ROYAUME DU MAROC
Université Mohammed V - RABAT - جامعة محمد الخامس – الرباط –
Faculté des sciences
كلية العلوم

CENTRE D'ETUDES DOCTORALES - SCIENCES ET TECHNOLOGIES

N° d'ordre 3740

THESE

En vue de l'obtention du : ***DOCTORAT***

***Centre de Recherche** : Biotechnologie Végétale et Microbienne, Biodiversité, et Environment*

***Structure de Recherche** : Biotechnologie et Physiologie Végétales*

***Discipline** : Biologie*

***Spécialité** : Bioinformatique*

Présentée et soutenue le 30/12/2022 par :

Bouchra CHAOUNI

Metagenomics analysis of the marine microbiome in Moroccan coastal lagoons

JURY

Rachida HASSIKOU,	PES,	Université Mohammed V-Rabat, Faculté des Sciences	Présidente
Bouabid BADAOUI,	PES,	Université Mohammed V-Rabat, Faculté des Sciences	Rapporteur/Examinateur
Rachid EL JAOUDI,	PES,	Université Mohammed V-Rabat, Faculté de Médecine et de Pharmacie	Rapporteur/Examinateur
Said BARRIJAL,	PES,	Université Abdelmalek Essaadi-Tanger, Faculté des Sciences et Techniques	Rapporteur/Examinateur
Hamid EL AMRI,	PES,	Institut Agronomique et Vétérinaire Hassan II-Rabat	Examinateur
Mohammed TIMINOUNI,	Dr,	Institut Pasteur du Maroc	Examinateur
Rajaa CHAHBOUNE,	PES,	Université Abdelmalek Essaadi-Tanger, Faculté de Médecine et de Pharmacie	Examinatrice
Linda AMARAL-ZETTLER,	PES,	Brown University, Marine Biological Laboratory	Co-encadrante
Hassan GHAZAL,	PH,	Centre National pour la Recherche Scientifique et Technique	Co-encadrant
Noureddine HAMAMOUCH,	PH,	Université Mohammed V-Rabat, Faculté des Sciences	Directeur de thèse

Année Universitaire : 2022/2023

✉ Faculté des Sciences, avenue Ibn Battouta, BP. 1014 RP, Rabat –Maroc
☎ 00212 (05) 37 77 18 76 🖷 00212(05) 37 77 42 61; http://www. fsr.um5.ac.ma

Dedication

This dissertation is dedicated to the memory of my sister **Souad Chaouni** and my father **Rachid Chaouni**. Although they were my inspiration to pursue my doctoral degree, they were unable to see my graduation. This is for them.

A special feeling of gratitude to my loving mother **Chama Elgarouaoui**, brothers, **Simohammed Chaouni** and **Youssef Chaouni**, whose words of encouragement and push for tenacity helped me succeed.

I also dedicate this dissertation to my in-law family who have supported me throughout the process, especially my mother-in-law **Hachchouma Armaz**. I will always appreciate all what you have done.

This thesis work is also dedicated to my husband, **Abdelmoughit Azdiddi**, who has been a constant source of moral, spiritual, emotional support, and encouragement at those times when it seemed impossible to continue. I am truly thankful for having you in my life.

A very special dedicate to my daughters, **Aya** and **Maria**. You provided the inspiration necessary for me to complete this process. I love you!

To all my dear friends, **Maria Isabel Aracil Hernandez**, **Fadwa Badrana**, **Yousra Hamdan**, **Siham Bounjoua**, **Abdellah Azami Idrissi**, and **Sofia Sehli**, who played such an important role along this journey, providing support during the challenges of graduation and life. Thank you.

Last but not least, lovely thanks are due to my laboratory partners and colleagues; **Pr. Francisco Rodriguez-Valera,** Pr. **Kristin Gunsalus**, Dr. **Kristen Hunter-Cevera**, Dr. **Hassenrück Christiane**, **Richard Fox**, and **Alan Twaddle**, and everyone I had met during my journey abroad, you have shown me endless support throughout the analysis of my thesis. I could not have done any of this without you. Thank you all.

A special Thanks to my colleagues from the ***Ocean Sampling Day*** Project and all the people and collaborators that have contributed to this work.

Acknowledgments

The present work was accomplished in the Laboratory of Plant and Microbial Biotechnology, Biodiversity and Environment, Faculty of Sciences, Mohammed V University in Rabat, Morocco, under the supervision of Professor **Noureddine Hamamouch** and cosupervision of Professors **Linda Amaral-Zettler** and **Hassan Ghazal**. This work is a collaborative project with the Josephine Bay Paul Center for Comparative Molecular Biology and Evolution at the Marine Biological Laboratory, Woods Hole, MA. USA, and the Ocean Sampling Day consortium at Max Planck Institute in Bremen, Germany.

First and foremost, I would like to praise and thank **Allah**, the almighty, who has granted me countless blessings, knowledge, and opportunities, so that I have been finally able to accomplish this thesis.

Then, I would like to thank my supervisor Pr. **Noureddine Hamamouch** for his constant support and kind advice throughout my PhD research studies. It was a real privilege and an honor for me to share his exceptional scientific knowledge and extraordinary human qualities.

I am also grateful to my cosupervisor Pr. **Hassan Ghazal**, whose constructive advice has been offered generously throughout the period of this study.

I warmly thank my cosupervisor Pr. **Linda Amaral-Zettler** for her genuine support and guidance, she has offered me throughout the span of this thesis project. She encouraged me to enlarge my network by meeting with and working alongside other scientists both locally and internationally. Her wisdom and strong background in metagenomics and bioinformatics have greatly contributed to the efficient execution of this project.

My honorable gratitude goes to Pr. **Rachida Hassikou**, for accepting to preside over the jury of my thesis defense. I am very grateful and thankful for the time you gave.

My deepest thanks go to Pr. **Bouabid Badaoui**, for honoring me by judging my thesis, providing valuable feedback, honest information, and thoughtful suggestions.

My warm gratitude goes to Pr. **Rachid El Jaoudi**, for accepting to report my thesis, and providing constructive comments.

My warm thanks go to Pr. **Said Barrijal,** for agreeing to report my thesis. I am thankful for the time you gave to evaluate my manuscript.

I am profoundly grateful to my General Pr. **Hamid El Amri**, Director of the Genetics Analysis Institute of the Royal Gendarmerie of Morocco, for challenging me when I needed challenging and supporting me when I needed supporting. Your input has been invaluable in helping me navigate some of the more emotionally challenging aspects of this thesis. Your creativity and wisdom are inspirational.

Thanks are due to Pr. **Mohammed Timinouni** for his time, questions and for being part of the defense jury.

Pr. **Rajaa Chahboun** for letting my defense be an enjoyable moment, and for your brilliant comments and suggestions, massive thanks to you.

Warm thanks go to the **H3ABionet**, and **MACECE Fulbright Joint-Supervision Program** for providing scholarships and financial grants that helped the accomplishment of international internships in good conditions.

Résumé

Les lagunes sont des écosystèmes marins fragiles qui sont considérablement affectés par des polluants anthropiques. Nous avons réalisé une caractérisation spatio-temporelle du microbiome de deux lagunes marocains, Marchica et Oualidia, classées comme sites Ramsar, la première sur le littoral méditerranéen et la seconde sur le littoral atlantique. Nous avons étudié leur diversité et abondance microbiennes à l'aide d'approches métagénomiques pendant l'été des années 2014 et 2015. Le microbiome bactérien était composé principalement de *Protéobactéries* (25%-53%, 29%-29%), *Cyanobactéries* (34%-12%, 11 %-0,53%), *Bactéroïdes* (24%-16%, 23%-43 %), *Actinobactéries* (7%-11%, 13%-7%) et *Verrucomicrobes* (4%-1%, 15%-14%) au niveau de Marchica et Oualidia en 2014 et 2015, respectivement. Fait intéressant, 48 souches ont été nouvellement signalées dans les écosystèmes lagunaires, tandis que huit virus inconnus ont été détectés uniquement dans la Marchica méditerranéenne. L'analyse statistique a montré une diversité microbienne plus élevée dans la lagune atlantique par rapport à la lagune méditerranéenn et une relation robuste entre la diversité alpha et les emplacements géographiques d'échantillonnage. De plus, les représentants de *Synechococcus* ont recruté un plus grand nombre de lectures de l'ARNr 16S à Marchica par rapport à Oualidia, où nous avons identifié 31 oligotypes de *Synechococcus* regroupés en 10 clades avec différents schémas de distribution. Cette toute première étude métagénomique sur les écosystèmes aquatiques marocains a enrichi le catalogue national des micro-organismes marins pouvant être utilisés comme candidats pour les propriétés de bioindication, le potentiel de biosurveillance, la valorisation des biotechnologies, la protection de la biodiversité et l'évaluation de la santé des lagons.

Mots clés: Métagénomique; microbiome marin; lagunes marocaines; Marchica, Oualidia; biodiversité; bioindicateurs.

Abstract

Lagoons are fragile marine ecosystems that are considerably affected by anthropogenic pollutants. We performed a spatiotemporal characterization of the microbiome of two Moroccan lagoons, Marchica and Oualidia, both classified as Ramsar sites, the former on the Mediterranean coast and the latter on the Atlantic coast. We investigated their microbial diversity and abundance using 16S rRNA amplicon- and shotgun-based metagenomics approaches during the summers of 2014 and 2015. The bacterial microbiome was composed primarily of *Proteobacteria* (25%-53%, 29%–29%), *Cyanobacteria* (34%–12%, 11%–0.53%), *Bacteroidetes* (24%–16%, 23%–43%), *Actinobacteria* (7%–11%, 13%–7%), and *Verrucomicrobia* (4%–1%, 15%-14%) in Marchica and Oualidia in 2014 and 2015, respectively. Interestingly, 48 strains were newly reported in lagoon ecosystems, while eight unknown viruses were detected in Mediterranean Marchica only. Statistical analysis showed higher microbial diversity in the Atlantic lagoon than in the Mediterranean lagoon and a robust relationship between alpha diversity and geographic sampling locations. Furthermore, *Synechococcus* representatives recruited a higher number of reads from the 16S rRNA in Marchica in comparison to Oualidia, where we identified 31 *Synechococcus* oligotypes that clustered into 10 clades with different distribution patterns. This first-ever metagenomics study on Moroccan aquatic ecosystems enriched the national catalog of marine microorganisms that could be used as candidates for bioindication properties, biomonitoring potential, biotechnology valorization, biodiversity protection, and lagoon health assessment.

Key words: Metagenomics; marine microbiome; moroccan lagoons; Marchica, Oualidia; biodiversity; bioindicators.

Abbreviations

OSD:	Ocean Sampling Day
DNA:	Deoxyribonucleic Acid
OTU:	Operational Taxonomic Unit
RAMSAR:	Wetlands of International Importance
Micro B3:	Marine Microbial Biodiversity, Bioinformatics, Biotechnology
mRNA:	Messenger Ribonucleic Acid
rRNA:	Ribosomal Ribonucleic Acid
BP:	Base Pair
NGS:	Next Generation Sequencing
HTS:	High Throughput Sequencing
DO:	Dissolved Oxygen
EC:	Electronic Conductivity
NO3-:	Nitrate
NO2-:	Nitrite
BOD:	Biological Oxygen Demand
COD:	Chemical Oxygen Demand
R1:	Forward Read
R2:	Reverse Read
VAMPS:	Visualization and Analysis of Microbial Population Structures

List of Figures

Figure 1. Marchica lagoon, northeastern Morocco.....32

Figure 2. Oualidia lagoon, Northwest Morocco.....32

Figure 3: Illumina preprocessing workflow.....34

Figure 4. Domain level classification of all metagenomic reads using MG-RAST (a) Marchica lagoon (b) Oualidia lagoon.....38

Figure 5. Cross comparison of 16S rRNA amplicon sequences from selected abundant high-level bacterial taxa from Marchica and Oualidia metagenomes.....40

Figure 6. Cross comparison of 16S rRNA amplicon sequences from 5 selected abundant high-level bacterial taxa from Marchica and Oualidia metagenomes to several saline metagenomes using R.....40

Figure 7. Proportion of reads mapping to viral contigs in OSD metagenomes. Legend: OSD24: Marchica lagoon.....47

Figure 8. Heatmap of viral sequence abundance (when covered at >75%) across OSD samples. Only metagenomes with at least one viral sequence covering >75% were included.....48

Figure 9. Functional category hit distribution annotated using MG-RAST Subsystems classification in Marchica lagoon (OSD24).....52

Figure 10. Functional category hit distribution annotated using MG-RAST Subsystems classification in Oualidia lagoon (OSD91).....52

Figure 11. Top functional gene abundance in Marchica using Prokka.....53

Figure 12. Top functional gene abundance in Oualidia using Prokka.....53

Figure 13. Correlation between functions and physicochemical parameters.....54

Figure 14. Cross comparison of 31 *Synechococcus* oligotypes in Marchica and Oualidia lagoons in 2014 and 2015.....57

Figure 15. Phylogenetic tree constructed from V4–V5 16S rRNA gene amplicon sequence representatives of the *Synechococcus* clade and representative sequences for the most abundant oligotypes, both aligned using Muscle (version 3.8.4, [169]). Shown is a neighbor-joining rooted tree generated using the Geneious Prime 2021.1 software package (Biomatters Ltd, Auckland, New Zealand).....59

Figure 16. A network analysis of the oligotypes present at each site was performed using Gephi [168]. Each dot indicates an oligotype present in at least one sample, and each edge on the network connects an oligotype to one or more samples. Colored circles represent oligotypes present in only one, two, three, or all samples.....63

Figure 17. Principal component analysis of *Synechococcus* oligotype relative abundance. The plot is generated using the relative abundance of each oligotype, T: Temperature, S: Salinity, and NO3-: Nitrate. Each point represents an oligotype. Colors represent the year of sampling; red for 2014 and blue for 2015. The shape of the point indicates the sampling site; rounded points refer to the Marchica Lagoon, and triangles refer to Oualidia. Circles represent the normal distribution of oligotypes; the red circle refers to 2014, and the blue circle refers to 2015.....64

List of tables

Table 1. Physical and chemical data measured in waters of Marchica and Oualidia lagoons at the time of sampling.....31

Table 2. Number of raw DNA sequence reads (R1 Forward and R2 Reverse) obtained from each Sterivex filter for both amplicons from Marchica and Oualidia water lagoons.....35

Table 3. Number of DNA sequence reads that passed through each step in the DADA2 tool pipeline.....36

Table 4. 16S rRNA amplicon sequences from selected highly abundant sequences assigned to a phylum for both sampling campaigns (2014 and 2015) in both lagoons.....41

Table 5. Microbial alpha diversity indices; normalization: none; counts min/max: 0–100%.....42

Table 6. 16S metagenome sequences assigned to a strain.....44

Table 7. 18S SSU rRNA amplicons from selected highly abundant sequences assigned to a phylum for both sampling campaigns (2014 and 2015) in both lagoons.....49

Table 8. 18S metagenomic rRNA from selected highly abundant sequences assigned to a genus for both sampling campaigns in 2014 and 2015 in both lagoons.....49

Table 9. Oligotyping details for V4-V5 data.....56

Table 10. Hypervariable regionsV4-V5 Oligotype observation matrix.....58

Table 11. Key database for the 31 Synechococcus oligotypes including their accession number on NCBI, source, clade or subclade affiliations, and sequence length.....60

Table 12. Percent sequence similarity between V4 and V5 representative sequences for each oligotype.....62

Table of Content

Dedication.....i
Acknowledgments.....ii
Résumé.....iv
Abstract.....v
Abbreviations.....vi
List of figures.....vii
List of tables.....ix
Table of content.....x
List of publications.....12
Chapter 1: Introduction.....13
1.1. Thesis Objectives.....15
1.2. Organizations and Contributions.....16
Chapter 2: Marine microbial metagenomics.....18
2.1. Marine ecosystem.....18
2.2. Marine microorganisms.....19
2.2.1. Example of marine microdiversity : Case of *Synechococcus*.....21
2.3. Coastal lagoons.....23
2.4. Moroccan coastal lagoons.....24
2.4.1. Marchica lagoon.....24
2.4.2. Oualidia lagoon.....25
2.5. Methods to study lagoon microbiomes.....26
2.5.1. Metagenomics.....26
2.5.2. Next generation sequencing.....27
2.5.3. Taxonomic characterization.....28
2.5.4. Functional analysis.....29
Chapter 3: Materials and Methods.....31
3.1. Sample collection.....31
3.2. 16S/18S Amplicon Sequencing.....32
3.3. Preparation of Shotgun Libraries for Metagenomic Samples and Sequencing.....33
3.4. Sequence data pre-processing.....33
3.5. Data processing and Statistical Analysis.....34

Chapter 4: Taxonomic Profiling of Microbial Communities in Marchica and Oualidia lagoons. 39
4.1. Environmental parameters..... 39
4.2. Community Composition..... 39
4.3. Species Diversity..... 41
4.4. Bacterial Diversity..... 42
4.5. Viral signature..... 47
4.6. Eukaryotic Diversity..... 48
Chapter 5: Functional Profiling of Microbial Communities in Marchica and Oualidia lagoons.. 50
5.1. Microbial DNA Sequences assembly and Annotation..... 50
5.2. Analysis Statistics..... 50
5.3. Top Functional Genes..... 51
5.4. Correlation between functions and physicochemical parameters..... 53
Chapter 6: Spatiotemporal patterns of Synechococcus oligotypes in Moroccan lagoonal environments..... 55
6.1. Sampling, data processing, and oligotyping analysis..... 55
6.2. Detection of Synechococcus oligotypes in environmental samples..... 57
6.3. Distribution of Synechococcus oligotypes..... 62
6.4. Environmental variables influencing oligotypes diversity..... 64
Discussion..... 66
Conclusion..... 77
References..... 79

List of Publications

-Kopf, A.; Chaouni, B.; Amaral-Zettler, L. et al. (2015). The ocean sampling day consortium. *GigaScience.* 4:27. DOI: 10.1186/s13742-015-0066-5.

-Chaouni, B.; Hamamouch, N.; Amaral-Zettler, L. et al. (2022). Moroccan Lagoon Microbiomes. *Water,* 14, 1715. DOI: https://doi.org/10.3390/w14111715.

-Chaouni, B.; Idrissi Azami, A.; Amaral-Zettler, L. et al. Spatio-temporal patterns of *Synechococcus* oligotypes in Moroccan lagoonal environments. *Sci Rep* 13, 110 (2023). https://doi.org/10.1038/s41598-022-27263-y.

Other publications:

-Khayi, S.; Chaouni, B.; et al. (2020). First draft genome assembly of the Argane tree (Argania spinosa). *F1000 Research*. DOI: https://doi.org/10.12688/f1000research.15719.2.

-Chaouni, B.; Idrissi Azami, A.; et al. (2021). Genomic evidence of multiple SARS-CoV-2 introductions into Morocco. *F1000 Research.*https://doi.org/10.12688/f1000research.24810.2.

