodiversity areas for conservation [53]. Marine microbial biogeography started to be unveiled by the Baas-Becking hypothesis,

which proposed that everything is everywhere, but the environment selects. In other words, the microorganisms have such high rates of dispersal that, the only factor that selects them are the environmental conditions[51]. Indeed, dispersal processes have a strong role in the bacterial community assembly[54] however, microorganism's dispersal rates can vary from very short distances to global range, depending on the hydrography of the oceans and the presence of spores and cysts[51]. Regarding the propagule' survival rates, these can vary due to their hardiness or due to the conditions that they might encounter, until reach a suitable habitat. The propagule, to establish itself in the new environment, must outcompete the local population that is probably better adapted to that location[51]. Population density also plays a role in colonization because higher density species have a higher chance to colonize. Meaning that the percentage of the successful propagules is increased in the higher density species, when compared with low abundance taxa[51]. By its turn, phenomenon of ocean´s hydrography like the ocean currents, will enable increased rates of dispersion [55]. Skunking and upwelling will also affect the dispersal vertically by mixing the communities at different depths, and horizontal currents can also mix seawater at a single depth level [53]. With this knowledge, the Baas-Becking hypothesis can be refuted because not all organisms can disperse globally, where some taxa are restricted to a specific geographic location[51].

Over the years, the information gathered suggested that the dispersal rates are not the only factors affecting spatial patterns. Community´ compositions are also influenced by contemporary environmental conditions and historical events[51]. The majority of the investigation on the area found a significant correlation between the microbial community composition and at least one environmental condition showing that selection imposed by the environmental pressures (e.g., nutrient, and light availability) and biotic interactions (e.g., viral lysis and grazing), have an important role shaping the microbial community[52]. Pasted events also shaped the microbial assemblages by processes of genetic drift or adaptation to past environments[51]. These historical events maintained the genetic isolation from the other communities mainly due to barriers like distance. This isolation left a legacy in the community´ composition that can endure for a long time, overwhelming sometimes the influence that contemporary environmental factors have on the distribution patterns[51]. As understood, distance plays a critical role in the community patterns because, as the geographic distance increases between two locations, the taxonomic similarity of the different communities decreases, and vice-versa. This phenomenon is called the distance–decay relationship[52]. Studies that compared different communities separated in a range of tens of thousands of kilometres, hypothesized that distance is influencing more strongly the community composition, whereas the environmental factors have a small impact[51]. By contrast, when the samples are collected within a small spatial scale of a few kilometres, the comparison of the community composition is greatly affected by the environmental conditions whereas differences related to the distance were not found [51]. And finally, studies, that were realized with intermediate spatial ranges of 10 to 3000 kilometres, proposed that both historical events and contemporary environmental conditions had a similar influence on the community's composition[51]. Currently is calculated that the overall significance of environmental variables is about 26.9%, contrasting with the geographic distance that can only explain 10.3 % of the microbial taxonomy diversity. Where both variables combined can significantly influence 49.7% of the community composition[52].

Furthermore, biogeography patterns can also be affected by evolutionary processes like mutation, natural selection, gene flow and genetic drift, specially at the intraspecies level[52]. For instance, mutations will increase the local genetic diversity across all the communities and therefore exacerbate the differences and the variance of the communities[52]. By contrast, ecological processes shape more intensely the interspecies diversity, hugely influencing the biogeography assemblages. Processes like speciation will add newly formed species to the community, increasing diversity. Selection imposed by the environmental conditions will affect the abundance of species based on their ability to thrive in a specific habitat[52]. The colonization that is driven by dispersal methods, will eventually establish with success new species in the community. And lastly, ecological drift will randomly change the populations' size due to the rates of births and deaths influencing the communities' survival rates[52].

Understanding these processes and how they interact will allow a more holistic view of the hole biogeography patterns. For instance, selection processes will exacerbate the distance–decay relationship. Also, the environmental conditions are often displayed spatially as a gradient and therefore, selection will act along that spatial gradient[52]. Thus, it is understandable that two communities further apart will share a lower environmental conditions similarity, and therefore the taxonomic similarities will also be decreased [52]. By its turn, two communities that are closer together will share more environmental conditions and therefore the similarity of the communities will also increase[52]. The distance–decay relationship, and selection will be counteracted by the dispersal processes that will approach the taxonomic patterns between distant communities [52]. As consequence, the distance–decay relationship must be weaker where dispersal is high, like under the influence of oceanic currents, and stronger where dispersal is more limited such as along disconnected water bodies [52]. In a particular study, where it was inspected the heterotrophic prokaryotes composition of the different areas of the oceans[56], the researchers found that at polar and temperate regions, 41% of microbial communities were similar. Moreover, both temperate and tropical sites have an overlap of 23% and when comparing the three regions together they present an 11% microbial similarity[56]. By contrast, there is a 24% microbial community that is restricted to the temperate region, and 8% to the polar region. Exclusive microbial taxa were not found in the tropical region[56]. In the end, it was concluded that the processes that increase similarity between communities have a higher impact on the global patterns, while the variables that drives the differences between communities can also display patterns at a global scale, but at a lesser degree.

Recently, anthropogenic disturbance of the global climate and the introduction of non-native species are serious factors that are shaping the global biodiversity, imposing a serious threaten to the current biogeographic patterns. These stressors have the ability to decrease the native populations and promote the colonization of non-native species. This process known as biotic homogenization, will increase the similarity at a global scale disrupting the local and specific ecosystem functions[53]. Dispersion rates can also be impacted by the increasing temperatures because ocean currents might change their present cycles[53].

In conclusion, is possible to understand that the spatial patterns are shaped by the interplay of multiple factors. However, factors like dispersal, distance, and environmental gradient conditions, seem to be the ones that will influence more significantly the microbial composition across space.

1.3.1.2 Changing over time

The diversity of the marine microbial community can also be greatly influenced by the temporal factors, that are studied at different levels. The first level of study is at the hours' range because, since microorganisms have a high rate of growth and replication, the microbial community composition can change rapidly, in a matter of hours[57]. For that reason, the variability between two samples, taken with the interval of one day, was about 5 to 10% higher than between replicate samples[57]. These changes in microbial composition can be mainly explained by the primary productivity cycles because, since this cycle is governed by a day-night periodicity, is expected a higher daytime productivity[57]. These dial changes in productivity will, directly and indirectly, shape the microbial composition in order for the community to best adapt to the day and night conditions[57]. Nevertheless, several other factors can influence daily patterns. Unpredictable processes like the weather, the hydrography processes, nutrients availability and biological interactions can also have a significant influence on the assemble of these communities[57]. For instance, climate events can increase the upwelling or sinking of nutrients. By its turn, nutrients can influence the rate at which the daily changes occur since they set the pace for growth and replication. In the oceans, the copiotrophic organisms that have access to a high supply of nutrients will replicate faster, contrasting with the oligotrophic organisms that have access to few nutrients[57]. Surprisingly, when samples are taken a few days apart, the variation is only about 0.2%, meaning that the community hourly changes are reversed, returning to a well-defined stable state within the local limits. These dynamics will lead to the temporal predictability where is possible to foresee the most likely microbial composition in a specific area for a given month, being much harder to predict it for a given day [57].

Since monthly community predictability is much higher, a large number of studies focused on inspecting the temporal variability from monthly to seasonal periods. The community´s changes respond to predictable events like variations in solar angle, associated with light intensity and water´s penetration. Additionally seasonal patterns like weather events, temperature frequencies, seasonal upwelling (related to the availability of nutrients and DOM), and ocean stratification will also affect the community composition over time. Other less obvious variables that can affect the communities seasonally are the day length, the variations of the input from land and the biological interactions [57]. All these seasonal parameters will shape in some way the microbial composition over the months however, there are some communities that are more influenced than others. The communities in the surface water are more intensely influenced by seasonal factors than the ones living in the darker deep waters, because more intense seasonal physical/chemical processes are being delivered to the surface water, like the delivery of sunlight and nutrients [57]. It has been also proved that diversity of the ocean surface, across a long time series, will show a distinct cyclical pattern due to seasonality, with the higher species richness being recorded at Winter and the lowest at summer [55]. Additionally, as the distance from tropics

increase, it will also increase the seasonal variability [57]. The seasonal events are also responsible for a burst in the atmospheric oxygen composition driven by the marine microorganism, where registers suggest that it is in spring that the oxygen peaks, gradually declining in the rest of the year [58]. This can be explained because an increase in primary production and a decrease in the rates of respiration are observed during spring and summer. This scenario is followed by a decrease in primary production and an increase in the rates of respiration during Autumn and Winter caused by the uptake of oxygen from the atmosphere [58]. These dynamics are highly correlated with the seasonality of light and temperature throughout a latitudinal gradient [59]. A practical example that relates tightly to these dynamics are the spring blooms of phytoplankton. Since these microorganisms received less light in the Winter, a lower density was observed [60]. However, when the spring at higher latitudes start to take place, the phytoplankton also started to be exposed to increased sunlight and therefore the blooms emerged due to the increased productivity. At lower latitudes, the blooms are just limited by the nutrients mixing because the sunlight is almost constant[60].

At the end of the spectrum is the interannual variability, which is mostly shaped by large temporal scale oscillations like the El Niño Southern Oscillation and the Pacific Decadal Oscillation [57]. The natural climate variability and recently the anthropogenic induced climate change, either global or local, are also factors that change communities over time[57]. Events like the general habitat degradation caused by pollution, alongside with overfishing or the introduction of invasive species are speeding up the communities´ natural adaptation, that otherwise took years[57].

As overview, the main conclusions drawn by studies of different magnitudes suggest that communities are more similar within the same season, with few years apart, than that same season with many years apart [57]. Particularly, communities tend to be more similar with one year apart, but the similarity decreases when the analysis are made with samples of 2, 3 and 4 years apart. From 4 years forward, similarity remains the same, shifting to a decline in similarity on the analysis of communities 10 years apart [57]. By contrast, when it is compared opposite seasons with 6 months apart, the community composition will present a significantly dissimilar composition [57]. Additionally, studies also conclude that a significant temporal variability occurs within seasons, followed by the intersessional and interannual variability. These findings suggest that less predictable factors like day-to-day weather changes, the complexity of oceans hydrogeography and biological interactions can also drive such variability within seasons [57]. Additionally seasonal events that shape light, temperature, and nutrients are the main drivers of the intersessional variability.

1.3.1.3 Environmental Conditions-Temperature

There are a huge range of environmental conditions that can change the composition of the microbial communities and temperature is one of them. Temperature will shape the species diversity at a global scale because it has a huge influence on the protein stability, which by its turn will regulate the rates of the chemical reactions and the metabolic pathways that take place in an organism [61,62]. This parameter will strongly pressure the composition of the microbial community because ultimately will affect basic

biological processes like the growth and replication rates [61]. Additionally, the natural temperature tolerance of the microorganism can also affect the distribution of the communities since temperature will select the best adapted organisms to the temperature´s environment [61].

Is now clear that climate change is promoting the increasing of the ocean´s temperature. For instance, this event will disrupt stratification by increasing it and therefore, limiting even further the nutrient's availability in oligotrophic regions[63]. The increase in temperature will also improve the production and respiration rates until a certain level, enhancing the biomass fluxes and therefore reinforcing the role of the producers and heterotrophs in the oceans[63]. Due to global temperature increasing, the communities will be rearranged spatially, in the latitudinal and/or altitudinal scales, in order to meet their optimal growth rate[61,64]. Microorganisms can also have the plasticity to evolve in the response to such pressures, possibility giving rise to new species [61].

1.3.1.4 Environmental Conditions-Light

Light is another environmental condition that is key for primary production because it energizes the photosystem II that by its turn will initiate the whole photosynthetic process[65]. Because of that, organisms that really on light to survive, like phytoplankton, exhibit a strong preference towards it. For instance, when Cyanobacteria are subjected to a dark environment for long periods, their survival rates decline immensely because grazing and viral lysis will naturally remove them[65]. Additionally, in the same study, the other members of the community, the non-cyanobacterial, showed mixed responses to light treatment, indicating that the majority of these species display minor responses to light presence and intensity, while a few members can gain an advantage from sunlight exposure[65]. However, there is a certain level of uncertainty in assessing if the effects are direct from light manipulation or indirect effects on the food web dynamics[65]. Following the same principles, the oligotrophic communities also exhibited a strong preference to light because they depend heavily on this source[65]. Light will have a strong influence in the communities living in the photic zone, while the populations living below that level will be less affected. It is in the photic zone that the majority of the primary production occurs and therefore, the variability of respiration is higher [66]. Nevertheless, it is also expected to encounter phytoplankton in the aphotic zone, since, at this level, they assume a heterotrophic metabolism because there is no light to promote photosynthesis[67]. These observations lead us to conclude that the carbon flux in the light-exposed area is greater than in the dark environment [66]. To summarize, light will have a significant impact on the marine community assemblages mainly influencing their taxonomic composition in the vertical gradient due to the range of light penetration. Light will promote phytoplankton to growth in these areas because it will enable photosynthesis to occur.

1.3.1.5 Environmental Conditions-Salinity

Sodium chloride is the major solute in the oceans and is understandable that microorganisms evolved to cope with this stress [68]. Across the millennia, several strategies have been developed to face different

osmotic stresses allowing the microorganisms to adapt to a certain environment with a given salinity concentration[68]. This means that the different salinities concentrations will influence the microbial composition because the microorganisms have different salinity´ sensitivity. According to the salinity concentration there are three different water bodies that can be distinguished. The freshwater, marine water and brackish water will accommodate different communities, each of them with their distinct taxonomic patterns due to the salinity conditions[69]. However, between water bodies, a concentration gradient of salt is present. For instance, this gradient can range from the saltier water in the ocean, until the freshwater inland[52]. Salinity will strongly influence the microbial biogeography distribution across different water bodies[52] leading to a clear taxonomic transition that is occurring according to this gradient [70]. These communities can be mainly distinguished by their changes in core metabolic pathways, which differ depending on the salinity concentrations that they are exposed to. Consequently, deep phylogenetic divisions between freshwater and marine communities are observed because different genes are required to deal with the different environmental conditions[71].

The oceans and freshwater salinity gradient´s are smoother and stable[72].However, estuaries represent a special case since they have a high changing rate on their salinity concentrations due to the constant freshwater flow and the tidal regime. The communities found here experience intermediate salinities resulting in a mixture of communities from the marine and from the freshwater environments [70]. Nevertheless, salinity is still one of the biggest barriers to the dispersal of microorganism from freshwater to the oceans, and vice-versa[70]. Salinity can also influence water densities that by its turn is responsible for the physical separation of water masses. The separation of these water masses will also separate the different microbial communities leading to an increased community heterogenicity[59]. Hyposalinity will also cause variability in communities' composition because the overall fitness is decreased, driven by the sub-optimal conditions[62]. Additionally, the community´s homogenization can also occur because phytoplankton can tolerate high levels of salinity when there are large quantities of nutrients available[69]. Moreover, salinity plays an important role in the sinking rates of nutrients and organic matter, affecting the communities' assemblage[72]. Once again, the increase in the global temperature driven by humanity will also affect the salinity gradients since the melting of the ice sheets and the increase in the evaporation rates, might result in the destabilization of the salinity gradients, affecting the communities´ composition[72]. To conclude, changes in the salinity levels will influence the community taxonomy because the salinity levels will select the species that can tolerance that concentration.

1.3.1.6 Environmental Conditions-Nutrients

The essential nutrient cycles in the ocean are often closely coupled with the marine microorganisms[73]. Carbon, nitrogen, and phosphorus are often considered very important for their metabolic pathways, but other nutrients like iron can also have an important impact on the communities. In 1934, Alfred Redfield found that the average concentrations of carbon, nitrogen, and phosphorus across the world´s oceans are nearly constant following the ratio 106 C:16 N:1 P (carbon: nitrogen: phosphorus) [73]. More recently the ratio was updated to include nutrients like iron, presenting the following values, 106 C: 16 N: 1 P: 0.0075 Fe (carbon: nitrogen: phosphorus: iron)[74]. The ratio underlines the magnitude of the nutrient

cycles that are taking place in the marine environment and therefore, any alteration in the ratio might indicate that the communities are suffering some sort of shifting or there is some event promoting this change. If these alterations are verified in the long term, it can have a significant impact on the microbial assemblages, leading to variations in the carbon pump and in the rates of bioavailable nutrients[73].

Phytoplankton on average, take up nutrients at Redfield ratio[74]. This affirmation must be taken with care because some phytoplankton species have their unique Redfield ratios. These species changed the typical resource allocation in order to increase the efficient to face a specific set of environmental conditions [75]. At the cellular level, the relative proportions of the different cellular compounds such as proteins and chlorophyll will change, resulting in the fluctuation of the Redfield ratio [75]. This led to the conclusion that the Redfield ratio is not universally optimal for the phytoplankton growth, but instead is an average that represents the phytoplankton diversity, growing under a variety of environmental conditions [75]. In the oligotrophic oceans, the absence of some nutrients like inorganic nitrogen, phosphorus, iron, and silica, will inevitably affect the community assemblages, limiting the growth of phytoplankton and consequently primary production [75]. In the oceans, there are some areas known as the high-nutrient low chlorophyll regions because, despite of the relative abundance of nitrate and phosphate it is observed a lower concentration of chlorophyll [74]. This pattern can be mainly explained by the lack of iron and silica in the environment. Outside of these areas, productivity is restricted by inorganic nitrogen[74]. As conclusion, when the nutrients are abundantly available, phytoplankton will grow rapidly [75]. This can be true for most of the organisms in the oceans because nutrient concentrations will strongly affect the microbial community's composition because they play a crucial role in their growth and development[54]. However, when humans discharge large quantities of nutrients to the environment, several processes might occur that lead to the disruption of the microbial composition.

1.3.1.7 Environmental Conditions-pH

Marine waters can often display a heterogeneous distribution of parameters, fluctuating spatially and temporally. However, regarding the pH values, they are overall stable around 8 to 8.5. The pH will affect the charge of the biological molecules that is essential to their proper structure and function[76]. Consequently, changes in these values can have an impact on the composition of the microbial communities. Species that are not able to resist alterations will eventually disappear, persisting and thriving the ones possessing better mechanisms to tolerate the pH changes[77]. In a case study where was raised pH artificially to 9.5, the phytoplankton increased their biomass mainly due to the decrease in the abundance of grazing agents that could not resist the pH alterations[77]. Recently, it was found that ocean acidification is occurring due to the rise in anthropogenic CO_2 emissions that are altering the stable pH of the oceans. Once again, microorganisms with mechanisms to tolerate a decrease in water pH will survive, where others will disappear since they cannot cope with the changes [76]. Ocean acidification will mainly affect essential processes like primary productivity and nitrogen fixation that rely on stable and properly charged proteins [78]. In conclusion, changes in pH will inevitably decrease the species richness of a specific community for both phototrophic and heterotrophic organisms[77], however,

predictions in the future community assemblages are difficult because is not yet understood the pH tolerance of each species[78].

1.3.1.8 Environmental Conditions-Depth

As discussed above, there are several environmental parameters that are greater at surface than at the bottom, like light or nutrients for instance. In contrast, the realm below 200 meters is a very hostile environment due to the lack of light, the presence of low temperatures, the short availability of nutrients and the increasing pressure according to the depth[79]. The hydrostatic pressure exerted in the microorganisms will modify their molecular assemblages and therefore a wide range of physiology traits will be influenced[79]. These modifications are selected from an embryonic stage to face this challenge[80]. Microorganisms that are more resilient to the pressure can propagate throughout the water column, whereas the ones less tolerant are confined to specific depths[80]. In short, the water column has a gradient of environmental conditions that change according with depth. This gradient will pressure the microbial communities to distribute across the water column in an optimal way.

1.3.1.9 Environmental conditions interaction

Each parameter described above is not affecting communities' composition by itself. In most cases, several parameters are influencing the taxonomy composition. In other words, the interaction of multiple parameters will be ultimately, the main factor that drives communities' assemblages. This interaction is often very complex since a lot of variables are at play. Because of that, when planning a study, it should be collected every possible parameter to be posteriorly analyzed. This practice will allow a broader picture over the environment, in order to better understand how the interactions of conditions lead to the current taxonomy composition. Nevertheless, in most cases, one environmental condition can significantly explain the majority of the variability in a community. For instance, salinity, temperature, and depth seem to be the most important parameters to influence the communities over a large spatial scale[59]. Temporally, the communities seem to respond mostly to seasonal parameters like temperature and nutrient concentrations. Salinity, temperature, light, and nutrients are the main parameters that influence communities' assemblages at depth [59].

1.4 Microbiome International Monitoring Programs

There are several monitoring campaigns around the world that are trying to characterize marine microbial communities. On this topic, it will be presented different marine monitoring campaigns that successfully meet their goals. All these monitoring programs try to evaluate the status of a certain marine ecological environment by understanding the microbial community's composition and distribution. To do that, all the recent monitoring programs rely on DNA sequencing because it offers an integrated ecosystem assessment. The high-throughput sequencing allows the identification of individuals and

populations in a more fast and more accurate way that, together with bioinformatics analysis, revolutionized the study of marine biology[81]. Additionally, this approach allowed to better understand the stress responses and the adaptation capacity of a specific community [81]. Currently, there are several projects that generate sequence-based biodiversity assessments[81]. The Ocean Sampling Day project[82], the Tara Oceans project[83] and the European Marine Omics Biodiversity Observation Network[84] are examples of monitoring programs focused on marine microbial assemblages.

The Ocean Sampling Day project is a collaborative global sequencing campaign that happens once a year since 2014. On the same day, researchers around the world collected a marine water sample in their region [82]. The dataset generated will be mainly used to understand the anthropogenic impact on the microbial community's biodiversity which could be used as an indicator for the Ocean Health Index [82]. In the Ocean Sampling Day project, five litters of coastal water are filtrated by a 0.22 um pore size Sterivex cartridge using a vacuum pump. The filters with the biomass are stored at -80 Celsius degrees and shipped to a central facility in order to retrieve the 16S/18S rRNA gene data sets[85].

The European Marine Omics Biodiversity Observation Network (EMO BON) a similar initiative, launched in the summer of 2021, that aims to generate a genomic diversity dataset over several years[84]. The access to this information will enable large research projects in fields like climate change, bioprospecting, and ocean science [84]. The methodology used in the EMO BON required a pre-filtration of the water with a 200 µm filter. The filtration of the sample is then carried by peristaltic pump that makes the water pass through a 3 µm polycarbonate membrane filter in the first step and a 0.2 µm polycarbonate membrane filter with in the second step. These two steps will allow to analyse two different fractions of the microbial communities, the organisms between 200 to 3 µm and the ones between 3 and 0.2 µm[84]. After filtration, the filters are flash frozen in nitrogen and kept at -80 Celsius degrees. The samples are shipped to a centralized facility in order to sequencing the metagenomes and metabarcoding of the microbial communities [84].

The Tara oceans was a multidisciplinary project that aimed to understand the complexity of the ocean life at a global scale. For the microbial communities, they aimed to better understand their diversity, functions, interactions, and phenotypic complexity over a planetary scale[83]. The sampling campaign started in 2009 and finished in 2013, where it was sampled most of the biogeographic and biogeochemical regions of the world [83]. The methodologies in the campaigns described above were analysed, and the most fitted methods were used in the present study. In the Tara oceans campaigns, the samples were collected both from the epipelagic waters, at the surface, and from the mesopelagic waters, down to 1000 meters[83]. The samples were pre-filtrated using a mesh of 200 µm pore size followed by a 20 µm filter, and additionally a 5 µm for protists[86]. Finally, the sample was filtered using a 3 µm pore size filter and cryopreserved to be later sequenced in a centralize facility by an Illumina device[86]. The classification was determined by the metagenomic reads of the 16S rRNA (miTAGs) and additionally, by obtaining the amplicon sequencing of the region V4-V5 region of the 16S rRNA gene (primers 515F-Y and 926R)[87]. Additionally, it was also analysed the metatranscriptomic and single-cell genomes.

In this field of study, the research carried in Portugal focused more on the microbiological water-quality of the coast [88,89] and river[90,91,], in order to ensure quality of the bathing areas[92] and protection of shellfish nursery zones[93]. Additionally, the studies also prioritized the description of the distribution of the larvae fish, like sardines[94], in the estuaries [95] and in the coast north of Portugal[96,97]. Additionally, past investigation also approached the distribution of the macrobenthic community associated with contamination of the Douro estuary[98,99]. Regarding the study of the marine microorganisms in Portugal, the research done is not vast. There are some previous studies that described the plankton interactions with the upwelling regimes[100,101]. In another kind of study, it was assessed the potential of the microbial communities for the hydrocarbon's degradation[102]. However, there are also a study which main goal was to access the protistan plankton diversity by deep sequencing, in the Portuguese Ria Formosa[103]. In Portugal, the monitoring programs that tried to describe the distribution and composition of the marine microbial communities are also not abundant. One rare example is the monitoring program that happened in the Tagus estuary, where the phytoplankton and the nutrients were recorded since 1999[104].

1.5 Dissertation goals

Marine microbial communities are key elements for the health of the planet; however, these communities are facing changes promoted by anthropogenic pressures. Due to that, monitoring programs are becoming essential to track the microbial communities´ composition across space and time. Around the world, such programs have already started, however in Portugal there are few studies that keep track of such important parameters. In order to fill this gap, the present thesis aims to evaluate and catalogue the biodiversity of marine microorganisms on the north coast of Portugal. The research took part in sampling of the coast NW of Portugal (integrated in the Atlântida campaign), plus the sampling of the Douro estuary. The samples from the Atlantic Ocean, plus the ones from the river will allow a much deeper and integrated understanding of the communities' dynamics in the northern region of Portugal. The monitorization of the NW coast will be prolonged in time, for at least 3 years, in the framework of the OCEAN3R project. The present study opened the initial phase of that project.

The present thesis aims to meet three main objectives. The first objective proposed is to determine the microbiome distribution across an estuary gradient. In this topic the Douro estuarine samples are going to be analysed spatially, across a transect and at depth, and temporally, between the Autumn and Winter season. The second objective proposed was to evaluate the impact of the tides in the estuarine microplankton communities in the Douro River. The samples taken in 2016, will be analysed and a posterior assessment will be done to understand the tide effect on communities' assemblages. And finally, the third goal is to evaluate the spatial biogeography of the microplankton communities in the northern coast of Portugal.

2. Materials and methods

2.1 Sampling Region characterization

The north region of Portugal is bathed by the waters of the Atlantic Ocean. Through this region, the coastal water circulation is mainly induced by the wind regime that is taking place[105], influenced by the Azores high-pressure system, which displays a robust seasonal variation [105]. Roughly, between April and September, Portugal´s current system is happening. This current flows southward, near the Portuguese shelf break[106], driven by the northern blowing winds that favours the upwelling of the cold Eastern North Atlantic Central Water of subpolar origin[106]. The upwelling of such cold waters will also bring a high abundance of nutrients that arise into the surface inshore[107]. By its turn, the Portugal coastal counter current (PCCC) happens between October and the end of March and is characterized by the downwelling-favourable season[106]. The PCCC is characterized by the transportation of the warm Eastern North Atlantic Central Water, of subtropical origin, southernly, to the Armorican Shelf off SW France. Nevertheless, upwelling events can occur frequently in the downwelling season and, vice-versa. These currents can penetrate and greatly influence the coastal areas however, their impact will depend on the intensity of the shelf winds and the continental runoff[106].

Specifically, the northern Portuguese shelf is flat with an average width of 45 km and 300 meters deep, in the upper shelf[108,109]. The most relevant topographic feature on the northern shelf is the Porto canyon which does not influence much the inner shelf region. Much more influence has the wind regime that creates a north-west swell in 80% of the year[105]. The northwest coast of Portugal holds a coastline that stretches for 100 kilometres, confined between Espinho, at the south, and Caminha, at the north. Additionally, the inshore zone is strongly influenced by the riverine discharges, especially from the Douro River [106]. For example, higher volumes of fresh water are released in the coastal areas in the Winter, lowering the salinity gradient [106]. The Lima and Minho Rivers contribute also to the overall discharges but on a smaller scale [106]. The Douro River is the third most important river in the Iberian Peninsula due to its 927 km length and because its hydrographic basin area is the largest in the Iberian Peninsula. The Douro River runs from East to West and drains off into the Atlantic Ocean, between Porto city (north bank) and Vila Nova de Gaia (south bank)[110] . This region is heavily populated, being the second-largest metropolitan area in Portugal[90]. The estuary is limited upstream by the Crestuma dam, located 21.6 km from the mouth of the estuary[111,112]. The Douro estuary is narrow and funnelled with a tidal range of 2 to 4 meters[113]. Freshwater flow is at his pick in the Winter, decreasing the volumes gradually when heading to summer [90]. The water quality in the estuary can change according to the wastewater treatment discharge regimes and due to agricultural and cattle-raising runoff. Discharges of eight wastewater treatment plants are poured into the Douro estuary [90,114]. Such variabilities in the flow regimes can have a physical and chemical impact on the estuarine water with direct consequences for the microbial communities[114].

2.2 Sampling stations and water collection

The present study joined the ATLANTIDA campaign (from ATLANTIDA project) for the sampling stage, where surface marine waters were collected along the NW coast of Portugal in predefined stations

displayed in the Table 1 and mapped in Figure 4. The campaign was carried out during the summer season, between 25th of June 2021, and 9th of September 2021.

The Douro River estuary was also object of study, where water samples were taken across an estuarine transect. In Figure 5 it was possible to visualize the locations of Douro estuary sampling sites that took place on the 12th of December 2016, in the Winter. During this campaign only surface water samples were taken at both high and low tide (Table 1). In Figure 6, the stations visualized were relatively to a posterior Douro campaign that took place on the 18th of October 2021, in the Autumn, and on the 2nd of March 2022, in the Winter. In these more recent campaigns samples were taken at surface (Between 0 and 1 meter of depth) and bottom, using a Niskin bottle (Table 1).

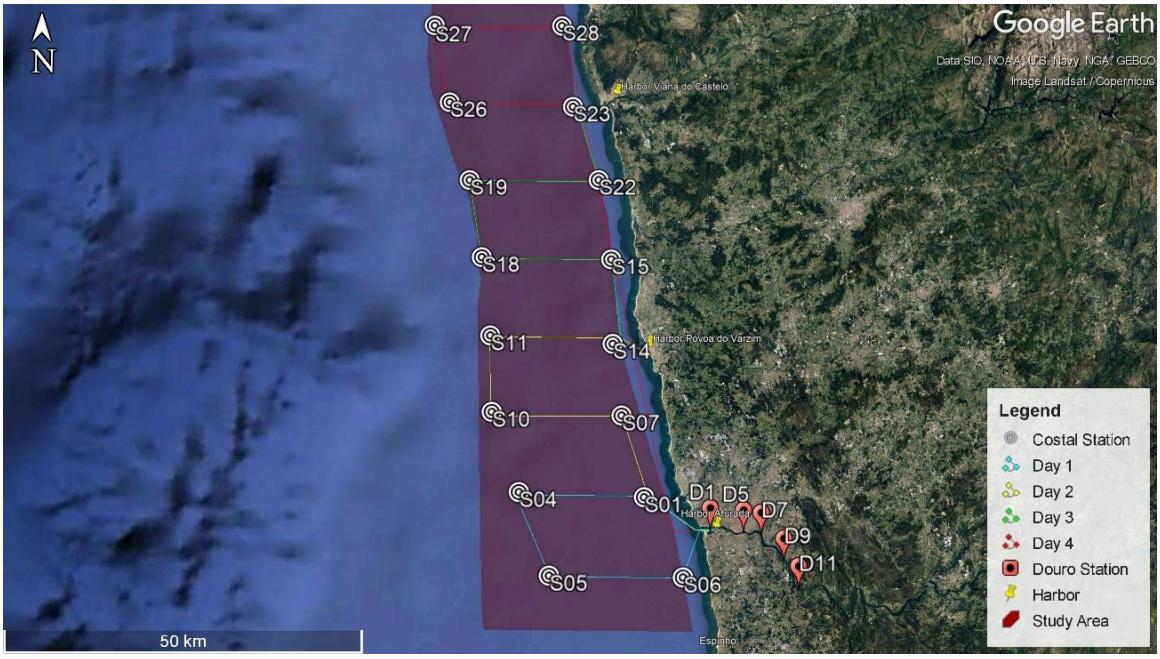


Figure 4: Topographic map of the north region of Portugal, where was possible to visualize the stations where the seawater samples were collected at the NW coast of Portugal. Additionally, and for contextualization, the stations of the 2021 and 2022 Douro campaigns were also represented.