-Ghazal, H.; **Chaouni, B.** et al. (2021). Plant genomics in Africa: Present and prospects. *The plant journal*. DOI: https://doi.org/10.1111/tpj.15272.

-El Jaddaoui, I.; **Chaouni, B.** et al. (2021). A Review on Current Diagnostic Techniques for COVID-19. *Expert Review of Molecular Diagnostics*. DOI: 10.1080/14737159.2021.1886927.

-Habib, N.; **Chaouni, B.**; et al. (2021). Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2): An overview of the viral genome structure and its origins. *Moroccan Journal Of Public Health*. Vol. 1, N° 1, 10-19.

-Sehli, S.; **Chaouni, B.** et al. (2022). Gut Microbiome 16S rRNA Gene Amplicon Taxonomic Profiling of Hospitalized Moroccan COVID-19 Patients. *Microbiol Resour Announc*. 11(7):e0025622. DOI: 10.1128/mra.00256-22.

-Touijer, H.; **Chaouni, B.** et al. (2019). Thermostable Cellulases from the Yeast Trichosporon sp. *Enzyme Research*. Article ID 2790414, 6 pages. https://doi.org/10.1155/2019/2790414.

Chapter 1: Introduction

Coastal lagoons are shallow water bodies, opened to a larger body of water, and in most cases protected from it by one or more bar of sand or coral reef submerged or partly exposed parallel to the shore (Cañedo-Argüelles *et al.* 2012). This barrier helps lagoons gain significantly distinct characteristics in comparison to the adjacent sea or ocean (Ghai *et al.* 2012). In addition, three contributors are essential for their formation: the basin and its neighboring areas, the interior, and sea accumulations (Soria *et al.* 2022). Coastal lagoons in the Mediterranean are generally not influenced by significant freshwater inflows or tides since they are very low, which keeps away from diel inputs commonly observed in oceanic saltmarshes (Ghai *et al.* 2012). However, they are increasingly isolated from the sea, leading to remarkable changes in salinity, nutrients, and dissolved oxygen. Another factor of importance is the depth. Due to the shallowness of the water body, light penetrates throughout the year throughout the entire water column, nutrients are rapidly recycled, and the rate of primary productivity is high (Trombetta *et al.* 2022). The water temperature is strongly affected by the atmospheric temperature (Trombetta *et al.* 2021), which is responsible for a winter and summer seasonal difference of 25 °C (Trombetta *et al.* 2019). It is alleged that depth, water temperature, and salinity in coastal lagoons are the most important factors shaping microbial biodiversity and its seasonal succession (Ghai *et al.* 2012; Trombetta *et al.* 2021). Furthermore, coastal lagoons are widespread on the Mediterranean, the Baltic Sea coasts, and a few Atlantic zones (i.e., Morocco, France, Portugal, Uruguay, Brazil). As a consequence of their placement between open seas and land, coastal lagoons are influenced by anthropogenic pressures due to agriculture, tourism, mining, domestic sewages, industrial effluents, and aquaculture. These variations are reflected in the high biodiversity related to important environmental ecosystem heterogeneity (microbial species diversity and coexistence). Hence, understanding the microbial structure provides insight into the impact the microbiome has on the coastal lagoon ecosystem.

The marine environment is colonized by a massive diversity of microorganisms, which play a pivotal role in primary productivity, nutrient recycling, and energy production and transfer (Buchan *et al.* 2014; Dolan *et al.* 2018; Yunfeng *et al.* 2022). This diversity is under the deep influence of physicochemical factors driving seasonal succession of microbes, which go through time series of cell proliferation and structural variation throughout seasonal blooms (Needham *et al.* 2016).

Microbial diversity in coastal lagoons has become the focus of an important number of metagenomics studies, which enhanced our understanding of the remarkable complexity and biodiversity of marine microbial populations **(Ghai *et al.* 2012; Trombetta *et al.* 2022; Chaouni *et al.* 2022; Walter *et al.* 2021; Yanez-Montalvo *et al.* 2020; Águila *et al.* 2022; Venter *et al.* 2004; Rusch *et al.* 2007; Pesant *et al.* 2015)**. However, coastal lagoons remain undersampled since they present many challenges, including their weather, dynamic turbulence, high amount of nutrients, water mixing, and closeness. Today, the composition of *Cyanobacteria* and their contribution to atmosphere and ocean oxygenation **(Mackey *et al.* 2017; Falkowski *et al.* 2022)**, the widespread abundance of SAR11 species **(Eggleston *et al.* 2016)**, vertical and horizontal patterns in bacterial diversity **(Li *et al.* 2018)**, and many other studies have well illustrated the microdiversity in the lagoon and marine world.

Understanding the microbiome of coastal lagoons is a flourishing field of research within the area of marine science. Currently, researchers are heavily focused on finding essential or "core" members of the microbiome **(Tan *et al.* 2015; Shade *et al.* 2012)**. A pure culture of microorganisms is needed to obtain an answer to simple questions, for instance, 'who is there" using the golden marker 16S rRNA gene or investigating the microbiomes as a whole to understand what functions they are capable of **(Kopf *et al.* 2016)**. Furthermore, the marine environment and its microbiome are undergoing major changes due to climate and anthropogenic effects **(Doney *et al.* 2012)**.

We chose two Moroccan coastal lagoon ecosystems for their importance, namely, Marchica and Oualidia, to describe their extant microbiome using metagenomics. Both lagoons are classified as Ecological and Biological Sites of Interest (SIBE) and Ramsar sites (Convention on Wetlands of International Importance). Oualidia is the site of intensive aquaculture production of the Pacific oyster and a migratory stopover for birds between Europe and Africa, whereas Marchica is the largest, most attractive lagoon of the south shore of the Mediterranean Sea in Morocco and a model of sustainable development. Both lagoons have suffered profound anthropogenic pressures from economic and agricultural actions in adjacent zones. Ostracods are the most famous bioindicators of pollution in Marchica **(Ruiz *et al.* 2006)**. Most conducted studies have only discussed phytoplankton composition. Phytoplankton blooms were contributed to by *Chaetoceros*, *Pseudonitzschia*, and *Nitzschia longissima*. Their abundance and seasonality characterized Marchica as a highly eutrophicated lagoon **(El Madani *et al.* 2011)**. The waters of the Oualidia Lagoon were dominated by diatoms, ranging from 70% to 98% of the algal

community. However, dinoflagellate relative abundance increased after small diatom cells diminished **(Natij *et al.* 2014)**. Lagoon–ocean exchanges and ecological factors directly influenced the richness and diversity of planktonic algae. In addition, the distribution of different species in the lagoon was directly impacted by the hydrodynamics of the lagoon and its shallow water conditions **(Natij *et al.* 2014)**. Moreover, in Oualidia, harmful species such as *Prorocentrum*, *Dinophysis*, and *Alexandrium* were detected at low concentrations and could cause toxicity accumulation in oysters **(Kopf *et al.* 2015)**.

As part of the Ocean Sampling Day project (OSD) to sample the world's marine environment **(Kopf *et al.* 2015)**, we followed both 16S-amplicon and WGS metagenomic sequencing approaches to characterize the microbial diversity of the Marchica and Oualidia Moroccan lagoons and compared them with neighboring open sea marine environments. The sampling in Marchica and Oualidia was performed on 21 June 2014 and 2015, the boreal summer solstice. This is the first time such a metagenomics approach has been applied in Morocco to study a marine ecosystem and initiate Morocco's marine microbial inventory. We hypothesize that the operational taxon unit (OTU) distribution differs between the 2014 and 2015 sampling campaigns; it is influenced by geography, and taxa diversity correlates with environmental parameters.

1.1. Thesis Objectives

Morocco is considered a hotspot of biodiversity waiting to be explored, with so few reports having been published on its microbial community structure using microbial culture. Therefore, in this study, we aim to provide a detailed explanation of the composition and metabolic potential of the microbial communities in Moroccan coastal lagoons using a metagenomics approach.

We chose coastal lagoons as a study model, namely, Marchica located on the Mediterranean coast and Oualidia on the Atlantic. By studying these lagoons, we cover the most representative ecological types of Moroccan coastal lagoons. These latter are both classified as ecological and biological sites of interest and are of intense tourism activities because of their geographical situation and economic interest. Oualidia is considered the "Oysters capital" of Morocco, while Marchica is designated as the largest, most beautiful, important lagoon of the south shore of the Mediterranean Sea and a model of sustainable development.

The objectives of this study are as follows:

i. Identify the microbial communities present in the marine Moroccan lagoons surface waters

ii. Study of the functional capabilities of microbial communities in Moroccan lagoons

iii. Establishing potential variation patterns in microbial communities across environmental and spatial gradients in Moroccan lagoons

The results obtained will contribute to a better understanding of the microbes that make up the communities within these coastal waters as well as their roles in the environment. This study will also allow us to gain insight into the relationship between microbial community patterns and geographical location. Furthermore, it promises the potential of discovering novel genes or activity that will contribute to existing models of study.

1.2. Organizations and Contributions

In chapter 1, we call attention to the importance of studying the marine environment and their microbiome, and we present the thesis's main objectives.

In chapter 2, we introduce the background concept that is used in this thesis. It includes five sections: marine ecosystems, marine microorganisms and their diversity, coastal lagoons, Moroccan coastal lagoons (e.g., Marchica and Oualidia), methods to study the lagoon microbiome and bioinformatics analysis.

In chapter 3, we present the general pipeline of the study, from sampling to DNA extraction, shotgun and amplicon sequencing, data and statistical analysis.

In chapters 4 and 5, we describe the extant microbiota of two Moroccan coastal lagoons, Marchica in the Mediterranean Sea and Oualidia in the Atlantic. The aim was to unmask the microbial diversity in an ecosystem where such information is scarce. These experiments suggest environmental factors that influence Moroccan marine microbial diversity. It also pointed to few microorganisms to be used as bioindicators for fragile systems such as lagoons.

In chapter 6, since the results in chapter 4 showed a dominance of the genus *Synechococcus* in the Marchica lagoon, we tried to identify *Synechococcus* clades in both lagoons using 16S rRNA gene oligotyping and all along with their seasonal patterns and spatial distribution. The

aim of this study is to enhance our understanding of the environmental and ecological factors that influence patterns of *Synechococcus* microbial community composition over space and time in lagoon waters.

Chapter 2: Marine microbial metagenomics

2.1. Marine ecosystem

Seas and oceans cover more than 70% of the Earth's surface and represent one of the largest living ecosystems on Earth **(Charette *et al.* 2010; Das *et al.* 2006)**. They host the majority of its biomass and contribute significantly to all global cycles of matter and energy. All life on Earth most likely originated from microbes in the sea.

This massive biome, stratified by physical conditions such as depth, salinity and temperature, gives birth to an extensive community further segregated into marine habitats, such as coral reefs or hydrothermal vents, with respect to their specific ecologies **(Gray *et al.* 1997)**. As a unit, however, marine communities play a shared role in regulating the major processes that take place in the ocean. With different groups of organisms carrying out their respective roles, a tightly interconnected web is established that runs the marine ecosystem like a well-oiled machine. Having the ocean account for 70% of the planet, these biogeochemical processes hence bear impact on a global scale and play a crucial role in its productivity—thus dubbing it the Earth's 'invisible rainforest' **(Falkowski *et al.* 2002; Karl *et al.* 2002)**.

In today's marine ecosystems, following billions of years of evolution, microbes such as bacteria, archaea, viruses, fungi and protists (including microalgae) dominate the living biomass. Recent rapid developments in molecular ecology, metagenomics and ecological modeling illustrate that microbes represent the most important biological group on Earth in terms of phylogenetic and functional diversity. In addition, interdisciplinary research has uncovered new and unexpected roles of microbes in the biogeochemical cycling of carbon, nitrogen, silica and iron and many other (trace) elements in our seas and oceans.

Marine microorganisms produce the organic matter and oxygen required to sustain life and facilitate the storage, transport, and turnover of key biological elements. Thus, microorganisms are the foundation of life and are of critical importance to the habitability and sustainability of our planet **(Glöckner *et al.* 2012)**.

The enormous microbial diversity also gives rise to a largely untapped amount of genetic information, bioactive compounds and biomaterials that could deliver important benefits and applications of societal interest, for example, to improve medical treatments, fisheries and

aquaculture applications, the supply of energy and for the development of industrial products and processes. However, despite the clear importance of marine microbes and the major opportunities they present, very little is known about marine microbial diversity, the enormous array of microbial types and their ecological functions and interactions. Moreover, the vast majority (90-99%) of marine microorganisms cannot be cultured under standard laboratory conditions, and their growth and physiology cannot, therefore, be studied in a way that has proven so successful throughout the 20th century for medically important microorganisms **(Glöckner *et al.* 2012)**.

With deeper insight into the structure and metabolism of these communities, in this thesis, we will thus be able to fill existing knowledge gaps and contribute to the advancement of existing study models.

2.2. Marine microorganisms

A large percentage of primary production contributes to the DOM in the ocean, a form of carbon readily utilized by its inhabitants. This source of carbon, however, typically comes in the form of particulate organic matter (POM) and must be made available to other organisms. This process is largely regulated by prokaryotes in the ocean, which produce enzymes capable of breaking down POM into simpler forms, consequently providing energy to other trophic levels **(Azam *et al.* 2007)**. Prokaryotes, made up of two major domains, Bacteria and Archaea, are the most abundant living organisms in the ocean, with more than 50% of the population inhabiting the pelagic layer **(Karner *et al.* 2001)**. One of the most abundantly found groups of bacteria in surface waters is the SAR11 clade, which makes up an estimated 18% of the total biomass found in upper ocean environments **(Viklund *et al.* 2011)**.

Other common surface water colonizers include *Prochlorococcus* and *Synechococcus*, both members of the *Cyanobacteria* class, which are light-harvesting prokaryotes typically found in photic regions **(Viklund *et al.* 2011; Malmstrom *et al* 2005; Zubkov *et al.* 2009)**. Archaea, on the other hand, are typically present in much smaller numbers but are found more abundantly in the deeper layers of the ocean **(Signori *et al.* 2014; Luria *et al.* 2014)**. However, certain groups are known to thrive in surface water environments, such as the Euryarchaeota clade **(Brown *et al.* 2009)**. Although extensive generalization cannot yet be made, prokaryotes have been observed to exhibit regional variation in terms of genomic and physiological properties. For example, coastal

waters appear to harbor a higher count of certain groups of bacteria, such as *Alphaproteobacteria* from the *Roseobacter* clade, compared to open ocean environments.

To successfully assimilate into these coastal environments, these microbes were found to have developed a form of genomic adaptation **(DeLong *et al.* 2005)**. One such example would be *Silicibacter pomeroyi*, a member of the *Roseobacter* clade, which possesses the genes necessary to metabolize not only dissolved organic carbon but also reduced inorganic compounds, such as carbon monoxide and sulfur—both of which were confirmed to be important substrates for its metabolism. Additionally, *S. pomeroyi* metabolism was also found to encode multiple transporters capable of the uptake of osmolytes such as DMSP and glycine betaine, both of which have been presumed to originate from marine algae, which are abundant in coastal environments **(Moran *et al.* 2004)**.

These microbes play an essential role not only in the cycling of main elements but also in the regulation of biogeochemical processes that play an important role in shaping other environmental factors such as atmospheric composition and climate trends **(DeLong *et al.* 2005)**. For example, biogeochemical processes involving the exchange of sulfur via the ocean-atmosphere boundary have been known to have a direct effect on climate patterns **(Strom *et al.* 2008)**.

Marine microbial communities involved in sulfur metabolism were found to produce the volatile compound dimethyl sulfide (DMS) as a product of the breakdown of precursor dimethylsulfoniopropionate (DMSP), which is commonly found in planktonic algae **(Strom *et al.* 2008; Reisch *et al.* 2011)**. The DMS molecules released into the atmosphere act as cloud condensation nuclei (CCN), which consequently promotes cloud formation and reduces the amount of solar radiation that reaches the Earth's surface **(Vallina *et al.* 2007; Welsh *et al.* 2000)**. Several species of marine microbes have been found to mediate the reaction by producing the enzymes necessary to cleave DMSP into the desired gas product **(Vila-Costa *et al.* 2010**), which can be a useful model for climate change studies. One such example would be the Roseobacter lineage, which was found to exhibit a positive correlation to DMSP dynamics in a study by **(González *et al.* 1997)**.

Through the analysis of heterotrophic bacterial communities within an algal bloom in North Atlantic waters using culture-independent methods, the percentage of Roseobacter rDNA was found to increase along with concentrations of DMSP and total dimethylated sulfur compounds

in surface waters and was hypothesized to be involved in the cycling of organic sulfur **(Fuhrman *et al.* 1993)**. While there still remains much to explore, the potential phylogenetic and metabolic diversity of these prokaryotes has been said to far exceed that of all eukaryotes combined. However, there still exists a large knowledge gap as to what makes up these microbial communities, which specific processes are more dominant, and how these factors vary between the different communities and environments. Recognizing this, scientists have thus focused on the taxonomic and metabolic diversity of these communities in an effort to gain a better perspective of the marine ecosystem as a whole and, to a larger extent, its effects on the variability of the global environment **(Fuhrman *et al.* 1993)**.

2.2.1. Example of marine microdiversity : Case of *Synechococcus*

Synechococcus is an important group of *Cyanobacteria* that contribute to global biogeochemical cycles **(Mazard *et al.* 2012)**. They offer an attractive system to explore bacterial taxa relationships, distribution, coexistence, ecology, and evolution **(Scanlan *et al.* 2002)**. The widespread distribution of this group can be attributed to its high degree of genetic diversity **(Scanlan *et al.* 2009)**. *Synechococcus* comprises some of the major forms of C*yanobacteria* that inhabit marine and freshwater environments **(Scanlan *et al.* 2009)**. Although their genetic diversity has been documented for these ecosystems, there is scant knowledge of the biodiversity, abundance, and distribution of this genus in coastal lagoons. The latter environments are highly productive and valued ecosystems while being morphologically and ecologically complex. Coastal lagoons are subject to more variable environmental conditions than the open sea. Due to relative isolation from the sea and their location within a hydrological catchment, these lagoons are more susceptible to changes in physicochemical parameters, leading to increasing salinity, a decrease in nutrient availability and concentration, and light spectral intensity. These factors influence the adaptation strategies of photosynthetic microorganisms, including *Synechococcus* species.