Figure 5: The topographic map of Douro estuary, showing the stations where the estuarine water was collected during the 2016 campaign.

The samples from the three campaigns were collected using a decontaminated bucket and and stored in the respective sample containers. All the material used for sampling was previously decontaminated with 10% hydrochloric acid (HCL), remained overnight in a horizontal position, following by multiple washes with MiliQ. The bottles were left to dry under the UV lights in the laminar flow cabinet. Once the bottles were dried, they were closed and only opened in the field where containers were also washed with the respective station´s water, before sample collection. The cleaning protocol was based on the EMO BON initiative[84]. For the coastal samples, the water collected was about two and half litters, while in the Douro estuary the samples had a volume of one litter. These differences in the volume collected were due to the biomass abundance in the two environments. Additionally, in each station, the multiparametric probe was submerged in order to save the CTD profile for posterior analysis. The probe measured a wide range of parameters like the depth of the station (meters), the temperature (°C), the salinity (PSU), and the pH (Annex 7-9).

Figure 6: The topographic map of Douro estuary, showing the stations where the estuarine water was collected for the 2021/2022 Douro campaigns.

Table 1: General information for each station where water samples have been collected during the coastal and Douro estuary campaigns.

Label	Station	Tide	Depth	Season	Coordinates	Sampling Day	Sampling Time
Estuary							
R1L	R1	Low	Surface	Winter	41.138903 -8.656658	22/12/2016	17:14
R2L	R2	Low	Surface	Winter	41.14005 -8.651247	22/12/2016	16:41
R3L	R3	Low	Surface	Winter	41.120503 -8.549803	22/12/2016	16:16
R4L	R4	Low	Surface	Winter	41.083728 -8.522553	22/12/2016	15:48
R5L	R5	Low	Surface	Winter	41.072975 -8.492239	22/12/2016	15:28
R1H	R1	High	Surface	Winter	41.138903 -8.656658	22/12/2016	10:00
R2H	R2	High	Surface	Winter	41.14005 -8.584581	22/12/2016	10:30
R3H	R3	High	Surface	Winter	41.120503 -8.549803	22/12/2016	11:07
R4H	R4	High	Surface	Winter	41.083728 -8.522553	22/12/2016	11:33
R5H	R5	High	Surface	Winter	41.072975 -8.492239	22/12/2016	11:53
D1S_A	D1	High	Surface	Autumn	41.14472 -8.6692	18/10/2021	13:12
D1B_A	D1	High	Bottom	Autumn	41.14472 -8.6692	18/10/2021	13:12
D5S_A	D5	High	Surface	Autumn	41.14071 -8.60725	18/10/2021	14:42
D5B_A	D5	High	Bottom	Autumn	41.14071 -8.60725	18/10/2021	14:42
D7S_A	D7	High	Surface	Autumn	41.12912 -8.57188	18/10/2021	15:20
D7B_A	D7	High	Bottom	Autumn	41.12912 -8.57188	18/10/2021	15:20
D9S_A	D9	High	Surface	Autumn	41.10799 -8.53365	18/10/2021	16:05
D9B_A	D9	High	Bottom	Autumn	41.10799 -8.53365	18/10/2021	16:05
D11S_A	D11	High	Surface	Autumn	41.07034 -8.5034	18/10/2021	16:46
D11B_A	D11	High	Bottom	Autumn	41.07034 -8.5034	18/10/2021	16:46
D1S_W	D1	High	Surface	Winter	41.14472 -8.6692	02/03/2022	13:40

D1B_W	D1	High	Bottom	Winter	41,14472 -8,6692	02/03/2022	13:40
D5S_W	D5	High	Surface	Winter	41,14071 -8,60725	02/03/2022	14:44
D5B_W	D5	High	Bottom	Winter	41,14071 -8,60725	02/03/2022	14:44
D7S_W	D7	High	Surface	Winter	41,12912 -8,57188	02/03/2022	15:32
D7B_W	D7	High	Bottom	Winter	41,12912 -8,57188	02/03/2022	15:32
D9S_W	D9	High	Surface	Winter	41,10799 -8,53365	02/03/2022	16:08
D9B_W	D9	High	Bottom	Winter	41,10799 -8,53365	02/03/2022	16:08
D11S_W	D11	High	Surface	Winter	41,07034 -8,5034	02/03/2022	16:50
D11B_W	D11	High	Bottom	Winter	41,07034 -8,5034	02/03/2022	16:50
Coastal							
S01	S01	Low	Surface	Summer	41.19498 -8.75939	25/06/2021	07:20
S04	S04	Low	Surface	Summer	41.19509 -8.97496	25/06/2021	09:42
S05	S05	Low	Surface	Summer	41.08724 -8.92186	25/06/2021	11:45
S06	S06	High	Surface	Summer	41.08695 -8.70313	25/06/2021	14:19
S07	S07	Low	Surface	Summer	41.28715 -8.80275	26/06/2021	07:34
S10	S10	Low	Surface	Summer	41.28784 -9.02265	26/06/2021	09:24
S11	S11	Low	Surface	Summer	41.38393 -9.02605	26/06/2021	10:40
S14	S14	Low	Surface	Summer	41.38457 -8.80731	26/06/2021	12:52
S15	S15	Low	Surface	Summer	41.485914 -8.822288	27/06/2021	07:05
S18	S18	Low	Surface	Summer	41.487032 -9.039212	27/06/2021	09:03
S19	S19	High	Surface	Summer	41.583603 -9.069748	24/09/2021	15:45
S22	S22	High	Surface	Summer	41.583534 -8.844626	24/09/2021	17:11
S23	S23	Low	Surface	Summer	41.684476 -8.888262	25/09/2021	08:09
S26	S26	Low	Surface	Summer	41.684891 -9.113227	25/09/2021	10:26
S27	S27	Low	Surface	Summer	41.782868 -9.133302	25/09/2021	11:36
S28	S28	Low	Surface	Summer	41.783265 -8.908013	25/09/2021	13:51

2.3 Filtration

After sample collection and storage water was filtrated to discard the liquid phase and recover the biomass present in the water sample. It is recommended that filtration take place as soon as possible [84.] In the Atlântida campaign, the conditions enabled to perform the filtration on board. Since lower biomass was expected in these samples, two and a half litters were filtered for each replica. The goal was to achieve two replicas, filtrating a total of five litters per station. In the Douro campaign, the filtrations took place back at the Ciimar´s laboratory where it was filtered two replicas per station, each with one litter since a higher concentration of biomass was expected. The protocol followed was designed to retrieve the prokaryotes´ and unicellular eukaryotes´ biomass. For that to happen, the water sample was passed through a 0.22 µm pore size Sterivex filter (Merck Millipore[82]). The system relied on a vacuum pump that was connected to a Manifold where the Sterivex cartridges were fixed. Through vacuum power, the water was sucked from the respective container into the Sterivex filter, to be filtered (Figure 7). Before this procedure, all the tubes were washed with the station´s water. Once the filtration was done, preservation of the samples was ensured by flash frozen the Sterivex cartridges in liquid nitrogen, and later stored at -80 Celsius degrees in a sterile labelled plastic bag [115].

Figure 7: The filtration system used for filtration, where the vacuum pump pulls the water contained in the bottles, through the Sterivex filters, concentrating the biomass present in the environment.

2.3 e-DNA Extraction

At this step, the particulate environmental DNA (e-DNA) retrieved in the Sterivex cartridges was isolated. To do that, the DNeasy® PowerWater Sterivex Kit (QIAGEN)[115] was selected and the manufacturer instructions followed. Briefly, the protocol started by detaching the cells from the membrane´s filter, into the solution that was previously poured. Next, the cells need to be broken to remove the DNA that was inside, for that a strong lysis buffer was used, along with thermal and mechanical procedures. Further steps will enable the discharge of non-DNA and other inorganic materials that were still present in the solution. Finally, the elution buffer will allow the collection of the purified DNA that was ready for downstream applications. The DNA must be saved in a frozen environment, with temperatures ranging from -15 to -80 °C.

2.4 DNA Quantification

The quantification step will allow to understand the quantity of DNA extracted from the previous step and realize if it was successful or not. The kit used to do the quantification was Invitrogen Qubit™ 4 Fluorometer (Thermo Fisher Scientific Inc.[116]). Briefly, the protocol started by the need to calibrate the Fluorometer with the standard solutions. Next, the samples were prepared and read in the fluorometer, following the manufacturer protocol [116]. This procedure was based on fluorimetry. In this case the Qubit Fluorometers will attach specifically to the DNA in the solution. After that, the dyes will only emanate fluorescence light if bound to their targets. The concentration of the fluorescence light will be posteriorly read by the Fluorometer. This procedure has some advantages over spectrophotometry, since it was more sensitive to the molecule of interest, even at low concentrations.

2.5 Polymerase chain reaction

The Polymerase chain reaction (PCR) is a molecular technique that will allow the amplification of a specific DNA from a pool of DNA. In the case of the present study, it was important to amplify specific regions of the 16S rRNA gene and 18S rRNA gene with a pool of environmental genomes from the water samples collected. The PCR protocol followed, started by the elaboration of the working solution, where 5 µL of the DreamTaq (Thermo Fisher Scientific Inc.[117]) were used for each sample, plus the negative control. DreamTaq was a ready-to-use solution containing DNA polymerase, an optimized buffer, MgCl2 and dNTPs. The working solution was completed with the addition of 3.5 µL of DNA free water, 0.5 µL of the forward and reverse primers, for each reaction. The primers used for the 16S rRNA gene amplification of the region V4-V5 were the pair "515F" and "926R" [87], and for the 18S rRNA gene amplification the "TAReuk454FWD1" and "TAReukREV3_modified" pair [118] were used to amplify the region V4. Next, 9.5 µl of the working solution was distributed through the necessary PCR tubes. At this point, 0.5 µl of the DNA of each sample was also added to the previously labelled PCR tubes. After brief centrifugation, the tubes were placed in the thermal cycler for PCR amplification. The amplification protocol started with 3 minutes of denaturation at 95˚C, followed by 35 cycles that start at 95˚C for 40 seconds, 52˚C for 40 seconds and finish at 72˚C for 10 seconds, according to Sousa et al. 2019. After the cycles were concluded, an extension stage for 5 minutes at 72˚C was run.

The PCR products were confirmed by gel electrophoresis in 1.5% agarose gels and visualized with SYBR Safe staining, under the excitation of UV light. Before running the PCR products in the gel samples and the negative control was mixed with 1.5µL of loading dye. Loading dye was a solution with high density that helps the loading of the samples into wells and additionally enables tracking of DNA migration across the gel. After that, the samples, the negative control, and the ladder (100 pb, Thermo Fisher Scientific Inc.) were ready to be loaded in the gel. The ladder was a solution with standard-sized molecular fragments that will help to determine the size of the PCR product after the electrophoresis. Electrophoresis was run for half an hour, with the parameters defined at 150 volts and 400 amperes. When the electrophoresis was finished, the gel was placed in the imaging system (Bio Rad Molecular Imager® Gel Doc™ XR System) and an image of the agarose gel was saved for posterior analysis.

2.6 Sequencing

In the sequencing step, the samples were sent to the Integrated Microbiome Resource (IMR) group, however the samples from 2016 were sent to the LGC group. In any case, the library preparation was accomplished by PCR amplification of specific regions of the ribosomal RNA genes. The amplification of the 16S ribosomal RNA gene was performed by using the primer pair 515F (5′-GTGYCAGCMGCCGCGGTAA-3′)[119] and, the recently revised 926R (5'-CCGYCAATTYMTTTRAGTTT-3'), specially created to target the V4-V5 hypervariable regions [120]. The Region V4 of the 18S ribosomal RNA gene was amplified using the forward primer TAReuk454FWD1 (5′-CCAGCASCYGCGGTAATTCC-3′) and TAReukREV3_modified (5′-ACTTTCGTTCTTGATYRATGA-3′)[118] .The primers were specially chosen to improve the coverage of desired organisms and to yield a

high-quality amplicon in order to optimize the phylogenetic resolution[121]. After the fragments were amplified, they were sequenced on Illumina MiSeq device due to its higher depth and accuracy in sequencing[115]. Finally, the amplicons libraries were processed as raw FASTQ format data for upstream and downstream analysis.

2.7 Inorganic nutrient analysis

The nutrient analysis allowed to understand how the microorganism communities responded to the nutrient's gradient. To do that, the water from each station was collected and filtered through a cellulose acetate membrane with 0.45 µm pore´s wide. This will ensure that the water was free from most of the living organisms and particles bigger than 0.45 µm, that could interfere with the analysis. Additionally, for each sample, a triplicate analysis was performed.

2.7.1 *Ammonia*

For the quantification of ammonia (NH_3) and ammonium (NH_4^+) in the samples, the Koroleff method was used[122]. This method is based on the fact that the ammonia, in alkaline solution will react with hypochlorite, creating monochloramine. By its turn, the phenol and hypochlorite in excess, in the presence of nitroprusside, will be catalysed into indophenol which displays an intense blue colour. In order to access the concentration in the samples, a standardization line with increasing ammonia concentration must be accomplished. For that, the secondary standard reagent (250 µM) was made from the dilution of the primary standard reagent (250 µM). Next, predefined quantities of distilled water and secondary standard reagent were mixed to obtain solutions with increasing concentrations of ammonia and ammonium. These solutions were used to build a standardization line where it will be possible to extrapolate the ammonia concentrations in the samples. The tubes containing the samples and the standards were incubated with Phenol/Nitroprussiate reagent and Hypochlorite reagent, vortexed, and left from 6 to 30 hours. After incubation, the samples and the standards were read in the spectrophotometer at λ=630 nm.

2.7.2 *Nitrates*

The protocol for the nitrates´ (NO_3^-) concentration, used in this study, was firstly described by Jones (1984)[123] and later adapted by Joye & Chambers (1993)[124]. This method is based upon the reduction of nitrates (NO_3^-) to nitrites (NO_2^-) triggered by the cadmium in a chloride ammonia solution. Similar to the previous protocol, the secondary standard reagent (250 µM) was made by dissolving the primary standard reagent (100 mM) in water. After that, the standardization solutions were made by adding the predefined values of water and secondary standard reagent to the tubes containing the chloride ammonia solution and spongy cadmium. The incubation period took about 1:30 hours, where the solutions were at constant agitation (100-200 rpm). After that, coloured nitrite reagent was added to

each tube, vortexed, and after 10 to 30 minutes of incubation, the solutions were read in the spectrophotometer at λ= 540 nm.

2.7.3 *Nitrites*

The Nitrites (NO_2^-) were quantified in this study by the methods described by Grasshoff et al. (2009)[125]. This method is centred around the reaction of nitrite with an aromatic amine (sulfanilamide), which give rise to a compost that will attach to a second aromatic amine. The result will be the accumulation of a pink compost, directly proportional to the nitrite concentration in the solution. The standardization line was made once more by mixing the secondary standard reagent with distilled water in preestablished concentrations. For the determination of the nitrites, coloured nitrite reagent was added, and the solutions were ready to incubate for 10 to 30 minutes. The spectrophotometer was set at λ= 540 nm for the reading.

2.7.4 *Phosphate*

The phosphates were also determined by Koroleff in the chapter "Determination of nutrients" in the book "Methods of Seawater Analysis" by Grasshoff et al. (2009)[125]. The reaction happens when the ascorbic acid reduces the phosphates ions, which results in an accumulation of a bluish complex. The standardization was once more done by the mixing of specific amounts of the secondary standard reagent with water. To determine the phosphorus present in the samples, the ascorbic acid solution was added to the tubes, vortex and immediately added mixed reagent. After all the tubes were vortex, the solutions were incubated for 20 minutes in the dark. Finally, the samples and the standards can be read in the spectrophotometer at λ= 880 nm.

2.7.5 *Silica*

Silica can be present dissolved in the water under the formula $Si(OH)_4$. The method used to determine the concentration of silica in the samples was based on the book "Methods of Seawater Analysis" by Grasshoff et al. (2009) [125]. The silica will react with ammonium molybdate, which will originate silicolybdic acid that will taint the solution of yellow. Since the yellow colour produced was not much intense, the ascorbic acid will react with the compost, turning the solution from a faded yellow to a strong blue. The procedure was similar to the ones above. For the standardization different tubes, with different concentrations of silica were produced. This was accomplished by adding preestablished quantities of primary standard reagent and distilled water. To determine the silica concentration, of molybdate acid were added to each tube. After a brief vortex, the solutions were incubated in the dark for 20 minutes. Next, 50 oxalic acid were added, quickly followed by ascorbic acid. The samples and the standards were incubated for 3 hours in the dark and after that, the solutions can be read in the spectrophotometer at λ= 810 nm.

2.8 Chlorophyll Analysis

The chlorophyll *a* analysis allowed to understand more deeply the abundance of the photosynthetic community in the water samples. To do that chlorophyll *a*, b, c need to be read in different wavelengths. Previously, water from each station was collected and filtered by a cellulose membrane with 0.45 µm pore size. Next, the filters were cut as tiny as possible. The small fragments were incubated with acetone at 90 % and left overnight. On the next day, the tubes were centrifuged at 4500rpm for 10 minutes for the supernatant to be safely poured into the cuvettes. The spectrophotometer will read the following wavelengths 750, 665, 663, 645, 630, 510 and 480 nm. After the reading of all the absorbances, 2 drops of HCL 50% were added to the cuvettes and the read of the wavelengths 750 and 665 nm, will take place. This will destroy the chlorophyll *a* present in the sample and allow to measure the phaeopigments in the samples.

2.9 Bioinformatic Analysis

2.9.1 *Upstream Analysis*

At this stage of the bioinformatic workflow, the filtering of the raw Illumina fastq files, into high-quality data must be assured to allow the proper analysis. The paired-end reads were treated using the DADA2 pipeline version 1.16.0[126], using standard parameters. This software was designed to return amplicon sequencing variants (ASVs), instead of the previously used molecular operational taxonomic units (OTUs). ASVs were amplicon sequences 100% similar, being able to be distinguished from other amplicon sequence variants by changes as little as one nucleotide[127]. The choice to follow ASV methods was based on evidence that shows better sensitivity and specificity to discriminate ecological patterns. They represent rRNA gene heterogeneities within microbial communities with higher fidelity than OTU data, and different ASV datasets can be directly compared[127].

Before initiation, this software requires that non-biological nucleotides were removed. Firstly, the DADA2 inspected the profiles of the sequencing quality of the reads, which provides the knowledge needed to trim the reads adequately. The trim interval must be chosen in a way that ensures the deletion of the portions that were poorly sequenced, but at the same time guarantees that the sequences were not cut extensively because this might result in the deletion of the merged zones. The pipeline also filtrated the sequences with low quality. After this step, the denoised forward reads were obtained along with the denoised reverse reads. The software will proceed to the merging of the forward reads with the reverse reads, creating the contig sequences. This can be done because both reads had a complementarity sequence portion, where they must overlap at least 12 bases for the merge to occur. In the next step, the chimaeras were removed. To do that, the program identified the most abundant sequences and remove the artificial sequences that were the ones that although have the same size, differ significantly from the main output, commonly called the chimaeras. In the end, the reads were assigned against a specialized database to identify the taxonomy of a given ASV. In the case of the 16S rRNA gene dataset,

the reads were assigned against the SILVA database (v. 138.1)[128], while in the 18S rRNA gene dataset, the sequences were assigned using the PR2 database (v. 4.14.0)[129]. Additionally, it was also necessary to remove undesirable lineages: sequences classified as “Chloroplasts”, “Mitochondria” and “Eukaryota”, in the 16S rRNA gene dataset and as “Metazoa”, “Fungi”, “Streptophyta”, “Ulvophyceae” and “Rhodophyceaea” for the 18S rRNA gene datase, were removed. In the end, the final output was a cleaned ASV vs. sample table with the correspondent taxonomy and abundance of each ASV per sample.

2.9.2 *Downstream Analysis*

The downstream analysis used the data generated previously by the DADA2, in the upstream analysis, to investigate the distribution and diversity of the microplankton community in the different datasets. This analysis relied mostly on the phyloseq package (v. 1.22.3)[130], that was loaded on the Rstudio, in order to execute their functions and tools. The visualization of the plots was attained by the ggplot2 R package (v. 3.3.6)[131]. This package was also used for the visualization of the environmental parameter’s gradients. In contrast, when it was analysed the environmental gradients for the stations from the NW coast, the Ocean Data View was used instead[132]. Next, the analysis inspected the rarefaction curves. The rarefaction curves were done for the 16S rRNA gene and 18S rRNA gene dataset since, it enables us to visualize if the sequencing was deep enough. If the curve reaches a plateau, it means that the biodiversity of a certain sample, was achieved. It was also weighted the relevance of performing a subsampling in order to enhance the comparison of the samples. After this step, the comparison between the diversity indexes will be more precise, however this procedure will also discharge representative sequences of the samples and it will also distort the true abundance of the samples. Because of that, subsampling was not performed in this study. After that, it was obtained the analysis charts for Alpha diversity, where the “Observed” ASVs and “Shannon” index were used. The first parameter measures richness by the number of ASVs present in each sample. The Shannon index will height both abundance and evenness of the species[133]. For beta-diversity analysis first, the Bray-Curtis distance calculated the dissimilarities between samples. These calculations were then visualized in an ordination plot, where the different clusters were arranged according to the similarity between samples. The ordination plot used was a principal components analysis (PCA). Additionally, a cluster dendrogram with logarithm transformation was done for better visualization of the dissimilarities between samples and the way they were hierarchically organized[134]. The phyloseq package also generated a bar plots graph in order to visualize the taxonomic profiles of the different samples. Additionally, a heatmap was also performed as alternative to visualize the taxonomic profiles. For each phylum in each sample, the heatmap attributes a colour. This colour will be darker if the phylum was more meaningful for the profile, and lighter if the phylum was not that important for the microbial taxonomic composition of the sample. In this analysis, it was performed the subsampling since it revealed to be more accurate. Finally, it was also analysed the correlation of the environmental factors with the biological samples. To do that, the function “envfit”, in vegan package, was used. For the same propose, a PERMANOVA analysis was

performed for each environmental parameter. The function used was the “adonis”, also present in the vegan package [135].

3. Results

3.1 Microbiome and biogeochemical distribution across the Douro River Estuary

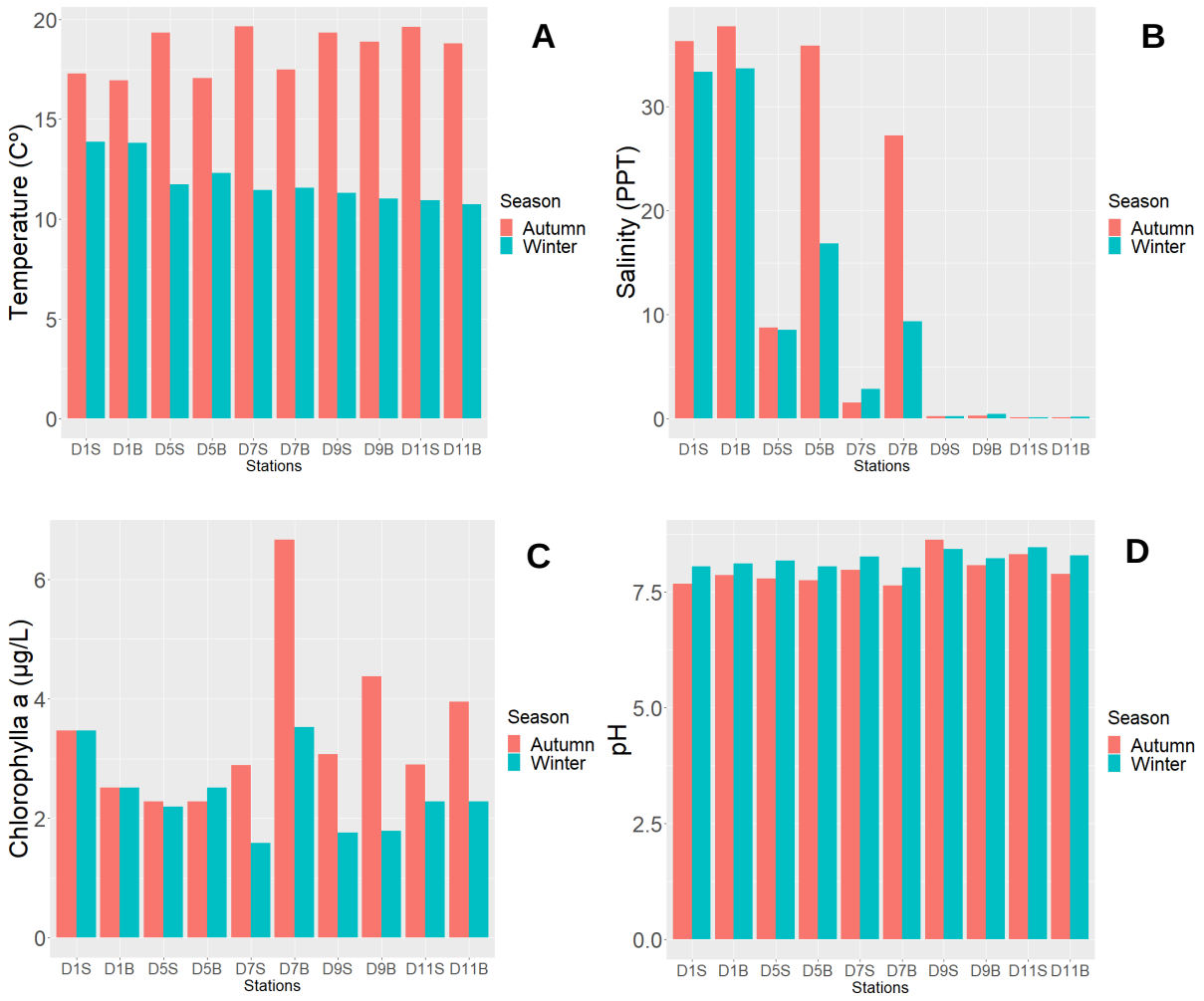
3.1.2 *Biogeochemical Gradients across the Douro River Estuary*

The following description of the biogeochemical gradients in the Douro estuary offer a general picture of the trends and patterns found throughout the sampling transect (stations location), depth (Surface and Bottom) and between the different seasons sampled (Autumn and Winter).

The temperature in the water column was higher in Autumn, in every station, when compared to the Winter stations (Figure 8a). In Autumn the values range between 19.64 ºC (D7S_A) and 16.94 ºC (D1B_A), and in Winter the values vary between 13.86 ºC (D1S_W) and 10.74 ºC (D11B_W) (Annex-7). Regarding depth in the Autumn season, it was noticed a generally higher temperature value at the surface (Figure 8a). Spatially, the temperature showed a slight decrease towards the downstream region in the Winter season (Figure 8a). When analysing the salinity, it was possible to notice that the highest values were registered in station D1 (Figure 8b), more exactly in sample D1B_A (37.67 PPT) (Annex-7). These values drop abruptly in station D5 at the surface (8.75 PPT), however, in the bottom (D5B) the values stayed high, especially in the Autumn (D5B_A: 35.81 PPT). The same pattern was noticed in station D7 with less amplitude of variation. At the upstream part of the estuary, in stations D9 and D11, at both surface and bottom, the salinity values were almost zero. Across seasons, the salinity variability was similar, with the exception of stations D5B and D7B where the salinity values were higher in Autumn. The pH doesn’t shift drastically between samples however, it was possible to notice a seasonal trend. The Winter samples show a higher pH when compared to the Autumn samples (Figure 8d), with the only expectation of station D9S (Annex-7). The chlorophyll *a* concentration was also analysed, where the values were higher in Autumn, especially in the bottom of stations D7, D9 and D11 (Figure 8c). The highest value was registered in Autumn in station D7 in the bottom (6.66 µg/L). Additionally, samples D1 and D5 showed similar values across depth and season. The lowest chlorophyll *a* value was registered in sample D7S_W (1.589 µg/L) (Annex-7).

Patterns of Inorganic nutrient distribution were also presented in Figure 8. The ammonia values, across the transect, start to decline progressively towards the upstream regions, being the highest value registered at the sample D1S_W (5,514 µM) and the lowest at the station D11S in the Winter (0,000 µM). Additionally, the ammonia concentrations were generally higher in the Winter. However, at the stations D11, there was no ammonia registered in the Winter (Figure 8e). Regarding the nitrites, the values were generally low but higher in the Autumn than in the Winter (Figure 8f). Across the transept was possible to notice an increase in the nitrites’ values from downstream to the upstream regions of the estuary. Regarding depth, nitrite concentrations were higher at surface than bottom for stations D5

and D7. Nitrate concentrations presented higher values in Winter compared with Autumn. Regarding the transect was possible to notice an increasing trend in the levels of nitrates towards the downstream stations, where the lowest value registered was in the sample D1B_A (0,801µM) and the highest at the station D11B ate the Winter (97,614 µM) (Figure 8g). The phosphates values were generally higher at the Autumn specially at the upstream regions. The only exceptions were found in the bottom of the stations D5 and D7, and in the sample D1S. Phosphates showed a totally different trend between seasons. In the Winter the values tend to decrease towards the upstream regions, while in the Autumn the opposite was true (Figure 8h). Regarding silica, was possible to notice higher values in the Winter than in the Autumn. Across the transect, the values tend to increase towards the upstream regions (Figure 8i). And regrading depth, once again, the stations D5 and D7 showed great variance between bottom and surface.