The emergence of next-generation sequencing approaches has provided astonishing insight into microbial biodiversity. Notably, the use of the gold standard gene marker for 16S rRNA now enables microbial diversity assessment across the globe in distinct seasons and locations depending on different environmental conditions **(Ahlgren *et al.* 2012)**. Recently, the use of high-resolution methods, such as oligotyping, has allowed researchers to investigate unexplained diversity within operational taxonomic units and uncover ecologically and biologically distinct

taxa **(Eren *et al.* 2013)**. Based on these techniques, seasonal and geographical behavioral patterns of *Synechococcus* strain abundances and distributions have been described in several studies **(Xia *et al.* 2015; Hunter-Cevera *et al.* 2016; Mackey *et al.* 2017; Nagarkar *et al.* 2021;** Needham ***et al.*** 2016**)**. Some studies reported higher abundances in the summer season (**Xia *et al.* 2015; Hunter-Cevera *et al.* 2016; Mackey *et al.* 2017**), whereas other studies have shown temporary blooms in spring or summer under eutrophic conditions **(Nagarkar *et al.* 2021; Robidart *et al.* 2012)**. In both cases, correlations between physicochemical parameters such as temperature, salinity, and nitrate concentration and *Synechococcus* abundances were reported **(Robidart *et al.* 2012; Rajaneesh *et al.* 2013; Flombaum *et al.* 2013; Paulsen *et al.* 2016)**. Spatial differences were also observed; *Synechococcus* strains were more abundant in coastal waters than in estuaries. Surprisingly, high *Synechococcus* cell abundance was observed even in polar oceans, which are thought to be devoid of them **(Paulsen *et al.* 2016)**. Twenty *Synechococcus* clades with different patterns of distribution have been described **(Hunter-Cevera *et al.* 2016)**. Indeed, many *Synechococcus* clades, including Clades I and IV, have been observed in temperate or polar waters, as well as coastal and higher latitude regions. The question of consistent co-occurrence (for instance, between Clades I and IV) remains partially unanswered, and the role of environmental parameters remains poorly understood.

Synechococcus strains have been classified into three major subclusters (5.1, 5.2, and 5.3) based on 16S rRNA gene phylogeny. Marker genes used to study the diversity of marine *Synechococcus* were essential to describe an important number of subclades, providing more accurate resolution, such as the internally transcribed spacer (ITS), catalase-peroxidase gene (cpeA), nitrate reductase gene (narB), global nitrogen regulator gene (ntcA), ribulose bisphosphate carboxylase large chain gene (rbcL), DNA-directed RNA polymerase subunit beta gene (rpoC1), and especially petB gene coding for cytochrome b6f. This latter gene helped identify more than thirty subclades **(Scanlan *et al.* 2009; Ahlgren *et al.* 2012)**. Furthermore, these markers divided the identified subclusters into more than twenty distinct genetic clades **(Ahlgren *et al.* 2012)**. Marine *Synechococcus* clades are significantly diverse in terms of depth, temperature, and nutrient availability requirements. Clade II is found in tropical offshore environments, with 'hot spots found in the nutrient-rich coastal upwelling of Morocco and causing seasonal blooms in the Red Sea and the Gulf of Aqaba **(Mazard *et al.* 2012)**. Clade III is predominantly found in tropical and subtropical warm waters **(Mazard *et al.* 2012)**, whereas Clades I and IV occur in nutrient-rich, temperate, and cold environments either nearshore or

offshore **(Mazard *et al.* 2012)**. Most of these clades (I, II, III, IV, and CRD1) belong to subcluster 5.1. Marine subcluster 5.2 has been observed in nearshore, coastal, and estuarine environments **(Mazard *et al.* 2012)**. Different clades may co-occur in similar ecological niches, with reports of as many as six clades found at once **(Ahlgren et al. 2012)**. Co-occurrence patterns are observed in coastal waters. Clades II and III cooccurred in the California Current during the spring prebloom period, while clades I and IV predominated when the bloom itself occurred. Clades V, VI, and X coexist in the Red Sea during transitional periods between mixing and stratification **(Mackey *et al.* 2017)**. Abundant *Synechococcus* clades are impacted by limiting factors such as light, nutrient availability, temperature, or viral infections **(Mackey *et al.* 2017)**. These factors might also change across seasons and over time scales of environmental changes, leading to clade coexistence **(Hunter-Cevera *et al.* 2016)**. When less abundant, they may persist at low but stable levels **(Pedrós-Alió *et al.* 2006; Sogin *et al.* 2006; Fuhrman *et al.* 2009)** and serve as reservoirs of genetic diversity **(Mackey *et al.* 2017)**.

2.3. Coastal lagoons

Coastal lagoons are shallow water bodies separated from the sea by a barrier, connected at least intermittently to the sea by one or more restricted inlets and usually oriented parallel to the shore. The formation of the barrier is crucial, as it allows lagoon waters to acquire significantly different characteristics compared to the nearby seawater. Mediterranean coastal lagoons are commonly not affected by significant tidal influences, as tides in the Mediterranean Sea are very low. This avoids the diel inputs of seawater that are common in oceanic salt marshes. Because of the relatively increased isolation from the sea and their location within a hydrological catchment, these lagoons also become more susceptible to changes in salinity, dissolved oxygen, and nutrient content, largely owing to the increased effect of evaporation in a restricted area, leading to increased salinity and deposition of various salts (e.g., calcium carbonate), as well as to the strongest influence of the surrounding land. Because of the high population density in coastal Mediterranean areas, these lagoons are usually impacted by agricultural, mining, tourism and general developmental activities, leading to the lagoon becoming a common sink for a wide variety of waste materials. These differences are also reflected in the organisms that inhabit these lagoons, which may largely contrast with those of the nearby marine environment. Such lagoons are very common environments in flat areas along the Mediterranean coasts and may range from small to very large in size **(Ghai *et al.* 2010)**.

2.4. Moroccan coastal lagoons

2.4.1. Marchica lagoon

The Nador lagoon also, known as the Marchica lagoon (from its Spanish name Mar-chica, “the little sea”), is the second largest lagoon complex of northern Africa, the broadest paralic environment of Morocco and the only lagoon located along the Mediterranean coast of this country **(Ruiz *et al.* 2006)**. This lagoon is protected by an “North West - South East" elongated spit (25 km length), only interrupted by an artificial inlet limited by two jetties that communicate it with the Mediterranean Sea. Most of the areas present water depths between 3 and 8 m, with internal hydrodynamics dominated by marine waters passing through the artificial inlet. The tidal regime is microtidal (0.35 m) and semidiurnal **(González-Regalado *et al.* 2005)**. The salinity range is typical of marine waters (38-40.9 mg/L), with low salinity values and high nutrient concentrations near the former wastewater treatment station of Nador **(Ruiz *et al.* 2006)**. Dissolved oxygen contents are highly variable (3.9–10.5 ml/L), with the highest values in the artificial inlet **(González-Regalado *et al.* 2005)**. The latter also causes salinity changes in the adjacent areas **(Ruiz *et al.* 2006).** This lagoon is subjected to increasing anthropogenic pressure owing to economic activities in neighboring areas. Heavy metals occur in the lagoon sediments largely due to anthropogenic activities in the area as urban effluents. Pollution indices show that the heavy metals pose an ecological risk and indicate that the Nador lagoon is from moderately to considerably polluted. Water quality parameters (salinity, pH, nutrients) of the dominant marine flows are altered by local fecal water effluents, urban discharges, and sewages derived from wastewater, residues originating in a slaughterhouse, as well as pesticides and fertilizers from the adjacent agricultural fields. The geochemical analyses carried out in surficial sediment samples showed very high concentrations of all metals studied near an old iron mine and moderate contents between Nador and its treatment station **(Maanan *et al.* 2014)**.

Ostracods are good bioindicators of these environmental impacts, with the presence of a highly brackish assemblage in quieter, more confined areas or the appearance of opportunistic species under hypoxic conditions **(Ruiz *et al.* 2006)**. Phytoplankton studies in the Marchica lagoon showed, among seven identified phytoplankton classes, a domination of *diatom and dinoflagellate* species. Frequent phytoplankton blooms are contributed by one to three species (*Chaetoceros*, *Pseudonitzschia*, *Nitzschia longissima*). The abundance and seasonality of

phytoplankton characterized the Nador lagoon as a highly eutrophicated environment **(Elmadani *et al.* 2011)**.

2.4.2. Oualidia lagoon

The Oualidia lagoon located on the Atlantic coast of Morocco is 7 km long, on average 0.5 km wide, and it exchanges waters with the ocean through a major inlet of approximately 150 m width and 2 m depth and, during spring tides, with a secondary, shallower inlet of approximately 50 m in width. An internal delta with a surface area of approximately 0.2 km^2 is normally found close to the inlet **(Carruesco, 1989)**. The hydrodynamics of the lagoon are subject to the rhythm of semidiurnal tides and influenced by coastal upwellings. The sedimentary cover has a spatial variation, sandy near the pass to muddy upstream. Freshening is observed at low tide upstream through small freshwater resurgences. The measurements of temperatures and salinities indicate a clear influence of the marine waters downstream of the lagoon and a decrease in salinity toward the upstream. This is related to the presence of many freshwater sources of continental origin and submarine freshwater resurgences inside the lagoon, which play a dominant hydrological role **(Hilmi *et al.* 2005)**. Hypersalinity was not recorded except in some puddles located in the schorre **(Bidet *et al.* 1982)**. Apparently, there are no rivers discharging into the lagoon, but several authors have mentioned the existence of underground freshwater seepage somewhere within the lagoon **(Beaubrun *et al.* 1989; Carruesco, 1989)**.

The breeding of the Japanese oyster *Crassostrea gigas* is practiced in five parks distributed along the lagoon, while the collection of *Ruditapes decussatus* is carried out in the northern part of the lagoon. The study of phytoplankton showed that *diatoms* were the dominant organisms at most times (70% to 98% of the phytoplankton population). Following the decline in small diatom species, *dinoflagellates* became dominant. Phytoplankton diversity and richness were more dependent on oceanic supply and environmental factors than on freshwater supplies into the lagoon. Lagoon hydrodynamics and low depth favor the homogenization of the vertical and horizontal distributions of species in lagoon waters. Potential harmful phytoplankton species identified belong to different genera, including *Alexandrium*, *Prorocentrum*, and *Dinophysis.* They occur at low concentrations and cause toxicity in oysters harvested from this lagoon **(Bennouna *et al.* 2000)**.

2.5. Methods to study lagoon microbiomes

2.5.1. Metagenomics

The field metagenomics can be defined as a culture-independent analysis of a 'collective set of genomes of mixed microbial communities' isolated from the environment **(Petrosino *et al.* 2009)**. Traditionally, the physiology and biochemistry of microbes were studied through conventional culture-based methods where live biomass was isolated from the environment and grown in laboratory conditions. However, this was restricted only to the more easily culturable species which are believed to make up approximately 1% of the global microbial population, resulting in a very limited description of the vast microbial communities that exist **(Jorgensen, 2006; Kennedy *et al.* 2010)**.

Since 1998, attention to metagenomics has gradually increased around the world, as witnessed by the growing number of related reports and even reviews in the PubMed database. Currently, two major types of metagenomics studies prevail in public literature databases in terms of research targets, one being the community survey that resolves community composition or dynamics using phylogenetic marker genes such as the 16S ribosomal RNA (rRNA) sequence and the other being the functional metagenomics survey that studies the metabolic potential embedded in coding genes obtained from an environment. Both types of studies usually relate ecological associations to environmental conditions.

The history of studying environmental microbes using culture-independent methods can be traced back to the application of the fosmid library, developed by DeLong's group, to clone large genomic DNA fragments into a fosmid vector maintained in an Escherichia coli surrogate host. Screening, sequencing, and analyzing those partial microbial genomes extend our understanding into previously untapped microbiological territory. The next boost to metagenomics research was the use of the whole genome shotgun (WGS) approach employing next generation sequencing to bypass the need for cloning and direct sequencing of metagenomic DNA fragments in a high-throughput manner. The massive decrease in sequencing costs in recent years further promises to dramatically increase access to genomic and metagenomic data.

Metagenomic strategies have not only provided us with near complete insights into the diversity of microbial communities, but have also given us access to in-depth studies of the metabolic pathways that are carried out within them **(Kennedy *et al.* 2010)**. Paired with the vast

array of computational analysis tools that have been made available, these new sequencing technologies provide a higher resolution description of whole communities that previously could not be achieved through traditional approaches alone **(Caporaso *et al.* 2011)**.

Metagenomic sequencing involves four main steps: the extraction of genomic DNA from a sample, NGS sequencing, and bioinformatics analysis, which consists of assembling the reads into longer sequences called contigs and then annotating and characterizing the sequences. Furthermore, it was shown that the functional potential of a microbiome was related to its environment **(Tringe *et al.* 2007)**. However, it is still difficult to link the different environmental conditions with microbial communities. Thus, the major challenge of metagenomics is how to study the impact of the environment on metabolic networks and microbial communities.

2.5.2. Next generation sequencing

For the past few decades, the field of genomics has relied heavily on traditional techniques such as Sanger sequencing and fluorescence-based electrophoresis for the purpose of genomic characterization. Additionally, known as dideoxy sequencing, the Sanger chain-termination method relies on the synthesis of a complementary DNA strand through the incorporation of 2'-deoxynucleotides (dNTPs) and fluorescently labeled 2',3'-dideoxynucleotides (ddNTPs), which serve to terminate the elongation process. Randomly terminated DNA strands are then separated by size using gel electrophoresis, after which the DNA sequence of the synthesized strand is revealed based on the terminal ddNTPs **(Morozova *et al.* 2008; Edwards *et al.* 2006)**. Since then, many sequencing technologies have built upon a similar concept, providing the option of sequencing long reads with a low error rate. However, despite the robustness and progressive advancement of the technology that it offers, modern Sanger-based sequencers have not been applicable to larger sequencing projects—its main limitations being the high cost and long processing time **(Morozova *et al.* 2001)**. As a result, new "next generation" sequencing technologies (NGS) have arrived at the forefront of genomics, offering higher, more cost-effective throughput. NGS—also known as massive parallel sequencing—creates microscale reactors by attaching DNA molecules to solid surfaces such as beads, which allows millions of sequencing reactions to run parallel to one another **(Reis-Filho, 2009)**.

NGS technologies have been made commercially available through a range of different platforms, such as 454 pyrosequencing (Roche Applied Science), SOLiD sequencing (Applied Biosystems), Illumina Solexa sequencing (Illumina, Inc.), and Heliscope single molecule

sequencing (Helicos, Inc.), each a variation of the same basic approach **(Morozova *et al*, 2008)**. In 2005, 454 Life Sciences launched the first NGS platform that relied on pyrosequencing technology, where DNA fragments are immobilized on beads and amplified through emulsion polymerase chain reaction (PCR). The company, which was eventually taken over by Roche Applied Science, introduced its most popular instrument, which was also known as 454 GS FLX Titanium later in 2008. This technology boasts longer read lengths of approximately 700 base pairs (bp) with an accuracy rate of 99.9%, as well as a significantly shorter processing time. However, despite its efficiency, the high cost of the equipment and the reagents required remains a major setback **(Liu *et al.* 2012)**.

Over the past few years, the sequencing industry has been taken over by Illumina, which offers a technology that relies on a sequencing-by-synthesis strategy. The DNA library, complete with adapter sequences, is denatured into single strands and ligated onto oligonucleotide primers that have been immobilized on the surface of a glass flow cell coated in acrylamide. The DNA strands are then put through bridge amplification in the presence of four different types of dNTPs, each fluorescently labeled, resulting in a signal captured for every dNTP molecule successfully incorporated **(Bentley *et al.* 2008)**. Sequencing instruments such as the Illumina Genome Analyzer and HiSeq2000 have brought high-throughput sequencing to new heights, with a vast array of applications having been developed on this platform **(Quail *et al.* 2012)**. Despite its relatively short read lengths, NGS has proven to be very useful in a multitude of applications, namely, in microbial identification. The inception of these new technologies has thus revolutionized not only the exploration of microbial genetic diversity but also the field of microbial ecology as a whole.

2.5.3. Taxonomic characterization

Taxonomic characterization is the science of classifying organisms into similar groups or taxa present in a sample from DNA or protein sequences; it includes classification, annotation and identification. Few methods are available for bacterial characterization; however, with the advent of bioinformatics analysis, several methods have been used. The most commonly used and known method is the 16S rRNA gene, which consists of aligning and comparing these genes with existing genes in different databases (RefSeq, GenBank, RDP, EMBL, GreenGenes, etc.). Metagenomic analysis also allows sometimes identifying the bacterial taxon from the DNA fragment with homology to genes annotated in databases or detection of phylogenetic markers in

those fragments. However, numerous fragments have not yet been identified, which is a well-known limitation of current metagenomic analysis. The 16S rRNA gene is a tool of choice for the study of phylogenetic relationships between different organisms and has been widely used in phylogenetic classification because it has the advantage of having highly conserved domains surrounding the variable domains, which can be performed using different software such as QIIME **(Caporaso *et al.* 2010)**, UPARSE **(Edgar *et al.* 2013)**, and MG-Rast **(Meyer *et al.* 2008)**. It consists of clustering the sequences into operational taxonomic units (OTUs) at a certain similarity threshold that can be fixed for each study. Then, those OTUs will be aligned to different databases to build a phylogenetic tree and taxonomic characterization. From the OTU and phylogenetic tree, the diversity can be estimated within and between samples. It provides information on species diversity, which means the measurement of species richness (how many species are there) and evenness of species (how evenly distributed the numbers of each species are) present in a sample.

2.5.4. Functional analysis

Functional analysis focuses on studying gene function directly from the extracted DNA of an ecosystem. The functional metagenome is a new approach to provide access to biological resources completely unexplored to date and to study the interaction metabolism between the host and the genes expressed by the metagenome **(Uchiyama *et al.* 2009)**. This approach allows the identification of functional genes and complete metabolic pathways and a better understanding of the mechanisms inherent to the health of the host, and it may also lead to the discovery of new genes of antibiotic resistance **(Langille *et al.* 2013)**, new pharmaceutical proteins, new enzymes secreted into the environment **(Uchiyama *et al.* 2009)**, new drug targets, etc. After sequencing, the sequences are compared to the known and present data in databases. Many software programs, such as PICRUSt (Phylogenetic Investigation of Communities by Reconstruction of Unobserved States) **(Langille *et al.* 2013)**, MEGAN (Metagenome Analyzer) **(Huson *et al.* 2007)**, MG-Rast (Metagenomics Rapid Annotation using Subsystem Technology) **(Meyer *et al.* 2008)**, HUMAnN (The HMP Unified Metabolic Analysis Network) **(Abubucker *et al.* 2012)** and IMG/M (Integrated microbial Genomes & Microbiomes), exist to assign a function to a sequence using genomic databases. One of the most common ways is to look for similarities between unknown sequences and sequences already annotated in different databases. Many databases containing functional metagenome information are available. Among the best known, UniProt (The Universal Protein Resource) (UniProt-Consortium 2012) includes several

protein databases and provides information on the sequence, structure and function of various proteins, and KEGG (Kyoto Encyclopedia of Genes and Genomes) **(Kanehisa *et al.* 2004)** is a set of databases of metabolic networks, complete genomes, enzymes, genes and proteins. The Pfam (Protein families) database **(Punta *et al.* 2012)** and TIGRFAMs **(Selengut *et al.* 2007)** use the concept of protein families. A protein family contains a set of proteins encoded by genes from the same common ancestor with similar functions. The COG (Clusters of Orthologous Groups of proteins) database **(Tatusov *et al.* 2003)** contains clusters of protein groups. Each group of proteins is composed of several orthologous proteins. Therefore, it appears that the functional metagenomic approach is fundamental to understanding how the microbiome interacts with its environment.

Chapter 3: Materials and Methods

3.1. Sample collection

Samples were collected from Marchica (N 35.11562, W 2.52803) and Oualidia (N 32.74675, W 9.036667) on 21 June 2014 and 2015, the boreal summer solstice (Figures 1 and 2). Approximately 20 L was collected using a 10% acid-washed bucket, sequentially filtered on five Sterivex filter units with a 0.22 μm pore size using a hand pump and stored at −80 °C. Metadata such as sample volume, depth, salinity, temperature, pH, latitude/longitude, dissolved oxygen (DO), turbidity, and conductivity (EC) were measured in situ using multivariate calibrated probes. Nitrate, nitrite (NO3-, NO2-), and phosphate concentrations were determined using the molecular UV/visible spectrophotometric method. The biochemical oxygen demand (BOD) for each sample was recorded for five days at 20 °C using an incubation flask (OxiTop ® Control measuring system). We measured chemical oxygen demand (COD) by applying a robust oxidizing reagent (potassium dichromate K2Cr2O7) under acidic conditions (Table 1).

Table 1. Physical and chemical data measured in waters of Marchica and Oualidia lagoons at the time of sampling.

		Units	Marchica		Oualidia	
			2014	2015	2014	2015
Parameters	Temperature	Celsius	26.5	27.2	21	20
	Salinity	ppt	35.98	35.96	27.24	29.41
	Electrical conductivity	mS	53.7	54.2	39.6	45.1
	pH	<7.6 to 8.4>	8.16	8.74	8.36	8.16
	Dissolved Oxygen	mg/L	8.33	8.50	7	7.5
	Phosphate	mg/L	0.03	0.04	0	0
	Turbidity	NTU	3.62	3.60	3	3.2
	Nitrates	mg/L	12.5	4.8	10.79	2.68
	Nitrites	mg/L	<0.3	0.11	<0.001	<0.05

Figure 1. Marchica lagoon, Northeastern of Morocco (N 35.11562, W 2.52803), (Image GeoEye from Google Earth, 2022).

Figure 2. Oualidia lagoon, Northwest of Morocco (N 32.74675, W 9.036667), (Image GeoEye from Google Earth, 2022).

3.2. 16S/18S Amplicon Sequencing

DNA extraction was performed using the isolation kit Power Water (MoBio, USA) as instructed by the manufacturer. Amplification of the 16S rRNA gene was performed using the primer pair 515F-Y (5′-GTGYCAGCMGCCGCGGTAA-3′) and 926R (5′-CCGYCAATTYMTTTRAGTTT-3′) **(Parada *et al.* 2016)**. The forward primer TAReuk454FWD1 (5′-CCAGCASCYGCGGTAATTCC-3′) and the reverse primer TAReukREV3_modified (5′-ACTTTCGTTCTTGATYRATGA-3′) were used for 18S rRNA gene amplification. Ovation Rapid DR Multiplex System 1–96 (NuGEN) was used to construct

Illumina libraries, pooled and size selected by preparative gel electrophoresis. The detailed protocol can be found under this link (https://github.com/MicroB3-IS/osd-analysis/wiki/Guide-to-OSD-2014-data_accessed on 1 December 2020). Sequencing was performed on an Illumina MiSeq using V3 Chemistry. Samples were sequenced in one MiSeq run (2 × 300 bp), generating 2 × 40,000 reads per sample.

3.3. Preparation of Shotgun Libraries for Metagenomic Samples and Sequencing

One half of the supplied DNA material per sample was sheared to approximately 500 bp fragments using a Covaris S220 sonicator. DNA was purified and concentrated by clean-up using MinElute columns (Qiagen). DNA concentrations were measured, and 100 ng (or less, if the sample contained insufficient amounts, successful libraries could be obtained from as little as 5 ng) was used to prepare Illumina libraries. Libraries were generated with the Ovation Rapid DR multiplex 1–96 system (NuGEN). Libraries were amplified using standard Illumina primers for 8 to 15 cycles with MyTaq (Bioline). Eighteen cycles were necessary to generate the library. Sequencing was performed on an Illumina MiSeq using V3 Chemistry. Samples were sequenced in eight MiSeq runs (2 × 300 bp) generating 2 × 1,000,000 reads per sample.

3.4. Sequence data preprocessing

Preprocessing is defined as the process in which a subset of the original raw sequence data is created that fulfills defined sequence quality criteria, suitable for further analyses. The workflow covers both amplicon (i.e. 16/18S rDNA) and shotgun (i.e. metagenome) data sequenced with Illumina MiSeq.

16S/18S rDNA amplicon data were demultiplexed from the sequencing center, reads were sorted, and primers were clipped using cutadapt, while adapters were clipped using trimmomatic, which released raw datasets. This version only includes initial quality control (QC) and demultiplexing. The ‘raw’ version is intended for advanced users who might either have their own preprocessing pipelines or prefer to use different tools and parameters. Later, reads were merged using Pear, trimmed by quality and filtered by length using trimmomatic, and then quality was checked using FastQC (Figure 3). After this step, we obtain workable datasets, which is the version resulting at the end of the preprocessing workflow. It will be used for all further analyses because it guarantees comparability of the results.

Metagenome (shotgun) data were already demultiplexed, adapters were clipped, and datasets were trimmed by quality, and filtered by length using trimmomatic (Figure 3).

The output of the preprocessing workflow is quality-controlled datasets that are ready for analysis (e.g., SILVAngs, MG-Traits, EMG (MG-Portal), among others).

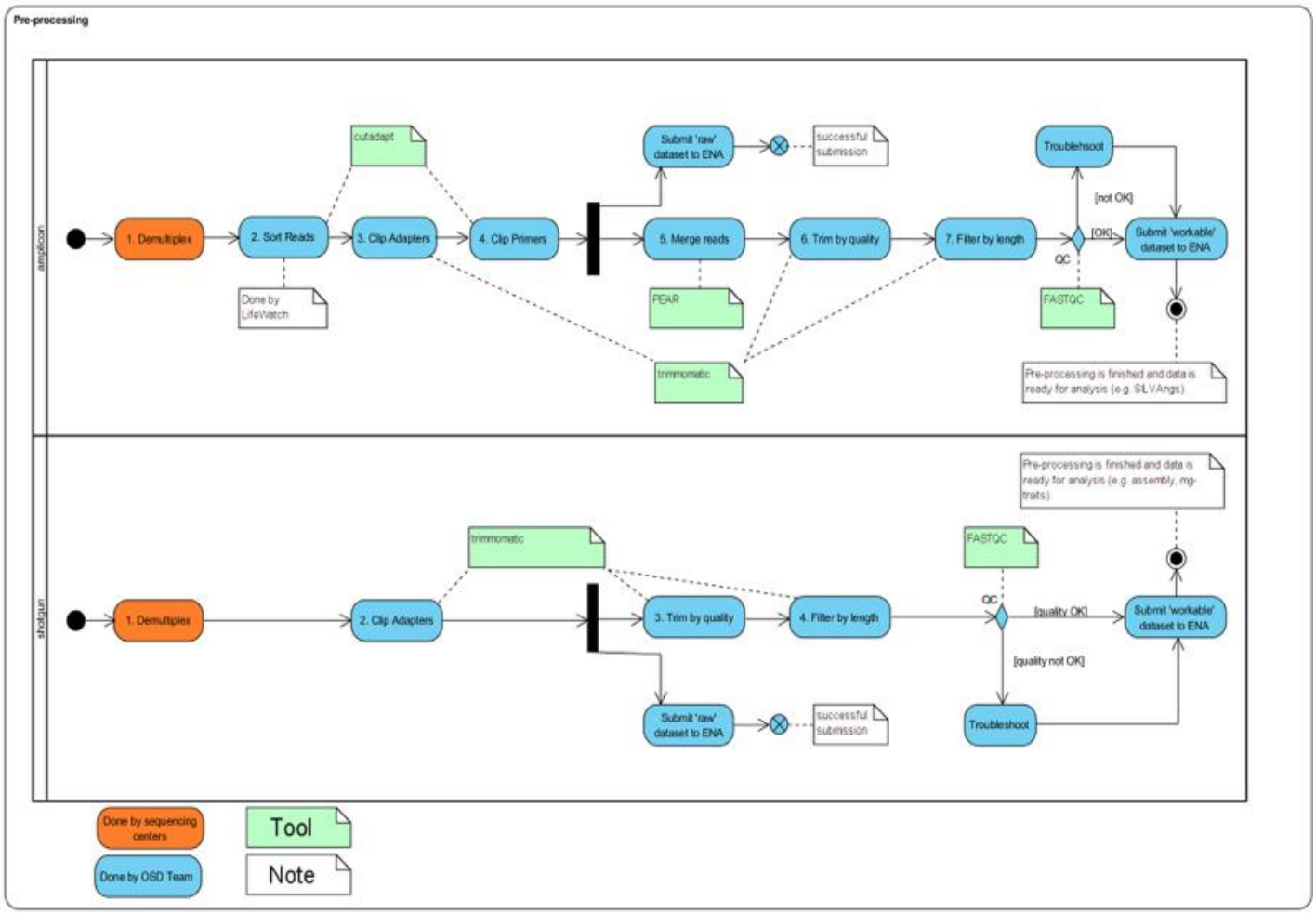


Figure 3: Illumina sequence data pre-processing workflow

Legend: For amplicon data, the output files per sample are: * raw: nonmerged, * workable: merged For shotgun data, the output files per sample are: * raw: nonmerged (used, e.g., for EMG), * workable, output files: merged (used, e.g., by mg-traits), nonmerged (used, e.g., for assemblies).

3.5. Data processing and Statistical Analysis

Sequence data were preprocessed following the OSD workflow (github.com/MicroB3-IS/osd-analysis/wiki/Sequence-Data-Preprocessing_ accessed on 1 December 2020), which

generated "workable" metagenome and amplicon fasta files. Raw files were stored at EBI (www.ebi.ac.uk/ena/data/view/ERX947554_accessed on 1 December 2020). Table 2 summarizes the number of raw 16S rRNA sequences collected from sterivex filters from both lagoons.

Amplicon sequences were processed with the VAMPS web service **(Huse *et al.* 2014)**, where taxonomy classification for 16S and 18S rRNA gene sequences was assigned in a Global Alignment for Sequence Taxonomy (GAST) proceeding **(Huse *et al.* 2008)** and SILVA rRNA gene reference database **(Quast *et al.* 2013)**. To crosscheck the identified genera, we used the DADA2 microbiome pipeline **(Callahan *et al.* 2016)**. Filtering parameters used the following settings: 200 bp for the forward (R1) and 190 bp for the reverse (R2) reads. All identical sequencing reads were combined into unique sequences, denoised, merged if the forward and reverse reads overlapped by at least 12 bases, and set as chimera-free. Similar sequences were grouped into distinct OTUs, and taxonomy was assigned to sequence variants using the SILVA reference database. Table 3 summarizes the number of reads that made it through each step in the DADA2 pipeline.

Table 2. Number of raw DNA sequence reads (R1 Forward and R2 Reverse) obtained from each Sterivex filter for both amplicons from Marchica and Oualidia water lagoons.

		Filter Size	# Reads	Total Length (Mb)	Average Read Length (bp)
Marchica	2014	0.22 µm	132,322	80.1	269
	2015	0.22 µm	76,482	46.6	271
Oualidia	2014	0.22 µm	179,448	108.2	268
	2015	0.22 µm	146,886	89.2	270
Total			535,138	324.1	

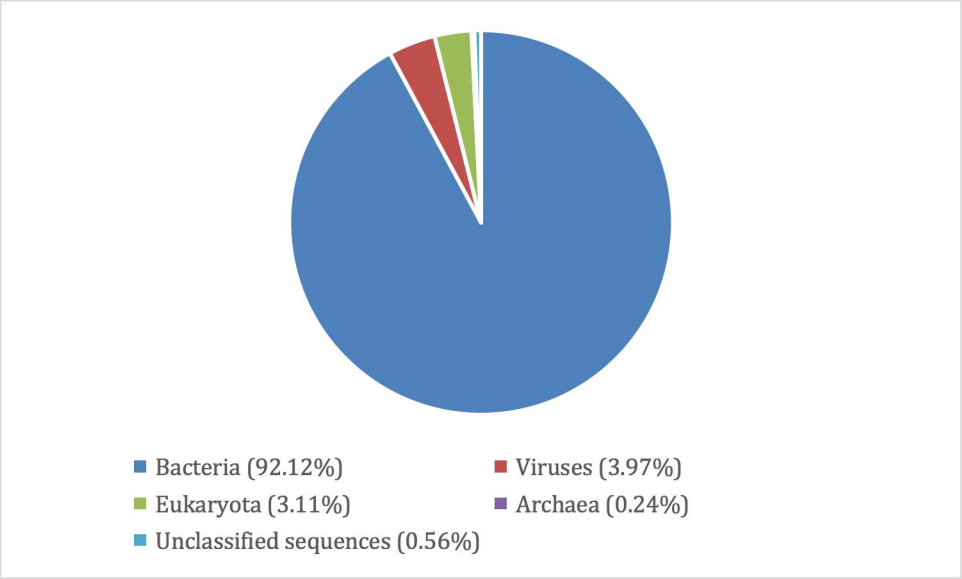
Table 3. Number of DNA sequence reads that passed through each step in the DADA2 tool pipeline.

Samples	Input	Filtered	DenoisedF	DenoisedR	Merged	Nonchim
OSD2414 (Marchica)	66,161	28,157	27,383	27,412	23,687	19,662
OSD2415 (Marchica)	38,241	18,570	18,124	18,198	16,957	16,355
OSD4714 (Venice)	44,664	21,137	20,199	20,354	17,763	15,331
OSD4715 (Venice)	60,996	31,535	30,156	30,205	26,166	23,088
OSD8114 (Ria Formosa)	53,394	22,538	22,059	21,935	18,895	16,205
OSD8115 (Ria Formosa)	90,023	45,355	44,652	44,584	41,038	37,323
OSD9114 (Oualidia)	89,724	43,757	41,329	41,707	34,329	29,064
OSD9115 (Oualidia)	73,443	33,878	32,946	32,968	28,967	23,401
OSD9314 (ElJadida)	41,323	20,930	19,889	20,052	16,612	13,680
OSD9315 (ElJadida)	62,923	31,790	31,088	31,164	27,840	25,242
OSD9414 (Saidia Marina)	61,932	28,805	27,557	27,544	22,009	18,190
OSD9415 (Saidia Marina)	48,501	20,674	20,169	20,156	18,152	16,940

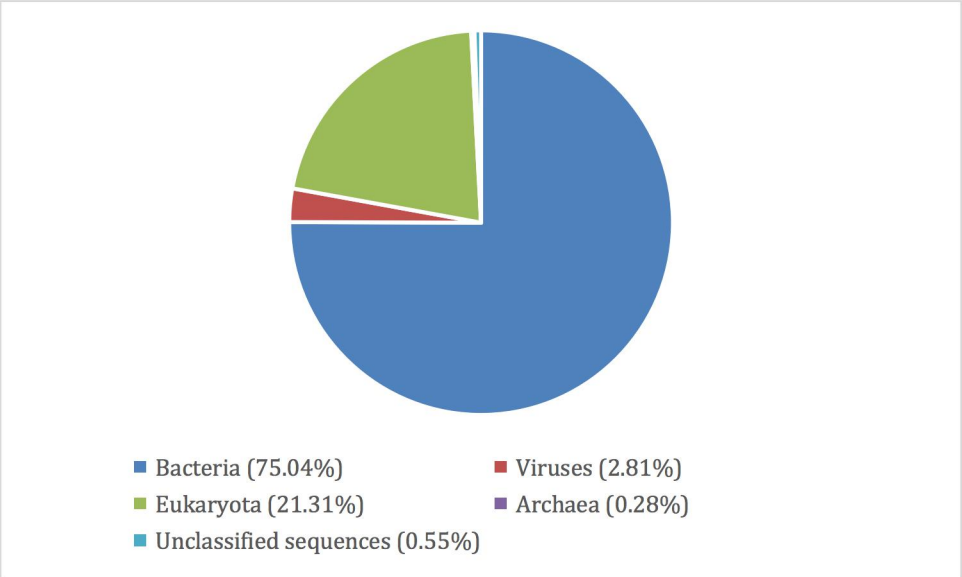
Metagenome sequence analysis was performed using MetaSpades **(Nurk *et al.* 2017)** to assemble trimmed forward (R1) and reverse (R2) reads and then MetaQuast **(Mikheenko *et al.* 2016)** to control assembly quality. A total of 196,153 contigs were produced for Marchica, labeled (OSD24, https://www.ebi.ac.uk/ena/browser/view/ERS667644_accessed on 1 December 2020), and 110,719 for Oualidia, labeled (OSD91, https://www.ebi.ac.uk/ena/browser/view/ERS667576_accessed on 1 December 2020). A total of 54,965 reads in OSD24 passed quality control (QC), while only 19,536 reads in OSD91 passed QC. We used Prodigal **(Hyatt *et al.* 2016)** for microbial genus, species, and strain level detection for sequence annotation, while we used MetaPhlAn2 **(Truong *et al.* 2015)** to obtain the relative abundance. We used MG-RAST **(Meyer *et al.* 2008)** for protein-coding gene prediction for the whole microbiome (Figure 4). Finally, we used Prokka **(Seemann *et al.* 2014)** to annotate the sequences that belong to the most abundant species of the five phyla *Proteobacteria*, *Bacteroides*, *Cyanobacteria*, *Verrucomicrobia*, and *Actinobacteria*.

Viral signal detection: Reads were mapped to a viral database, including 1575 complete viral genomes from RefSeq and 6322 contigs from the TOV_43 dataset **(Brum *et al.* 2015)**. Read recruitment values were calculated as part of the bowtie2 **(Langmead *et al.* 2012)** output. The number of reads and aligned nucleotides were computed using a custom Python script employing Biopython. Coverage values for all sequences were calculated using a Python script modified to normalize for metagenome sizes. All other calculations were performed using the Python modules pandas and numpy, and images were generated using a combination of Seaborn, matplotlib, and palettable python modules. The virus-sample heatmap was produced with the package pheatmap for drawing heatmaps in R using default hierarchical clustering for the viral sequences (rows) and manual clustering of samples by geographical zone (columns). Geo-coordinates were confirmed using OSD GitHub (https://github.com/MicroB3-IS/osd-analysis/wiki/Guide-to-OSD-2014-data_accessed on 30 October 2020).

Statistical analyses: Graphics and ecological statistical analyses for the relative abundance of specific phyla were performed using the R Statistical package R v. 3.0.2 (www.R-project.org/ accessed on 15 November 2020). Visualization and analysis of inter- and intra community comparisons comprise the realization of community heatmaps, skyline plots, pie charts, dendrograms, refraction, and diversity indices using R packages (phyloseq, ggplot2, ape, pheatmap, vegan, jsonlite, Rcolorbrewer) and python3 modules scipy, numpy, and cogent.



(a) Marchica



(b) Oualidia

Figure 4. Domain level classification of all metagenomic reads using MG-RAST [23]. (a) Marchica lagoon (b) Oualidia lagoon.