G H

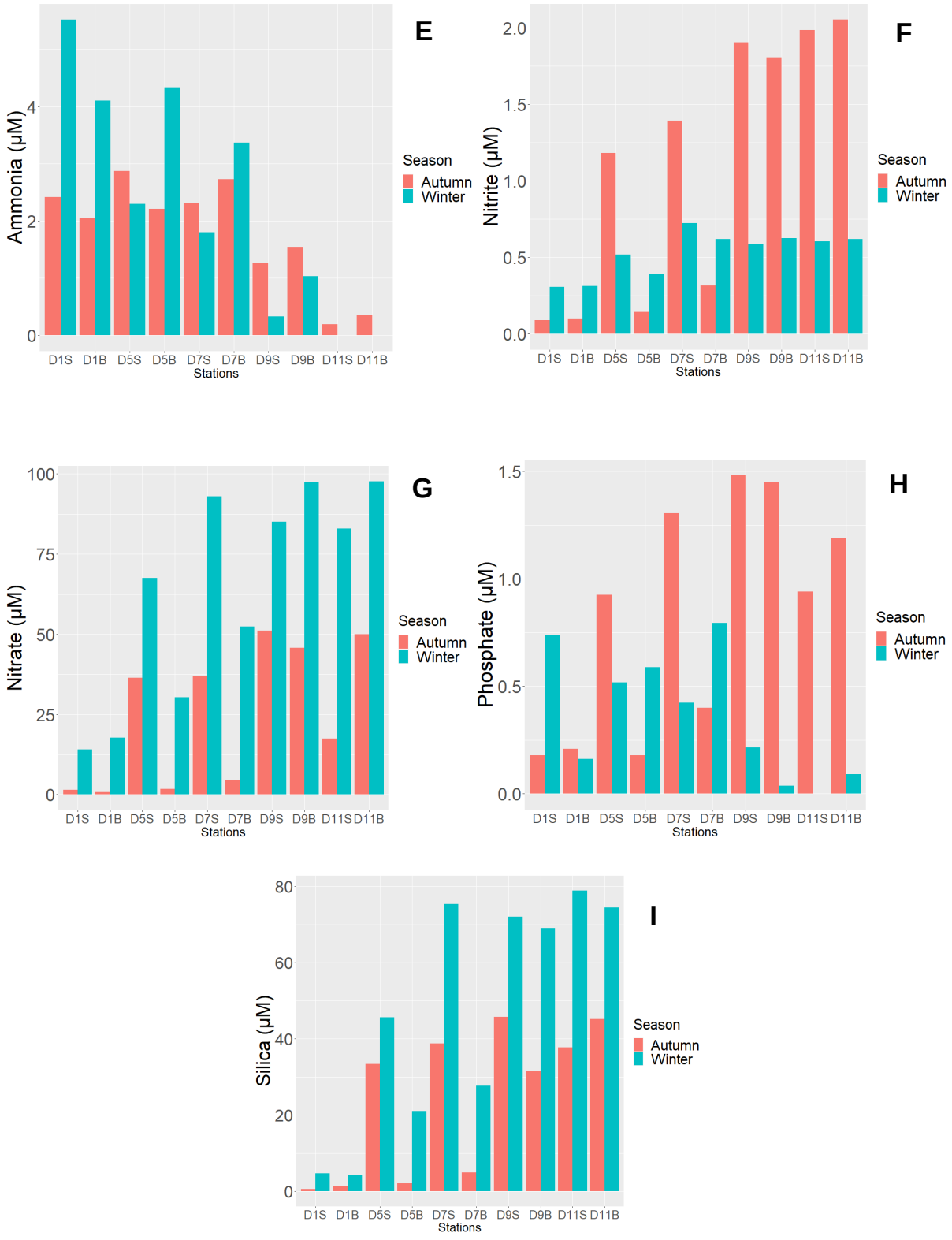


Figure 8: Environmental gradients of the different variables measured (Y) in the Douro estuary sampling stations (X) collected at 2021/2022. Bars with the red colour were referring to the Autumn camping while the blue bars were referring to the Winter campaign**. a)** Temperature (Cº) variability across stations; **b)** Salinity (PPT) variability across stations; **c)** Chlorophylls (µg/L) variability across stations; **d)** pH variability across stations; **e)** ammonia (µM) variability across stations; **f)** nitrites (µM) variability across stations; **g)** nitrates (µM) variability across stations; **h)** phosphate (µM) variability across stations; **i)** silica (µM) variability across stations.

3.1.2 *Planktonic Microbiome gradients*

In the 16S rRNA gene dataset the Winter sample D11B_W was removed from the analysis due to low sequencing depth. Excluding this sample, the 16S rRNA gene amplicon sequencing yielded 542 620 raw reads and the18S rRNA gene amplicon sequencing yielded 1 968 584 raw sequences. The sequences went through the upstream analysis, where the DADA2 package filtered, denoised, merged and removed the chimeras. After all these steps, the final sum of sequences for the 16S rRNA gene dataset was 236 834, with a loss of 56.35 % from the initial sequences (Annex-1). The sum of the high-quality sequences in the 18S rRNA gene dataset was 1 101 196, yielding a loss of 44.06 % from the initial raw reads (Annex- 2).

3.1.2.1 Alpha and Beta diversity

In Figure 9, the rarefaction curves for both 16S rRNA gene and 18S rRNA gene datasets were presented. As it was possible to notice, the plateau was reached in every sample, and therefore it was possible to ensure that the biodiversity in all samples was fully covered.

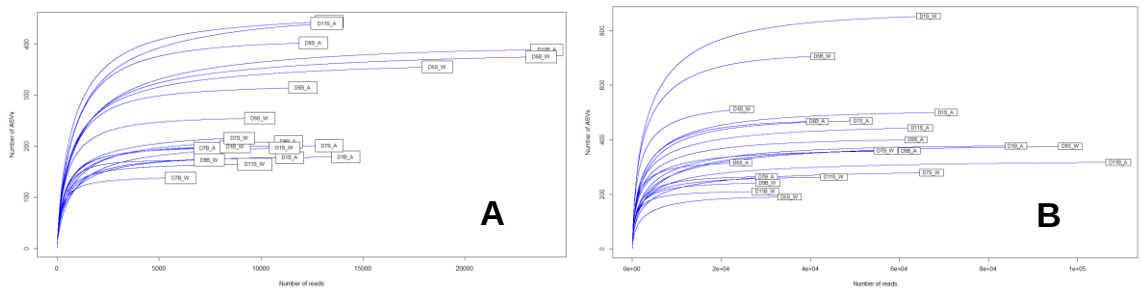


Figure 9: Rarefaction curves of the Douro samples collected in 2021/2022; where the Figure 9a displays the 16S rRNA gene dataset, while the 18S rRNA gene dataset was showed on the Figure 9b.

To study the Alpha diversity of the prokaryotic and unicellular eukaryotic communities, the observed ASVs and the Shannon index were analysed. In the Figure 10 it was possible to notice that the samples D5S_A and D11S_A were the ones with higher prokaryotic diversity and abundance. The Shannon index showed a general increase gradient of prokaryotic diversity from the estuary mouth to the upstream region of the estuary. In the Winter, the prokaryotic diversity was generally higher in the most downstream part of the transept, decreasing towards the upstream region. This pattern was more pronounced in the Autumn than in the Winter. Regarding depth, the stations D5 and D9, in both seasons, display a great variation in observed ASVs and in the Shannon index between surface and bottom, being the samples at surface more diverse. Additionally, D1 at the Winter also displays a great diversity variation between surface and bottom.

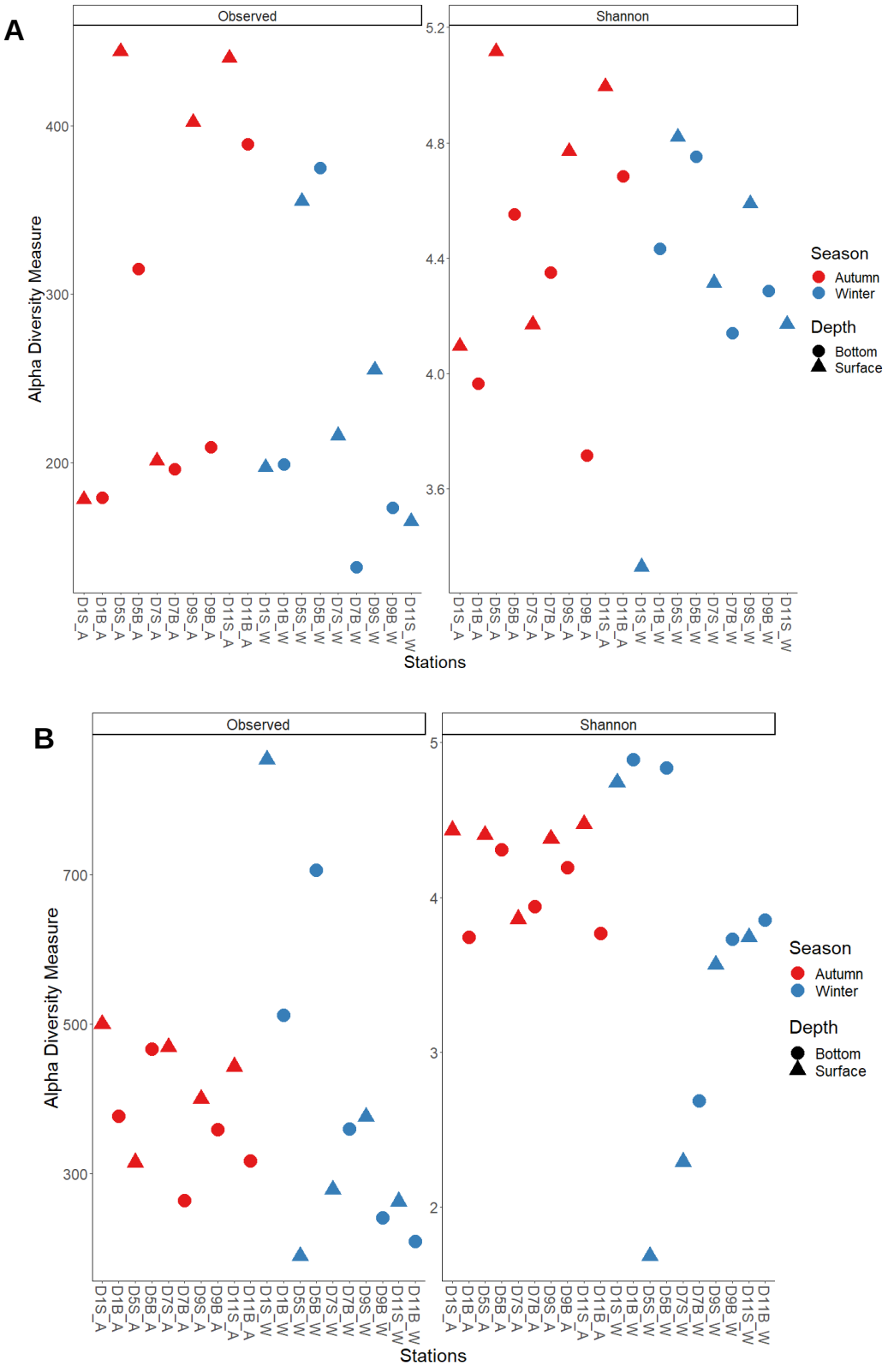
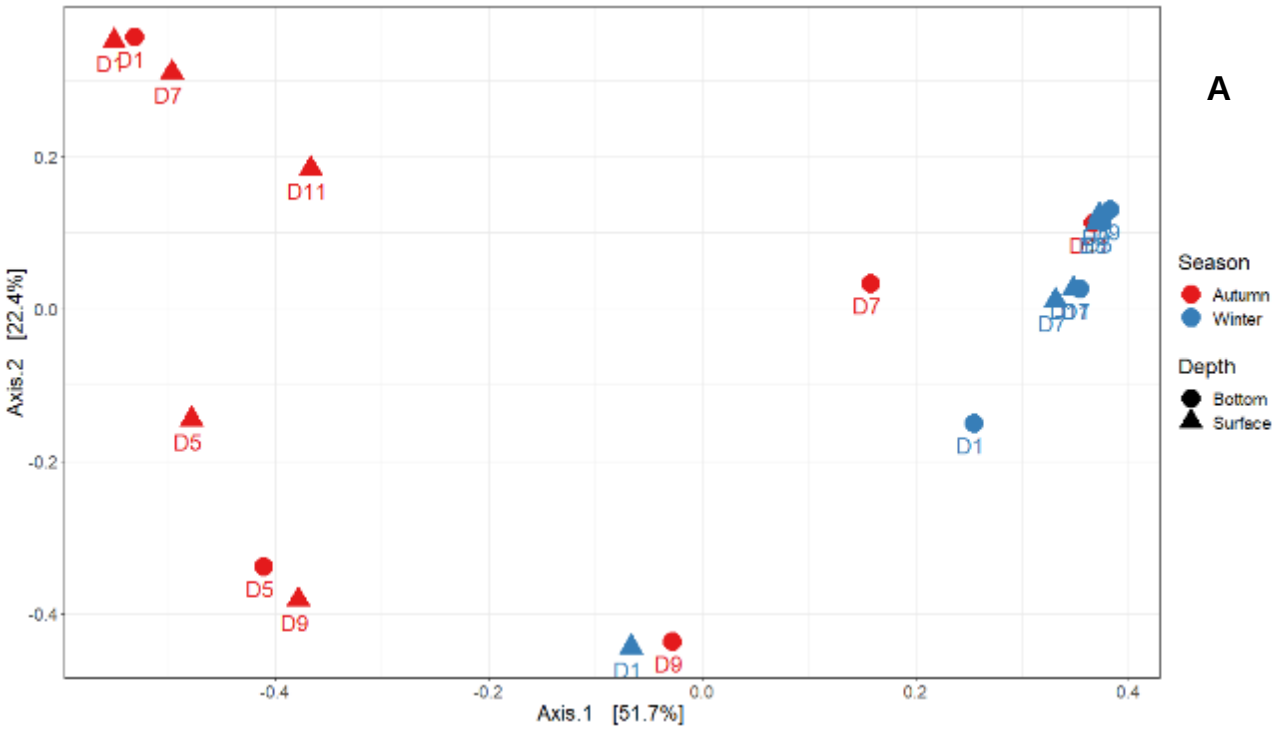


Figure 10: Alpha diversity (number of observed ASVs and Shannon index) of the Douro samples, collected in 2021/2022, for the different sampling depths (Bottom and Surface) and season (Autumn and Winter). The Figure 10a, was generated using the 16S rRNA gene dataset and the Figure 10b the 18S rRNA gene dataset.

Alpha diversity for the unicellular eukaryotic community was presented in Figure 10b, it was possible to visualize that the samples with higher diversity and number of observed ASVs were the D1S_W, D1B_W and D5B_W. The Alpha diversity metrics for the unicellular eukaryotes of the rest of the samples were in the same range, being the D5S_W, D7S_W and D7B_W the samples with lower diversity, according to the Shannon index. Regarding the seasonal variability, the Autumn and Winter samples showed generally similar levels of unicellular eukaryotic diversity. Variation with depth also shows no particular pattern, with the exception of D5S_W and D5B_W samples that display great variability in both number of observed ASVs and the Shannon index between bottom and surface samples. If compared both prokaryotic and unicellular eukaryotic diversity, it was possible to notice higher Observed ASVs values for the unicellular eukaryotic community.

The prokaryotic beta diversity was assessed by performing a multidimensional scaling plot, and for additional visualization, a hierarchical cluster was also performed (Figure 11). This analysis allowed to visualize a clearly prokaryotic community's separation. Results showed that samples from the Winter campaign clustered together, although the samples from the Autumn campaign did not cluster as tightly (Figure 11a). However, in the hierarchical cluster analysis (Figure 11b), it was possible to verify that most of the samples from these two seasons cluster separately, with the exception of some bottom stations from Autumn campaign (D7B_A and D11B_A) that were most probably representative of a different water mass (Annex 7).



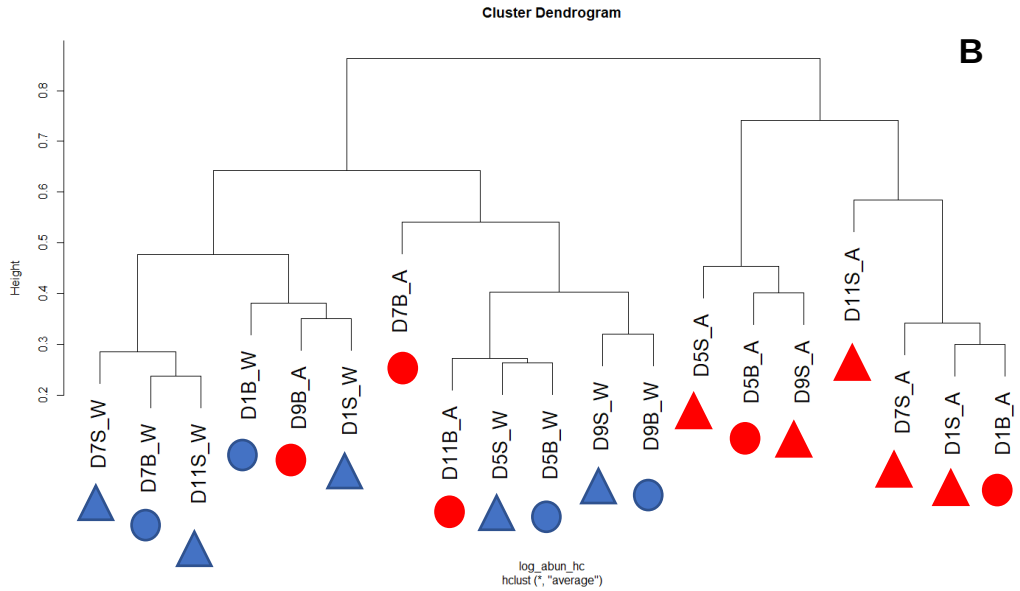


Figure 11: Beta diversity analysis of the 16S rRNA gene Douro dataset, collected during 2021/2022; The Figure 11a PCoA ordination plot with the different depths (surface and bottom) distinguished by different shapes (circle and triangle), and the different seasons (Autumn and Winter) were notable with a different colour (red and blue); In Figure 11b, hierarchical cluster analysis.

For the 18S rRNA gene dataset it was also possible to notice four clusters (Figure 12 a). A close analysis showed that the samples from Winter and Autumn were more disperse in the plot, showing that the unicellular eukaryotic communities were not as well separated by season, as the prokaryotic communities (Figure 12). Additionally, when analysing the hierarchical cluster, it was possible to notice two major groups, and within each cluster there were two subgroups well separated by season. The patterns noticed in these clusters show that the samples from the stations near the mouth of the estuary form a cluster, for both seasons; while the stations more upstream form another cluster (Figure 12b). Regarding depth it was possible to notice that the samples D1S_A and D5S_A were more similar than D1B_A. Additionally, the sample D1S_W clusters more strongly with D5B_W than with D1B_W. As the transect moves upstream in the Autumn and Winter, the samples D9 and D11 from surface and bottom cluster together.

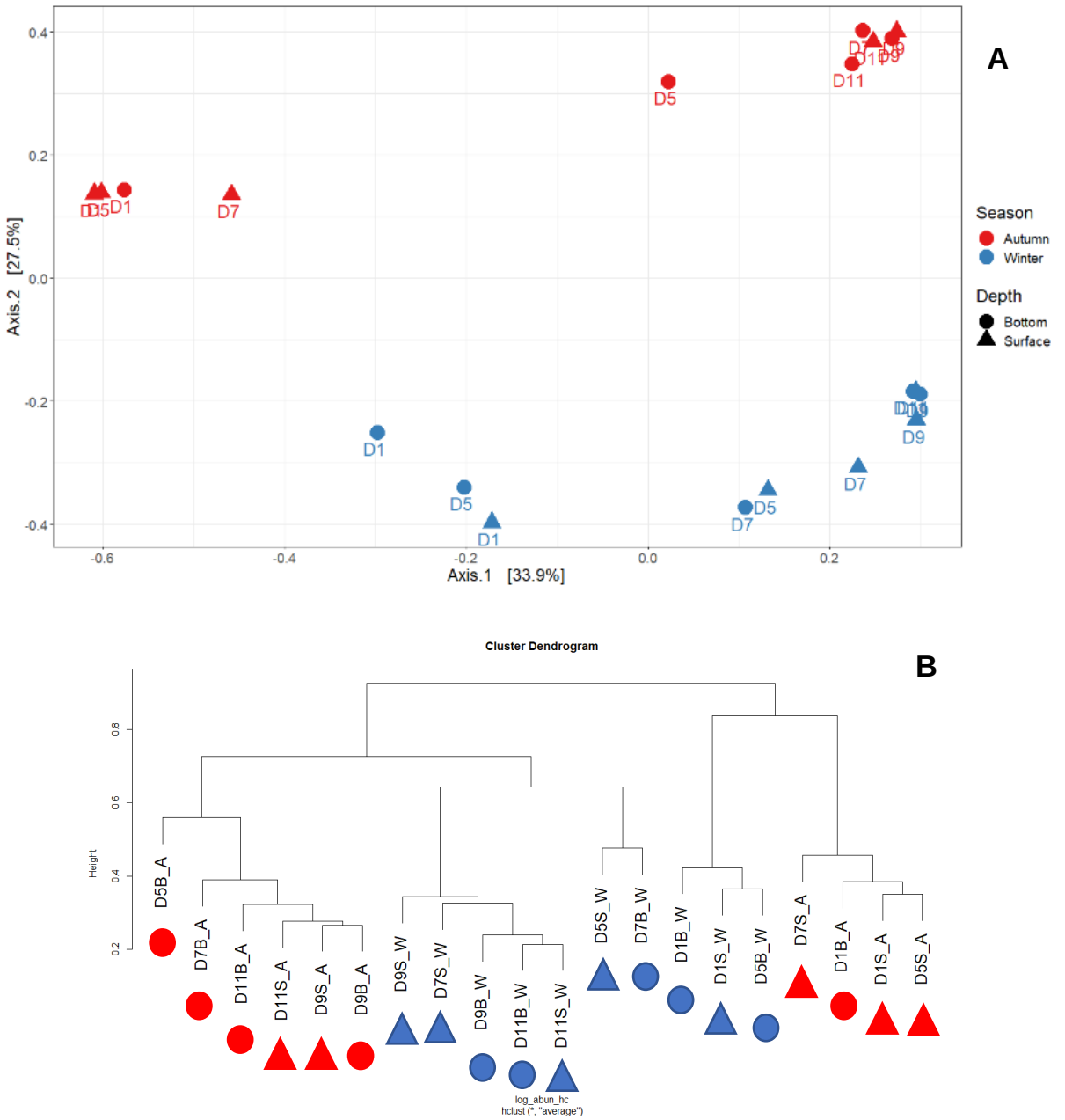


Figure 12: Beta diversity analysis of the 18S rRNA gene Douro dataset, collected during 2021/2022; The Figure 12a PCoA ordination plot with the different depths (surface and bottom) distinguished by different shapes (circle and triangle), and the different seasons (Autumn and Winter) were notable with a different colour (red and clue); In the Figure 12b, hierarchical cluster analysis.

3.1.2.2 Taxonomic composition

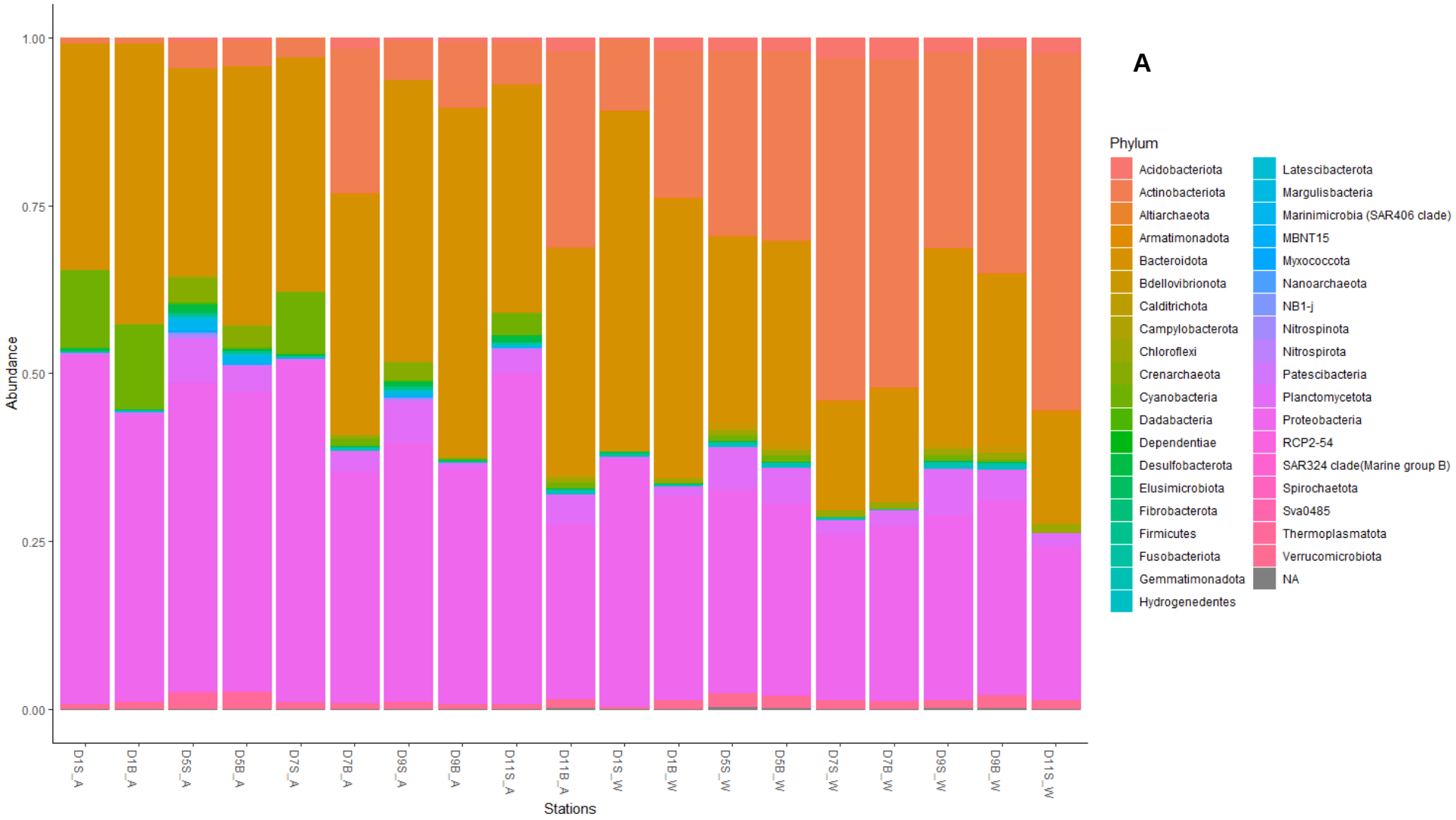
The visualization of the taxonomic profiles was accomplished by plotting each sample with its corresponding phyla profile. If we take an overlook at the prokaryotic community profiles (Figure 13a), was possible to notice that, in almost all samples, the taxonomic pattern changes greatly between seasons and differ little when comparing surface with bottom samples (Figure 13a). The most relatively abundant and cosmopolitan prokaryotic phyla were Proteobacteria, Bacteroidota and Acidobacteriota. Regarding Acidobacteriota, it was possible to visualize that their relative abundance was greater in the

Winter compared with Autumn and tends to increase as the transect moves upstream. By their turn, Proteobacteria concertation decreased slightly in Winter compared with the Autumn. Cyanobacteria were also more representative in the Autumn, mainly in the downstream stations, than in Winter (Figure 13a). In the distribution of unicellular eukaryotic phyla were also mostly different between seasons compared with depth. The samples were defined by three main phyla, the Stramenopiles, the Hacrobia and the Alveolata, that were present between seasons and depths, and across the transept. The Alveolate phylum appeared generally more abundant in the samples from the Autumn dataset. On the contrary, in the Winter samples, a decrease in the Alveolate levels was followed by an increase in the Stramenopiles levels, especially in station D5. And finally, the Hacrobia phylum tends to be also more abundant from the station D7 to station D11, with exceptions for D1S_A and D5S_A (Figure 13b). Regarding depth, the station with more variation was the D5 in the Winter (Figure 13b).

3.1.2.3 Environmental drivers of microbiome distribution

We attempted to understand which were the environmental parameters that shaped more significantly the microbial community's distribution. For that, PERMANOVA analysis were done for each environmental condition (Temperature, Salinity, pH, Chlorophyll, Ammonia, Nitrite, Nitrate, Phosphate, Silica) (Annex-10 in the supplementary material). From this analysis it was possible to notice that the most significant conditions that shaped the prokaryotic communities were the temperature (R2=0.2290, p-value=0.001) and nitrate concentration in the water column (R2=0.133, p-value=0.043). For the unicellular eukaryotic communities, the most relevant variables were pH (R2=0.127, p-value=0.010), temperature (R2=0.237, p-value=0.001), salinity (R2=0.124, p-value=0.019), ammonia (R2=0.1424, p-value=0.014), nitrite (R2=0.132, p-value=0.015) and nitrate (R2=0.159, p-value=0.002) concentrations. All these parameters where able to reject the null hypothesis and therefore there was evidence that microplankton community changed between samples, with the different concentration of the specific nutrients reported above.

Next, a CCA analysis was performed to correlate the environmental parameters with the microbial community's data sets. Regarding the prokaryotic communities, was it possible to notice that the factor that correlates with the samples more significantly was the temperature. However, for the unicellular eukaryotes, not only temperature (R2= 0.5416, p-value =0.001) shaped the distribution of these communities but also the salinity (R2= 0.4545, p-value =0.018), and the concentration of ammonia (R2= 0.5032, p-value =0.007), nitrite (R2= 0.4762, p-value =0.013) and nitrate (R2= 0.3372, p-value =0.045) significantly correlated with the distribution of these planktonic communities. Nevertheless, in the CCA analysis (Figure 14) variability of the prokaryotic communities between Winter samples were found to be also influenced by the silica, pH, and nitrates levels, while Autumn samples were mostly influenced by temperature, salinity, and phosphates (Figure 14a). The CCA analysis for the unicellular eukaryotic communities (Figure 14b), showed a strong influence of the salinity, ammonia, and temperature in the samples closer to the downstream areas (D1 and D5), while the samples in the most upstream regions (D7, D9 and D11) were more influenced by the inorganic nutrient concentrations (nitrates, nitrites, phosphates, and silica).


A
Abundance
1.00
0.75
0.50
0.25
0.00
Stations
D1S_A
D1B_A
D5S_A
D5B_A
D7S_A
D7B_A
D9S_A
D9B_A
D11S_A
D11B_A
D1S_W
D1B_W
D5S_W
D5B_W
D7S_W
D7B_W
D9S_W
D9B_W
D11S_W
Phylum
Acidobacteriota
Actinobacteriota
Altiarchaeota
Armatimonadota
Bacteroidota
Bdellovibrionota
Calditrichota
Campylobacterota
Chloroflexi
Crenarchaeota
Cyanobacteria
Dadabacteria
Dependentiae
Desulfobacterota
Elusimicrobiota
Fibrobacterota
Firmicutes
Fusobacteriota
Gemmatimonadota
Hydrogenedentes
Latescibacterota
Margulisbacteria
Marinimicrobia (SAR406 clade)
MBNT15
Myxococcota
Nanoarchaeota
NB1-j
Nitrospinota
Nitrospirota
Patescibacteria
Planctomycetota
Proteobacteria
RCP2-54
SAR324 clade(Marine group B)
Spirochaetota
Sva0485
Thermoplasmatota
Verrucomicrobiota
NA

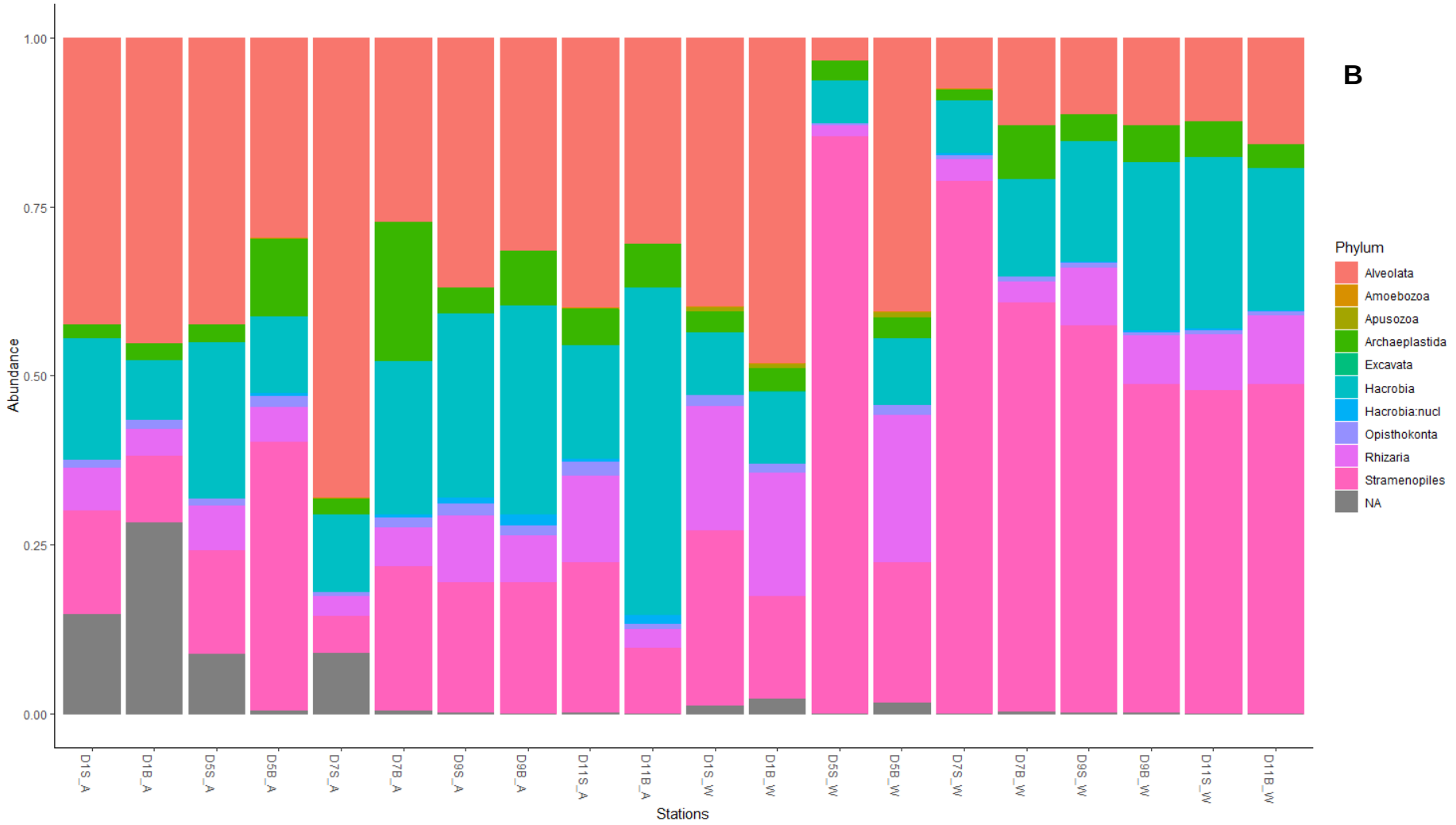


Figure 13: Taxonomic diversity graph of the Douro 18S rRNA gene dataset, collected at 2021/2022, where was represented the taxonomic profiles in a bar plot. Figure 13a, represents the 16S rRNA gene dataset and Figure 13b the 18S rRNA gene dataset.