Chapter 4: Taxonomic Profiling of Microbial Communities in Marchica and Oualidia lagoons

4.1. Environmental parameters

The measurements of physicochemical parameters are shown in (Table 1). The water temperature of the Marchica (~27 °C) and Oualidia (~20.5 °C) lagoons varied significantly with the sampling locations. The electrical conductivity (EC) shows a clear evolution, with a maximum in the summer of 2015 in Oualidia (45.1 ms/cm) and a minimum in 2014 (39.6 ms/cm). In Marchica, it varied between 53 ms/cm and 54 ms/mc, which was similarly observed in 2015 by **(Aknaf *et al.* 2017)**. Salinity, as represented by EC, was much higher in Marchica (36–40 ppt) than in the Mediterranean Sea (24–32 ppt), whereas Oualidia waters showed a lower salinity value (27–29 ppt) regardless of their connection to the Atlantic where salinity can reach above 36 ppt. Waters of Marchica and Oualidia were mildly alkaline (pH ~ 8.3), which is the optimum for marine bacterial growth **(Padan *et al.* 2005)**. The recorded DO concentrations in Marchica (8.5 mg/L) and Oualidia (7.5 mg/L) indicated that surface waters were tolerably oxygenated (6.15 to 9.02 mg/L) **(Aknaf *et al.* 2017)**. In addition, both lagoons were slightly turbid (3 nephelometric turbidity units (NTU)). In Marchica, nitrate concentrations reached 12.5 mg/L in 2014 and dropped to 4.8 mg/L in 2015, while in Oualidia, they followed the same pattern, varying between 10.79 mg/L in 2014 and 2.68 mg/L in 2015. The concentrations of phosphate in the surface waters of Marchica and Oualidia lagoons were 0.03–0.04 mg/L and 0–0 mg/L, respectively, in 2014 and 2015.

4.2. Community Composition

To gain insight into microbial diversity in Moroccan lagoons, we used bioinformatics tools to analyze the metagenome and amplicon sequence data of the Marchica and Oualidia sites produced from water samples after the sampling campaigns of the 21st of June 2014 and 2015 (Figures 5 and 6, Table 4). 16S rRNA gene classification using the Silva reference database revealed a high percentage of bacteria in both lagoons. On average, bacteria accounted for 90% of the total prokaryotes in Marchica and ~70% in Oualidia, outnumbering Archaea. We only found <1% archaeal gene sequences in both sites assigned to *Euryarchaeotes*. In total, ten phyla were identified using MG-RAST, not including the unknown class, while VAMPS and DADA2 identified 27 phyla, which formed 62 classes, 129 orders, 260 families, 799 genera, and 1379 candidate species for the 2014 and 2015 sampling in Marchica and Oualidia lagoons. Based on

VAMPS and DADA2, the five phyla that were the most abundant in both lagoons, Marchica and Oualidia, respectively, were *Proteobacteria* (53.62%, 29.18%), *Bacteroidetes* (16.46%, 43.49%), *Cyanobacteria* (0.53%, 34.35%), *Verrucomicrobia* (1.75%, 15.82%), and *Actinobacteria* (7.42%, 13.98%).

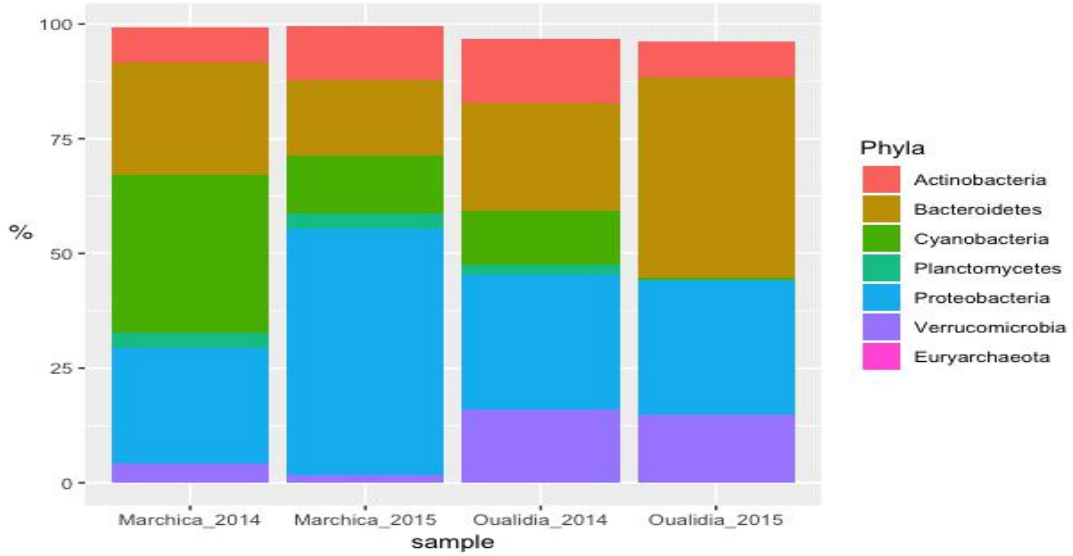


Figure 5. Cross comparison of 16S rRNA amplicon sequences from selected abundant high-level bacterial taxa from Marchica and Oualidia metagenomes.

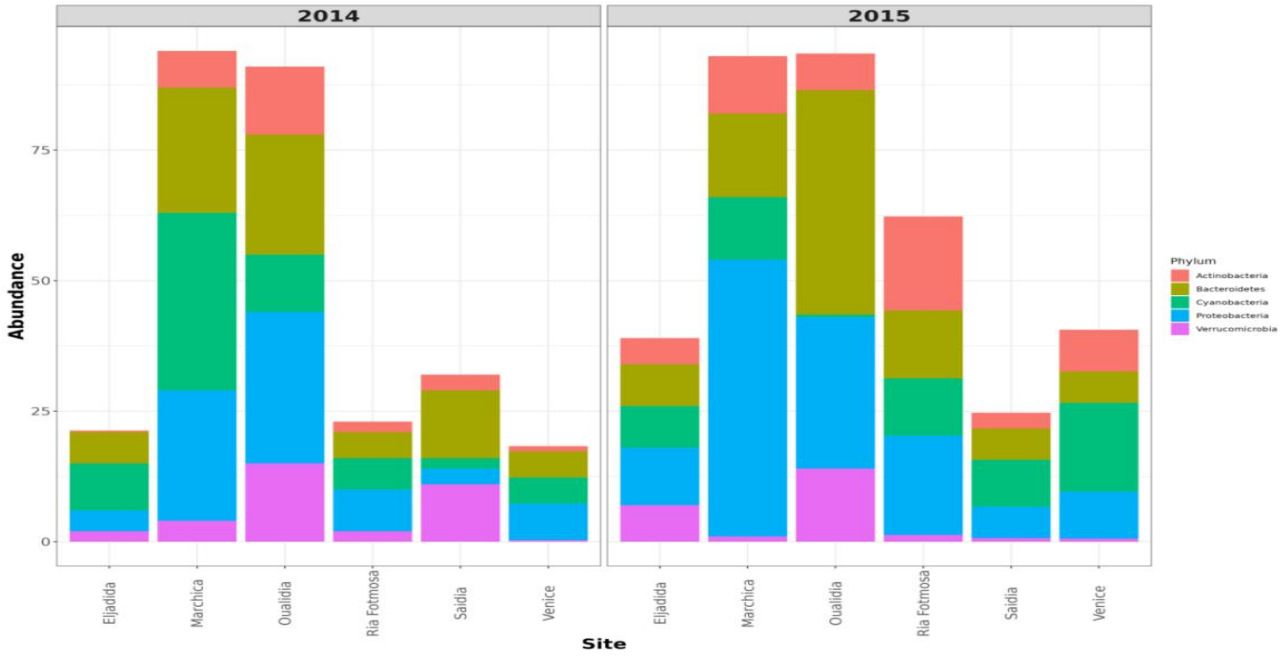


Figure 6. Cross comparison of 16S rRNA amplicon sequences from selected 5 abundant high-level bacterial taxa from Marchica and Oualidia metagenomes to several saline metagenomes using R.

Table 4. 16s rRNA amplicon sequences from selected high abundant sequences assigned to a phylum for both sampling campaigns 2014 and 2015 in both lagoons.

Marchica 2014			Oualidia 2014		
Taxa	#Hits	%	Taxa	#Hits	%
Actinobacteria	4793	7.42	Actinobacteria	11,670	13.98
Bacteroidetes	15,896	24.62	Bacteroidetes	19,660	23.55
Cyanobacteria	22,178	34.35	Cyanobacteria	9721	11.65
Planctomycetes	2190	3.39	Planctomycetes	1904	2.28
Proteobacteria	16,310	25.26	Proteobacteria	24,559	29.42
Verrucomicrobia	2712	4.20	Verrucomicrobia	13,202	15.82
Euryarchaeota	3	0	Euryarchaeota	19	0.02
Marchica 2015			Oualidia 2015		
Taxa	#Hits	%	Taxa	#Hits	%
Actinobacteria	4408	11.75	Actinobacteria	5613	7.80
Bacteroidetes	6173	16.46	Bacteroidetes	31,312	43.49
Cyanobacteria	4751	12.67	Cyanobacteria	384	0.53
Planctomycetes	1217	3.25	Planctomycetes	142	0.20
Proteobacteria	20,109	53.62	Proteobacteria	21,011	29.18
Verrucomicrobia	656	1.75	Verrucomicrobia	10,740	14.92
Euryarchaeota	10	0.03	Euryarchaeota	1	0

4.3. Species Diversity

Following OTU clustering, we calculated the Shannon and Simpson diversity indices from 88,482 bacterial sequences to determine species relative abundances at site and time. By definition, species richness is related to the count of the different species found in a sample. If the number of species present in “Sample A” is greater than that in “Sample B”, then “Sample A” is richer than “Sample B”. The diversity increases when species richness also increases **(Kort, *et al.* 2019)**. The Shannon value likewise increases as the species number increases **(Magurran**

et al. **2003)**. However, the Simpson index rises when the diversity drops **(Simpson *et al.* 1949)**. On the other hand, both ACE and Chao1 were conceived to infer richness based on abundance data (importance of rare OTUs). Chao1 gives more weight to the low-abundance species, while ACE informs whether abundant species are present or absent **(Kim *et al.* 2017)**. Based on Vamps, calculating alpha diversity assessed by observed richness, Shannon and Simpson indices, ACE, and Chao1 revealed higher bacterial diversity in Oualidia in comparison to Marchica in both summer 2014 and 2015 (Table 5).

Table 5. Microbial alpha diversity indices; normalization: none; counts min/max: 0–100%.

Dataset	Observed Richness	ACE	Chao1	Shannon	Simpson
Marchica Bv4v5 (OSD24, 2014)	462	677.66	656.16	4.18	0.86
Marchica Bv4v5 (OSD24, 2015)	316	490.39	497.5	4.13	0.87
Oualidia Bv4v5 (OSD91, 2014)	1144	1591.09	1605.32	6.13	0.94
Oualidia Bv4v5 (OSD91, 2015)	704	1053.14	1022.78	4.59	0.89

4.4. Bacterial Diversity

Proteobacteria

This group is considered the largest and most diverse **(Yilmaz *et al.* 2016)**. The functional attributes of *Proteobacteria* are linked to mechanisms such as denitrification and amino acid biosynthesis **(Gupta *et al.* 2000; Dang *et al.* 2000)**. They formed the largest part of the bacterial group in both Marchica (25.2%, 2014; 53.6%, 2015) and Oualidia (29.4%, 2014; 29.1%, 2015) (Figures 4 and 5, Table 4). α-Proteobacteria was the most predominant group at both sites, with a significant increase in 2015 in Marchica and a small increase in Oualidia (46% in Marchica; 21.59% in Oualidia) compared to 2014 (14.53% in Marchica; 18.5% in Oualidia). Marine *Alphaproteobacteria* are known to be more abundant in the epipelagic zone of coastal waters **(Yu *et al.* 2015)**. *Rhodobacteraceae* was the largest represented family in *α-proteobacteria* in Marchica (10.8%, 2014; 9.8%, 2015) and Oualidia (10.5%, 2014; 19.3%, 2015). It is dominant in different marine regions, namely, the Mediterranean, Atlantic, and Pacific Oceans **(Gilbert *et al.* 2011)**, including areas low in nutrients but with a high rate of primary production **(Hartsock *et al.* 2011)**. Both *α-Proteobacteria* and *Cyanobacteria,* known for nitrate elimination, were

cosmopolitan, which is consistent with former reports **(Vincent *et al.* 1994)**. In *Gammaproteobacteria*, the relative abundance of *Alteromonadaceae* was higher in Marchica during 2014 (7.81%) than in 2015 (2.04%) but stable in Oualidia (2.31%, 2014; 2.44%, 2015). Shotgun metagenomic analysis of 2014 datasets showed the dominance of *Candidatus Pelagibacter* Rappé in Marchica (11%) and Oualidia (7%). This genus is known amid the broadly represented bacteria in open oceans **(Zubkov *et al.* 2009)**. In Mediterranean Saidia Marina (OSD94), the majority of alphaproteobacterial reads were ascribed to *Candidatus Pelagibacter* (10%), as well as in Venice Lagoon (OSD47) (7%). This was similarly observed in Mar Menor, a lagoon on the Mediterranean Spanish coast, where the dominance of *Candidatus Pelagibacter* was reported in 2012 (43%) **(Ghai *et al.* 2012)**. In contrast, at the Atlantic site El Jadida (OSD93), we observed only a minority of reads attributed to *Candidatus Pelagibacter* (1%), while none were present in Ria Formosa Lagoon (OSD81), located south of Portugal on the Atlantic coast (ERS667579, https://www.ebi.ac.uk/ena/browser/view/ERS667579 ,accessed on 1 December 2020). Furthermore, analysis of both lagoon metagenomes in 2014 showed remarkable differences at the strain level (Table 6). For instance, in *the alphaproteobacteria group, Rickettsia conorii* 7 Ogata—causative agent of Mediterranean spotted fever **(Solano-Gallego *et al.* 2006)** was the most dominant in both sites, but with a slightly higher proportion in Marchica (10.49%) than Oualidia (7.42%), followed by *Orientia tsutsugamushi* Boryong Cho (Marchica, 3.75%; Oualidia, 1.95%). A small proportion of reads were assigned to *Anaplasma marginale* Maries Brayton in Marchica (2.21%) compared to Oualidia (8.28%) and *Anaplasma phagocytophilum* HZ Hotopp (6% in Oualidia, 3% in Marchica), in contrast to the *gammaproteobacteria* strain *Xylella fastidiosa* Temecula Sluys, which was more abundant in Marchica (4.19%) compared to Oualidia (2.62%) (Table 6). Interestingly, we only observed the strain *Anaplasma phagocytophilum* HZ in our comparative Moroccan Mediterranean site Saidia Marina (ERS667573, https://www.ebi.ac.uk/ena/browser/view/ERS667573 accessed on 1 December 2020) (5%) and Atlantic site Eljadida (ERS667574, https://www.ebi.ac.uk/ena/browser/view/ERS667574_accessed on 1 December 2020) (5%) as well as in the Mediterranean lagoon Venice (ERS667621, https://www.ebi.ac.uk/ena/browser/view/ERS667621_accessed on 1 December 2020) (4%) and Atlantic Portuguese lagoon Ria Formosa (4%) (Table 6).

Table 6. 16S metagenome sequences assigned to a strain.

				Oualidia Lagoon		Marchica Lagoon	
Taxa	Genus	Species	Strain	#Hits (bp)	%	#Hits (bp)	%
Alphaproteobacteria	Anaplasma	marginale	Maries	3,521,109	8.28	2,149,110	2.21
Alphaproteobacteria	Rickettsia	conorii	Malish 7	3,079,827	7.24	1,016,387	10.49
Alphaproteobacteria	Anaplasma	phagocytophilum	HZ	1,683,742	3.96	3,623,917	3.74
Gammaproteobacteria	Xylella	fastidiosa	Temecula1	1,501,655	3.53	3,857,299	3.98
Gammaproteobacteria	Xenorhabdus	nematophila	1_9_0_6_1	1,117,290	2.62	4,068,450	4.19
Alphaproteobacteria	Orientia	tsutsugamushi	Boryong	829,248	1.95	3,639,647	3.75
Gammaproteobacteria	Escherichia	colia	UMN026	490,326	1.15	1,533,000	1.58
Betaproteobacteria	Burkholderia	rhizoxinica	4_5_4	430,554	1.01	1,112,721	1.15
Alphaproteobacteria	Brucella	canis	2_3_3_6_5	283,501	0.66	592,700	0.61
Alphaproteobacteria	Erythrobacter	litoralis	HTCC2594	277,397	0.65	725,363	0.75
Betaproteobacteria	Ralstonia	solanacearum	PSI07	241,328	0.56	614,227	0.63
Alphaproteobacteria	Rhizobium	sp	N_G_R_2_3_4	217,740	0.51	305,658	0.31
Gammaproteobacteria	Shigella	dysenteriae	Sd197	124,696	0.29	346,084	0.35
Gammaproteobacteria	Marinobacter	aquaeolei	VT8	664,804	1.56	2,030,270	2.09
Cyanobacteria	Prochlorococcus	marinus	9_3_1_3	1,387,027	0.3	3,261,370	3.36
Cyanobacteria	Nostoc	azollae	0_7_0_8	1,340,033	3.15	5,328,515	5.5
Cyanobacteria	Synechococcus	sp	C_C_9_6_0_5	406,828	0.95	3,636,034	3.75
Cyanobacteria	Synechococcus	sp	JA_2_1_3	287,486	0.67	1,310,559	1.35
Bacteroidetes	Candidatus -Amoebophilus	asiaticus	5_a_2	2,748,208	6.46	10,026,340	10.34
Bacteroidetes	Bacteroides	fragilis	9_3_4_3	956,572	2.25	2,783,907	2.87
Verrucomicrobia	Methylacidiphilum	infernorum	V4	655,181	1.54	1,576,020	1.62
Verrucomicrobia	Akkermansia	muciniphila	8_3_5	517,529	1.21	571,583	0.58
Actinobacteria	Rothia	dentocariosa	1_7_9_3_1	2,513,106	5.91	3,917,261	4.04
Actinobacteria	Mycobacterium	leprae	TN	1,072,610	2.52	1,013,872	1.04
Actinobacteria	Tropheryma	whipplei	2_7	1,010,398	2.37	1,929,685	1.99
Actinobacteria	Rhodococcus	jostii	RHA1	199,761	0.47	277,060	0.28
Actinobacteria	Catenulispora	acidiphila	4_4_9_2_8	194,558	0.45	614,181	0.63

Cyanobacteria

Cyanobacteria are naturally occurring photosynthetic prokaryotes that can be found in almost all bodies, both fresh and saltwater. These organisms are considered the dominant primary producers on Earth, contributing more than 25% of photosynthesis worldwide **(Bullerjahn *et al.* 2014)**. *Cyanobacteria* can grow into large colonies visible to the naked eye, often referred to as a harmful cyanobacterial algal bloom (HAB) **(Cooke *et al.* 2005).** The most striking characteristic of Marchica Lagoon was a large number of cyanobacterial sequences (comprising nearly 34.5% of all 16S rRNA sequences in 2014, (see Figure 4 and Table 4) compared to Oualidia Lagoon (11.65%). This frequency significantly decreased in 2015 (12.6% in Marchica, 0.53% in Oualidia). The classification of the 16S rRNA metagenomic sequences indicated the presence almost exclusively of the *Synechococcus* Nägeli genus, which was highly abundant in Marchica (32%) compared to Oualidia (0.07%) in 2014. In 2015, the frequency dropped to 22% in Marchica and 0.04% in Oualidia. The top organisms identified as cyanobacterial candidates in the 2014 WGS metagenome were *Synechococcus* strains (e.g., *sp. CC9605* Salasar and *sp. JA213*). The *sp. CC9605* strain occupies a crucial spot at the food chain base (https://www.UniProt.org/proteomes/UP000002711_accessed on 15 December 2020), whereas *sp. JA213* is not described. In addition, we were able to identify a cyanobacterial strain in Marchica, namely, *Nostoc azollae* 0708 Ran (5%), but not in Oualidia. This strain was also identified on the Mediterranean coast of Saidia Marina, Venice Lagoon, El Jadida Atlantic coast, and Ria Formosa Lagoon (Table 6). The *Prochlorococcus* Penny genus was absent in Oualidia and presented only at ~0.1% in Marchica. Differences emerged at the organismal level as the analysis of the whole metagenome allowed; at the strain level, the identification of *Prochlorococcus marinus* 9313 Rocap species was the same in both lagoons, with the same proportion of 3%.