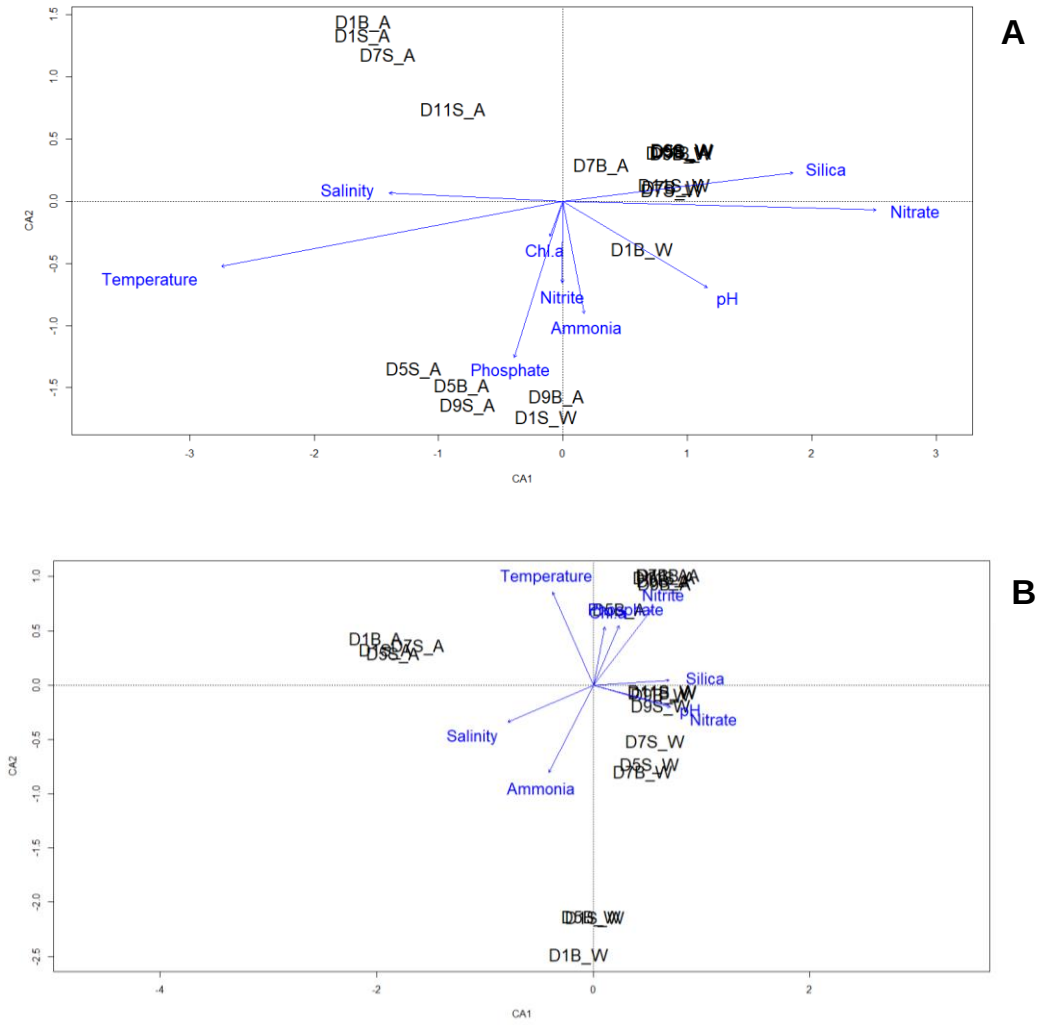


Figure 14: CCA ordination plot for 16S rRNA gene rRNA (Figure 14a) and 18S rRNA gene (Figure 14b) data sets. The length of the environmental arrows was indicative of the degree of influence of the samples.

3.2 Tides impact on microbiome and biogeochemical distribution in the Douro River Estuary

3.2.1 Biogeochemical Gradients across the Douro River Estuary

Here we described how the environmental data varies within high and low tide, across Douro estuary, during the Winter season (December 2016). In Figure 15 it was possible to visualize that salinity was always higher in the stations that were near the estuary mouth in both tides (Annex-8). Regarding temperature, was it possible to notice a difference of 0.8 ºC between tides, in the stations R1H and R1L. In the rest of the samples the temperatures were similar (Figure 15). The nitrate variation was also noticed between tides, especially in the samples R4H, R4L and R5H, R5L. The nitrite does not seem to display any relevant pattern however, wider variations between tides happened in the stations R1 and R5 (Figure 15). Regarding the phosphate levels, from the station R2 to R5, they were always higher in

the low tide comparing with the high tide. The concentrations of the ammonia were quite stable between both tides, with the exception of R1 where high range of variation was registered with higher values at high tide (Annex-8). Regarding pH, the values were similar, ranging from 7.72 to 7.49, however in station R1 values varied greater between tides (Figure 15).

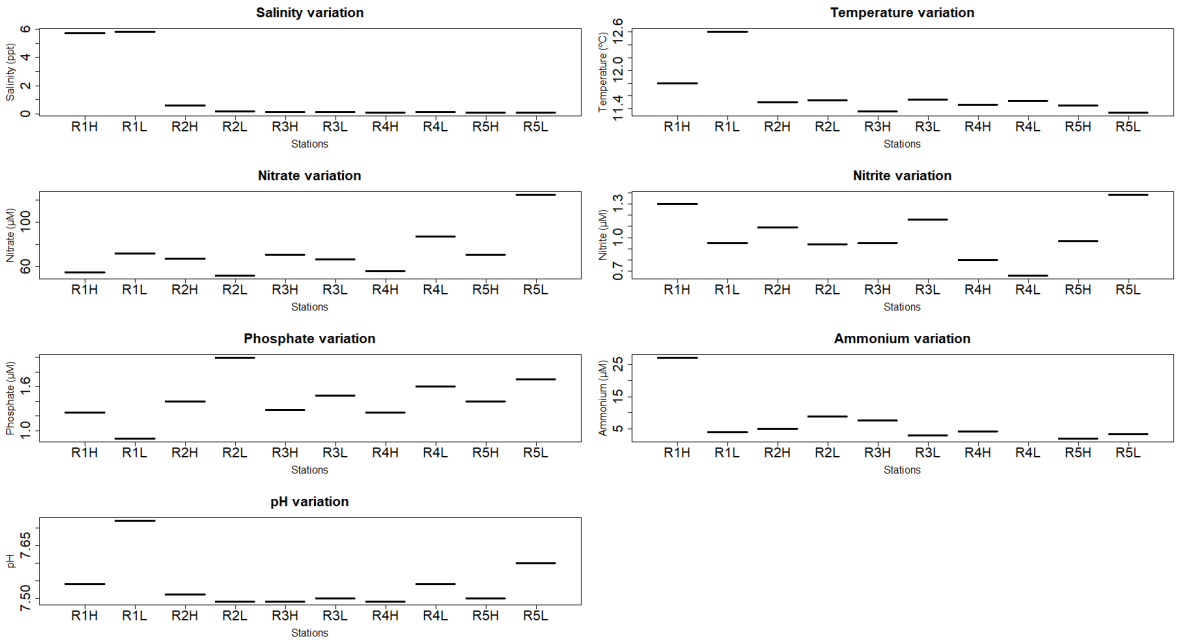


Figure 15: Environmental gradients of the Douro estuary taken in 2016, where in the X axis was present each station sampled at different tide (H-High Tide; L-Low Tide); the Y axis represents the numeric scale where was measured the concentration of the temperature (Cº), Salinity (PPT), Nitrate (µM), Nitrite (µM), Phosphate (µM), Ammonia (µM), and pH.

3.2.2 Planktonic Microbiome gradients between estuarine tides

The 2016 Douro campaign provided a total of 801 836 and 771 489 raw sequences, from the 16S rRNA gene and 18S rRNA gene amplicon sequencing, respectively. The reads were posteriorly filtered and denoised; the forward and reverse reads were merged, and the chimeras were removed. After all the upstream analysis, the total sum of ASVs was 427 415, meaning that it was lost 46.69% of the original 16S rRNA gene sequences (Annex-3). The 18S rRNA gene dataset suffered a similar process where the final cleaned reads were 471894, yielding a percentage of 38.83 sequences lost (Annex-4). Next, the rarefaction curves were done for the 16S rRNA gene and 18S rRNA gene datasets and, as it was possible to notice in Figure 16, that the plateau was reached in every sample, and therefore it was possible to ensure that the biodiversity in all samples was covered.

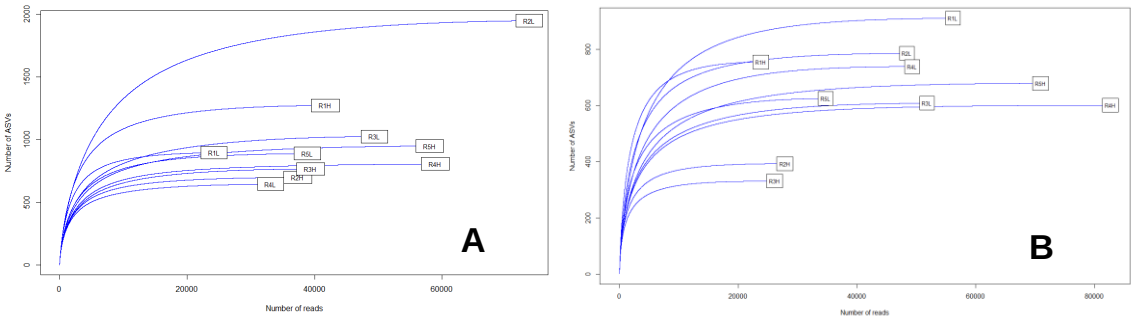


Figure 16: Rarefaction curves of the Douro samples collected in 2016; where the Figure 16a displays the 16S rRNA gene dataset, while the 18S rRNA gene dataset was showed on the Figure 16b.

3.2.2.1 Alpha and Beta diversity

The Alpha diversity of each sample was analysed by the number of observed ASVs, alongside with the Shannon index. In the Figure 17a it was possible to visualize the prokaryotic diversity. In this dataset the samples R1H, R1L and R2L presented higher diversity levels. The rest of the samples had low Alpha diversity metrics, increasing slightly the Shannon index values toward upstream regions. Regarding tide variation, the only remarkable station that displayed a higher Alpha diversity metrics fluctuation was the station R2. The rest of the stations did not display such a wider variation between tides. By its turn, the unicellular eukaryotic communities showed similar patterns, since R1H and R2L showed to be a higher Shannon diversity. The station R1L showed lower diversity in the Shannon index, but high ASVs counts in the richness analyses. Regarding the tides variability in alpha diversity were especially noticed in the R1 and R2 stations. Additionally, it was possible to observe lower Alpha diversity metrics on the unicellular eukaryotic communities, when compared with the prokaryotic communities (Figure 17).

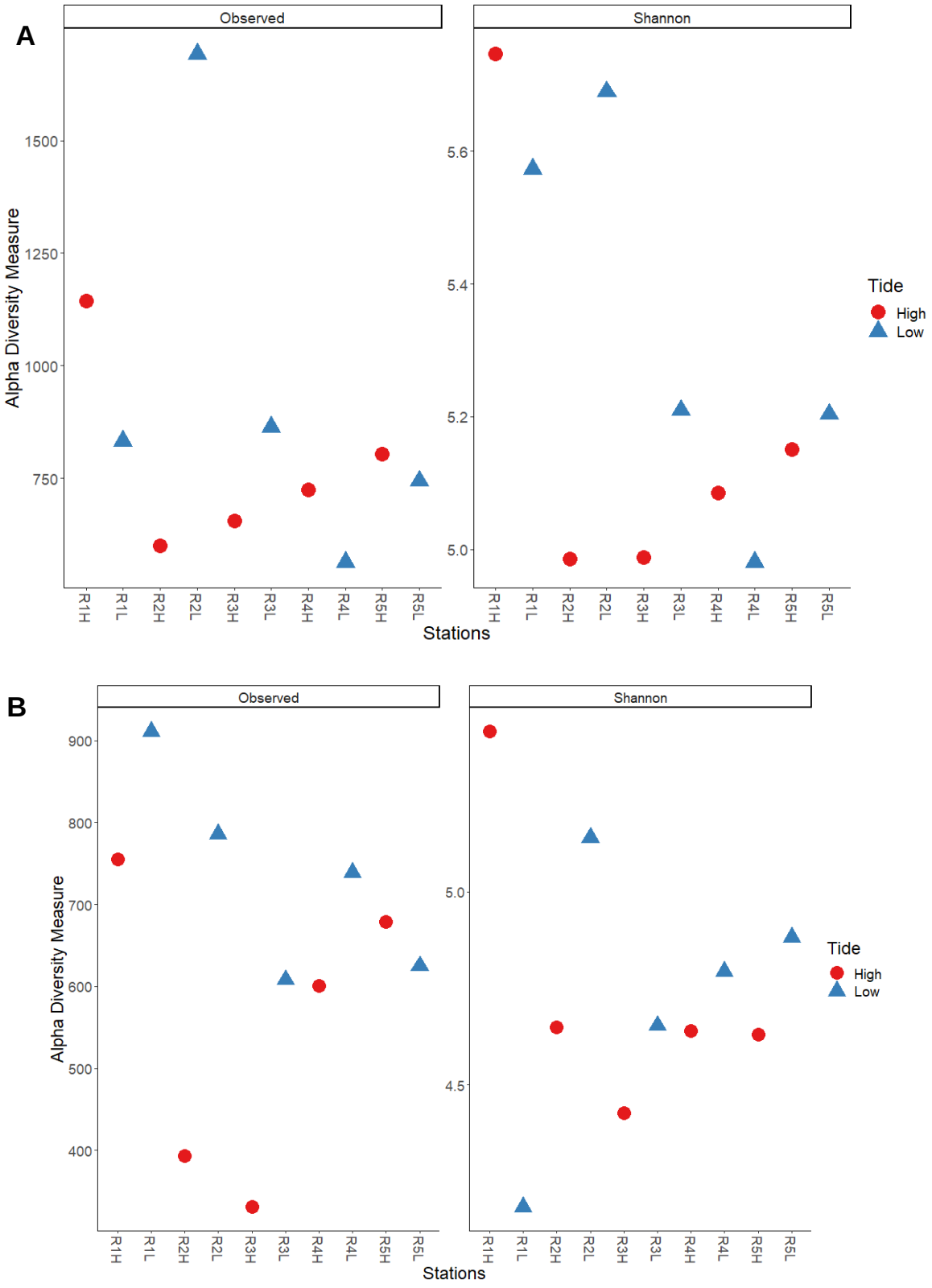


Figure 17: Alpha diversity (Observed and Shannon measure) of the Douro samples, collected in 2016, where the different sampling tides (High and Low) have different colour and shape; The Figure 17a, represents the 16S rRNA gene dataset and the Figure 17b the 18S rRNA gene dataset.

For the beta diversity, it was analysed the ordination plot for the prokaryotic (16S rRNA gene dataset) (Figure 18a) and unicellular eukaryotic communities (18S rRNA gene dataset) (Figure 18b). The results were similar for both microbial communities showing that the stations R1H, R1L and R2L strongly differ in terms of community structure. Also, the samples closer to the mouth of the estuary show a higher variation between tides since R1H, R1L, R2H and R2L cluster separately. On the contrary, the samples more upstream (R3, R4, R5) cluster together showing the increase similarity in the microbial composition on these parts of the transect. In addition, the samples showed a clear gradient caused by tides in the prokaryotic communities (16S rRNA gene dataset) and in the unicellular eukaryotic communities (18S rRNA gene dataset). This gradient extends itself from the mouth of the estuary (R1 and R2), where large differences between high and low tide were noticed, to the Crestuma (R4 and R5), where the more homogeneous communities were observed between tides (Figure 18).

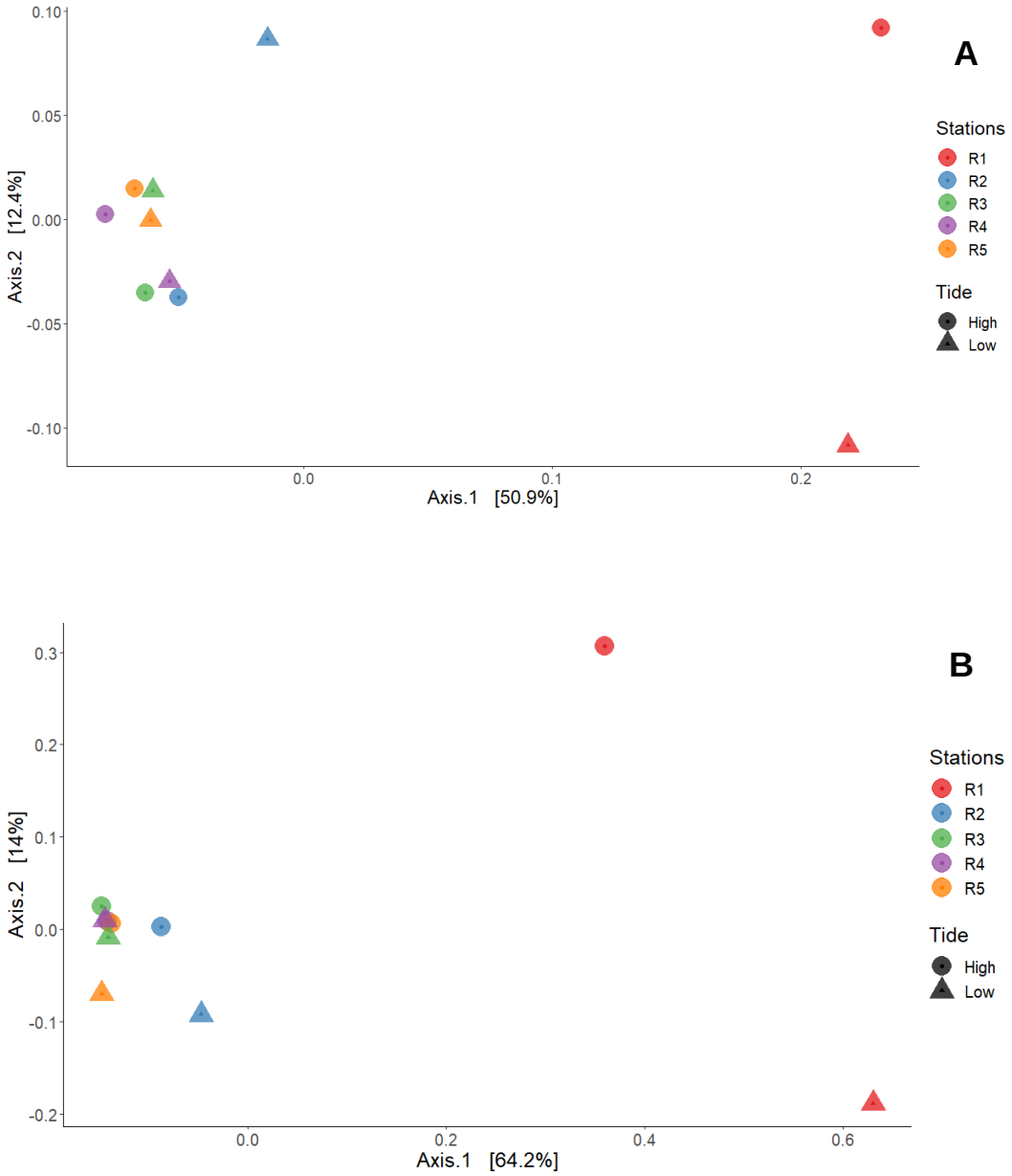
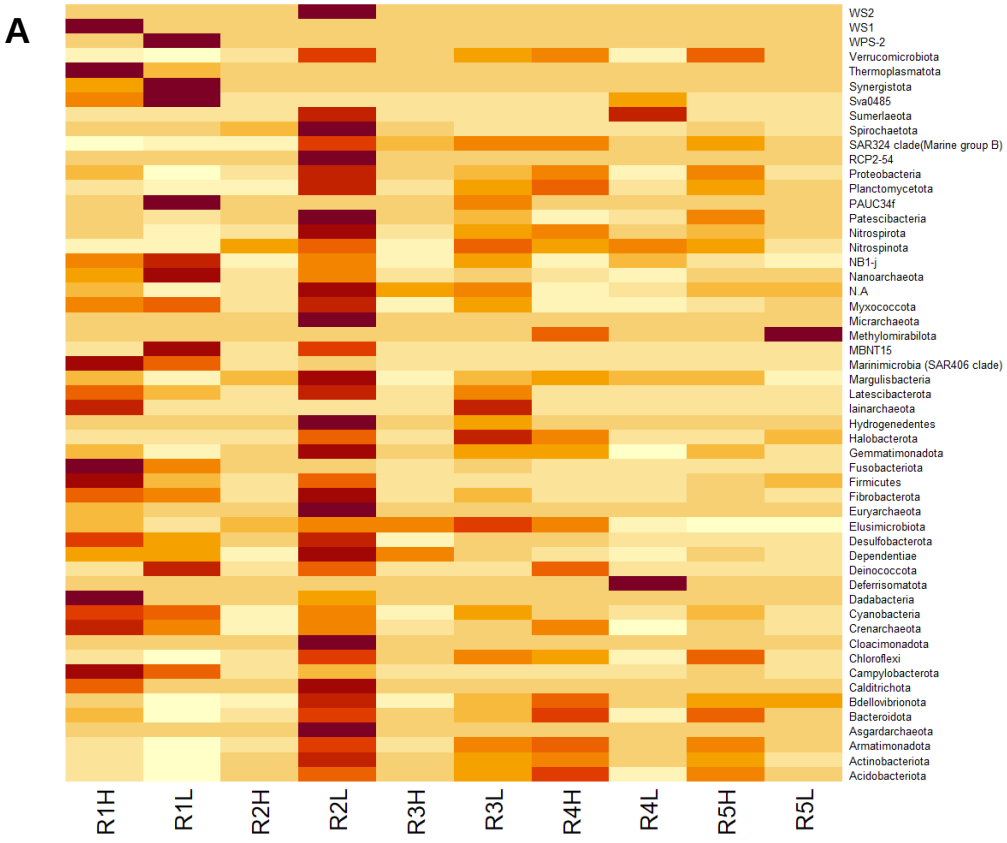


Figure 18: Beta diversity analysis of the Douro samples collected during 2016, for 16S rRNA gene (Figure 18a) and 18S rRNA gene (Figure 18b) datasets. The analysis was graphically represented on the PCoA ordination

plot, where the different stations have different colours, and the tides (High and low) were distinguished by different shapes (circle and triangle).

3.2.2.2 Taxonomic Composition

The Figure 19a and 19b display the relative abundance of prokaryotic and unicellular eukaryotic phyla identified in the different samples, respectively. As it was possible to notice in both datasets, the samples R1H and R1L presented a distinct phyla profile, meaning that the microbial taxonomic composition at high taxonomic lever differs between tides. The same was true for the samples R2H, R2L and in lesser degree for the samples R3H, R3L. Upstream of the station R3, the taxonomic profiles between tides start to become more homogeneous, meaning that the stations R4 and R5, were showing a more similar taxonomic composition. The homogeneity levels were higher at the unicellular eukaryotic communities than at the prokaryotic communities.



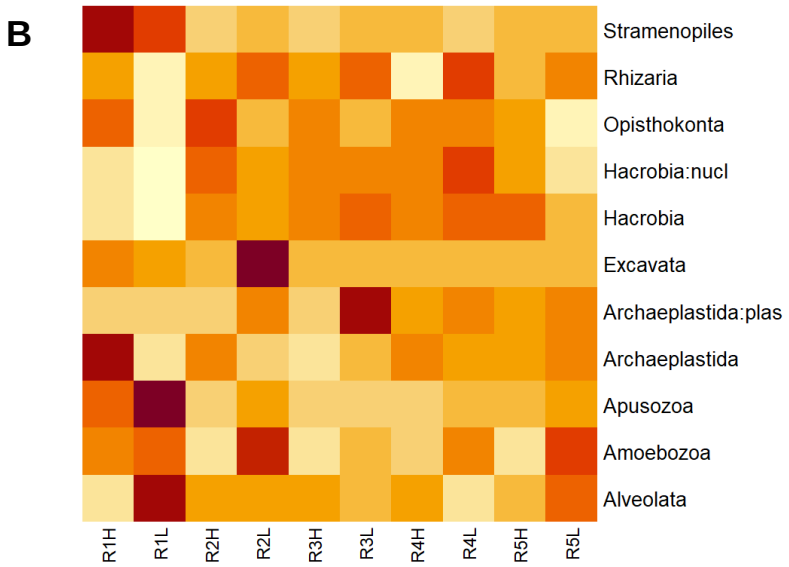


Figure 19: Heatmaps of the relative abundance of the different prokaryotic (Figure 19a) and unicellular eukaryotic (Figure 19b) phyla along the different Douro estuary stations in Winter (2016) at low and high tides.

3.2.2.3 Environmental drivers on microbiome distribution

The relationship between the environmental and the distribution of the microbial communities, was accomplished by PERMANOVA analysis and by the correlation of the environmental vectors with the estuarine microbiome. The PERMANOVA analysis showed that across all the transect, the tides did not display any statistically significant influence in the prokaryotic communities (R2=0.063, p-value= 0.764) and in the unicellular eukaryotic communities (R2=0.061, p-value= 0.690) (Annex 11).

Temperature (R2=0.4365, p-value= 0.011), salinity (R2=0.4643, p-value= 0.004) and pH (R2=0.2730, p-value= 0.006) were a significant condition that shaped the prokaryotic communities (Annex-11 a.). By its turn, the unicellular eukaryotic communities were mostly influenced by the temperature (R2=0.4010, p-value= 0.007), salinity (R2=0.4573, p-value= 0.001) and pH (R2=0.2968, p-value= 0.031) concentrations (Annex-11 b.). The rest of the parameters where not able to reject the null hypothesis and therefore there was no evidence that those specific parameters significantly changed the microbial communities between samples.

The analysis made by the “envfit” function, showed that salinity concentration significantly shaped the microplankton communities, influencing both prokaryotic (R2=0.6090, p-value=0.048) and unicellular eukaryotic communities (R2=0.6110, p-value=0.041). The samples more affected by salinity were the R1L and R1H, while the rest seems to be influenced by other factors (Figure 20). Additionally, in the unicellular eukaryotic communities (18S rRNA gene dataset) it was also possible to notice that temperature, was also a significant environmental factor (R2=0.9042, p-value=0.032). Once again, this condition shaped more intensely the R1L and R1H samples (Figure 20).

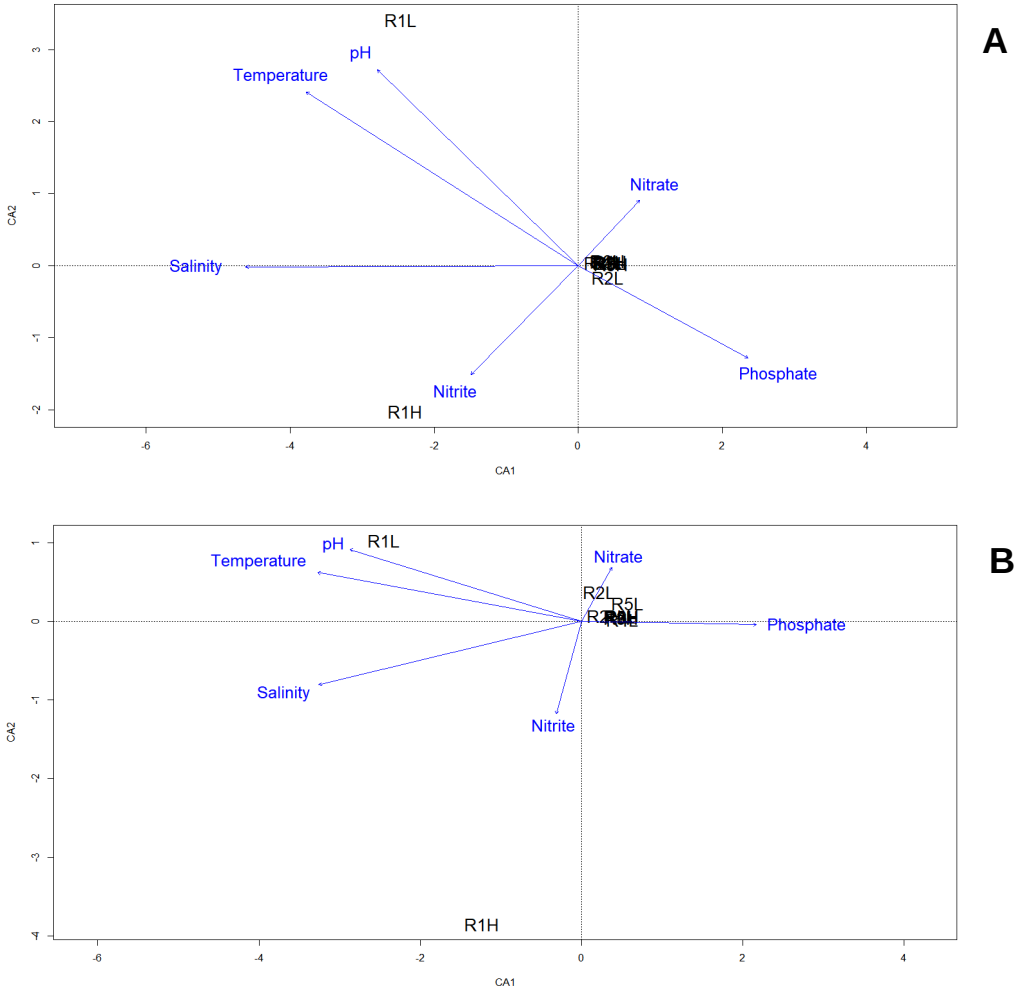


Figure 20: Correlation of the environmental conditions for the 16S rRNA gene (Figure 20a) and 18S rRNA gene (Figure 20b) data sets in a CCA ordination plot.

3.3 Microbiome distribution and water column biogeochemical characterization along the NW coast of Portugal

3.3.1 Biogeochemical Gradients across the NW coast of Portugal

The characterization of the coast´s biogeochemical gradients was displayed in Figure 21. In the north region of the transect was registered the highest water column surface temperature (Figure 21a) at stations S27 (17,68 ºC) and S26 (16,83 ºC). In the south the temperature rose again in the stations S05 (16,35 ºC) and S06 (15,55 ºC). These two high temperature spots were registered twelve miles from the coast, while the lowest values were registered two miles from the coast in the stations S01 (14,21 ºC) and S23 (14,29 ºC). Higher values of water surface salinity were registered in the stations away from the coast (Figure 21b), with the highest value being registered at station S19 (40.9 PPT). The lowest values were confined to the station S15 (30,69 PPT) near the coast, alongside with the sample S28 (32,39 PPT) (Annex-9). In its turn, the highest pH values were registered at the station S01 (9,86) and

station S15 (9,41), both located two miles from the coast. The lowest values were placed at the region north, more specifically at the stations S27 (6,98) and S28 (6,91) (Annex-9). The lowest chlorophyll *a* values were registered at stations S01 (0.26 µg/L), S06 (0,34 µg/L), S07 (0,26 µg/L) and S15 (0.77 µg/L) (Annex-9). All these stations were located near the coast in the southern region of the transect (Figure 21d). The highest chlorophyll *a* concentration was observed at station S28 (5.5 µg/L), near the coast in the north zone. Most of other high chlorophyll *a* were registered in stations away from the coast, like stations S04 (4,67 µg/L) and S18 (3,94 µg/L) (Annex-9).

Regarding the inorganic nutrient concentrations measured, the ammonia (Figure 21e) values were general low, mostly away from the coast and in the north area of the transect ranging from 1,20 µM (S15) to 3,57 µM (S07) (Annex-9). However, it was in the region south near the coast, that the values rise substantially to 7,78 µM at the station S06, 5,80 µM at the station S01, and 4,35 µM at the station S14 (Annex-9). Nitrites values were higher at the stations near the coast and low at the stations away from the coast (Figure 21f). Within the stations near the coast, the highest value was the one registered in the station S23 (0,78 µM). From that station, in every direction, a gradient was formed where the values were increasingly lower. The silica values were high at three well defined places, in the station S23 (9,81 µM), S15 (9,83 µM), S01 (5,60 µM) and S06 (10,21 µM), all of them near the coast (Figure 21g). The lowest values were all registered away from the coast ranging from 0.23 µM (S05) to 1,86 µM (S19) (Annex-9). The nitrates values were also higher near the coast, and lower away (Figure 21h). The higher values were registered in stations S23 (15,06 µM) and S06 (14,68 µM), and the lowest values were away from the coast, with station S27 displaying the lowest value (0.87 µM). Phosphate values were also higher near the coast and lower away from it (Figure 21i). The highest value was registered in the south, at station S06 (1,09 µM). The values remained high near the coast, ranging from 0,75 µM, at stations S01 and S14, to 0,26 µM at station S28. Twelve miles from the coast, the samples S26 and S27, were the ones with the lowest values of phosphates (0.00 µM) (Annex-9).

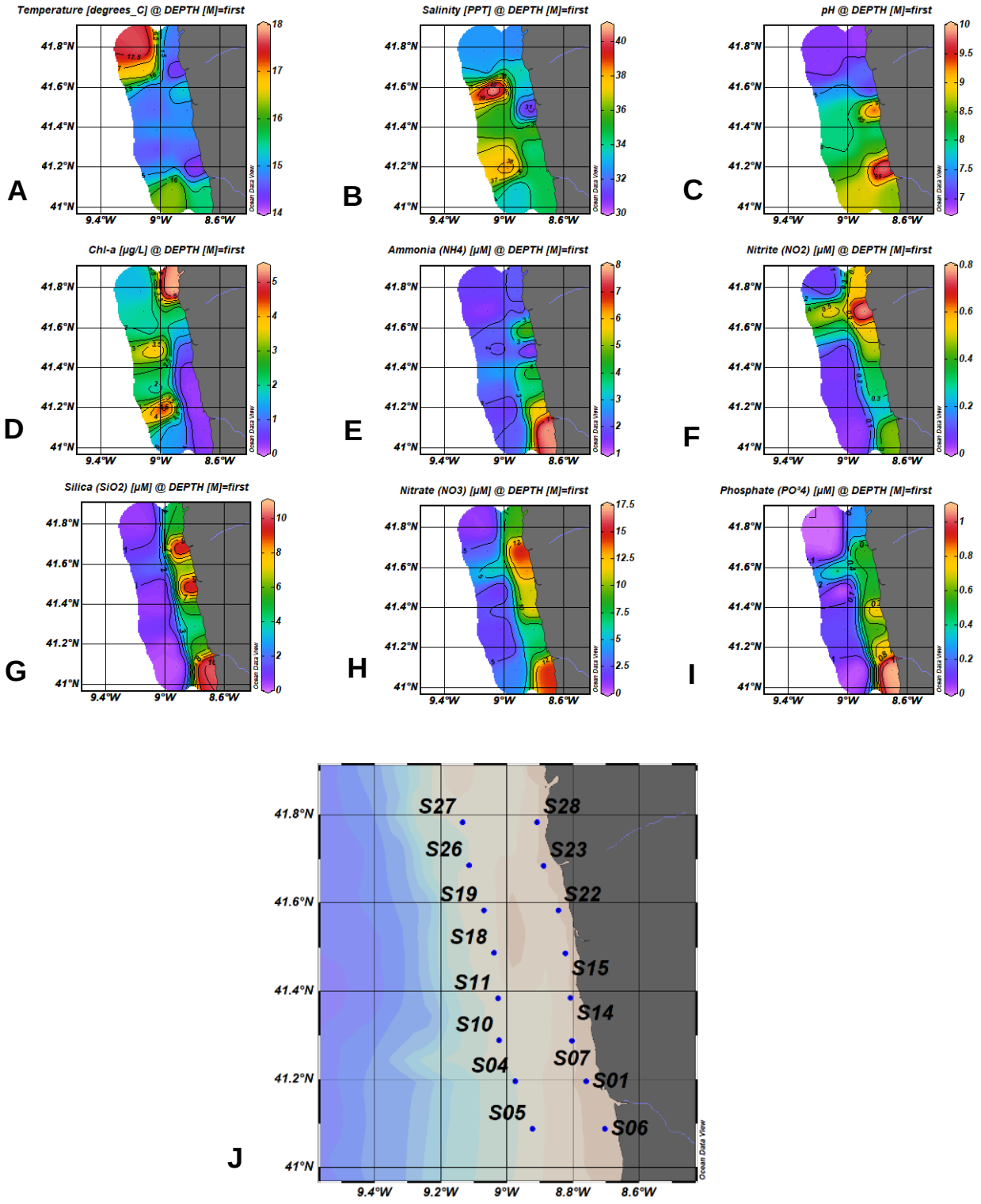


Figure 21: Map of the spatial distribution of the environmental parameters measured (Y-axis), across predefined sampling stations (X-axis) in the Portuguese NW coast. a) Plot of the temperature (Cº) parameter; b) Plot of the salinity (PPT) parameter; c) Plot of the pH parameter; d) Plot of the chlorophyl *a* (µg/L) parameter; e) Plot of the ammonia (µM) parameter; f) Plot of the nitrites (µM) parameter; g) Plot of the silica (µM) parameter; h) Plot of the nitrates (µM) parameter; i) Plot of the phosphate (µM) parameter; j) Map where was represented the location of the different stations.

3.3.2 Planktonic Microbiome gradients along the NW coast of Portugal

The 16S rRNA gene dataset provided a total 324 197 raw sequences, that yield a total of 170 320 high quality sequences, meaning that it was lost 47,46% of the initial raw data (Annex-5). In the 18S rRNA gene dataset, the initial sequences provided by the NGS (New generation sequencing) was 1 276 515. From that, 36,53 % was lost, remaining just 810 162 sequences, these clean sequences were used in the downstream analysis (Annex-6). The rarefaction curves were also done for the 16S rRNA gene rRNA and 18S rRNA gene dataset. As it was possible to notice in the Figure 22, the plateau was reached in every sample, and therefore it was possible to ensure that the biodiversity in all samples was fully covered. However, in the 18S rRNA gene dataset, sample S27 was excluded due to insufficient sequencing depth to describe unicellular eukaryotic diversity.

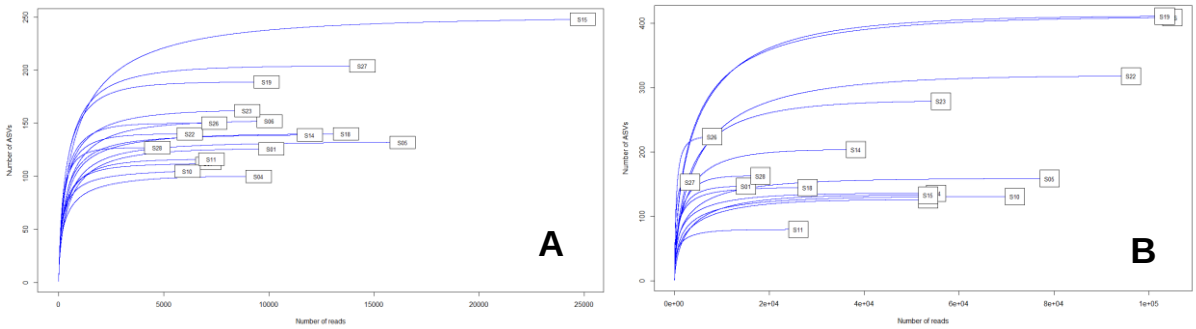


Figure 22: Rarefaction curves for both 16S rRNA gene (Figure 22a) and 18S rRNA gene datasets (Figure 22b) generated from samples from the NW coast of Portugal.

3.3.2.1 Alpha and Beta diversity

The alpha diversity of samples from the NW coast of Portugal was analysed by using the Observed ASVs sequences and the Shannon index (Figure 23 a). The graph showed an overall increase in the diversity, where the lowest Shannon values were registered south, in the stations S01, S04 and S05. From there, the values gradually increase towards the north region, being the samples S19 and S27, both away from the coast, the ones with the highest Shannon value. The analysis of the unicellular eukaryotic communities (18S rRNA gene dataset) (Figure 23 b) demonstrate that the lowest Observed ASVs were present in the station S11. By its turn, the samples S07, S15 and S22, all near the coast, show the lowest Shannon values. On the contrary, the stations S19 and S06 have the highest Observed ASVs, while the sample S26 was the one with higher Shannon value.

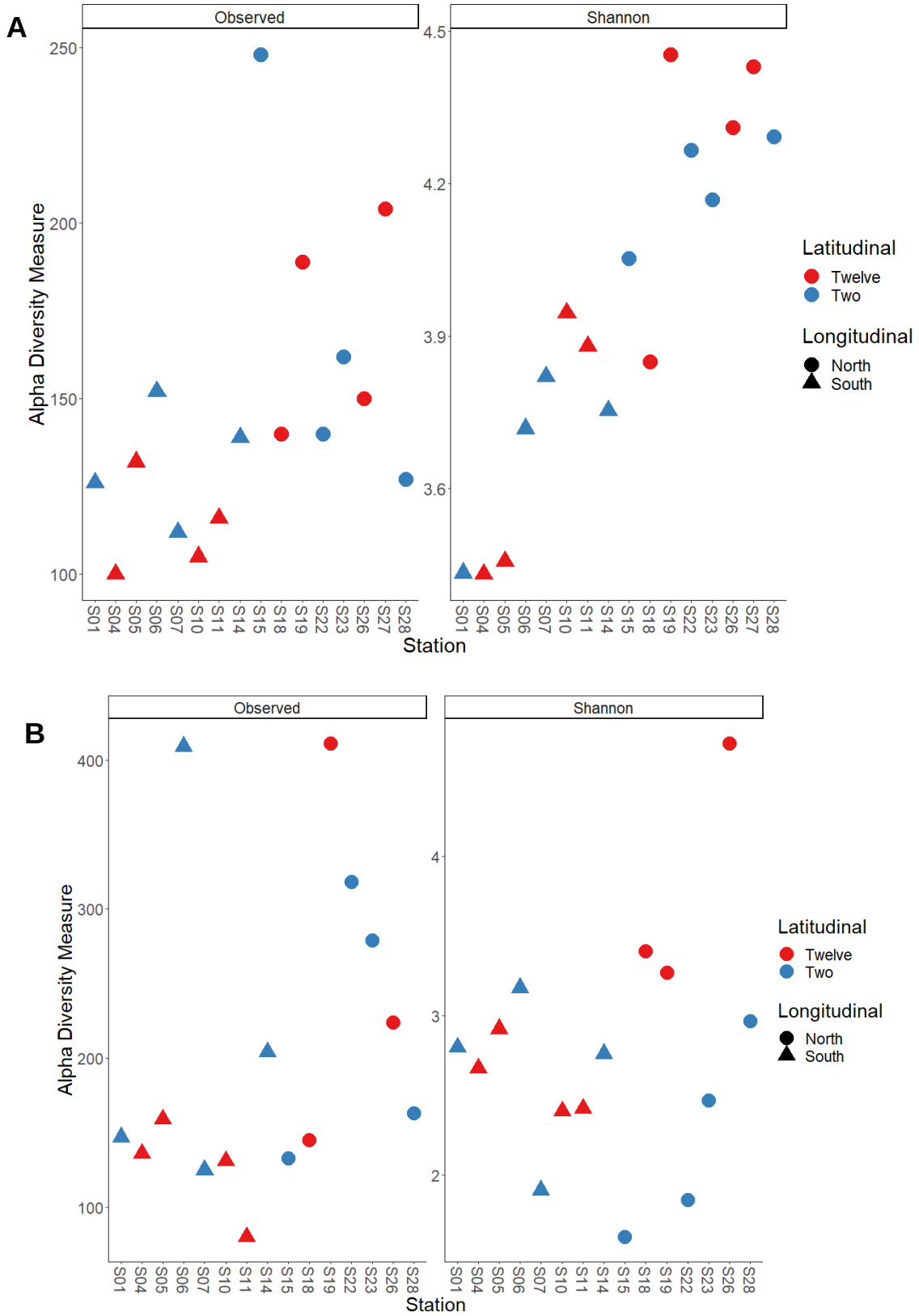
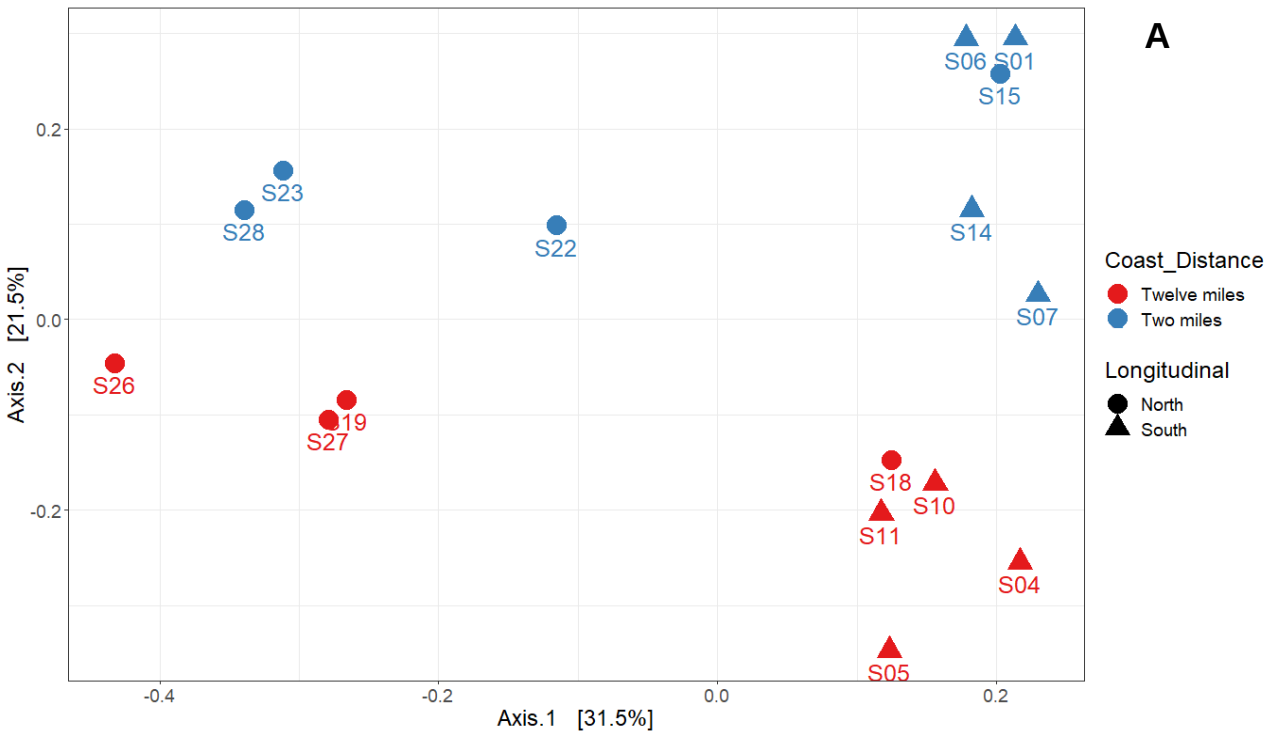


Figure 23: Alpha diversity (number of observed ASVs and Shannon index) of the samples collected in NW coast of Portugal; the samples were coloured (red and blue) according to the distance from the coast (Two or twelve miles from the coast) and different shapes (circle and triangle) where given to the samples from the different longitudinal regions (North and South). The Figure 23a, represents the 16S rRNA gene dataset and the Figure 23b the 18S rRNA gene dataset.

For beta diversity visualization, the PCoA ordination plot and the dendrogram were generated for both prokaryotic and unicellular eukaryotic communities. In Figure 24, it was possible to notice that the prokaryotic communities were separated in two clusters according to their different longitudinal amplitudes. The first cluster was composed by the samples S19 to S28, and the second was formed by the samples S01 to S18. Within each of these clusters was possible to notice two subgroups that were separated based on the distance to the coast (Figure 24b). On the Figure 25, it was possible to visualize that the unicellular eukaryotic communities (18S rRNA gene dataset) from the samples analysed were distributed within two clusters. The first cluster was formed by the samples S01, S06, S14 and S15, and the second cluster formed by the rest of the samples. With this arrangement was possible to notice that there was not a clear longitudinal pattern, like the one seen for the prokaryotic communities. However, when analysing the cluster dendrogram in Figure 25b, the cluster with the samples S01, S06, S14 and S15 was also well defined however, this cluster was more strongly associated with the samples from the stations S04, S05, S07, S10, S11 and S18, than the samples the samples S19, S22, S23 and the S28, all from the north region (Figure 25b).



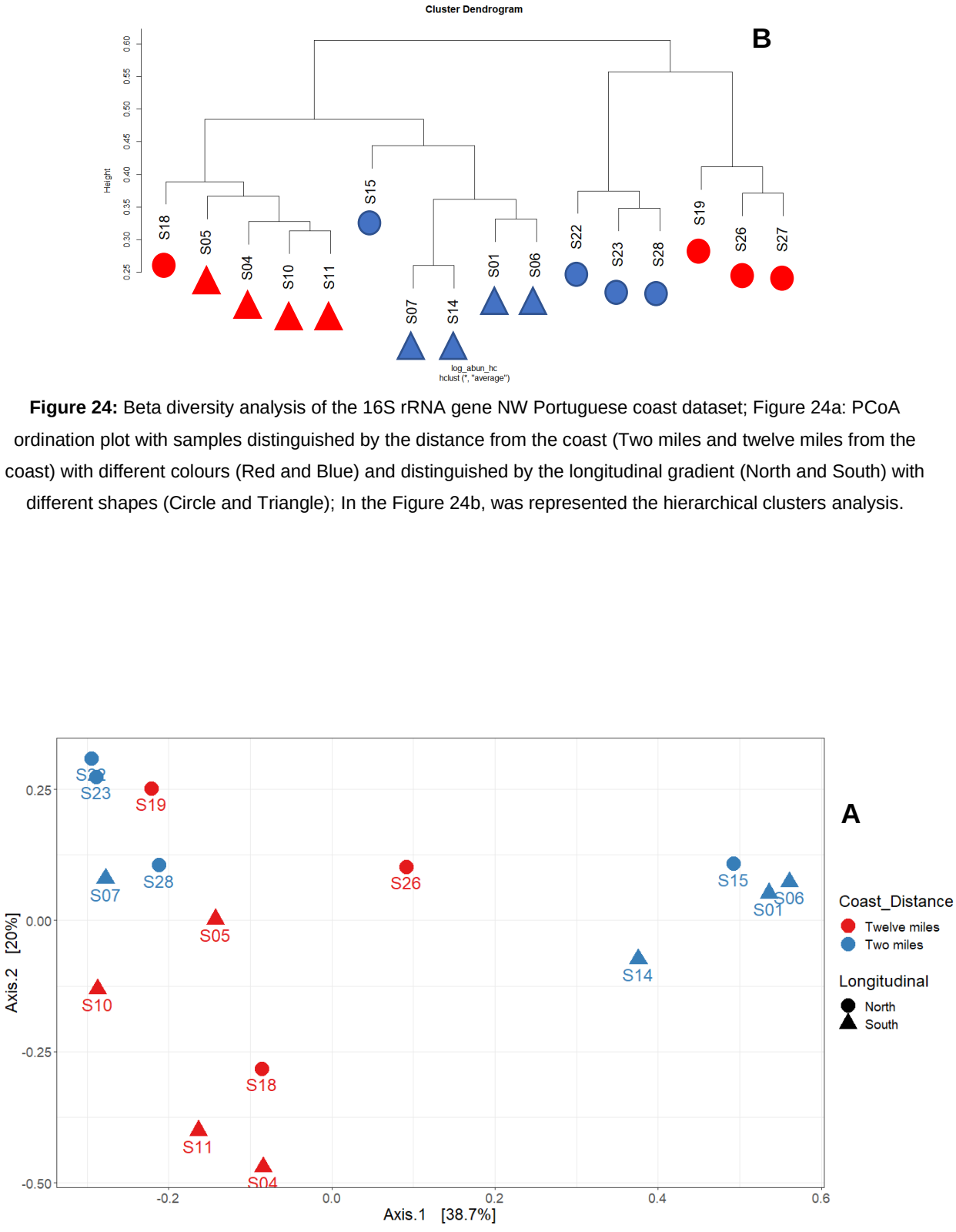


Figure 24: Beta diversity analysis of the 16S rRNA gene NW Portuguese coast dataset; Figure 24a: PCoA ordination plot with samples distinguished by the distance from the coast (Two miles and twelve miles from the coast) with different colours (Red and Blue) and distinguished by the longitudinal gradient (North and South) with different shapes (Circle and Triangle); In the Figure 24b, was represented the hierarchical clusters analysis.

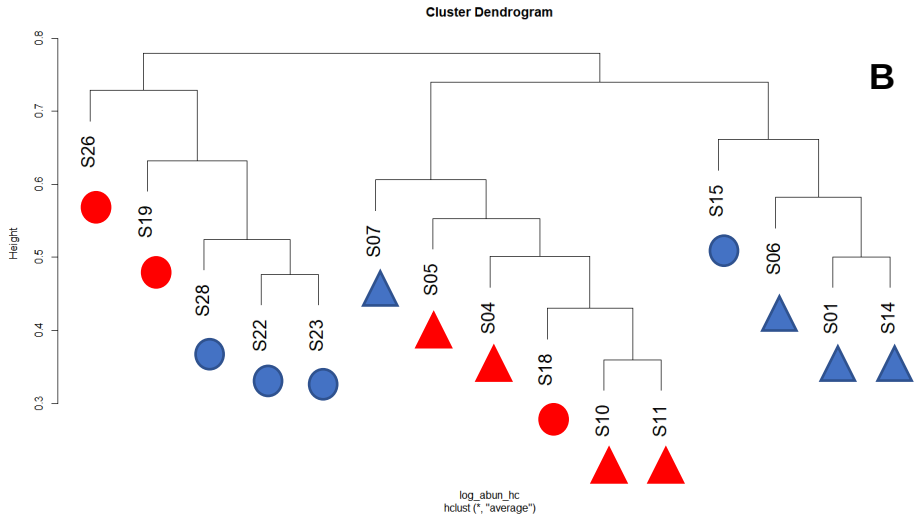


Figure 25: Beta diversity analysis of the 18S rRNA gene NW Portuguese coast dataset; Figure 25a: PCoA ordination plot with samples distinguished by the distance from the coast (Two miles and twelve miles from the coast) with different colours (Red and Blue) and distinguished by the longitudinal gradient (North and South) with different shapes (Circle and Triangle); In the Figure 25b, was represented the hierarchical clusters analysis.

3.3.2.2 Taxonomic composition

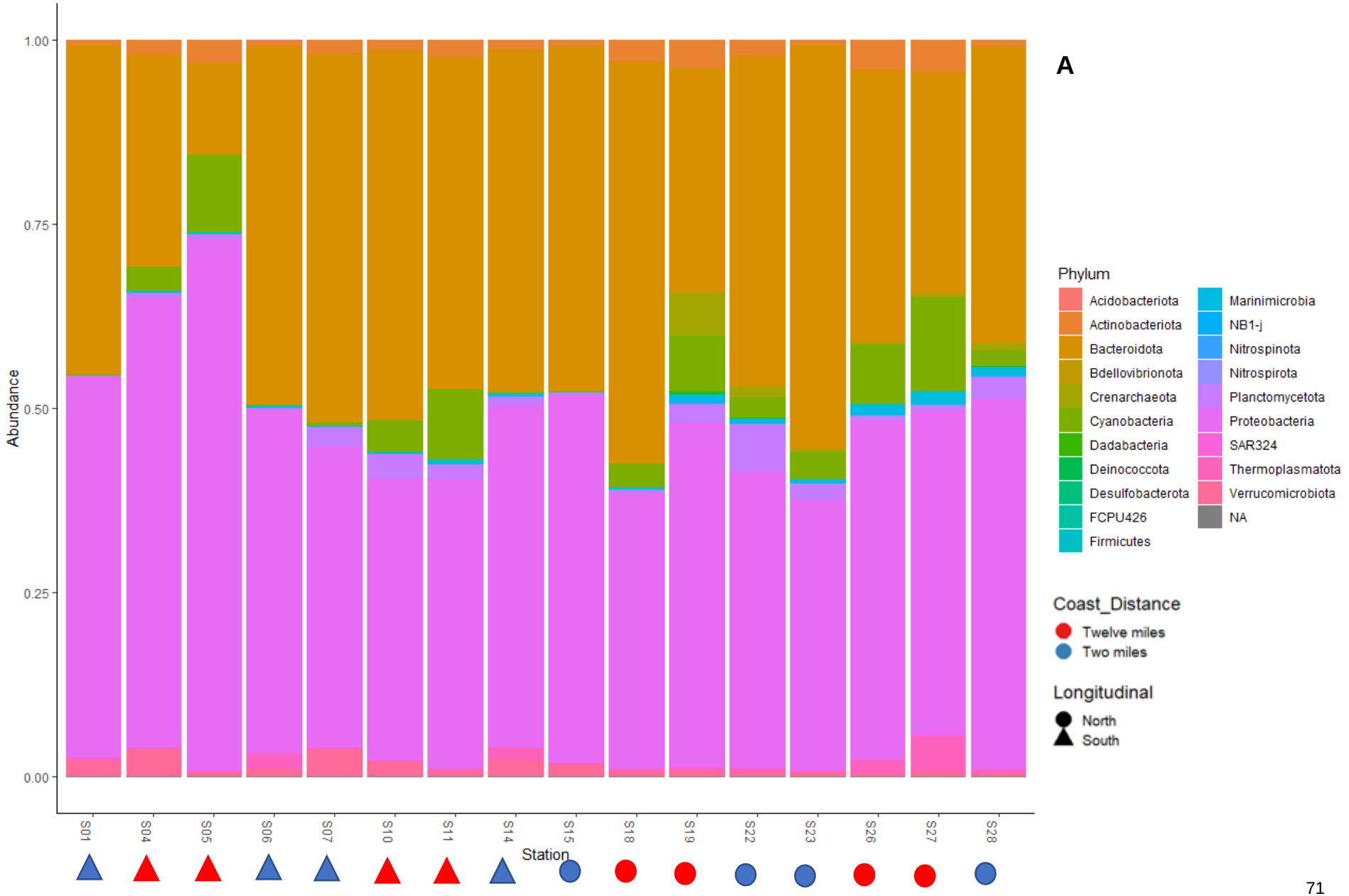
The taxonomic composition of the prokaryotic communities (Figure 26a) consisted mainly of the phyla Proteobacteria and Bacteriodota. The samples with higher abundance of Proteobacteria were the S05 and S15, and the one with higher abundance of Bacteriodota was the sample S18. When the phyla relative abundance profiles were analysed more closely, it was possible to notice on the stations closer to the coast were characterized by lower relative abundance of Cyanobacteria than the stations located 12 miles from the coast. Additionally, the samples with higher Cyanobacteria levels were the samples S27 and S05. It was also possible to notice that between the sample S14 and S28, the levels of Marinimicrobia increase since they were almost inexistent tin the rest of the samples. The phyla Actinobacteriota and Planctomycetota do not show a specific pattern, however the Verrucomicrobiota seems to be more abundant from S01 to S15, than in the north regions. For the unicellular eukaryotic communities (18S rRNA gene dataset) (Figure 26b), the most abundant phylum in all samples was the Alveolata, being the sample S15, almost exclusively composed by this phylum. The phylum Stramenopiles was also very abundant on the samples S10, S11, S04 and S28. The phyla Rhizaria and Hacrobia did not show a specific pattern however, the Hacrobia phylum was very abundant in sample S26. The phylum Opisthokonta was more abundant on the samples more south (From S01 to S18) and less as the transect moves north (S19 to S28).

3.3.2.3 Environmental drivers on microbiome distribution

The PERMANOVA results, the prokaryotic community's distribution along the NW coast of Portugal were also strongly influenced by environmental parameters like pH (R2=0.2290, p-value=0.001),

ammonia (R2=0.1218, p-value=0.039), nitrite (R2=0.1830, p-value=0.003), nitrate (R2=0.1308, p-value=0.043), and phosphate (R2=0.1545, p-value=0.011) (Annex-12 a). Additionally, for the unicellular eukaryotic communities, showed that the environmental parameters pH (R2=0.1474, p-value=0.020), chlorophyl a (R2=0,1351, p-value=0.028), silica (R2=0.1479, p-value=0.026) and phosphate (R2=0.1519, p-value=0.023), were statistical significantly in shaping these group of microbial communities (Annex-12 b.) All these parameters where able to reject the null hypothesis and therefore there was evidence that the coastal microbiome community changed between samples, with the different concentration of the specific nutrients reported above.

In the “envfit” analysis, some of the environmental parameters showed a correlation with the biological data. In the 16S rRNA gene dataset the pH (R2= 0.5161; p-value= 0.009), the nitrites (R2= 0.3423; p-value= 0.042), phosphate (R2= 0.3986; p-value= 0.030), and silica (R2= 0.4842; p-value= 0.007) were the most important parameters shaping variability in prokaryotic community distribution. The pH and phosphate seem to influence more the stations S01, S05 and S06 (Figure 27a). In the 18S rRNA gene dataset the parameters most significant correlated were the nitrates (R2= 0.5923; p-value= 0.004), the phosphate (R2= 0.5636; p-value= 0.010), and silica (R2= 0.6247; p-value= 0.002). Additionally, in Figure 27b was possible to notice that these factors were influencing more the stations S01, S06, S14 and S15 (Figure 27b).