Bacteroidetes

This group is more widespread in marine environments together with coastal waters, sediments, hydrothermal vents, and polar regions **(Fernández-Gómez *et al.* 2013)**. Most members of the *Bacteroidetes* phylum picked from the lagoon water samples belonged to the *Flavobacteriales* order (Marchica: 22% of bacterial reads in 2014; 15% in 2015; Oualidia: 20% in 2014; 20% in 2015), which is mainly composed of *Cryomorphaceae* and *Flavobacteriaceae*. Marine *Flavobacteria* clades can adopt different strategies to coexist **(Gómez-Pereira *et al.***

2010). Most hits (Marchica: 11% in 2014 and 8% in 2015, Oualidia: 5% in 2014 and 22% in 2015) belonged to the unassigned F*lavobacterium* genus, according to 16S rRNA analysis, except for the genus *Formosa* Ivanova, which was more abundant in Marchica (5% in 2014 and 2.5% in 2015) than in Oualidia (0.36% in 2014 and 2.14% in 2015). This genus occurs in marine territories with high organic matter levels, for instance, associated with algae, fecal pellets, and invertebrates **(Mann *et al.* 2013)**. The 2014 WGS metagenomic analysis revealed two strains: *Candidatus Amoebophilus asiaticus* 5a2 Schmitz-Esser (10.34% in Marchica; 6.46% in Oualidia) and *Bacteroides fragilis* 9343 Coyne (2.87% and 2.25%, respectively).

Verrucomicrobia

Verrucomicrobia appears most regularly in polar and temperate zones **(Cardman *et al.* 2014)**. They are also found in marine animals and plants as symbionts/parasites **(Bünger *et al.* 2020)**. They feature interesting characteristics, such as the presence of genes homologous to eukaryotic tubulins **(Jenkins *et al.* 2002)** and methane oxidation capacity in low pH environments **(Dunfield *et al.* 2007)**. In the 2014 and 2015 datasets, they are largely represented by the genus *Roseibacillus* Yoon. In Oualidia, they showed a significant abundance (15.18%, 2014; 14.5%, 2015) compared to Marchica (3.46%, 2014; 0.86%, 2015). The OTU fraction assigned to *Roseibacillus* varied considerably across the Mediterranean and Atlantic sites we selected from the OSD database for this study. In Saidia Marina (OSD94) (7%, 2014; 0.5%, 2015) and Venice Lagoon (OSD47) (0.08%, 2014; 0.3%, 2015), abundances followed the same pattern as Marchica. Differences were more pronounced in the Atlantic sites, as the abundance of *Roseibacillus* dropped in El Jadida (OSD93) to (1%, 2014; 3%, 2015). The same was observed in Ria Formosa (OSD81) (1.5%, 2014; 0.3%, 2015). Analyses at the strain level allowed the identification of two strains, *Methylacidiphilum infernorum* V4 Dunfield and *Akkermansia muciniphila* 835 Van Passel (Table 6). *Methylacidiphilum infernorum* holds most of the key metabolic pathways of amino acid biosynthesis **(Hou *et al.* 2008)**.

Actinobacteria

Actinobacteria was the least abundant phylum in both lagoons, but the proportions varied across the datasets (Marchica: 7.42%, 2014; 11.75%, 2015; Oualidia: 13.98%, 2014, 7.80%, 2015) (Table 4). In our study, it was represented by the order *Micrococcales* (Marchica 3% of total *Actinobacteria*, 2014; 0.8%, 2015; Oualidia: 15%, 2014; 7%, 2015). The marine *Actinobacteria* group was first reported on the southern California coast and Bermuda Island

(Ventura *et al.* 2007). Currently, they are described as omnipresent taxa of marine bacterioplankton **(Shivlata *et al.* 2015)** and may form up to 10–60% of the bacterioplankton group **(Pinto *et al.* 2021)**. Two major strains belonging to the *Micrococcales* order were identified in the two studied lagoons, *Rothia dentocariosa* 17,931 Eisenberg and *Tropheryma whipplei* 27 Bentley (Table 6), both known as causative human disease strains **(Shivlata *et al.* 2015)**. Another bacterium belonging to the *Corynebacteriales* order was identified, namely, *Mycobacterium leprae* TN Cole, a pathogenic microorganism causing leprosy **(Pinheiro *et al.* 2011).**

4.5. Viral signature

Looking at the OSD metagenomics dataset in 2014 as a whole revealed 42 viral genomes/contigs that were confidently identified (more than 75% of the genome/contig covered) across 31 OSD samples. Most of these viruses corresponded to known cyanophages from NCBI or "unknown Caudovirales" assembled from the Tara Oceans samples corresponding to coastal sites and/or the Mediterranean Sea. Here, we are interested in Moroccan lagoons. A coverage filter was applied, with a viral contig/genome being considered "detected" only if at least 75% of its bases were covered, which narrowed the number of viral sequences found. Reads in Marchica mapped at 2.7% against the combined ViralRefSeq and TOV_43 contigs (https://www.ivirus.us/data, accessed on 15 November 2020) (Figure 8). A "zoom-in" of these data revealed eight viruses in only Marchica Lagoon, none of which are known (Figure 7), suggesting the specificity of the Marchica virome.

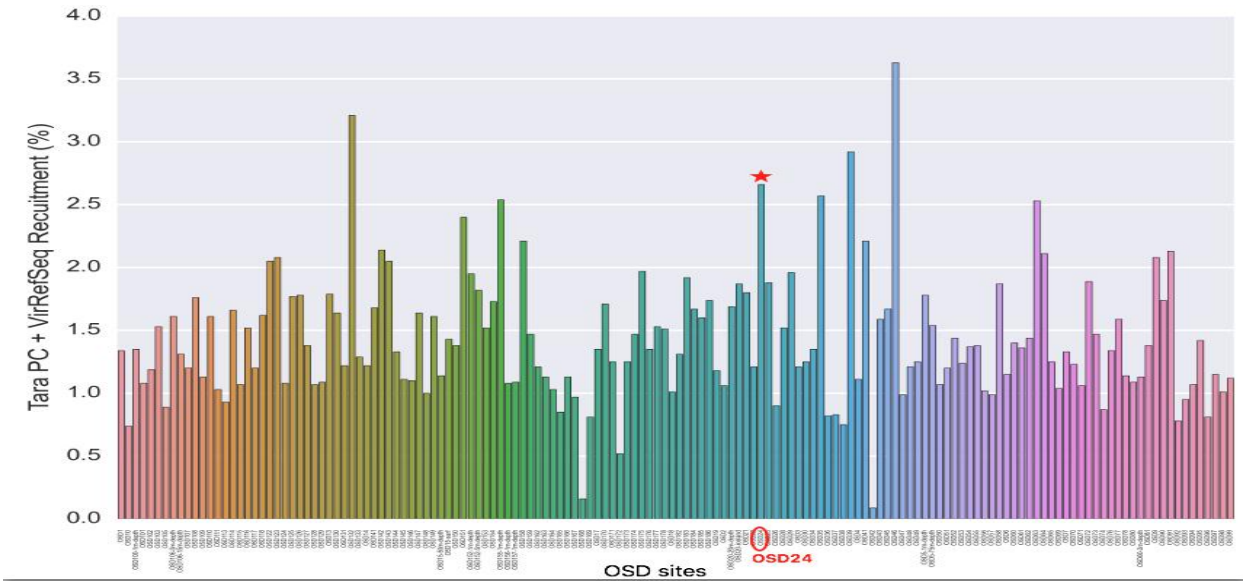


Figure 7. Proportion of reads mapping to viral contigs in OSD metagenomes. Legend: OSD24: Marchica lagoon.

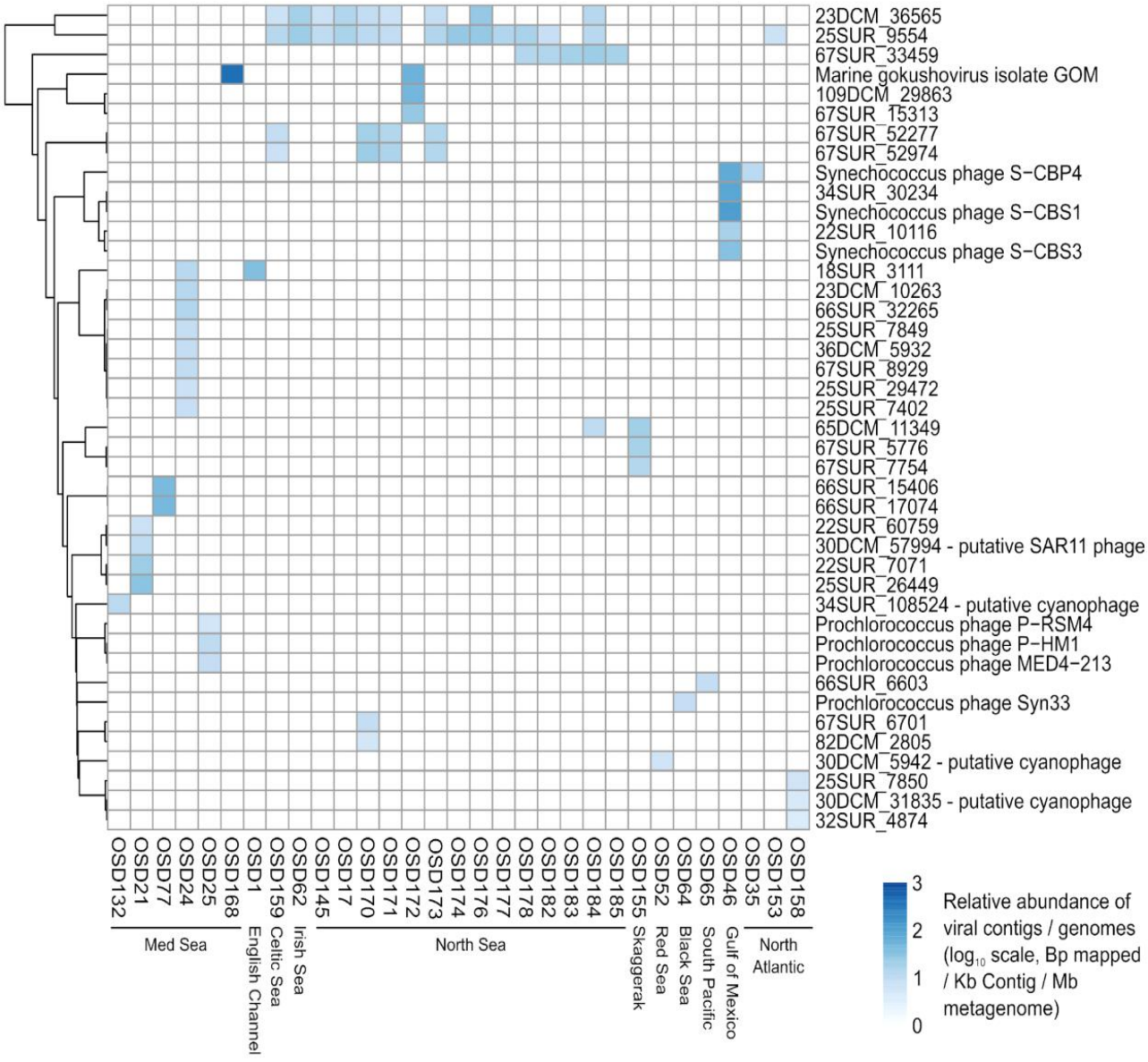


Figure 8. Heatmap of viral sequence abundance (when covered at >75%) across OSD samples. Only metagenomes with at least one viral sequence covering >75% were included.

4.6. Eukaryotic Diversity

The 18S rRNA sequences accounted for 11,220 (14%) in Marchica and 33,442 (28%) in Oualidia during the 2014 sampling campaign, whereas 36,243 and 31,511 eukaryotic reads were obtained in 2015 (49% and 30% of total small subunit ribosomal RNA (SSUs), respectively). The main identified eukaryotic phyla were *Dinophyta*, *Ochrophyta*, and *Chlorophyta* (Tables 7 and 8). *Gyrodinium* Hulbert (25%, 2014; 33%, 2015), *Pseudo-Nitzschia* Peragallo (15%, 2014;

2%, 2015), and *Tetraselmis* Stein (4%, 2014; 0.1%, 2015) were detected in Marchica samples in important relative abundances compared to other genera. In Oualidia, the most abundant organisms revealed by 18S rRNA were the *Ostreococcus* Courties genus (0%, 2014; 16%, 2015), *Pelagodinium* HJ Spero (12%, 2014; 0%, 2015), and *Thalassiosira* Cleve (8%, 2014; 25%, 2015).

Table 7. 18S SSU rRNA amplicons from selected highly abundant sequences assigned to a phylum for both sampling campaigns 2014 and 2015 in both lagoons.

	Marchica Lagoon				Oualidia Lagoon			
	2014		2015		2014		2015	
Taxa	#Hits	%	#Hits	%	#Hits	%	#Hits	%
Dinophyta	4233	37.7	22,505	62.09	7617	22.7	1723	5.4
Ochrophyta	2725	24.2	6128	16.9	10,848	32.4	13,302	42.2
Chlorophyta	777	6.9	939	2.5	1827	5.4	6806	21.6

Table 8. 18S metagenomic rRNA from selected high abundant sequences assigned to a genus for both sampling campaigns 2014 and 2015 in both lagoons.

Marchica Lagoon						Oualidia Lagoon					
		2014		2015				2014		2015	
Taxa	Genus	#Hits	%	#Hits	%	Taxa	Genus	#Hits	%	%#Hits	%
Dinophyta	Gyrodinium	2844	25.3	12,160	33.5	Dinophyta	Pelagodinium	4051	12.1	20	0.06
Ochrophyta	Pseudo-Nitzschia	1776	15.8	752	2	Ochrophyta	Thalassiosira	2862	8.5	7979	25.3
Chlorophyta	Tetraselmis	503	4.4	61	0.1	Chlorophyta	Ostreococcus	8	0.02	5240	16.6

Chapter 5: Functional Profiling of Microbial Communities in Marchica and Oualidia lagoons

5.1. Microbial DNA Sequences and Annotation

An estimated total of 1 million sequencing data was generated for each of the 4 sites. Adapter and duplicate sequences were removed from the raw data as explained in chapter 3, as well as reads with low quality, when the phred score is less than 20. Sequencing reads were assembled using metaspades **(Nurk *et al.* 2017)** and contigs were uploaded directly to MG-RAST **(Meyer *et al.* 2008)** where another round of QC was performed.

Sequences were screened for potential coding elements using BLASTX search tool **(Altschul *et al.* 1997)**, then mapped against SEED database, with 10-5 expect value (E) cutoff, 60% minimum identity cutoff, and a minimum alignment length cutoff of 15. Based on the data obtained from the MG-RAST server, protein encoding genes were classified according to their respective functions and mapped against SEED Subsystems to suggest the possible metabolic pathways and enzymes encoded within the genome. After annotation, the sequencing information was exported into R where the reads were sorted according to their different functional categories for further analysis. Based on read abundance counts, pie and bar charts were generated for the top functions found in each location, as well as for the biogeochemically relevant genes such as those involved in nitrogen recycling, and other metabolisms of interest.

5.2. Analysis Statistics

A shotgun metagenomics approach was applied to the collected and filtered samples on 21 June 2014 from Marchica and Oualidia lagoons to examine the functional diversity of the microbial communities. In Marchica, of the sequences that passed the quality control test (QC), 603 sequences possessed rRNA genes, 120,555 sequences (61.46%) had predicted proteins with known functions, and 74,995 sequences (38.23%) had predicted proteins of unknown functions. In Oualidia, 1452 sequences contained rRNA genes, 31,957 sequences (30.01%) had predicted proteins with known functions, and 73,081 sequences (68.63%) had predicted proteins with unknown functions.

5.3. Top Functional Genes

The top gene-encoded functions were similar in both locations, mainly comprising genes required for amino acid, carbohydrate, and protein metabolism. This shows that microorganisms dominating surface water environments could share essential genes for their survival and adaptation in these ecosystems **(Hewson *et al.* 2009)**. The metabolic distribution appeared to follow a similar pattern throughout all sample sets regarding composition and fraction size. In Marchica, the reads were dominated by features responsible for carbohydrate metabolism, which made up 11.9% of the total reads, followed by amino acid and derivative metabolism (10%), miscellaneous (8%), and protein metabolism reads, which were tied at 7% (Figure 9). Although they follow similar functional patterns, slightly fewer reads were assigned to metabolic functions in Oualidia (57%) than in Marchica (62%). Carbohydrates accounted for 11%, followed by amino acids and derivatives and protein metabolism, with 9% of total reads in Oualidia (Figure 10). Genes encoding essential biogeochemical processes were also present but in low proportions (1%). Functional gene properties were further assigned to microbes. For example, in Marchica, 25% of *Rickettsia conorii* 7 Ogata (belonging to the *Proteobacteria* phylum), 35% of *Nostoc azollae* 0708 Ran (*Cyanobacteria* phylum), 75% of *Methylacidiphilum infernorum* V4 Dunfield (*Verrucomicrobia* phylum), 30% of *Candidatus Amoebophilus asiaticus* 5a2 Schmitz-Esser (*Bacteroidetes* phylum), and 70% of *Rothia dentocariosa* 17931 Eisenberg (*Actinobacteria* phylum) proteins have a metabolic function (Figure 11). For comparison, in Oualidia, 15% of *Candidatus Amoebophilus asiaticus* 5a2 Schmitz-Esser (*Proteobacteria* phylum), 65% of *Anaplasma marginale* Maries Brayton (*Proteobacteria* phylum), 60% of *Nostoc azollae* 0708 Ran (*Cyanobacteria* phylum), 25% of *Methylacidiphilum infernorum* V4 Dunfield (*Verrucomicrobia* phylum), and 30% of *Rothia dentocariosa* 17931 Eisenberg (*Actinobacteria* phylum) genes have a metabolic function (Figure 12).