A
Abundance
1.00
0.75
0.50
0.25
0.00
S01
S04
S05
S06
S07
S10
S11
S14
S15
S18
S19
S22
S23
S26
S27
S28
Station
Phylum
Acidobacteriota
Actinobacteriota
Bacteroidota
Bdellovibrionota
Crenarchaeota
Cyanobacteria
Dadabacteria
Deinococcota
Desulfobacterota
FCPU426
Firmicutes
Marinimicrobia
NB1-j
Nitrospinota
Nitrospirota
Planctomycetota
Proteobacteria
SAR324
Thermoplasmatota
Verrucomicrobiota
NA
Coast_Distance
Twelve miles
Two miles
Longitudinal
North
South

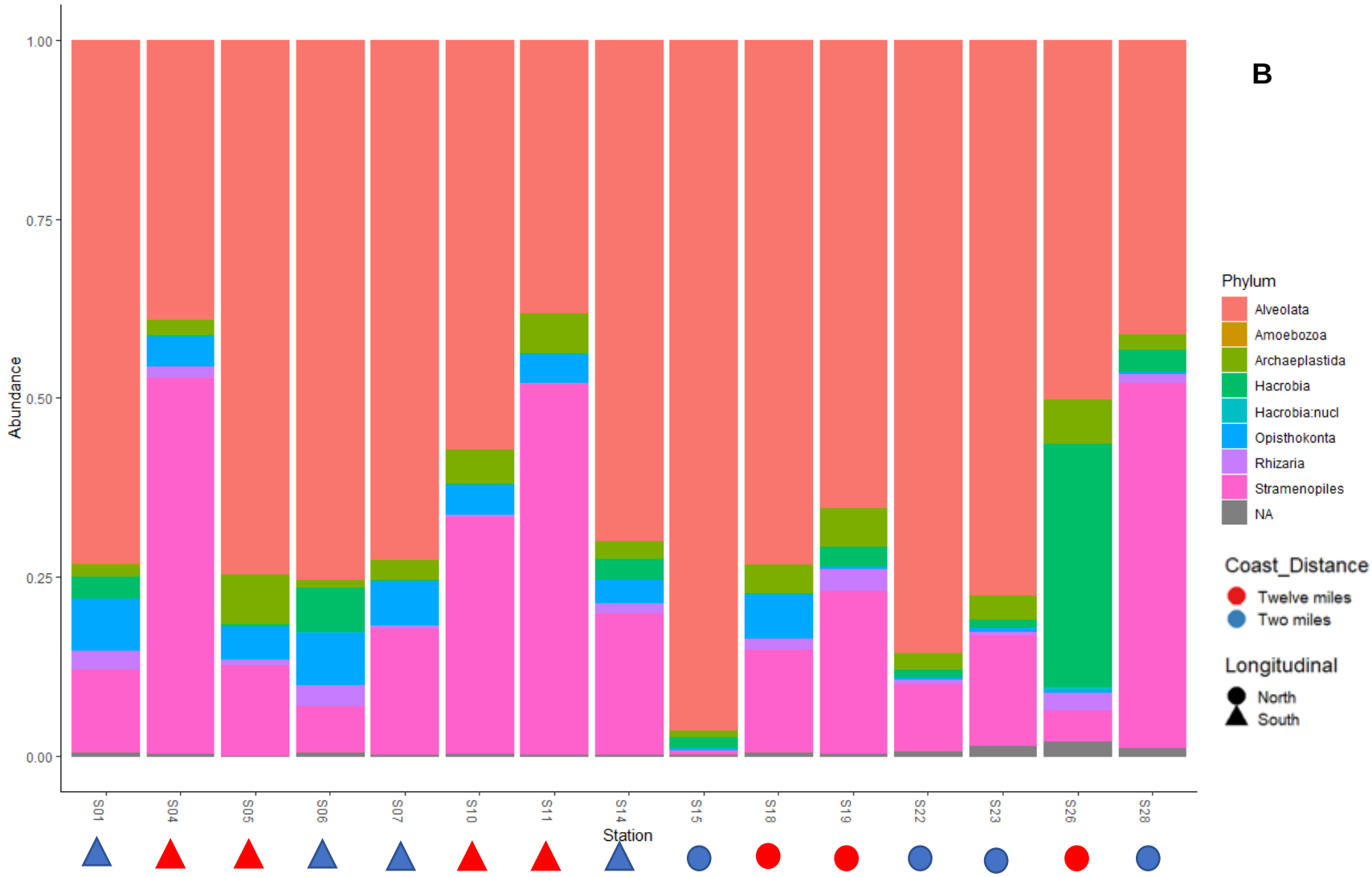


Figure 26: Distribution of the relative abundance of the different prokaryotic (Figure 26a) and unicellular eukaryotic (Figure 26b) phyla along the different NW Portuguese coast stations.

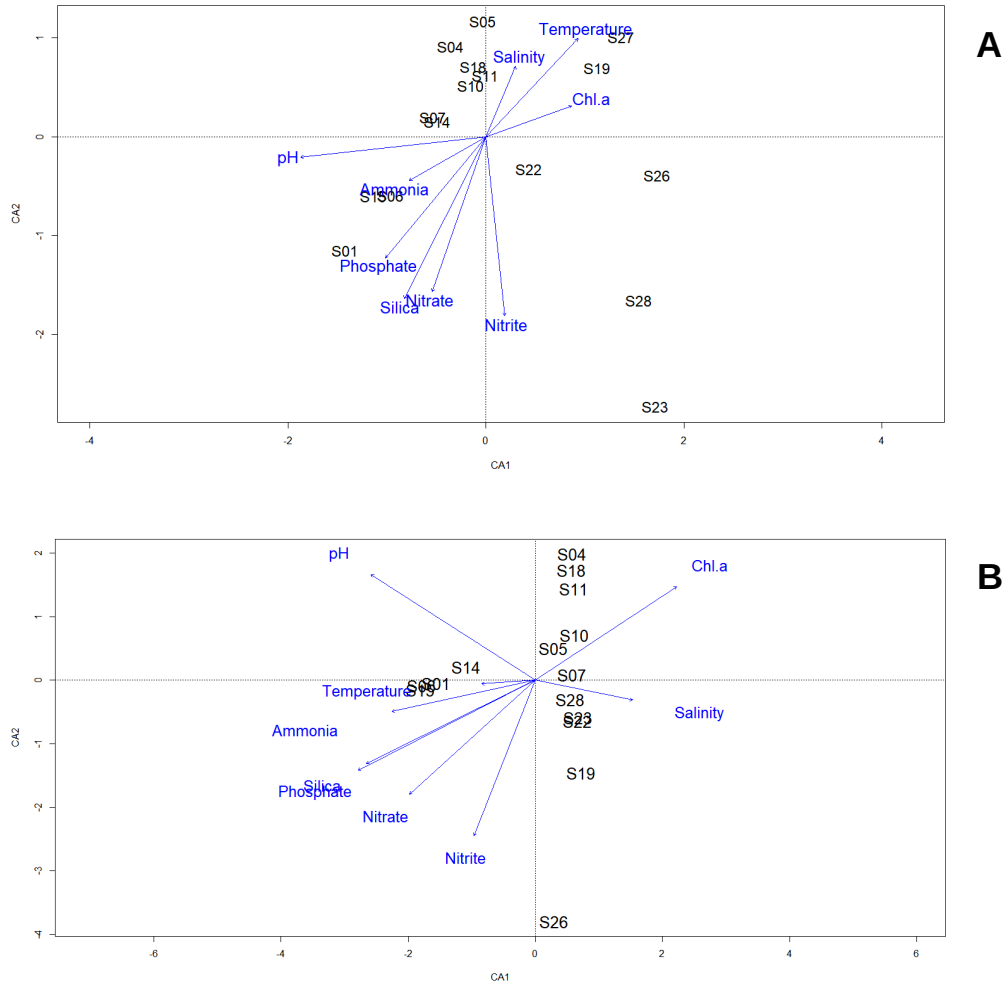


Figure 27: CCA ordination plot for the 16S rRNA gene (Figure 27a) and 18S rRNA gene (Figure 27b) datasets. The length of the environmental arrows was indicative of the degree of influence of the samples.

4. Discussion

4.1 Microbial communities' distribution across the Douro estuary

4.1.1 *Estuarine Microbiome Horizontal distribution*

The prokaryotic communities (16S rRNA gene dataset) differences driven by distance between stations, were assessed by PERMANOVA. Form this analysis, it was understood that the distance between stations along the Douro estuarine transect do not significantly affect the prokaryotic communities (R^2=0.1983 p-value=0.605) (Annex-10 a). Such results support the theory that in a small transect where the distance between the stations was small, distance itself was not the main significant factor that drives the prokaryotic diversity[52]. Although not supported statistically, the alpha diversity analysis showed a particular spatial pattern for the prokaryotic microbiomes. Additionally, in the 2016 Douro campaign, it was possible to notice that the upstream and downstream regions were clustered separately. This

division lead us to conclude that prokaryotic communities were influenced differently by the upstream and downstream regions, although without statistical support.

By its turn, the unicellular eukaryotic (18S rRNA gene dataset) microbial communities showed significant differences across the different stations. These findings were supported by the PERMANOVA analysis, that showed that unicellular eukaryotic from different stations were statistically significant (R^2=0.306, p-value=0.043) (Annex-10 b). Supporting these results, the beta diversity showed a clear distinction between the upper and lower areas of the estuarine transect. These results lead to the conclusion that the distance between the estuarine stations significantly influence the distribution of unicellular eukaryotic microbial communities. This means that the similarity of the unicellular eukaryotic communities decreases as geographic distance increases, even between the small distances where the samples were collected along the Douro estuarine transect. The same conclusions were obtained by Liu et al. (2013)[136], arguing that these findings could be explained by the better dispersal ability that bacteria possess, since they were smaller than microbial eukaryotes. Additionally, since eukaryotes were more complex than prokaryotes, they tend to get their appendages and feeding apparatuses damaged more frequently when carried by the watercourses[137,138]. This will result in lower rates of colonization and therefore the dispersal of these communities may be constrained to a small geographic area.

As discussed above, in the case of the 16S rRNA gene dataset the environmental parameters temperature, nitrates, and silica significantly contributed to differences in the prokaryotic communities across the estuarine transect. In the case of the 18S rRNA gene dataset the environmental parameters that significantly differentiate the community structure across the estuarine transect were the pH, temperature, salinity and the concentrations of ammonia, nitrites, and nitrates. From these results it was possible to understand that the unicellular eukaryotic microbial communities were influenced by a higher number of environmental conditions than the prokaryotic communities. This can be mainly explained by the fact that eukaryotes were more complex than prokaryotic cells, meaning that they will require more resources (e.g., nutrients) to perform their functions[139]. Among these parameters, salinity was known to strongly influence the spatial abundance and composition of the prokaryotic and unicellular eukaryotic communities across the environmental gradients of an estuarine transect [140–142]. These results were not completely aligned with ours since, salinity only significantly shaped the unicellular eukaryotic communities and the 2016 tidal microbiome dataset, across the Douro estuarine transect. The salinity influence in the surface communities can be explained by the distribution of the microorganism according to their salinity sensitivity. A specific example was found by Bouvier et al. (2002) [143], showing that salinity levels above 17 PPT will decrease the levels of zooplankton and therefore increase the phytoplankton abundance [144], shaping the communities.

Additionally, parameters like the availability of ammonia, nitrates, nitrites can also influence the microplankton communities' distribution[145]. In our study, the nitrogen forms like nitrates and nitrites were mainly present in the upstream regions. Zhou et al. (2021) [140], noticed similar patterns in the Pearl River estuary and Azevedo et al. (2003)[146] in the Douro estuary. In our study, it was observed that the river water flow was a main source of nitrates, nitrites, phosphates, and silica to the upstream regions. Such

nutrient distributions can be explained by the nitrogenous fertilizers that runoff from agricultural lands. In its turn, the downstream regions that have lower influence from the upstream will display a higher salinity, and a higher concentration of ammonia. Azevedo et al. (2003)[146] also reported an increase ammonia concentration in the downstream regions, suggesting that such nutrient distribution can be explained by the direct discharges of organic matter from municipal waste waters.

4.1.2 *Estuarine Depth Distribution*

The depth distribution of the microbial communities in the Douro estuary was also object of study. PERMANOVA analysis showed that depth do not significantly influence prokaryotic (R^2=0.035, p-value=0.652) and unicellular eukaryotic communities (R^2=0.046, p-value=0.489) (Annex-10) distribution in the Douro estuary. To corroborate these results, the diversity metrics analysis did not show any relevant pattern regarding depth. Similar results were found by Seguro et al. (2015)[147] in the Gulf of Nicoya. However, the influence of depth was noticed by Ohore et al. (2022)[148] in the estuary of the Rongjiang River. Such results were expected where the distance between the surface and the bottom were more relevant. Depth has more influence in vertical microplankton distribution when significant environmental variation in nutrients availability, pressure, and temperature are present.

Nevertheless, there were other factors that can influence the vertical microbial distribution. From previous studies, it was known that the Douro River was stratified mainly from middle to lower estuary [146]. This stratification was mainly influenced by salinity and water flows patterns that were regulated by the tide's regimes [146]. The present results were obtained during high tide when the saltwater entered more strongly in the Douro estuary. This phenomenon promoted the stratification further into the estuary. In our results it was possible to notice a higher stratification patterns in the stations D5 and D7, since such variability was noticed in the salinity, silica, nitrates, and nitrites parameters. This means that, although not statistically supported, in the middle of the estuary, the communities were more influenced by depth since a stronger environmental gradient between surface and bottom was registered. This can be also explained in part due to the higher depth registered in these stations, (D5=12.92 meters, D7=11.45 meters), while the others range between 7.428 meters (D1) and 8.026 meters (D9). The stations closer to the mouth of the estuary (D1 and D5), and the ones closer to the Crestuma (D9 and D11) showed a more homogeneous environmental distribution across depth and, in part because of that, the communities were also more homogeneous. Regarding chlorophyll a distribution, it was expected lower values in the bottom, due to lower levels of light penetration and phytoplankton biomass[149]. This trend was generally confirmed across all the transect, mainly in the Autumn, at stations D1, D7, D9 and D11.

4.1.3 *Influence of Estuarine Tides*

The influence of tides in the surface microbial communities of the Douro estuary showed to be not significant when a PERMANOVA analysis was performed, for both prokaryotic (R^2=0.061, p-value=0.69)

and unicellular eukaryotic communities (R^2=0.062, p-value=0.764) (Annex-11). The same results were obtained by Aquino et al. (2014)[150] in the Capibaribe River estuary.

Nevertheless, there were some trends that can be discussed. Mainly, it was possible to demonstrate, as expected, that tides influence more strongly the surface communities near the mouth estuary, especially in the stations R1 and R2, for both unicellular eukaryotic and prokaryotic communities. This tide instability in the downstream region of the estuary was supported by the beta diversity analysis since samples from the same station collected at high and low tide cluster very distantly. Additionally, on these stations it was also possible to notice higher alfa diversity values, especially for the prokaryotic communities. Studies showed that in general, the planktonic communities can have a positive effect in the diversity levels due to the high physiological adaptability to salinity fluctuations and the lack of inter-specific competition, corroborating our results[151]. Additionally, these results can also be explained by the entry and exit of water masses between tides at R1 and R2 stations. At high tide the intrusion of the saltwater masse from the ocean into the estuary was occurring. At low tide the water from the ocean cannot penetrate as deep into the estuary, being registered a less extent of saltwater intrusion. This event causes the intrusion of halo-tolerant communities at high tide from the ocean, and at low tide the community changes to a more limno-tolerant [152,153]. These changes were mainly caused by salinity concentration [154]. Knowing this, it was possible to understand that the microbial populations in the estuary mouth were composed by a mix of river and coastal communities [155].

Understandably, tides don't have major effects on the upstream communities of the Douro estuary. This can be explained, because the influence of tides was not significant in this region and therefore, the environmental conditions remain more homogeneous, not influencing microbial community's structure at the surface. Nevertheless, tides can have a stronger influence at the lower and middle part of the estuary at the bottom, as discussed previously in section 4.1.2.

4.1.4 *Estuarine Microbiome Distribution with Season*

The prokaryotic communities were also shaped significantly by seasonality. This claim was supported by the alpha diversity analysis, since the two seasons showed different patterns, and also by beta diversity analysis, where the prokaryotic communities from Autumn and Winter were found to be dissimilar forming two separated clusters. To increase even further the robustness of the results, PERMANOVA was performed, yielding a significant statistical value (R^2=0.2928, p-value=0.002) (Annex 10 a), corroborating the influence of the seasons in structuring planktonic prokaryotic microplankton. In agreement, the unicellular eukaryotic communities also showed that season was a statistically significant since the PERMANOVA yield a p-value bellow 0.005 (R^2=0.2523, p-value=0.001) (Annex 10). Nevertheless, the alpha and beta diversity analysis, performed for the unicellular eukaryotic dataset, showed small differences and less clear clusters differentiation between seasons, when compared with the prokaryotic communities.

The overall conclusions were supported by the results presented by Staroscik et al. (2004)[156], since he reported a strong seasonality in the bacterioplankton community in the temperate Narragansett Bay. In

this study, the bacterial communities were at their pikes of abundance at spring and fall. In our study, the expected decrease in the abundance from Autumn to Winter was not clearly noticed. In the USA, the bacterial communities in the temperate Delaware Estuary were also influenced by seasons, as it was described in the study published by Osterholz et al. (2018)[157]. The results published in this study showed that the microbial community were more dissimilar between August and November than between communities from the same month. Even in the tropics, the estuarine microbial communities showed to be influenced by seasons. Kaestli et al. (2017)[158] , in the tropical Darwin Harbor estuary, reported that the dry and wet seasons had an influence on the bacteria composition. In the monsoonal wet season, the periodicity of the rainfall decreases the salinity levels, while in the dry season the temperature decreases and the salinity levels rise due to low rainfall [158]. Similar results were taken by Reis et al. (2019)[159] in the Curuperé estuary, since the parameters pH, temperature, rainfall, and salinity were also related to the community's seasonal variability. Zhou et al. (2021)[140] found, in a highly urbanized Pearl River estuarine, a strong influence of seasonality, more than the spatial influence, in the microbiome distribution. In his study, it was found an increased diversity value in the dry seasons, when compared with the wet season. Additionally, they also found a higher diversity fluctuation in the dry season than in the wet season [140].

Although seasonality was shaping the communities, the seasons were not a condition by themselves. Ultimately seasons were defined by the periodic changes in the environmental conditions which were influencing the changes in the structure of microbial communities. In this study, it was possible to notice that the majority of the environmental parameters registered have a degree of seasonality change. Conditions like temperature, pH, nitrates, phosphates were at higher levels in Autumn while nitrites and silica were at higher levels in Winter. The higher concentration of nutrients in Autumn can be due to the low rainfall that was registered in this period, which led to a decrease in the river flow and consequently the decrease of the dilution of the nutrients in the estuary[136,140]. The higher concentrations of nitrates, nitrites, ammonia, and phosphates can be an inhibitor and promoter of microbial community development. On the one hand, it inhibits the microorganisms that were too sensitive to the high concentrations of a particular nutrient and, on the other hand, was a promoter of organisms that were adapted to consume that nutrient effectively[160]. Additionally, temperature also plays an important role in shaping the microplankton communities because each organism has its own optimal temperature range that, if met, can lead to a considerable increase in microbial activity[156,161,162]. Since different organisms have different optimal temperatures, the composition of a specific community can change according to this parameter [163,164]. The temperature distribution was highly seasonal meaning that it will influence the community's structure in a seasonal pattern, especially in the temperate regions.

Additionally, seasons can also have huge importance on tidal regimes. Azevedo et al. (2008)[146] study such interplay in the Douro estuary. The author found that in summer the flow was generally decreased due to low rainfall and, it was expected that the salinity increases in the estuary, as well as the light availability and therefore, temperature increases. This phenomenon will promote the decline of chlorophyll *a,* and phosphate [146]. This can be explained mainly because the freshwater phytoplankton was incapable to grow in such saline waters that extended across the estuary in this period[146]. As the

Autumn arises, the salinity gradients decrease, and the phytoplankton can become more abundant since salinity decreases. At Winter, the chlorophyll *a* decreases probability due to low phosphate that was registered in the Winter at the downstream regions. This phenomenon was more prominent at the upstream regions since they were more propitious to salinity fluctuations due to river flows or tide movements[146].

4.2 Spatial Microbiome Distribution Across the NW Coast of Portugal

The diversity metrics support the fact that the longitudinal and latitudinal gradients of the NW coast of Portugal influence greatly the prokaryotic community's distribution. The alpha diversity showed a clear pattern towards the north region of the coastal transect. The beta diversity also showed a strong longitudinal and latitudinal division of the microbial communities. To support these findings, PERMANOVA was carried out, where it was confirmed that indeed, the Coast Distance (R^2=0.1585, p-value=0.011) and the longitudinal gradients (R^2=0.1691, p-value=0.008) have a significant statistical impact on the structure of prokaryotic communities' assemblages (Annex-12 a). Contrary to what was observed for the prokaryotic communities, the unicellular eukaryotic communities did not show the same robust patterns in diversity metrics. Strengthening these observations, PERMANOVA outcomes did not reveal any significant influence on the community's distribution across the different samples caused by coastal distance (R^2=0.1217, p-value=0.058) or longitudinal gradients (R^2=0.1032, p-value=0.150) (Annex-12 b). The main findings that can be taken from these results were that the prokaryotic communities were much more influenced by distance-decay relationships than the unicellular eukaryotic communities[52].Similar conclusions were drawn by Fortunato et al. (2012) [155] and Du et al. (2013) [165] , in the Oregon to Washington coast and South China coast, respectively, showing that indeed, robust prokaryotic spatial patterns can be found across a coastal transect. Contrary to our results, Sporng et al. (2020)[166] reported a clear spatial pattern in the eukaryotic communities across the coastal German Bight. These strong spatial patterns were commonly explained by the hydrological processes like the upwelling, the magnitude of mixing by river flow, the winds regimes, the fluctuation of tides, etc. Because of that, nearshore coastal communities sampled at surface (<35 km offshore) were more heterogeneous, than surface populations sampled 60 km from the coast, being these communities more homogeneous [167]. In other words, it was more easily noticed, horizontal gradients of microbial diversity and communities' structure along the coast, than in the open sea[167]. These differences between coast to offshore, happened gradually mainly caused by the increase of hydrodynamic gradients and the impact of offshore waters with the proximity of the coast[168]. The physical processes were also one of the explanations for the variation in the coastal bacterioplankton found from the Oregon and Washington coasts, given by Fortunato et al. (2011) [167].

The physical processes were also tightly related with the environmental gradients across the coastal areas. The PERMANOVA results, for the unicellular eukaryotic communities, showed that the environmental parameters pH (R^2=0.1474, p-value=0.020), chlorophyl *a* (R^2=0,1351, p-value=0.028), silica (R^2=0.1479, p-value=0.026) and phosphate (R^2=0.1519, p-value=0.023), were statistical

significantly in shaping the unicellular eukaryotic communities. Additionally, the prokaryotic communities were also strongly influenced by environmental parameters like pH (R^2=0.2290, p-value=0.001), ammonia (R^2=0.1218, p-value=0.039), nitrite (R^2=0.1830, p-value=0.003), nitrate (R^2=0.1308, p-value=0.043), and phosphate (R^2=0.1545, p-value=0.011). This means, that biogeochemical environmental parameters had a crucial role in shaping the planktonic microbiome, however the unicellular eukaryotic and prokaryotic microbiomes were influenced differently, mainly due to their metabolic differences[139].

Further analysis of the environmental parameters was done and, it was possible to notice specific environmental gradients along the coastal transect. For instance, the salinity presented higher values in stations located twelve miles from the coast. This can be explained due to the influence of the river discharges that dilute the salinity concentration[169]. In general, salinity was often considered an significant parameter in shaping the coastal communities [52] .In a study performed by Fortunato et al. (2012)[155], where it was sampled the coastal environment between Oregon and Washington, it was concluded that salinity contributes strongly to physically separate water masses, and therefore strongly influencing the microbial communities with depth [155,170]. However, in our study, coastal samples were only collected in the surface of water column, and thus the influence of the water column depth gradient, caused by the different water masses, was not considered.

In its turn, the temperature patterns described above can be linked to the high complex upwelling systems in the NW coast of Portugal[169]. Temperature was also tightly connected with light delivery, that by its turn, light, was also known to promote chemical and biochemical processes that affect the diversity of the coastal microbiomes[162]. As a result, the primary production on the coastal areas increases, leading to an increase in diversity and productivity [165]. When the conditions were optimal, the production of organic carbon exceeds the consumption of the herbivores and it becomes available to the detritus organism, resulting in an increased diversity level [171].

Additionally, was also possible to notice an increase of ammonia, nitrites, nitrates, silica, and phosphates availability in stations located two miles from the coast. These values were closely related to the river discharges that happened more strongly at the station S06 in the mouth of Douro River, S14 in the mouth of Ave River, between S15 and S22 in the mouth of Cávado River, and at the S23 in the mouth of Lima River. Douro River has a huge influence on the levels of silica, nitrates, ammonia, and phosphates, which were discharged into the coastal area. By its turn, the Lima River was discharging also great quantities of nitrates, nitrites, and silica. Additionally, important changes in the pH were noticed near the Leixões harbour. Similar nutrients patterns were obtained in the Pearl River coastal discharging areas[172] and in the Oregon coast[173], corroborating the huge influence of river plumes in the coastal areas.

4.3 Microbial Taxonomic Composition Across Coastal and Estuarine Water Masses

The general taxonomic profiles of prokaryotic communities reported in this thesis were similarly described in analogous environments by Du et al. (2013)[165], Zhang et al. (2021)[174], Nolde et al. (1998)[175], Azevedo et al. (2008) [114], Yi et al. (2020)[176], Farnelid et al. (2016)[177], Wang et al. (2020)[178].

The Proteobacteria taxon was found in every single sample and was the most abundant phylum across all the coastal and estuarine stations[179]. Proteobacteria was one of the most diverse phyla, that can obtain energy from different sources. Some members of the phylum can perform photosynthesis, being often called "the phototrophic purple bacteria and their relatives" [180]. Bacteroidota was also very abundant in the northwest coast of Portugal and in the Douro estuary, being possible to detect worldwide[165]. This phylum was usually characterized by the ability to degrade various biopolymers like chitin, cellulose, and pectin[181]. Additionally, Planctomycetes was also involved in the degradation of different polymeric organic matter [174]. This function was essential for a balanced environment because for instance, when algal blooms appear, these phyla, start to degrade algal complex biopolymers. The degradation of the algal blooms will allow the availability of the organic carbon to Proteobacteria and therefore, promoting the carbon cycle, while at the same time controlling the algal blooms growth[182]. Actinobacteria was also a cosmopolitan phylum, found in both NW Portugal coast and Douro estuary. Although at lower quantities, Verrucomicrobia was also a ubiquitous phylum found across the marine, and freshwater datasets being also known to appear in soil [174].

Cyanobacteria were primarily distributed in the surface waters of both marine and freshwater environment [175], and were detected in all seasons [178]. In our study, Cyanobacteria were detected in all coastal samples, but not on all the estuarine samples. In the estuary, Cyanobacteria levels were higher at the mouth of the estuary and lower at the middle and upstream part of Douro estuary. These findings can be explained by the fact that some Cyanobacteria members increase their numbers as temperature and salinity increases[183]. On the contrary, when salinity starts to decrease, also decrease Cyanobacteria abundance. Regarding seasons, it was also possible to notice a decrease in their levels in the Winter season. This can be explained by the lower temperature and salinity values, when compared with the Autumn. In the costal dataset, primary production depicted by chlorophyll *a* and Cyanobacteria analysis showed generally, higher levels in the stations twelve miles from the coast and increasing numbers towards the region north of the transect. Cyanobacteria also play an important role in the nitrogen cycle, especially in the N fixation, alongside with other phyla like the planctomycetes [184]. Additionally, other phyla like Proteobacteria, Verrucomicrobia, Acidobacteria, Firmicutes, Planctomycetes and some eukaryotes like diatoms, also play an important role in several other steps in the nitrogen cycle [184,185].

The taxonomic profiles of the unicellular eukaryotic communities were mostly represented by the Alveolate and the Stramenopiles phyla since they appear in all samples across a wide range of ecosystems. The Alveolate phylum include the former Dinoflagellata, Ciliophora, and Apicomplexa groups[186]. The Dinoflagellata group was characterized by the high diversity of species in the marine plankton, mostly photosynthetic[187]. The algal forms were known for the formations of the algal blooms and consequently the production of toxins that can harm entire ecosystems, while the non-algal were mainly predators of other microorganisms or parasites[187]. The Stramenopiles phylum consists of a large variety of usually flagellate algae containing complex plastids[187]. The unicellular forms and the colonial

microalgae produce a complex cell covering made of silica. Within this phylum there were also some photosynthetic groups that can cause toxic algal blooms and “red tides” [188].

5. Conclusions and Future Perspectives

The marine and estuarine microbial communities are key players in marine and estuarine ecosystems. However, their community dynamics is still poorly understood particularly in the region of north of Portugal. The present thesis shed the first light on the characterization in terms of distribution, diversity and taxonomy of the prokaryotic and unicellular eukaryotic communities in such relevant ecosystems. The study took in consideration a wide range of environmental variables to investigate the physical and biogeochemical parameters that may influence the distribution and diversity of the prokaryotic and unicellular eukaryotic planktonic communities in a coastal gradient and in an estuarine gradient.

Several preliminary conclusions can be draw from the present thesis. Spatially, the unicellular eukaryotic communities, inhabiting the surface waters of the Douro estuary, were increasingly dissimilar as the distance between stations increases. In its turn, the prokaryotic communities, inhabiting the surface waters of the Douro estuary, showed a less pronounced dissimilarity as the distance between stations increases. In addition, the Douro estuary tides did not significantly influence the microbial communities located in the area of the estuary more influenced by the river. However, the microbial communities from the downstream station were highly influenced by tides showing dissimilar communities at high and low tide. Results from Douro estuary also revealed that depth do not significantly influencing the microbial communities in the Douro estuary at high tide. The microbial communities from the two sampling water column depths (surface and bottom) do not significantly differ in terms of structure, suggesting a homogeneous water column with respect to the prokaryotic and unicellular eukaryotic communities’ distribution. Regarding temporal patterns, Autumn, and Winter seasons significantly shaped microbial communities of the Douro estuary, although prokaryotic communities were the ones that showed to be more influenced with season. Concerning the environmental parameters, it was found that temperature, nitrates, salinity, and pH concentrations, significantly shaped the taxonomic composition of the estuarine prokaryotic communities. In its turn, the estuarine unicellular eukaryotic communities were mainly influenced by temperature, salinity, pH, ammonia, nitrites, and nitrates concentrations.

Preliminary conclusions for the microbiomes that inhabiting the surface waters of the NW coast of Portugal were also taken. Spatially, it was found that the prokaryotic communities were increasingly dissimilar as the distance between stations increases. In its turn, the unicellular eukaryotic communities, inhabiting the surface waters of the NW coast of Portugal, showed a less pronounced dissimilarity as the distance between stations increases. In addition, the taxonomic composition of the coastal prokaryotic communities was mainly influenced by ammonia, nitrates, nitrates, phosphate, silica, and pH concentrations, and the coastal unicellular eukaryotic communities were mainly influenced by pH, phosphate, and silica concentrations.

Regarding the taxonomic profiles, Proteobacteria, Bacteriodota, Actinobacteria, Verrucomicrobia prokaryotic phyla were found as most abundant in every sample in both NW Portuguese coast and in the Douro estuary. While the unicellular eukaryotes taxonomic profiles were mainly characterized by the presence of Alveolata and Stramenopiles phyla across both marine and estuarine environments.