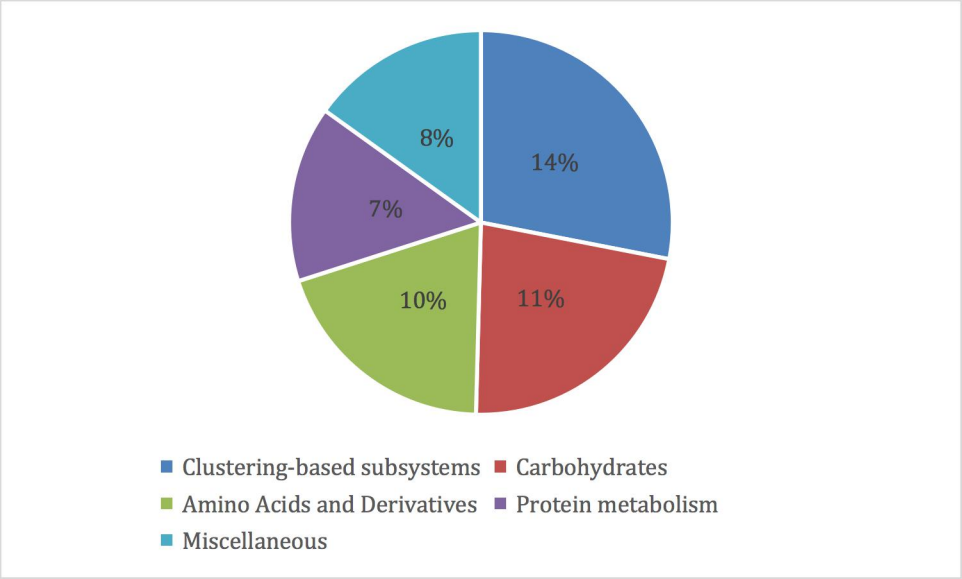


Figure 9. Functional category hits distribution annotated using MG-RAST Subsystems classification in Marchica lagoon (OSD24).

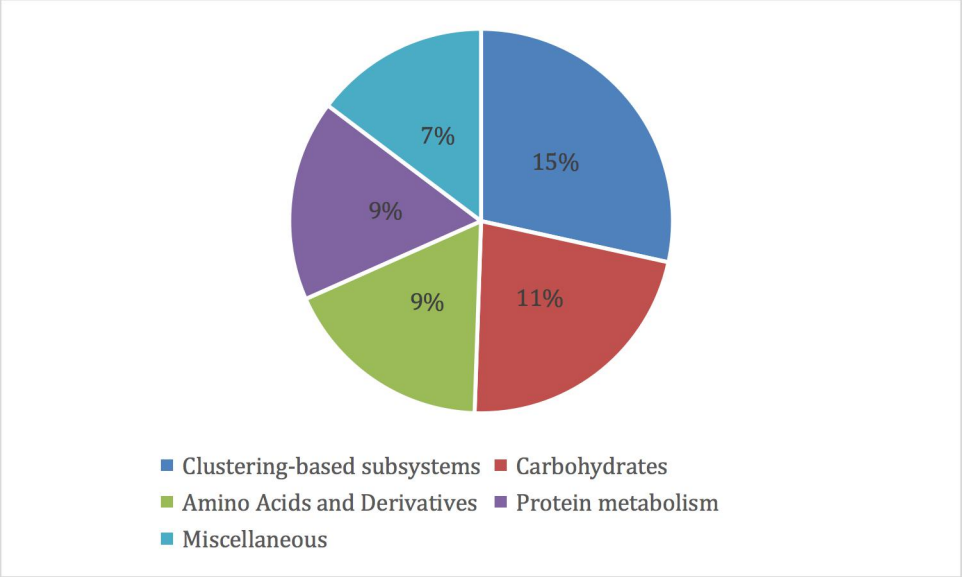


Figure 10. Functional category hits distribution annotated using MG-RAST Subsystems classification in Oualidia lagoon (OSD91).

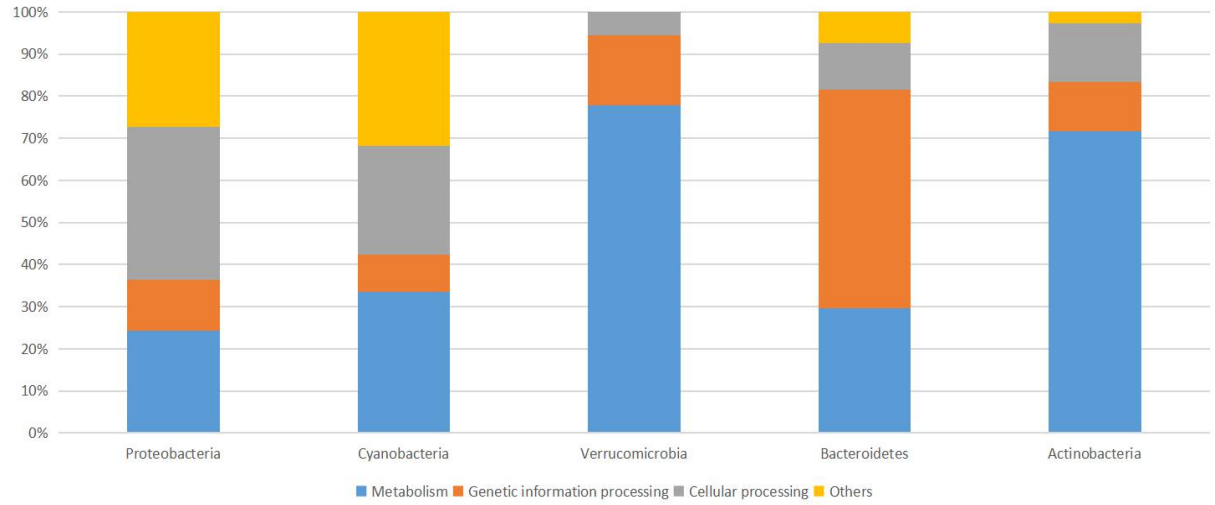


Figure 11. Top functional genes abundance in Marchica using Prokka [24].

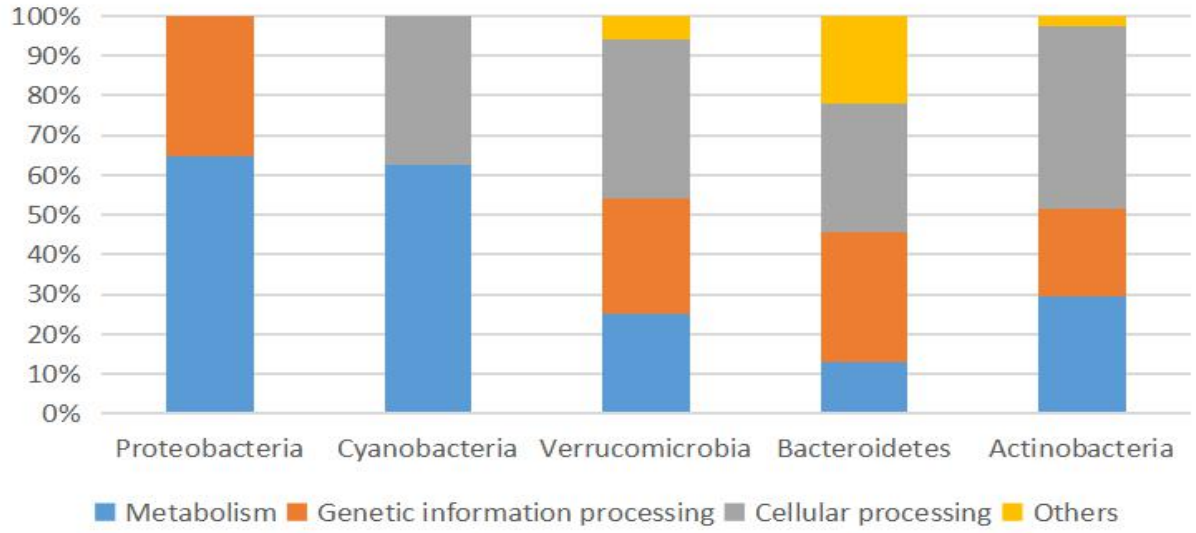


Figure 12. Top functional gene abundance in Oualidia using Prokka [24].

5.4. Correlation between functions and physicochemical parameters

The correlation between detected gene functions and physical–chemical parameters was calculated using the R package ggally **(Schloerke *et al.* 2020)**. The analysis was based on nine water properties (pH, temperature, conductivity, turbidity, oxygen, salinity, nitrate, nitrite, and phosphate) and the type of gene function (metabolic, genetic, and cellular process). Seven out of the nine measured physicochemical parameters, namely, temperature, salinity, conductivity, oxygen, phosphate, nitrates, and nitrites, were positively correlated with metabolic functions (correlation = 0.025) and negatively correlated with genetic and cellular processes (correlation =

−0.263 and −0.2, respectively); this correlation is almost the same among all previously cited parameters and functions. pH was positively correlated with genetic functions and cellular processes (correlation = 0.263 and 0.2, respectively) and negatively correlated with metabolic functions (correlation = −0.025) (Figure 13).

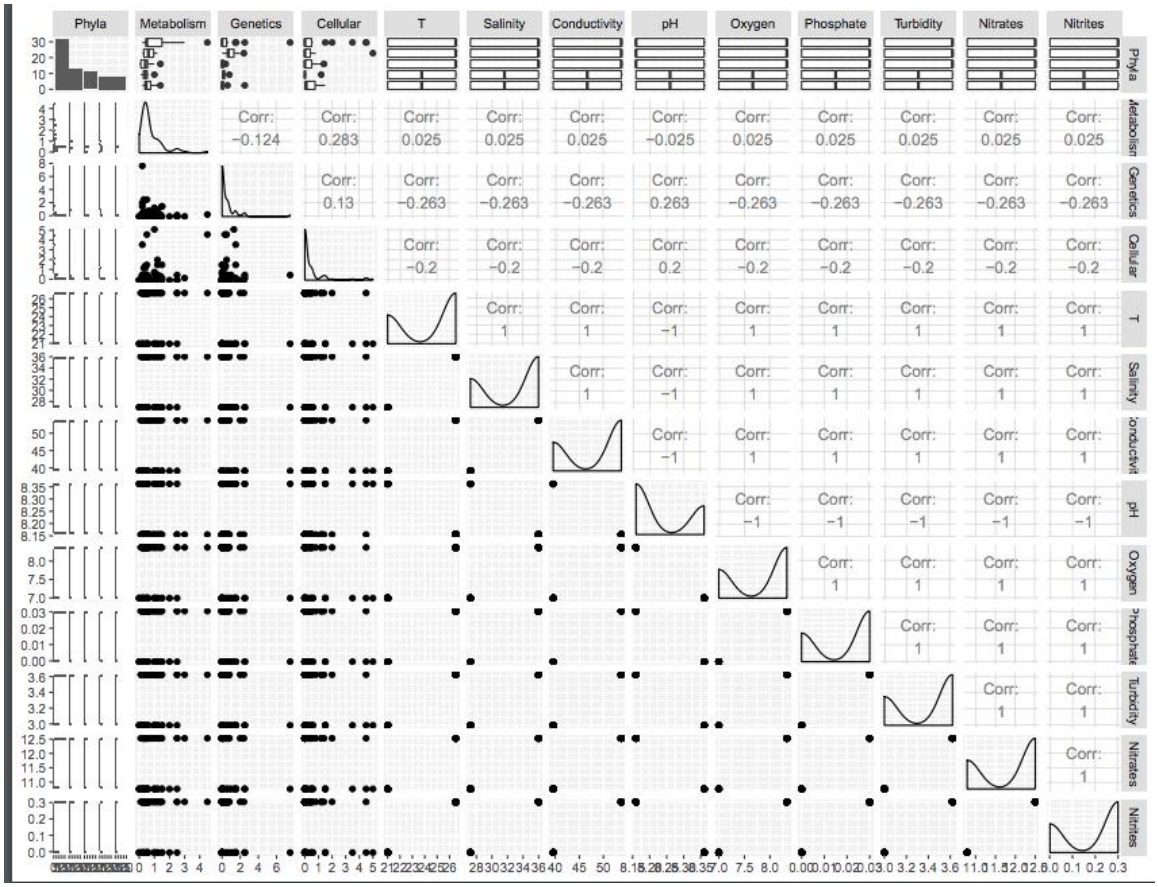


Figure 13. Correlation between functions and physicochemical parameters.

Chapter 6: Spatiotemporal patterns of *Synechococcus* oligotypes in Moroccan lagoonal environments

6.1. Sampling, data processing, and oligotyping analysis

Sample collection, dna extraction, and sequencing are described in chapter 3. We used VAMPS **(Huse *et al.* 2014)** to process 16S rRNA gene sequences, where taxonomy assignment was performed using Global Alignment for Sequence Taxonomy (GAST) and the SILVA rRNA gene reference database **(Quast *et al.* 2012)**. The obtained files include the reference ID, the taxonomy assigned, and the source of the taxonomy.

Oligotyping: For *Synechococcus* investigation, we used 16S rRNA gene oligotyping as described in **(Eren *et al.* 2013)**. This method is based on a supervised algorithm that identifies microdiversity using 16S rRNA gene sequences. Oligotyping is unlike regular taxonomic classification based on available reference databases, available sequences or cluster analysis based on the selection of the similarity threshold. This technique tackles the taxonomic resolution limitation by finding the most information-rich nucleotide positions (i.e., oligotypes). Sequences identified as *Synechococcus were extracted* from the Vamps database. Of the 22,387 sequences identified as *Synechococcus*, 17,941 remained after quality filtration and Pynast alignment. The mean length of *Synechococcus* reads was 254 bp. Next, we removed the uninformative gaps in the resulting aligned sequences using the "o-trim-uninformative-columns-from-alignment" script. Subsequently, we calculated the entropy of each nucleotide position within the oligotype package. After the initial calculation of Shannon entropy using the "analyze-entropy" script, we ran 16S rRNA oligotyping for the S*ynechococcus* genus until each oligotype had converged. Uninformative nucleotide positions were excluded. Seven nucleotide positions were used in total to define each oligotype, and to minimize the impact of sequencing errors on oligotyping results, we used a "minimum substantive abundance" criterion (M) of 5; thus, an oligotype was not included if the most common sequence for that type occurred less than five times. To reduce the noise, each oligotype was required to appear in at least one sample but was not required to comprise a certain percentage of reads or represent a minimum number of reads in all samples combined. We removed any oligotypes that did not meet these criteria from the analysis. The final number of quality-controlled oligotypes revealed by the analysis was 31 and represented 95% of the total *Synechococcus* reads. For each oligotype, the oligotyping pipeline chose the most abundant unique read as the representative sequence to be used for downstream

analyses. Upon completion of oligotyping analysis, the resulting "observation matrices" are concatenated to generate a single "observation matrix" for our V4-V5 dataset (Table 11). These observation matrices report counts, which are the number of reads assigned to each oligotype in each sample. We then converted counts to percent abundances within each sample and used these normalized relative abundances for subsequent analyses. We searched the most biologically relevant representative sequence of our oligotypes using blastn version 2.2.26 to assign taxonomy for each oligotype. We kept default parameters, except 'per. identity 100' to have hits with 100% sequence identity reported.

Table 9. Oligotyping details for V4-V5 data.

Genus	All reads	Av. Read Length	After QC Trim	Av. QC Read Length	After Oligotyping QC	MinSubsAb	Sensitivity	Total Num Oligos
Synechococcus	22,387	354	17,941	273	17,154	256	0.01	31

Oligotype network analysis: We performed network analysis using Gephi software, version 0.9.2**,** to determine the distribution of all *Synechococcus* oligotypes from both lagoons using a force-directed graph algorithm (ForceAtlas2 in Gephi software). Every dot identifies an oligotype present in at least one sampling site, and each edge on the network connects an oligotype to one or more sampling sites.

Clade identification: We designated a clade for each oligotype's representative sequence by matching this latter to a key reference database of 16S rRNA gene sequences from cultured *Synechococcus* **(Xia *et al.* 2015)**. *Synechococcus* sequences were downloaded from NCBI GenBank, clade classifications were obtained from the following sources **(Ahlgren *et al.* 2012; Xia *et al.* 2015)**. We added the representative sequences for each oligotype to the *Synechococcus* database, and we aligned them with Muscle (version 3.8.4, **(Edgar *et al.* 2004))**. We used exact matches between each oligotype *Synechococcus* sequence and the *Synechococcus* sequence database to infer clade designation.

Statistical Analyses: To group oligotypes statistically, we computed a principal component analysis (PCA) using the R package "ggfortify" with respect to a sample matrix of *Synechococcus* oligotype reads normalized to total *Synechococcus* reads for that sample. Each

oligotype was projected onto the first two PCs of the matrix. To investigate the environmental correlates of oligotype grouping, multiple regressions of each of the first two PCs were computed against the three environmental factors, which are the water temperature, salinity, and the concentration of nitrate.

6.2. Detection of Synechococcus oligotypes in environmental samples

A total of 535,138 16S rRNA amplicon sequence reads were generated from the microbial populations of four samples collected at two stations on the summer solstice, 21 June in 2014 and 2015. The relative *Synechococcus* read number compared to the total microbial community varied between both sampling sites during both time points (Figure 19 & Table 12).