The sampling effort described in the present study is not representative of coastal microbial dynamics since the complex and variable parameters like hydrological and environmental conditions, are constantly changing. Such variability will certainly increase even more due to climate change and pollution. Because of that, it would be important to continue the sampling campaigns for several years since, as the time-series becomes larger, the results will become more robust and representative, and the microbiome dynamics and patterns will be accessed in a more robust data set. To tackle this issue, the OCEAN3R and ATLANTIDA projects are schedule to continue the sampling effort for a few more years, on the Douro Estuary and the NW Portuguese coast, in order to generate a long-term microbiome monitoring data set for both regions. The data generated from these long-term monitoring programs will be highly relevant for future understanding on how seasonal and yearly cycles impact the dynamics of coastal and estuarine microbiomes taking into account the climate change forces.

Additionally, it will be also important to increase the physical, chemical and biogeochemical parameters measured (e.g. dissolved organic nitrogen and carbon, dissolved oxygen, dissolved iron, rainfall, water fluxes dynamics, light incidence) to understand more accurately the environmental controls of the coast microbial dynamics. In the future, for a better understating estuarine tidal dynamic, it is crucial to provide sampling in the same campaign from both surface/bottom and at low/high tides in order to fully understand the stratification dynamics imposed by tides. Additionally, to fully access the microbial dynamics in the Portuguese coastal areas, the study of depth must be included in further monitoring campaigns. Finally, future studies must include functional characterization of these estuarine and coastal microplankton communities by using metagenomic and metatranscriptomics analysis in order to understand the metabolic role of these enormous diversity influence ecosystem functioning and stability.

Overall, the present study disclosed important information about the microplankton that inhabits the coastal region of Portugal, for which there is almost no data available for the microbial community's diversity and distribution. The preliminary result in the present thesis demonstrates that the temporal and spatial patterns, governed by environmental factors, significantly shaped the microbial communities in the north coastal region of Portugal.

The present thesis delivers for the first time new molecular data on the diversity of prokaryotes and unicellular eukaryotes for the Douro estuary and coastal zone associated with it. This research will help to enhance our understanding of such important ecosystems, providing a theoretical foundation for the marine ecological health management, that ultimately will allow the maintenance of the human activities in the coastal and estuarine regions.

6. References

1. Madigan, M. T., Martinko, J. M., Bender, K. S., Buckley, D. H., & Stahl, D. A. (2016). Microbiologia de Brock-14ª Edição. Artmed Editora.

2. Bertrand, J. C., Brochier-Armanet, C., Gouy, M. & Westall, F. For three billion years, microorganisms were the only inhabitants of the earth. in *Environmental Microbiology: Fundamentals and Applications* 25–71 (Springer Netherlands, 2015). doi:10.1007/978-94-017-9118-2_4.

3. Meyer, K. M., Kump, L. R. & Ridgwell, A. Biogeochemical controls on photic-zone euxinia during the end-Permian mass extinction. *Geology* **36**, 747–750 (2008).

4. Hallam, A. *A review of Mesozoic climates*. *J. geol. Soc. London* vol. 142 http://jgs.lyellcollection.org/ (1985).

5. Raymo, M. E. *GLOBAL CLIMATE CHANGE: A THREE MILLION YEAR PERSPECTIVE*.

6. Chris Stringer. Out of Ethiopia. *Nature 423.6941* 693–695 (2003).

7. Ripple, W. J., Wolf, C., Newsome, T. M., Barnard, P. & Moomaw, W. R. *World Scientists' Warning of a Climate Emergency SCIENTIST SIGNATORIES FROM 153 COUNTRIES (LIST IN SUPPLEMENTAL FILE S1)*. vol. 70 https://academic.oup.com/bioscience (2020).

8. Institute of Actuaries, C. *Research Paper: Climate Change and Resource Sustainability - An Overview for Actuaries*. (2015).

9. *Climate Change 2021 Working Group I contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change Summary for Policymakers*.

10. Pörtner, H.-O. *et al. The Ocean and Cryosphere in a Changing Climate A Special Report of the Intergovernmental Panel on Climate Change Edited by*. (2019).

11. Barbier, E. B. *et al. The value of estuarine and coastal ecosystem services*. *Ecological Monographs* vol. 81 (2011).

12. O'higgins, T. G., Ferraro, S. P., Dantin, D. D., Jordan, S. J. & Chintala, M. M. *Habitat Scale Mapping of Fisheries Ecosystem Service Values in Estuaries*. vol. 15 www.ecologyandsociety.org/vol15/iss4/art7/ (2010).

13. Harrison, R. M. Pollution: causes, effects and control. in *harrison2015pollution* (Royal society of chemistry, 2015).

14. Vikas, M. & Dwarakish, G. S. Coastal Pollution: A Review. *Aquat Procedia* **4**, 381–388 (2015).

15. Al-Musharafi, S. K., Mahmoud, I. Y. & Al-Bahry, S. N. Heavy Metal Pollution from Treated Sewage Effluent. *APCBEE Procedia* **5**, 344–348 (2013).

16. *Eutrophication Th e Role of in the Global Proliferation of Harmful Algal Blooms*. www.ut.ee/~olli/eutr/ (2005).

17. Khan, M. N. & Mohammad, F. Eutrophication: Challenges and solutions. in *Eutrophication: Causes, Consequences and Control* vol. 2 1–15 (Springer Netherlands, 2014).

18. van Beusekom, J. E. E. Eutrophication. in *Handbook on Marine Environment Protection* 429–445 (Springer International Publishing, 2018). doi:10.1007/978-3-319-60156-4_22.

19. Thornton, P. K., Ericksen, P. J., Herrero, M. & Challinor, A. J. Climate variability and vulnerability to climate change: A review. *Global Change Biology* vol. 20 3313–3328 Preprint at https://doi.org/10.1111/gcb.12581 (2014).

20. Pepper, I. L., Rensing, C. & Gerba, C. P. *ENVIRONMENTAL MICROBIAL PROPERTIES AND PROCESSES*. (2004).

21. Stal, L. J. & Cretoiu, M. S. *The marine microbiome: An untapped source of biodiversity and biotechnological potential*. *The Marine Microbiome: An Untapped Source of Biodiversity and Biotechnological Potential* (Springer International Publishing, 2016). doi:10.1007/978-3-319-33000-6.

22. Witman, S. World's biggest oxygen producers living in swirling ocean waters. *Journal of Geophysical Research: Oceans* (2017).

23. Pierella Karlusich, J. J., Ibarbalz, F. M. & Bowler, C. Exploration of marine phytoplankton: From their historical appreciation to the omics era. *Journal of Plankton Research* vol. 42 595–612 Preprint at https://doi.org/10.1093/plankt/fbaa049 (2020).

24. Jeffries, T. C. *et al.* Bacterioplankton dynamics within a large anthropogenically impacted urban estuary. *Front Microbiol* **6**, (2016).

25. Hays, G. C., Richardson, A. J. & Robinson, C. Climate change and marine plankton. *Trends in Ecology and Evolution* vol. 20 337–344 Preprint at https://doi.org/10.1016/j.tree.2005.03.004 (2005).

26. José, J., Karlusich, P., Ibarbalz, F. M. & Bowler, C. Phytoplankton in the Tara Ocean. (2019) doi:10.1146/annurev-marine-010419.

27. Abida, H. *et al.* Bioprospecting marine plankton. *Marine Drugs* vol. 11 4594–4611 Preprint at https://doi.org/10.3390/md11114594 (2013).

28. Everett, J. D. *et al.* Modeling what we sample and sampling what we model: Challenges for zooplankton model assessment. *Frontiers in Marine Science* vol. 4 Preprint at https://doi.org/10.3389/fmars.2017.00077 (2017).

29. Fenchel, T. MARINE PLANKTON FOOD CHAINS. *Annu Rev Ecol Syst* **19**, 19–38 (1988).

30. Potter, I. C., Chuwen, B. M., Hoeksema, S. D. & Elliott, M. The concept of an estuary: A definition that incorporates systems which can become closed to the ocean and hypersaline. *Estuar Coast Shelf Sci* **87**, 497–500 (2010).

31. Azam, F., Fenchel, T., Field, J. G., Gray, J. S., Meyer-Reil, L. A., & Thingstad, F. (1983). The ecological role of water-column microbes in the sea. Marine ecology progress series, 257-263.

32. Fenchel, T. The microbial loop - 25 years later. *J Exp Mar Biol Ecol* **366**, 99–103 (2008).

33. Finkel, Z. v. *Does Phytoplankton Cell Size Matter? The Evolution of Modern Marine Food Webs*. (2007) doi:10.1016/B978-0-12-370518-1.50016-3.

34. Pierce, R. W., Turner, J. & Turner, J. T. *Ecology of plankton ciliates in marine food webs Massachusetts Water Resources Authority View project Arctic Microzooplankton View project Ecology of Planktonic Ciliates in Marine Food Webs*. *Reviews in Aquatic Sciences* vol. 6 https://www.researchgate.net/publication/246005560 (1992).

35. Bauer, J. E. *et al.* The changing carbon cycle of the coastal ocean. *Nature* vol. 504 61–70 Preprint at https://doi.org/10.1038/nature12857 (2013).

36. Carlson, C. A., del Giorgio, P. A. & Herndl, G. J. Microbes and the dissipation of energy and respiration: From cells to ecosystems. *Oceanography* **20**, 89–100 (2007).

37. Azam, F. & Malfatti, F. Microbial structuring of marine ecosystems. *Nature Reviews Microbiology* vol. 5 782–791 Preprint at https://doi.org/10.1038/nrmicro1747 (2007).

38. Sanders, R. *et al.* The Biological Carbon Pump in the North Atlantic. *Prog Oceanogr* **129**, 200–218 (2014).

39. Friedlingstein, P., Dufresne, J.-L., Cox, P. M. & Rayner, P. How positive is the feedback between climate change and the carbon cycle? *Tellus B: Chemical and Physical Meteorology* **55**, 692–700 (2003).

40. Friedlingstein, P. *et al.* Positive feedback between future climate change and the carbon cycle. *Geophys Res Lett* **28**, 1543–1546 (2001).

41. Henson, S. A. *et al.* A reduced estimate of the strength of the ocean's biological carbon pump. *Geophys Res Lett* **38**, (2011).

42. Zehr, J. P. & Kudela, R. M. Nitrogen cycle of the open ocean: From genes to ecosystems. *Ann Rev Mar Sci* **3**, 197–225 (2011).

43. Hatzenpichler, R. Diversity, physiology, and niche differentiation of ammonia-oxidizing archaea. *Appl Environ Microbiol* **78**, 7501–7510 (2012).

44. Delwiche, C. and B. B. A. Denitrification. *Annu Rev Microbiol* **30**, 241–262 (1976).

45. Mulholland, M. R. & Fu, F. X. *Nutrient cycles and marine microbes in a CO2-enriched ocean*. *Oceanography* vol. 22 https://digitalcommons.odu.edu/oeas_fac_pubs (2009).

46. Paytan, A. & McLaughlin, K. The oceanic phosphorus cycle. *Chem Rev* **107**, 563–576 (2007).

47. Wallmann, K. Phosphorus imbalance in the global ocean? *Global Biogeochem Cycles* **24**, (2010).

48. Tréguer, P. J. & de La Rocha, C. L. The world ocean silica cycle. *Ann Rev Mar Sci* **5**, 477–501 (2013).

49. Conley, D. J., Schelske, C. L. & Stoermer, E. F. *Modification of the biogeochemical cycle of silica with eutrophication*. *Source: Marine Ecology Progress Series* vol. 101 (1993).

50. Douvere, F. & Ehler, C. N. The importance of monitoring and evaluation in adaptive maritime spatial planning. *J Coast Conserv* **15**, 305–311 (2011).

51. Martiny, J. B. H. *et al.* Microbial biogeography: Putting microorganisms on the map. *Nature Reviews Microbiology* vol. 4 102–112 Preprint at https://doi.org/10.1038/nrmicro1341 (2006).

52. Hanson, C. A., Fuhrman, J. A., Horner-Devine, M. C. & Martiny, J. B. H. Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nature Reviews Microbiology* vol. 10 497–506 Preprint at https://doi.org/10.1038/nrmicro2795 (2012).

53. Holman, L. E. *et al.* Animals, protists and bacteria share marine biogeographic patterns. *Nat Ecol Evol* **5**, 738–746 (2021).

54. Mo, Y. *et al.* Biogeographic patterns of abundant and rare bacterioplankton in three subtropical bays resulting from selective and neutral processes. *ISME Journal* **12**, 2198–2210 (2018).

55. Gilbert, J. A. *et al.* Defining seasonal marine microbial community dynamics. *ISME Journal* **6**, 298–308 (2012).

56. Dolan, J. R. An introduction to the biogeography of aquatic microbes. *Aquatic Microbial Ecology* **41**, 39–48 (2005).

57. Fuhrman, J. A., Cram, J. A. & Needham, D. M. Marine microbial community dynamics and their ecological interpretation. *Nature Reviews Microbiology* vol. 13 133–146 Preprint at https://doi.org/10.1038/nrmicro3417 (2015).

58. Ecol, A. M., Sherr, E. B. & Sherr, B. F. *Temporal offset in oceanic production and respiration processes implied by seasonal changes in atmospheric oxygen: the role of heterotrophic microbes*. *AQUATIC MICROBIAL ECOLOGY* vol. 11 (1996).

59. Fortunato, C. S., Herfort, L., Zuber, P., Baptista, A. M. & Crump, B. C. Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *ISME Journal* **6**, 554–563 (2012).

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

61. Verde, C., Giordano, D., Bellas, C. M., di Prisco, G. & Anesio, A. M. Polar Marine Microorganisms and Climate Change. in *Advances in Microbial Physiology* vol. 69 187–215 (Academic Press, 2016).

62. Kaur-Kahlon, G. *et al.* Response of a coastal tropical pelagic microbial community to changing salinity and temperature. *Aquatic Microbial Ecology* **77**, 37–50 (2016).

63. Sarmento, H., Montoya, J. M., Vázquez-Domínguez, E., Vaqué, D. & Gasol, J. M. Warming effects on marine microbial food web processes: How far can we go when it comes to predictions? *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 365 2137–2149 Preprint at https://doi.org/10.1098/rstb.2010.0045 (2010).

64. O'Brien, P. A., Morrow, K. M., Willis, B. L. & Bourne, D. G. Implications of ocean acidification for marine microorganisms from the free-living to the host-associated. *Frontiers in Marine Science* vol. 3 Preprint at https://doi.org/10.3389/fmars.2016.00047 (2016).

65. Schwalbach, M. S. and B. M. and F. J. A. Impact of light on marine bacterioplankton community structure. *Aquatic microbial ecology* **39**, 235–245 (2005).

66. Cottrell, M. T., Michelou, V. K., Nemcek, N., DiTullio, G. & Kirchman, D. L. Carbon cycling by microbes influenced by light in the Northeast Atlantic Ocean. *Aquatic Microbial Ecology* **50**, 239–250 (2008).

67. Yamaguchi, A. *et al. Structure and size distribution of plankton communities down to the greater depths in the western North Pacific Ocean*. *Deep-Sea Research II* vol. 49 (2002).

68. Galinski, E. A. & Trüper, H. G. Microbial behaviour in salt-stressed ecosystems. *FEMS Microbiol Rev* **15**, 95–108 (1994).

69. Dube, A., Jayaraman, G. & Rani, R. Modelling the effects of variable salinity on the temporal distribution of plankton in shallow coastal lagoons. *Journal of Hydro-Environment Research* **4**, 199–209 (2010).

70. Fortunato, C. S. & Crump, B. C. Microbial gene abundance and expression patterns across a river to ocean salinity gradient. *PLoS One* **10**, (2015).

71. Dupont, C. L. *et al.* Functional tradeoffs underpin salinity-driven divergence in microbial community composition. *PLoS One* **9**, (2014).

72. Telesh, I., Schubert, H. & Skarlato, S. Life in the salinity gradient: Discovering mechanisms behind a new biodiversity pattern. *Estuarine, Coastal and Shelf Science* vol. 135 317–327 Preprint at https://doi.org/10.1016/j.ecss.2013.10.013 (2013).

73. Weber, T. S. & Deutsch, C. Ocean nutrient ratios governed by plankton biogeography. *Nature* **467**, 550–554 (2010).

74. Bristow, L. A., Mohr, W., Ahmerkamp, S. & Kuypers, M. M. M. *Current Biology Nutrients that limit growth in the ocean*. *R474 Current Biology* vol. 27 (2017).

75. Marine microorganisms and global.

76. Lund, P. A. *et al.* Understanding How Microorganisms Respond to Acid pH Is Central to Their Control and Successful Exploitation. *Front Microbiol* **11**, (2020).

77. Pedersen, M. F. and H. P. J. Effects of high pH on a natural marine planktonic community. *Mar Ecol Prog Ser* **260**, 19–31 (2003).

78. Das, S. & Mangwani, N. Ocean acidification and marine microorganisms: responses and consequences. *Oceanologia* **57**, 349–361 (2015).

79. Tamburini, C., Boutrif, M., Garel, M., Colwell, R. R. & Deming, J. W. Prokaryotic responses to hydrostatic pressure in the ocean - a review. *Environmental Microbiology* vol. 15 1262–1274 Preprint at https://doi.org/10.1111/1462-2920.12084 (2013).

80. Pradillon, F. & Gaill, F. Pressure and life: Some biological strategies. in *Life in Extreme Environments* vol. 9781402062858 341–355 (Springer Netherlands, 2007).

81. Goodwin, K. D. *et al.* DNA sequencing as a tool to monitor marine ecological status. *Frontiers in Marine Science* vol. 4 Preprint at https://doi.org/10.3389/fmars.2017.00107 (2017).

82. Kopf, A. *et al.* The ocean sampling day consortium. *GigaScience* vol. 4 Preprint at https://doi.org/10.1186/s13742-015-0066-5 (2015).

83. Sunagawa, S. *et al.* Tara Oceans: towards global ocean ecosystems biology. *Nature Reviews Microbiology* vol. 18 428–445 Preprint at https://doi.org/10.1038/s41579-020-0364-5 (2020).

84. Santi, I. *et al. European Marine Omics Biodiversity Observation Network (EMO BON) Handbook European Marine Omics Biodiversity Observation Network (EMO BON) 1 European Marine Omics Biodiversity Observation Network (EMO BON)-Handbook*. www.embrc.eu (2021).

85. *Ocean Sampling Day Handbook*. http://www.pangaea.de (2016).

86. Pesant, S. *et al.* Open science resources for the discovery and analysis of Tara Oceans data. *Sci Data* **2**, (2015).

87. Ibarbalz, F. M. *et al.* Global Trends in Marine Plankton Diversity across Kingdoms of Life. *Cell* **179**, 1084-1097.e21 (2019).

88. Teodoro, A., Pais-Barbosa, J., Piqueiro, F. & Aguiar, R. Quantitative and qualitative coastal water quality parameters monitoring using field data and aerial photography: Porto (Portugal) beaches. in *Earth Resources and Environmental Remote Sensing/GIS Applications* vol. 7831 78311M (SPIE, 2010).

89. Bordalo, A. A. Microbiological water quality in urban coastal beaches: The influence of water dynamics and optimization of the sampling strategy. *Water Res* **37**, 3233–3241 (2003).

90. Couto, C. M. C. M. *et al.* Assessment of Douro and Ave River (Portugal) lower basin water quality focusing on physicochemical and trace element spatiotemporal changes. *J Environ Sci Health A Tox Hazard Subst Environ Eng* **53**, 1056–1066 (2018).

91. Bordalo, A. A., Teixeira, R. & Wiebe, W. J. A water quality index applied to an international shared river basin: The case of the Douro River. *Environ Manage* **38**, 910–920 (2006).

92. Amorim, E., Ramos, S. & Bordalo, A. A. Relevance of temporal and spatial variability for monitoring the microbiological water quality in an urban bathing area. *Ocean Coast Manag* **91**, 41–49 (2014).

93. Vale, P., Botelho, M. J., Rodrigues, S. M., Gomes, S. S. & Sampayo, M. A. de M. Two decades of marine biotoxin monitoring in bivalves from Portugal (1986-2006): A review of exposure assessment. *Harmful Algae* **7**, 11–25 (2008).

94. Stratoudakis, Y., Bernal, M., Borchers, D. L. & Borges, M. F. *Changes in the distribution of sardine eggs and larvae off Portugal, 1985-2000*.

95. Ramos, S., Cowen, R. K., Paris, C., Ré, P. & Bordalo, A. A. Environmental forcing and larval fish assemblage dynamics in the Lima River estuary (northwest Portugal). *J Plankton Res* **28**, 275–286 (2006).

96. Azeiteiro, U. M., Bacelar-Nicolau, L., Resende, P., Gonçalves, F. & Pereira, M. J. Larval fish distribution in shallow coastal waters off North Western Iberia (NE Atlantic). *Estuar Coast Shelf Sci* **69**, 554–566 (2006).

97. dos Santos, A. *et al.* Diel vertical migration of decapod larvae in the Portuguese coastal upwelling ecosystem: Implications for offshore transport. *Mar Ecol Prog Ser* **359**, 171–183 (2008).

98. Mucha, A. P., Teresa, M., Vasconcelos, S. D. & Bordalo, A. A. *Macrobenthic community in the Douro estuary: relations with trace metals and natural sediment characteristics*. www.elsevier.com/locate/envpol.

99. Mucha, A. P., Vasconcelos, M. T. S. D. & Bordalo, A. A. Spatial and seasonal variations of the macrobenthic community and metal contamination in the Douro estuary (Portugal). *Mar Environ Res* **60**, 531–550 (2005).

100. Pérez, F. F. *et al.* Plankton response to weakening of the Iberian coastal upwelling. *Glob Chang Biol* **16**, 1258–1267 (2010).

101. Queiroga, H., Silva, C., Sorbe, J. C. & Morgado, F. Composition and distribution of zooplankton across an upwelling front on the northern Portuguese coast during summer. *Hydrobiologia* **545**, 195–207 (2005).

102. Gouveia, V., Almeida, C. M. R., Almeida, T., Teixeira, C. & Mucha, A. P. Indigenous microbial communities along the NW Portuguese Coast: Potential for hydrocarbons degradation and relation with sediment contamination. *Mar Pollut Bull* **131**, 620–632 (2018).

103. Filker, S., Gimmler, A., Dunthorn, M., Mahé, F. & Stoeck, T. Deep sequencing uncovers protistan plankton diversity in the Portuguese Ria Formosa solar saltern ponds. *Extremophiles* **19**, 283–295 (2015).

104. Tracana, A. & Brotas, V. Monitoring Phytoplankton and Nutrients in Tagus Estuary, Portugal, for 20 years. *Front Mar Sci* **5**, (2018).

105. Drag0, T. *et al. Some evidences of northward fine sediment transport in the northern Portuguese continental shelf*.

106. Álvarez-Salgado, X. A. *et al.* The Portugal coastal counter current off NW Spain: New insights on its biogeochemical variability. *Progress in Oceanography* vol. 56 281–321 Preprint at https://doi.org/10.1016/S0079-6611(03)00007-7 (2003).

107. Sordo, I., Barton, E. D., Cotos, J. M. & Pazos, Y. An inshore poleward current in the NW of the Iberian Peninsula detected from satellite images, and its relation with G. catenatum and D. acuminata blooms in the Galician Rias. *Estuar Coast Shelf Sci* **53**, 787–799 (2001).

108. Vitorino, J., Oliveira, A., Jouanneau, J. M. & Drago, T. *Winter dynamics on the northern Portuguese shelf. Part 1: physical processes*. *Progress in Oceanography* vol. 52 www.elsevier.com/locate/pocean (2002).

109. Drag0, T. *et al. Some evidences of northward fine sediment transport in the northern Portuguese continental shelf*.

110. Cachao, M., Guerreiro, C. & Cachão, M. *Calcareous nannoplankton as a tracer of the marine influence on the NW coast of Portugal, over the last 14000 years Calcareous nannoplankton as a tracer of the marine influence on the NW coast of Portugal over the last 14 000 years Braarudosphaera bigelowii-Global Reassessment of Evolutionary Paleogeography and*

Paleoecology View project DUSTTRAFFIC View project. https://www.researchgate.net/publication/235771178 (2005).

111. Azevedo, I. C., Duarte, P. M. & Bordalo, A. A. Pelagic metabolism of the Douro estuary (Portugal) - Factors controlling primary production. *Estuar Coast Shelf Sci* **69**, 133–146 (2006).

112. Mendes, R. *et al.* Observation of a turbid plume using MODIS imagery: The case of Douro estuary (Portugal). *Remote Sens Environ* **154**, 127–138 (2014).

113. Boothroyd, J. C. *Mesotidal Inlets and Estuaries*.

114. Azevedo, I. C., Duarte, P. M. & Bordalo, A. A. Understanding spatial and temporal dynamics of key environmental characteristics in a mesotidal Atlantic estuary (Douro, NW Portugal). *Estuar Coast Shelf Sci* **76**, 620–633 (2008).

115. Ruppert, K. M., Kline, R. J. & Rahman, M. S. Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation* vol. 17 Preprint at https://doi.org/10.1016/j.gecco.2019.e00547 (2019).

116. Thermo Fisher Scientific Inc. Qubit™ 4 Fluorometer User Guide. Preprint at (2021).

117. R. S. Weyant et al. Effect of ionic and nonionic detergents on the Taq polymerase. *BioTechniques* 308–309 (1990).

118. Piredda, R. *et al.* Diversity and temporal patterns of planktonic protist assemblages at a Mediterranean Long Term Ecological Research site. *FEMS Microbiol Ecol* **93**, (2017).

119. Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME Journal* **6**, 1621–1624 (2012).

120. Apprill, A., Mcnally, S., Parsons, R. & Weber, L. Minor revision to V4 region SSU rRNA 806R gene primer greatly increases detection of SAR11 bacterioplankton. *Aquatic Microbial Ecology* **75**, 129–137 (2015).

121. Parada, A. E., Needham, D. M. & Fuhrman, J. A. Every base matters: Assessing small subunit rRNA primers for marine microbiomes with mock communities, time series and global field samples. *Environ Microbiol* **18**, 1403–1414 (2016).

122. Koroleff, F. Determination of ammonia. *Methods of seawater analysis* (1983).

123. Jones, M. N. *NITRATE REDUCTION BY SHAKING WITH CADMIUM ALTERNATIVE TO CADMIUM COLUMNS*. *~ter Res* vol. 18 (1984).

124. Joye, S. and C. R. Nitrogen exchange between microvegetated intertidal sediments and the overlying water column. in *Estuarine Research Federation Annual Meeting* 58 (1993).

125. Grasshoff, K. and K. K. and E. M. *Methods of seawater analysis*. (2009).

126. Callahan, B. J. *et al.* DADA2: High-resolution sample inference from Illumina amplicon data. *Nat Methods* **13**, 581–583 (2016).

127. Callahan, B. J., McMurdie, P. J. & Holmes, S. P. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *ISME Journal* **11**, 2639–2643 (2017).

128. McLaren, M. R. , & C. B. J. Silva 138.1 prokaryotic SSU taxonomic training data formatted for DADA2 [Data set]. Zenodo. (2021).

129. Guillou, L. *et al.* The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. *Nucleic Acids Res* **41**, (2013).

130. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, (2013).

131. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. . *Springer-Verlag New York* (2016).

132. Schlitzer, R. Ocean Data View, version 4.4. 2." 2014-03-31). (2011).

133. Shannon, C. E. & Weaver, W. *THE MATHEMATICAL THEORY OF COMMUNICATION*. (1949).

134. Buttigieg, P. L. & Ramette, A. A guide to statistical analysis in microbial ecology: A community-focused, living review of multivariate data analyses. *FEMS Microbiol Ecol* **90**, 543–550 (2014).

135. Jari Oksanen, F. G. B. M. F. R. K. P. L. D. M. P. R. M. R. B. O. G. L. S. P. S. M. H. H. S. E. S. and H. W. Vegan: Community Ecology Package. R package version 2.5-7. (2020).

136. Liu, L., Yang, J., Yu, X., Chen, G. & Yu, Z. Patterns in the composition of microbial communities from a subtropical river: Effects of environmental, spatial and temporal factors. *PLoS One* **8**, (2013).

137. External control of bacterial community structure in lakes.

138. Bergström, A. K. & Jansson, M. Bacterioplankton production in humic Lake Ortrasket in relation to input of bacterial cells and input of allochthonous organic carbon. *Microb Ecol* **39**, 101–115 (2000).

139. David P.. Clark, P. N. J. , & M. R. . M. *Molecular Biology*. (Academic Press, 2019).

140. Zhou, L. *et al.* Environmental filtering dominates bacterioplankton community assembly in a highly urbanized estuarine ecosystem. *Environ Res* **196**, (2021).

141. Kieft, B. *et al.* Microbial community structure-function relationships in Yaquina Bay estuary reveal spatially distinct carbon and nitrogen cycling capacities. *Front Microbiol* **9**, (2018).

142. Wang, H., Zhang, C., Chen, F. & Kan, J. Spatial and temporal variations of bacterioplankton in the Chesapeake Bay: A re-examination with high-throughput sequencing analysis. *Limnol Oceanogr* **65**, 3032–3045 (2020).

143. Bouvier, T. C., del Giorgio, P. A., Wasniak, T., Sunberg, K. & Alexander, J. *Compositional changes in free-living bacterial communities along a salinity gradient in two temperate estuaries*. *Limnol. Oceanogr* vol. 47 www.cbos.org (2002).

144. Cunha, M. A., Dias, J. M., Almeida, M. A., Lopes, J. F. & Alcântara, F. *Fluxes of bacterioplankton between a tidal estuary and the sea: returning to the 'Outwelling Hypothesis'*. (2003).

145. Gobler, C. J., Burson, A., Koch, F., Tang, Y. & Mulholland, M. R. The role of nitrogenous nutrients in the occurrence of harmful algal blooms caused by Cochlodinium polykrikoides in New York estuaries (USA). *Harmful Algae* **17**, 64–74 (2012).

146. Azevedo, I. C., Duarte, P. M. & Bordalo, A. A. Understanding spatial and temporal dynamics of key environmental characteristics in a mesotidal Atlantic estuary (Douro, NW Portugal). *Estuar Coast Shelf Sci* **76**, 620–633 (2008).

147. Seguro, I. *et al.* Seasonal changes of the microplankton community along a tropical estuary. *Reg Stud Mar Sci* **2**, 189–202 (2015).

148. Ohore, O. E. *et al.* Vertical characterisation of phylogenetic divergence of microbial community structures, interaction, and sustainability in estuary and marine ecosystems. *Science of the Total Environment* **851**, (2022).

149. Diehl, S. Phytoplankton, light, and nutrients in a gradient of mixing depths: theory. *Ecology* **80**, 386–398 (2002).

150. Microphytoplankton community and environmental variables in an urban eutrophic estuary (Capibaribe River, Northeast Brazil). *EVELINE P. AQUINO, GISLAYNE C. P. BORGES, MARCOS HONORATO-DA-SILVA, JOSÉ Z. O. PASSAVANTE, MARIA G. G. S. CUNHA* (2014).

151. Telesh, I., Schubert, H. & Skarlato, S. Life in the salinity gradient: Discovering mechanisms behind a new biodiversity pattern. *Estuarine, Coastal and Shelf Science* vol. 135 317–327 Preprint at https://doi.org/10.1016/j.ecss.2013.10.013 (2013).

152. Prieur, D. , T. M. , R. A. , C. S. , M. G. , & B. B. Evolution of bacterial communities in the Gironde Estuary (France) according to a salinity gradient. in *Estuarine, Coastal and Shelf Science 24.1* 95–108 (1987).

153. Hyun, J. H. , C. J. K. , C. K. H. , Y. E. J. , & K. M. K. Tidally induced changes in bacterial growth and viability in the macrotidal Han River estuary, Yellow Sea. in *Estuarine, Coastal and Shelf Science 48.2* 143–153 (1999).

154. Qasim, S. Z. Indian estuaries (Allied publication Pvt. Ltd. Heriedia Marg, Ballard estate, Mumbai). (2003).

155. Fortunato, C. S., Herfort, L., Zuber, P., Baptista, A. M. & Crump, B. C. Spatial variability overwhelms seasonal patterns in bacterioplankton communities across a river to ocean gradient. *ISME Journal* **6**, 554–563 (2012).

156. Andrew M. Staroscik, D. C. S. Seasonal patterns in bacterioplankton abundance and production in Narragansett Bay, Rhode Island, USA. *AQUATIC MICROBIAL ECOLOGY* **35**, 275–282 (2004).

157. Osterholz, H., Kirchman, D. L., Niggemann, J. & Dittmar, T. Diversity of bacterial communities and dissolved organic matter in a temperate estuary. *FEMS Microbiol Ecol* **94**, (2018).

158. Kaestli, M. *et al.* Spatial and temporal microbial patterns in a tropical macrotidal estuary subject to urbanization. *Front Microbiol* **8**, (2017).

159. Fernanda Nogueira dos REIS, L. F. B. F. C. P. L. T. L. R. G. D. da V. N. F. A. C. de M. MICROPHYTOPLANKTON DYNAMICS IN CURUPERÉ ESTUARY AT THE AMAZONIAN MANGROVE ECOSYSTEM. *Instituto de Pesca*.

160. Beaver, J. R. & Crisman, T. L. *The trophic response of ciliated protozoans in freshwater lakes"*. *Limnol. Oceanogr* vol. 27 (1982).

161. Kerfoot, W. C., Robbins, J. A. & Weider, L. J. *A new approach to historical reconstruction: Combining descriptive and experimental paleolimnology*. *Limnol. Oceanogr* vol. 44 (1999).

162. Pomeroy, L. R. , & W. W. J. Temperature and substrates as interactive limiting factors for marine heterotrophic bacteria. *Aquatic Microbial Ecology* **23**, 187–204 (2001).

163. Lepère, C. , B. D. , J. L. , D. I. , & D. D. Structure and regulation factors of eukaryotic picoplankton in lacustrine ecosystems. in *ppl Environ Microbiol, 72,* 2971-2981. (2006).

164. Suikkanen, S. , P. S. , E.-O. J. , L. M. , L. S. , & B. Climate change and eutrophication induced shifts in northern summer plankton communities. *PLoS One* (2013).

165. Du, J. *et al.* Temporal and Spatial Diversity of Bacterial Communities in Coastal Waters of the South China Sea. *PLoS One* **8**, (2013).

166. Sprong, P. A. A. *et al.* Spatial dynamics of eukaryotic microbial communities in the German Bight. *J Sea Res* **163**, (2020).

167. Fortunato, C. S. & Crump, B. C. Bacterioplankton Community Variation Across River to Ocean Environmental Gradients. *Microb Ecol* **62**, 374–382 (2011).

168. Ghiglione, J. F. , L. M. , & L. P. Spatial and temporal scales of variation in bacterioplankton community structure in theNW Mediterranean Sea. *Aquatic Microbial Ecology* **40**, 229–240 (2005).

169. Klemas, V. Remote sensing of coastal plumes and ocean fronts: Overview and case study. *J Coast Res* **28**, 1–7 (2012).

170. Wang, Y. *et al.* Functional gene diversity and metabolic potential of the microbial community in an estuary-shelf environment. *Front Microbiol* **8**, (2017).

171. Mou, X., Sun, S., Edwards, R. A., Hodson, R. E. & Moran, M. A. Bacterial carbon processing by generalist species in the coastal ocean. *Nature* **451**, 708–711 (2008).

172. Dai, M. *et al.* Effects of an estuarine plume-associated bloom on the carbonate system in the lower reaches of the Pearl River estuary and the coastal zone of the northern South China Sea. *Cont Shelf Res* **28**, 1416–1423 (2008).

173. Sigleo, A. C., Mordy, C. W., Stabeno, P. & Frick, W. E. Nitrate variability along the Oregon coast: Estuarine-coastal exchange. *Estuar Coast Shelf Sci* **64**, 211–222 (2005).

174. Zhang, R. *et al.* High-throughput sequencing reveals the spatial distribution variability of microbial community in coastal waters in Shenzhen. *Ecotoxicology* **30**, 1429–1436 (2021).

175. Nold, S. C. & Zwart, G. *Patterns and governing forces in aquatic microbial communities*. *Aquatic Ecology* vol. 32 (1998).

176. Yi, J., Lo, L. S. H. & Cheng, J. Dynamics of Microbial Community Structure and Ecological Functions in Estuarine Intertidal Sediments. *Front Mar Sci* **7**, (2020).

177. Farnelid, H. M., Turk-Kubo, K. A. & Zehr, J. P. Identification of associations between bacterioplankton and photosynthetic picoeukaryotes in coastal waters. *Front Microbiol* **7**, (2016).

178. Wang, H., Zhang, C., Chen, F. & Kan, J. Spatial and temporal variations of bacterioplankton in the Chesapeake Bay: A re-examination with high-throughput sequencing analysis. *Limnol Oceanogr* **65**, 3032–3045 (2020).

179. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes.

180. Imhoff, J. F., Rahn, T., Künzel, S. & Neulinger, S. C. Photosynthesis is widely distributed among Proteobacteria as demonstrated by the phylogeny of PufLM reaction center proteins. *Front Microbiol* **8**, (2018).

181. Kirchman, D. L. The ecology of Cytophagaâ€“Flavobacteria in aquatic environments. *FEMS Microbiol Ecol* **39**, 91–100 (2006).

182. Newton, R. J., Jones, S. E., Eiler, A., McMahon, K. D. & Bertilsson, S. A Guide to the Natural History of Freshwater Lake Bacteria. *Microbiology and Molecular Biology Reviews* **75**, 14–49 (2011).

183. Li, J. *et al.* Spatial and seasonal distributions of bacterioplankton in the Pearl River Estuary: The combined effects of riverine inputs, temperature, and phytoplankton. *Mar Pollut Bull* **125**, 199–207 (2017).

184. Pajares, S. & Ramos, R. Processes and Microorganisms Involved in the Marine Nitrogen Cycle: Knowledge and Gaps. *Frontiers in Marine Science* vol. 6 Preprint at https://doi.org/10.3389/fmars.2019.00739 (2019).

185. Semedo, M. *et al.* Depth Profile of Nitrifying Archaeal and Bacterial Communities in the Remote Oligotrophic Waters of the North Pacific. *Front Microbiol* **12**, (2021).

186. Dettner, K. *4.09—Chemical Defense and Toxins of Lower Terrestrial and Freshwater Animals. Comprehensive Natural Products II; Liu, H.-WB, Mander, L., Eds,* . (2010).

187. Kliman, R. M. *Encyclopedia of evolutionary biology. Academic Press,* . (2016).

188. Johnson, E. A. , and J. T. Lecomte. The haemoglobins of algae. In Advances in microbial physiology . *Academic Press* **67**, 177–234 (2015).

7. Supplementary material

Annex 1- Cleaning Process of the 16S rRNA gene sequences of the samples collected in the Douro estuary in 2021 and 2022 (The ASVs classified as Eukaryota/Chloroplast/Mitochondria were removed).

Samples	input	filtered	denoisedF	denoisedR	merged	nonchim	noEuk.	noMito.	noCloro.
D1B_A	23548	17456	16483	16712	13573	13102	13102	13033	11405
D1S_A	26217	19648	18830	18963	16256	15837	15837	15785	14152
D1B_W	34852	25662	22894	24223	16363	15708	15708	15670	13337
D1S_W	28575	21003	19236	19890	15616	15084	15084	15046	12040
D5B_A	25441	19263	18119	18550	15533	14823	14823	14786	13347
D5S_A	15825	11906	10986	11269	8130	7766	7766	7758	7379
D5B_W	30405	22358	20218	21133	15038	14226	14226	14210	12531
D5S_W	34002	25714	24721	25025	22154	20399	20399	20353	11321
D7B_A	30145	21738	19384	20275	14352	13801	13801	13792	13160
D7S_A	47669	36374	34649	35216	28354	26113	26113	26080	23968
D7B_W	25591	19095	17932	18453	15021	13507	13507	13507	11125
D7S_W	31096	24103	23298	23565	20919	18930	18930	18901	8715
D9B_A	36911	27416	25984	26445	21380	20333	20333	20293	18602
D9S_A	45362	33557	32152	32598	26936	25606	25606	25574	23707
D9B_W	26446	19589	18691	19008	16169	15124	15124	15100	8912
D9S_W	18464	13030	12465	12636	11017	10529	10529	10511	6046
D11B_A	20944	15647	14604	14993	11530	10976	10976	10972	9935
D11S_A	16208	11987	11376	11443	9135	8827	8827	8796	7462
D11S_W	24919	18434	17687	17932	14981	13976	13976	13942	9690

Annex 2- Cleaning Process of the 18S rRNA gene sequences of the samples collected in the Douro estuary in 2021 and 2022 (The ASVs classified as Metazoa/ Fungi/ Streptophyta/ Ulvophyceae/ Rhodophyceae were removed).

Samples	Input	Filtered	denoisedF	denoisedR	Merged	Nonchim	no Metazoa	no Fungi	no Strepto.	no Ulvop.	no Rhodo.
D1B_A	149143	114875	113955	113920	109667	104026	70644	70475	70470	70470	70470
D1S_A	185634	144189	143387	143128	140062	129171	86350	86171	86171	86171	86171
D1B_W	42792	32641	31360	31863	28608	27407	24797	24728	24723	24656	24656
D1S_W	104930	80860	79243	79720	74015	73063	67363	66872	66872	66508	66508
D5B_A	53300	40924	40239	40423	38702	36792	24515	24407	24407	24384	24384
D5S_A	93425	72772	71545	71730	63577	60656	52798	43454	43435	41555	41555
D5B_W	72393	55165	53249	54063	48492	47038	43574	42968	42959	42784	42784

D5S_W	54366	42582	41761	42042	40272	38798	37106	35594	35585	35345	35345
D7B_A	103595	79522	77918	78364	74477	69073	52878	51491	51491	51413	51413
D7S_A	53384	41082	40393	40512	38084	35998	33749	30344	30344	30115	30115
D7B_W	95078	71891	70669	71049	67001	65187	62944	57094	57056	56936	56936
D7S_W	100694	79161	78204	78383	75971	70859	69976	67468	67413	67201	67201
D9B_A	119534	88893	87910	87967	83075	78835	74362	63767	63740	63740	63740
D9S_A	104561	79646	78834	78793	74116	70382	68963	62269	62222	62215	62215
D9B_W	47361	35967	35299	35496	34075	32539	32126	30494	30436	30436	30436
D9S_W	158812	122129	120946	120830	116678	106444	105735	98587	98364	98364	98364
D11B_A	118181	89067	88057	88023	82622	78518	74628	64703	64666	64666	64666
D11S_A	188841	140860	140284	139791	132233	120896	112194	109140	109139	109139	109139
D11B_W	50844	39115	38487	38584	36533	33012	32034	29908	29861	29861	29861
D11S_W	71716	55209	54696	54538	52782	49251	48609	45302	45237	45237	45237

Annex 3- Cleaning Process of the 16S rRNA gene sequences of the samples collected in in the Douro estuary in 2016 (The ASVs classified as Eukaryota/Chloroplast/Mitochondria were removed).

Samples	**input**	**filtered**	**denoisedF**	**denoisedR**	**merged**	**nonchim**	**noChlo.**	**noMito.**	**noEuk.**
R1H	83153	66189	53397	54179	43814	41670	40078	39498	39498
R2H	63122	51074	44773	45121	38812	37336	35717	34970	34970
R3H	68475	55564	47407	47818	41017	39370	37781	37003	37003
R4H	93083	78427	74488	75154	62209	58913	56819	55899	55899
R5H	93826	77898	69822	70237	60670	58035	55292	54046	54046
R1L	48638	39693	31830	32550	25558	24218	23531	23279	23276
R2L	134070	105431	90916	91389	76845	73667	71325	69794	69786
R3L	86808	70707	60673	61292	51270	49377	47495	46369	46369
R4L	59147	46663	40136	40447	34419	33106	32038	31540	31540
R5L	71514	57370	49034	49468	40477	38866	36381	35028	35028

Annex 4- Cleaning Process of the 18S rRNA gene sequences of the samples collected in the Douro estuary in 2016 (The ASVs classified as Metazoa/ Fungi/ Streptophyta/ Ulvophyceae/ Rhodophyceae were removed).

Samples	**input**	**filtered**	**denoisedF**	**denoisedR**	**merged**	**nonchim**	**no Fungi**	**No Meta**	**No Strept.**	**No Ulvo.**	**No Rhodo.**
R1H	54697	44672	35547	37496	29817	29056	28097	24178	24153	23820	23820
R2H	42165	35991	32250	33102	29320	28870	28249	27895	27860	27860	27860
R3H	43056	34080	30466	31351	27735	27247	26687	26120	26120	26120	26120
R4H	120162	109027	106921	107839	96809	86100	83174	82754	82738	82720	82720
R5H	103030	90173	84731	86104	77711	74665	72617	70983	70940	70940	70940
R1L	95297	83330	75583	77038	65995	62023	59550	57047	57026	56176	56176
R2L	76329	64952	60226	61019	54925	54026	52216	48449	48435	48423	48423

R3L	79727	68777	62740	63883	56381	54643	53201	51834	51815	51806	51806
R4L	83622	68788	61999	63371	53911	52142	50776	49312	49304	49301	49301
R5L	73404	50630	42739	43538	38174	36757	35741	34748	34728	34728	34728

Annex 5- Cleaning Process of the 16S rRNA gene sequences of the samples collected in the NW coast of Portugal (The ASVs classified as Eukaryota/Chloroplast/Mitochondria were removed).

Samples	**input**	**filtered**	**denoisedF**	**denoisedR**	**merged**	**nonchim**	**noEuka**	**noCholo**	**noMito**
S01	19124	14387	13908	14014	11262	10139	10139	10116	10116
S04	21909	15975	15603	15593	12971	12185	12185	9571	9512
S05	30628	21965	21455	21450	18425	17841	17841	16470	16361
S06	18309	13452	12827	12924	10711	10206	10206	10034	10026
S07	11835	8573	8278	8278	7497	7336	7336	7124	7117
S10	11187	8353	8073	8058	7491	7390	7390	6149	6131
S11	15328	11145	10784	10811	10146	9941	9941	7379	7266
S14	20863	15389	15021	14952	12772	12304	12304	11966	11948
S15	47724	34681	33673	33755	27500	25656	25656	24931	24925
S18	24041	18074	17658	17573	15022	14347	14347	13667	13661
S19	18173	12906	12499	12417	10858	10445	10445	9976	9891
S22	11372	8306	7855	7903	6770	6677	6677	6248	6237
S23	18367	13260	12603	12703	9926	9413	9413	8987	8969
S26	13935	10273	9894	9852	7973	7794	7794	7423	7396
S27	26242	19104	18749	18619	16360	15848	15848	14542	14446
S28	9853	7229	6992	6991	6441	6328	6328	4790	4694

Annex 6- Cleaning Process of the 18S rRNA gene sequences of the samples collected in the NW coast of Portugal (The ASVs classified as Metazoa/ Fungi/ Streptophyta/ Ulvophyceae/ Rhodophyceae were removed).

	Input	**Filtered**	**denoisedF**	**denoisedR**	**Merged**	**nonchim**	**No Metazoa**	**No Fungi**	**No Strepto.**	**No Ulvo.**	**No Rhodo.**
S01	37860	28613	28264	28342	26294	26188	15342	15067	15067	15067	15067
S04	75407	59117	58784	58702	57583	57267	55249	55153	55153	55153	55153
S05	103800	82852	82517	82276	80919	79621	79374	79052	79052	79052	79052
S06	163422	124209	123412	123097	114953	113109	105768	105204	105199	104923	104923
S07	171701	129480	129211	128956	128065	127638	53871	53439	53424	53424	53424
S10	97152	75733	75514	75328	74639	74394	71977	71746	71741	71741	71741
S11	39434	30832	30669	30683	30383	30341	26214	26130	26119	26119	26119
S14	59983	45808	45393	45440	42972	42804	38316	38277	38277	38277	38277
S15	78135	60635	59952	59948	58691	55314	53672	53480	53472	53472	53472
S18	51225	40517	40273	40224	39573	39422	28047	28029	28022	28015	28015
S19	140387	109585	109065	108917	107099	105631	103659	103320	103320	103311	103311

S22	133828	104172	103532	103630	101185	100517	96612	96209	96205	96180	96180
S23	80922	62783	62249	62302	60487	60309	56741	56185	56185	56160	56160
S26	13585	9394	8964	9169	8092	7983	7977	7898	7893	7893	7893
S27	5123	3976	3670	3788	3365	3288	3281	3275	3275	3275	3275
S28	24551	19122	18894	18946	18383	18306	18240	18106	18106	18100	18100

Annex 7- Table with the environmental data and the CTD parameters measured in the Douro estuary campaign in 2012 and 2022.

Stations	**pH**	**Temperature (°C)**	**Salinity (PPT)**	**Chl-a (µg/L)**	**Ammonia (NH^4)(µM)**	**Nitrite (NO^2)(µM)**	**Nitrate (NO^3)(µM)**	**Phosphate (PO^3_4)(µM)**	**Silica (SiO^2)(µM)**
D1S_A	7.67	17,29	36,27	3,473	2,414	0,090	1,466	0,180	0,647
D1B_A	7.86	16,94	37,67	2,515	2,050	0,097	0,801	0,210	1,438
D5S_A	7.79	19.33	8,75	2,284	2,874	1,182	36,474	0,926	33,459
D5B_A	7.75	17,06	35,81	2,28	2,208	0,144	1,822	0,180	2,094
D7S_A	7.98	19,64	1,56	2,89	2,303	1,393	36,902	1,306	38,810
D7B_A	7.64	17,5	27,2	6,66	2,731	0,318	4,537	0,400	5,005
D9S_A	8.62	19,34	0,22	3,07	1,257	1,906	51,126	1,481	45,742
D9B_A	8,07	18,87	0,29	4,38	1,543	1,806	45,777	1,452	31,557
D11S_A	8.31	19,63	0,12	2,900	0,196	1,986	17,439	0,940	37,716
D11B_A	7.89	18,79	0,13	3,953	0,354	2,053	50,041	1,189	45,204
D1S_W	8,05	13,86	33,32	3,473	5,514	0,308	14,000	0,739	4,769
D1B_W	8,11	13,82	33,66	2,515	4,104	0,314	17,710	0,161	4,332
D5S_W	8,18	11,75	8,52	2,191	2,295	0,520	67,580	0,519	45,574
D5B_W	8,05	12,31	16,82	2,513	4,333	0,394	30,294	0,588	21,058
D7S_W	8,26	11,46	2,88	1,589	1,804	0,725	93,037	0,423	75,289
D7B_W	8,02	11,56	9,37	3,524	3,368	0,619	52,342	0,794	27,704
D9S_W	8,43	11,31	0,23	1,759	0,333	0,586	85,074	0,216	72,025
D9B_W	8,22	11,04	0,47	1,792	1,038	0,625	97,547	0,037	69,064
D11S_W	8,46	10,93	0,15	2,284	0,000	0,606	82,897	0,000	78,823
D11B_W	8,29	10,74	0,17	2,284	0,000	0,619	97,614	0,092	74,431

Annex 8- Table with the environmental data and the CTD parameters measured in the Douro estuary campaign in 2016.

Stations	**Tide**	**Temperature (Cº)**	**Salinity (PPT)**	**Nitrate (µM))**	**Nitrite (µM)**	**Phosphate (µM)**	**Ammonium (µM)**	**pH**
R1H	High	11.8	5.72	54.44	1.3	1.25	27.03	7.54
R2H	High	11.503	0.61	67.35	1.09	1.4	5.09	7.51
R3H	High	11.36	0.12	70.85	0.95	1.28	7.64	7.49
R4H	High	11.46	0.11	55.6	0.8	1.25	4.19	7.49
R5H	High	11.45	0.10	70.37	0.97	1.4	2.01	7.50
R1L	Low	12.6	5.8	71.8	0.95	0.89	3.99	7.72
R2L	Low	11.53	0.17	51.78	0.94	1.99	8.89	7.49

R3L	Low	11.54	0.12	66.59	1.16	1.48	2.96	7.50
R4L	Low	11.52	0.12	86.98	0.66	1.6	N.A	7.54
R5L	Low	11.34	0.10	124.77	1.38	1.7	3.48	7.60

Annex 9- Table with the environmental data and the CTD parameters measured NW coast of Portugal.

Station	Temperature (°C)	Salinity (PPT)	pH	Chl-a (µg/L)	Ammonia (NH4) (µM)	Nitrite (NO2) (µM)	Nitrate (NO3) (µM)	Phosphate (PO^34) (µM)	Silica (SiO2) (µM)
S01	14,21	34,7	9,86	0,26	5,80	0,24	7,20	0,75	5,60
S04	14,94	38,36	8,44	4,67	1,67	0,06	2,07	0,11	0,26
S05	16,35	33,4	8,82	1,33	2,05	0,04	3,63	0,02	0,23
S06	15,55	34,64	8,62	0,34	7,78	0,46	14,68	1,09	10,21
S07	14,67	34,22	8,41	0,26	3,57	0,33	6,33	0,56	4,02
S10	14,55	37,23	7,98	1,82	2,12	0,07	1,85	0,14	0,97
S11	14,95	35,28	7,97	2,46	2,81	0,06	2,67	0,12	0,47
S14	14,95	35,84	7,97	0,34	4,35	0,34	11,48	0,75	5,37
S15	14,86	30,69	9,41	0,77	1,20	0,49	10,96	0,52	9,83
S18	14,7	35,32	8,00	3,94	1,91	0,07	1,45	0,01	0,57
S19	14,74	40,9	7,56	1,87	2,16	0,30	5,93	0,42	1,86
S22	15,3	33,34	7,07	1,59	4,87	0,53	13,39	0,50	6,15
S23	14,29	33,8	7,33	2,21	1,89	0,78	15,06	0,56	9,81
S26	16,83	32,64	6,94	1,97	1,42	0,55	2,88	0,00	1,00
S27	17,68	32,67	6,98	1,61	1,77	0,06	0,87	0,00	0,55
S28	14,76	32,39	6,91	5,5	2,06	0,55	8,99	0,26	5,03

Annex 10- Table presenting the values of the PERMANOVA analysis for the Douro 16S rRNA gene (Annex 10a.) and 18S rRNA gene (Annex 10b.) dataset, collected in 2021 and 2022.

a)	Df	SumofSqs	R^2	F	p-value	Signif.
Station	4	1,056	0,198	0,886	0,605	
Residual	14	4,266	0,802			
Total	18	5,321	1,000			
Season	1	1,559	0,293	7,041	0,002	*
Residual	17	3,763	0,707			
Total	18	5,321	1,000			
Depth	1	0,188	0,035	0,622	0,652	
Residual	17	5,133	0,965			
Total	18	5,321	1,000			
Temperature	1	1,219	0,229	5,052	0,001	*
Residual	17	4,102	0,771			
Total	18	5,321	1,000			

b)	Df	SumofSqs	R^2	F	p-value	Signif.
Station	4	2,0361	0,30625	1,6554	0,043	*
Residual	15	4,6122	0,69375			
Total	19	6,6483	1			
Season	1	1,6776	0,25233	6,0747	0,001	*
Residual	18	4,9707	0,74767			
Total	19	6,6483	1			
Depth	1	0,3107	0,04673	0,8824	0,489	
Residual	18	6,3376	0,95327			
Total	19	6,6483	1			
Temperature	1	1,5793	0,23754	5,6079	0,001	*
Residual	18	5,069	0,76246			
Total	19	6,6483	1			

Salinity	1	0,319	0,060	1,083	0,329	
Residual	17	5,003	0,940			
Total	18	5,321	1,000			
pH	1	0,343	0,064	1,170	0,297	
Residual	17	4,979	0,936			
Total	18	5,321	1,000			
Chl.a	1	0,260	0,049	0,873	0,511	
Residual	17	5,061	0,951			
Total	18	5,321	1,000			
Ammonia	1	0,141	0,027	0,463	0,788	
Residual	17	5,180	0,973			
Total	18	5,321	1,000			
Nitrite	1	0,224	0,042	0,746	0,839	
Residual	17	5,098	0,958			
Total	18	5,321	1,000			
Nitrate	1	0,709	0,133	2,614	0,043	*
Residual	17	4,612	0,867			
Total	18	5,321	1,000			
Phosphate	1	0,222	0,042	0,739	0,863	
Residual	17	5,100	0,958			
Total	18	5,321	1,000			
Silica	1	0,681	0,128	2,493	0,056	
Residual	17	4,641	0,872			
Total	18	5,321	1,000			

Salinity	1	0,8296	0,12479	2,5664	0,013	*
Residual	18	5,8187	0,87521			
Total	19	6,6483	1			
pH	1	0,8474	0,12746	2,6294	0,01	*
Residual	18	5,8009	0,87254			
Total	19	6,6483	1			
Chl.a	1	0,6422	0,09659	1,9246	0,055	
Residual	18	6,0061	0,90341			
Total	19	6,6483	1			
Ammonia	1	0,9471	0,14246	2,9902	0,014	*
Residual	18	5,7012	0,85754			
Total	19	6,6483	1			
Nitrite	1	0,8791	0,13223	2,7427	0,015	*
Residual	18	5,7692	0,86777			
Total	19	6,6483	1			
Nitrate	1	1,0577	0,15909	3,40053	0,002	*
Residual	18	5,5906	0,84091			
Total	19	6,6483	1			
Phosphate	1	0,6409	0,0964	1,9203	0,059	
Residual	18	6,0074	0,9036			
Total	19	6,6483	1			
Silica	1	0,9107	0,13698	2,857	0,012	
Residual	18	5,7376	0,86302			
Total	19	6,6483	1			

Annex 11- Table presenting the values of the PERMANOVA analysis for the Douro 16S rRNA gene (Annex 11a.) and 18S rRNA gene (Annex 11b.) dataset collected in 2016.

a)	Df	SumofSqs	R^2	F	p-value	Signif.
Station	4	0,840	0,612	1,974	0,055	
Residual	5	0,532	0,388			
Total	9	1,372	1,000			
Tide	1	0,086	0,063	0,534	0,764	
Residual	8	1,287	0,937			
Total	9	1,372	1,000			
Temperature	1	0,550	0,401	5,357	0,004	*
Residual	8	0,822	0,599			
Total	9	1,372	1,000			
Salinity	1	0,628	0,457	6,742	0,002	*
Residual	8	0,745	0,543			
Total	9	1,372	1,000			
pH	1	0,407	0,297	3,378	0,027	*
Residual	8	0,965	0,703			
Total	9	1,372	1,000			

b)	Df	SumofSqs	R^2	F	p-value	Signif.
Station	4	0,298	0,555	1,559	0,138	
Residual	5	0,239	0,445			
Total	9	0,536	1,000			
Tide	1	0,033	0,061	0,522	0,69	
Residual	8	0,503	0,939			
Total	9	0,536	1,000			
Temperature	1	0,234	0,437	6,199	0,013	*
Residual	8	0,302	0,563			
Total	9	0,536	1,000			
Salinity	1	0,249	0,464	6,936	0,005	*
Residual	8	0,287	0,536			
Total	9	0,536	1,000			
pH	1	0,240	0,448	6,506	0,011	*
Residual	8	0,296	0,552			
Total	9	0,536	1,000			

Nitrite	1	0,183	0,133	1,229	0,296	Nitrite	1	0,034	0,064	0,548	0,678
Residual	8	1,190	0,867			Residual	8	0,502	0,936		
Total	9	1,372	1,000			Total	9	0,536	1,000		
Nitrate	1	0,121	0,088	0,773	0,496	Nitrate	1	0,051	0,094	0,833	0,421
Residual	8	1,252	0,912			Residual	8	0,486	0,906		
Total	9	1,372	1,000			Total	9	0,536	1,000		
Phosphate	1	0,299	0,218	2,229	0,072	Phosphate	1	0,146	0,273	3,006	0,06
Residual	8	1,073	0,782			Residual	8	0,390	0,727		
Total	9	1,372	1,000			Total	9	0,536	1,000		

Annex 12- Table presenting the values of the PERMANOVA analysis for the costal samples of the 16S rRNA gene (Annex 12a.) and 18S rRNA gene (Annex 12b.) dataset.

a)	Df	SumofSqs	R^2	F	p-value	Signif.	**b)**	Df	SumofSqs	R^2	F	p-value	Signif.
Coast_Distance	1	0,49316	0,15859	2,6388	0,011	*	Coast_Distance	1	0,5218	0,12172	1,8016	0,058	
Residual	14	2,61648	0,84141				Residual	13	3,7654	0,87828			
Total	15	3,10964	1				Total	14	4,2872	1			
Longitudinal	1	0,52584	0,1691	2,8492	0,008	*	Longitudinal	1	0,4425	0,10322	1,4963	0,15	
Residual	14	2,5838	0,8309				Residual	13	3,8447	0,89678			
Total	15	3,1094	1				Total	14	4,2872	1			
Temperature	1	0,36094	0,11607	1,8384	0,072		Temperature	1	0,3349	0,07812	1,1017	0,376	
Residual	14	2,74869	0,88393				Residual	13	3,9523	0,92188			
Total	15	3,10964	1				Total	14	4,2872	1			
Salinity	1	0,18773	0,06037	0,8995	0,514		Salinity	1	0,3279	0,07648	1,0766	0,379	
Residual	14	2,92191	0,93963				Residual	13	3,9593	0,92352			
Total	15	3,10964	1				Total	14	4,2872	1			
pH	1	0,71224	0,22904	4,1592	0,001	*	pH	1	0,6323	0,14748	2,2489	0,02	*
Residual	14	2,3974	0,77096				Residual	13	3,6549	0,85252			
Total	15	3,10964	1				Total	14	4,2872	1			
Chl.a	1	0,32511	0,10455	1,6346	0,1		Chl.a	1	0,5795	0,13516	2,0317	0,028	*
Residual	14	2,78452	0,89454				Residual	13	3,7077	0,86484			
Total	15	3,10964	1				Total	14	4,2872	1			
Ammonia	1	0,37899	0,12188	1,9431	0,039	*	Ammonia	1	0,3827	0,08926	1,274	0,25	
Residual	14	2,73064	0,87812				Residual	13	3,9045	0,91074			
Total	15	3,10964	1				Total	14	4,2872	1			
Nitrite	1	0,56907	0,183	3,1359	0,003	*	Nitrite	1	0,5344	0,12465	1,8513	0,064	
Residual	14	2,54057	0,817				Residual	13	3,7528	0,87535			
Total	15	3,10964	1				Total	14	4,2872	1			
Nitrate	1	0,40688	0,13085	2,1076	0,043	*	Nitrate	1	0,5536	0,12913	1,9276	0,051	
Residual	14	2,70276	0,86915				Residual	13	3,7336	0,87087			
Total	15	3,10964	1				Total	14	4,2872	1			
Phosphate	1	0,48054	0,15453	2,5589	0,011	*	Phosphate	1	0,6514	0,15195	2,3293	0,023	*
Residual	14	2,6291	0,84547				Residual	13	3,6358	0,84805			
Total	15	3,10964	1				Total	14	4,2872	1			
Silica	1	0,46941	0,15095	2,4891	0,011	*	Silica	1	0,6341	0,1479	2,2565	0,026	*

Residual	14	2,64023	0,84905	Residual	13	3,6531	0,8521
Total	15	3,10964	1	Total	14	4,2872	1