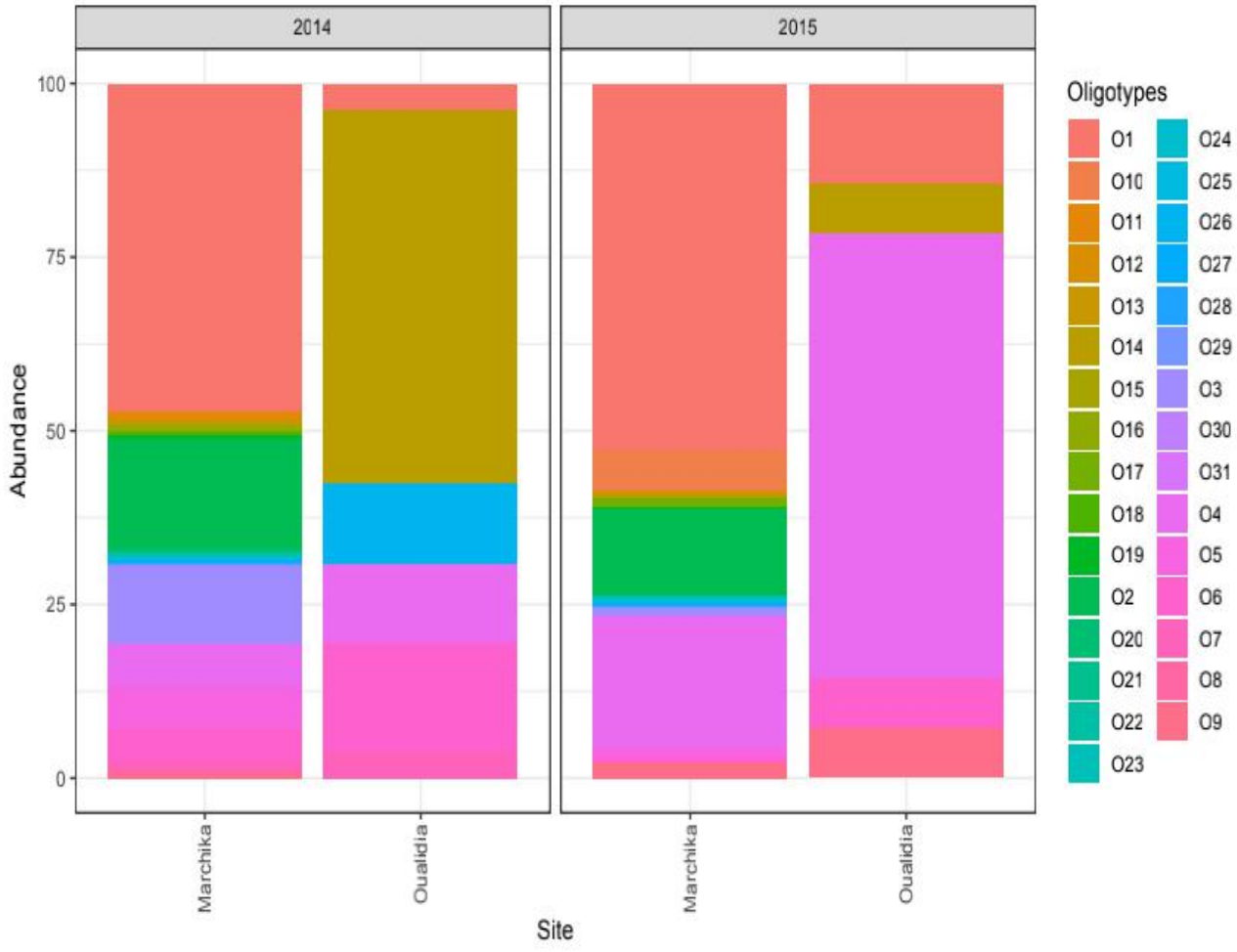


Figure 14. Cross comparison of 31 *Synechococcus* oligotypes in Marchica and Oualidia lagoons in 2014 and 2015. Legend: O: *Synechococcus* sequence oligotypes

Table 10. V4-V5 Oligotype observation matrix.

**OSD2414: Marchica Lagoon 2014 sampling; *OSD2415: Marchica Lagoon 2014 sampling.*

**OSD9114: Oualidia lagoon 2014 sampling; *OSD9115: Oualidia lagoon 2015 sampling.*

Oligotype	Oligotype ID	Sites (abundance count)				Sites (abundance percent)			
		OSD2414	OSD2415	OSD9114	OSD9115	OSD2414	OSD2415	OSD9114	OSD9115
TTAATCT	O1	7290	878	1	2	47.19	52.66	3.84	14.28
GACTCTC	O2	2509	208	0	0	16.24	12.47	0	0
TTCATCT	O3	1727	15	0	0	11.18	0.89	0	0
TTAAGCT	O4	974	319	3	9	6.30	19.13	11.53	64.28
GACTCCT	O5	874	20	0	0	5.65	1.19	0	0
TTAATTC	O6	750	13	4	1	4.85	0.77	15.38	7.14
TTCATTC	O7	177	0	1	0	1.14	0	3.84	0
ATAATTC	O8	119	0	0	0	0.77	0	0	0
TTAATTT	O9	76	37	0	1	0.49	2.21	0	7.14
GTCTCCT	O10	16	98	0	0	0.10	5.87	0	0
TTCTCTC	O11	86	1	0	0	0.55	0.05	0	0
TTAAGTC	O12	74	5	0	0	0.47	0.29	0	0
AGAATTC	O13	70	0	0	0	0.45	0	0	0
AAAATTC	O14	44	11	14	1	0.28	0.65	53.84	7.14
GACATCT	O15	67	1	0	0	0.43	0.05	0	0
TTAACTC	O16	65	0	0	0	0.42	0	0	0
AGCATCT	O17	41	21	0	0	0.26	1.25	0	0
ATAATCT	O18	53	0	0	0	0.34	0	0	0
TTAACCT	O19	46	3	0	0	0.29	0.17	0	0
TTATCTC	O20	45	0	0	0	0.29	0	0	0
TAAATCT	O21	43	2	0	0	0.27	0.11	0	0
TTCTCCT	O22	43	1	0	0	0.27	0.05	0	0
TTACTCT	O23	36	8	0	0	0.23	0.47	0	0
TTGATCT	O24	37	3	0	0	0.23	0.17	0	0
TTAGTCT	O25	29	5	0	0	0.18	0.29	0	0
TAAATTC	O26	24	7	3	0	0.15	0.41	11.53	0
GACTTCT	O27	29	1	0	0	0.18	0.05	0	0
TTAATCC	O28	26	3	0	0	0.16	0.17	0	0
TCAATCT	O29	25	4	0	0	0.16	0.23	0	0
GAAATCT	O30	29	0	0	0	0.18	0	0	0
CTAATCT	O31	23	3	0	0	0.14	0.17	0	0

We identified 31 *Synechococcus* oligotypes, which was equivalent to 95% of all *Synechococcus* reads analyzed. The most abundant *Synechococcus* reads were used for downstream analyses depending on the sampling location and date. Our phylogeny confirmed

that *Synechococcus* strains are classified into ten different clades based on representative V4–V5 16S rRNA sequences (Figure 20).

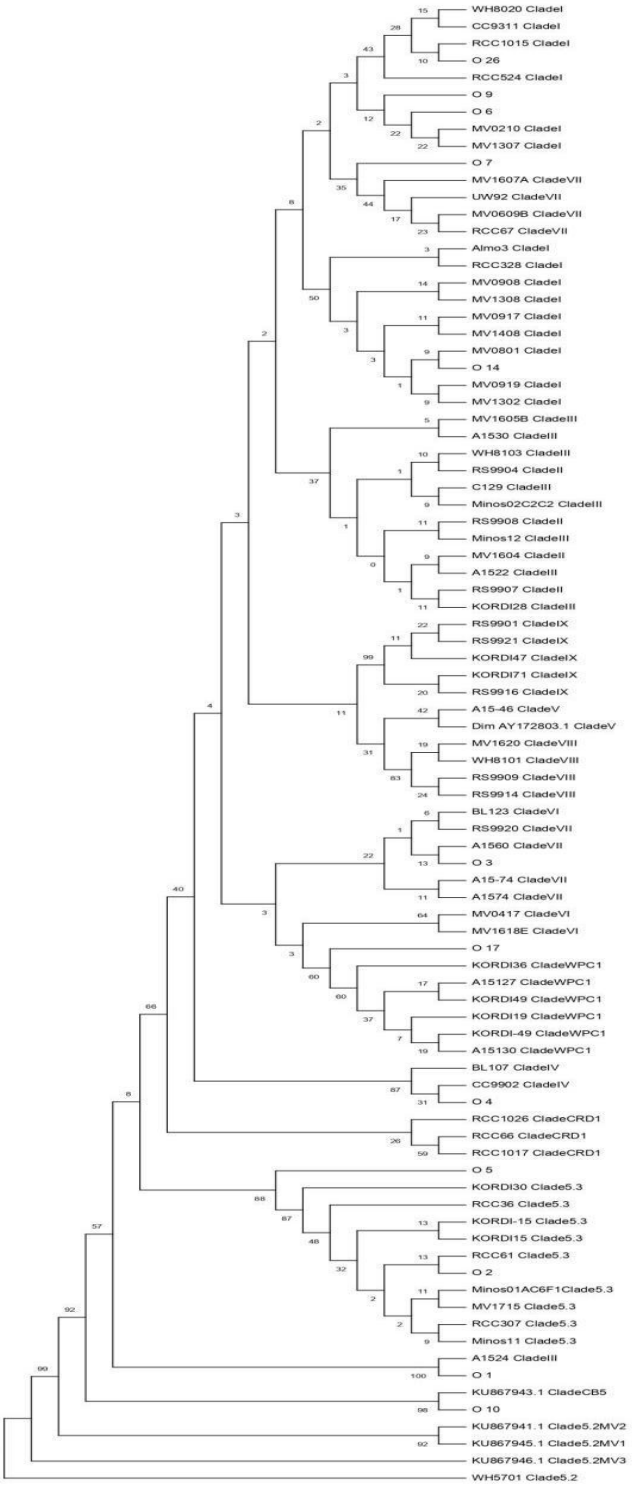


Figure 15. Phylogenetic tree constructed from V4–V5 16S rRNA gene amplicon sequence representatives of *Synechococcus* clade and representative sequences for the most abundant

oligotypes, both aligned using Muscle (version 3.8.4, **(Edgar *et al.* 2004)**). Shown is a neighbor-joining rooted tree generated using the Geneious Prime 2021.1 software package (Biomatters Ltd, Auckland, New Zealand).

The identified clades were separated into three distinct subclusters (5.1, 5.3 and 5.2). Ten oligotypes belonged to clade III (subclade 5.1A) (O1, O15, O19, O21, O23, O24, O25, O29, O30, O31), six belonged to clade I (subclade 5.1B) (O6, O8, O9, O14, O16, O26), another six belonged to Clade 5.3 (O2, O5, O11, O20, O22, O27), two belonged to clade IV (subclade 5.1A) (O4, O12), another two belonged to clade VII (subclade 5.1B) (O3, O7), and only one oligotype belonged to clades II (subclade 5.1A) (O28), CB5 (subclade 5.2) (O10), WPC1 (subclade 5.1) (O17), VIII (subclade 5.1B)(O18), and IX (subclade 5.1B) (O13) (Tables 13).

Table 11. Key database for the 31 *Synechococcus* oligotypes including their accession number on NCBI, source, clade or subclade affiliations, and sequence length

Oligotype ID	Strain name	Accession number	Description	Source	Subclade or Clade ID best match	Clade ID	Length	Number of bp different from the best match
O1	WH8103	AF311293.1	*Synechococcus*sp. WH 8103 16S ribosomal RNA gene partial sequence	Ahlgren & Rocap 2012	III	III	1438	0 (exact match)
O2	KORDI-15	FJ497743.1	*Synechococcus* sp. KORDI-15 16S ribosomal RNA genepartial sequence	Choi & Noh 2009	Subcluster 5.3	5.3	1227	0 (exact match)
O3	A15-74	JF306716.1	*Synechococcus* sp. A15-74 16S ribosomal RNA genepartial sequence	Mazard et al. 2011	VII	VII	1099	0 (exact match)
O4	CC9902	CP000097	*Synechococcus* sp. CC9902 chromosomecomplete genome	Ahlgren & Rocap 2012	IV	IV	1479	0 (exact match)
O5	RCC61	JF306685.1	*Synechococcus* sp. RCC61 16S ribosomal RNA genepartial sequence	Mazard et al. 2011	SC 5.3	5.3	1148	1
O6	MV0210	KU867926.1	*Synechococcus* sp. MV0210 16S ribosomal RNA gene	Hunter-Cevera et al. 2016	IE	I	1148	0 (exact match)
O7	UW92	JQ421031.1	*Synechococcus* sp. UW92 16S ribosomal RNA genepartial sequence	Ahlgren & Rocap 2012	VII	VII	1019	1
O8	MV1308	KU867940.1	*Synechococcus* sp. MV1308 16S ribosomal RNA gene	Hunter-Cevera et al. 2016	IC	I	1148	1
O9	MV1307	KU867924.1	*Synechococcus* sp. MV1307 16S ribosomal RNA gene	Hunter-Cevera et al. 2016	IE	I	1148	1
O10	MV0605E	KU867943.1	*Synechococcus* sp. MV0605E 16S ribosomal RNA gene	Hunter-Cevera et al. 2016	CB5	CB5	1148	0 (exact match)
O11	KORDI-15	FJ497743.1	*Synechococcus* sp. KORDI-15 16S ribosomal RNA genepartial sequence	Choi & Noh 2009	Subcluster 5.3	5.3	1227	1
O12	CC9902	CP000097	*Synechococcus* sp. CC9902 chromosomecomplete genome	Ahlgren & Rocap 2012	IV	IV	1479	1

O13	KORDI-71	FJ497742.1	*Synechococcus* sp. KORDI-71 16S ribosomal RNA genepartial sequence	Choi & Noh 2009	IX	IX	1145	0 (exact match)
O14	MV1308	KU867940.1	*Synechococcus* sp. MV1308 16S ribosomal RNA gene	Hunter-Cevera et al. 2016	IC	I	1148	0 (exact match)
O15	WH8102	NC_005070	*Synechococcus* sp. WH8102 complete genome	Ahlgren & Rocap 2012	III	III	1464	2
O16	WH8016	AY172834.1	*Synechococcus* sp. WH 8016 16S ribosomal RNA genepartial sequence	Mazard et al. 2011	I	I	1439	1
O17	KORDI-49	FJ497747.1	*Synechococcus* sp. KORDI-49 16S ribosomal RNA genepartial sequence	Choi & Noh 2009	WPC1	WPC1	1380	1
O18	WH8101	AF001480.1	AF001480 *Synechococcus* WH8101 16S ribosomal RNA genepartial sequence	Ahlgren & Rocap 2012	VIII	VIII	1437	1
O19	WH8103	AF311293.1	*Synechococcus* sp. WH 8103 16S ribosomal RNA genepartial sequence	Ahlgren & Rocap 2012	III	III	1438	1
O20	KORDI-15	FJ497743.1	*Synechococcus* sp. KORDI-15 16S ribosomal RNA genepartial sequence	Choi & Noh 2009	Subcluster 5.3	5.3	1227	2
O21	MV1605B	KU867931.1	*Synechococcus* sp. MV1605B 16S ribosomal RNA gene	Hunter-Cevera et al. 2016	III	III	1148	1
O22	KORDI-15	FJ497743.1	*Synechococcus* sp. KORDI-15 16S ribosomal RNA genepartial sequence	Choi & Noh 2009	Subcluster 5.3	5.3	1227	2
O23	WH8103	AF311293.1	*Synechococcus* sp. WH 8103 16S ribosomal RNA genepartial sequence	Ahlgren & Rocap 2012	III	III	1438	1
O24	WH8103	AF311293.1	*Synechococcus* sp. WH 8103 16S ribosomal RNA genepartial sequence	Ahlgren & Rocap 2012	III	III	1438	1
O25	WH8103	AF311293.1	*Synechococcus* sp. WH 8103 16S ribosomal RNA genepartial sequence	Ahlgren & Rocap 2012	III	III	1438	1
O26	WH8020	AY172835.1	*Synechococcus* sp. WH 8020 16S ribosomal RNA genepartial sequence	Ahlgren & Rocap 2012	I	I	1226	0 (exact match)
O27	KORDI-15	FJ497743.1	*Synechococcus* sp. KORDI-15 16S ribosomal RNA genepartial sequence	Choi & Noh 2009	Subcluster 5.3	5.3	1227	2
O28	MV0519B	KU867928.1	*Synechococcus* sp. MV0519B 16S ribosomal RNA gene	Hunter-Cevera et al. 2016	II	II	1149	1
O29	WH8103	AF311293.1	*Synechococcus* sp. WH 8103 16S ribosomal RNA genepartial sequence	Ahlgren & Rocap 2012	III	III	1438	1
O30	WH8103	AF311293.1	*Synechococcus* sp. WH 8103 16S ribosomal RNA genepartial sequence	Ahlgren & Rocap 2012	III	III	1438	1
O31	WH8103	AF311293.1	*Synechococcus* sp. WH 8103 16S ribosomal RNA genepartial sequence	Ahlgren & Rocap 2012	III	III	1438	1

The eight abundant oligotypes in our dataset (i.e., represented by >100 reads) shared more than 95% V4-V5 sequence similarity with each other (Table 14).

Table 12. Percent sequence similarity between V4 and V5 representative sequences for each oligotype.

	O1	O2	O3	O4	O5	O6	O7	O8
O1	-	96.07	99.44	99.44	96.63	99.44	98.88	99.72
O2		-	96.26	95.99	99.47	96.79	96.79	96.52
O3			-	98.93	96.79	98.93	99.47	99.2
O4				-	96.52	98.93	98.4	99.2
O5					-	96.26	96.26	96.52
O6						-	99.47	99.73
O7							-	99.2
O8								-

Synechococcus comprised a higher fraction of the microbial population and showed higher relative abundances in the summer of 2014 in Marchica than in 2015. We placed the eight abundant oligotypes within a phylogenetic tree that included known *Synechococcus* strains (Figure 2). Table 1 shows higher read counts of *Synechococcus* oligotypes in the Marchica Lagoon (n = 15,447), (n=1,667) in the summers of 2014 and 2015, respectively, compared to Oualidia (n = 26), (n = 14) during both the summers of 2014 and 2015.

Interestingly, Oligotype 1 was strongly represented in both sampling lagoons (Table 11), where it comprised a larger segment of the overall *Synechococcus* community. In contrast to the 2014 summer community in Marchica, a few *Synechococcus* oligotypes decreased in 2015; for instance, O7 changed from 177 *Synechococcus* reads to 0, and O8 changed from 119 *Synechococcus* reads to 0. In Oualidia, we noticed the presence of clades III (O1), IV (O4), I (O6, O14, O26), and VII (O7) in 2014 in contrast with 2015, when clade VII was absent and only 5 *Synechococcus* oligotypes were identified: O1, O4, O6, O9 and O14.

6.3. Distribution of Synechococcus oligotypes

Although some oligotype distribution patterns within each sampling site clearly displayed some differences over both space and time, many oligotypes were shared as well. Network analysis allowed visualization of the specificity of the oligotypes and how they were distributed in Mediterranean Marchica and Atlantic Oualidia lagoons and further investigation of which factors influenced this distribution (Figure 21).

We identified oligotypes that were either unique to a single sampling site, shared by two sites, or present in all sites (Figure 21). Site-specific oligotypes accounted for the largest fraction

(25/31), whereas oligotypes shared by the 4 sampling sites made up a small portion of the total number (4/31). Most oligotypes were shared between both 2014 and 2015 samples of Mediterranean Marchica, in addition to some overlap with Atlantic Oualidia. Only six oligotypes were found in Mediterranean Marchica in 2014 (Figure 3). Furthermore, the distribution of co-occurring oligotypes was clearly different in Marchica and Oualidia during the summers of 2014 and 2015. Cooccurring *Synechococcus* oligotypes tended to be well connected to each other (Table 3), showing strong correlations (e.g., O4 (clade IV) and O6 (clade I), = 0.98, O4 (clade IV) and O8 (clade I), = 0.99).

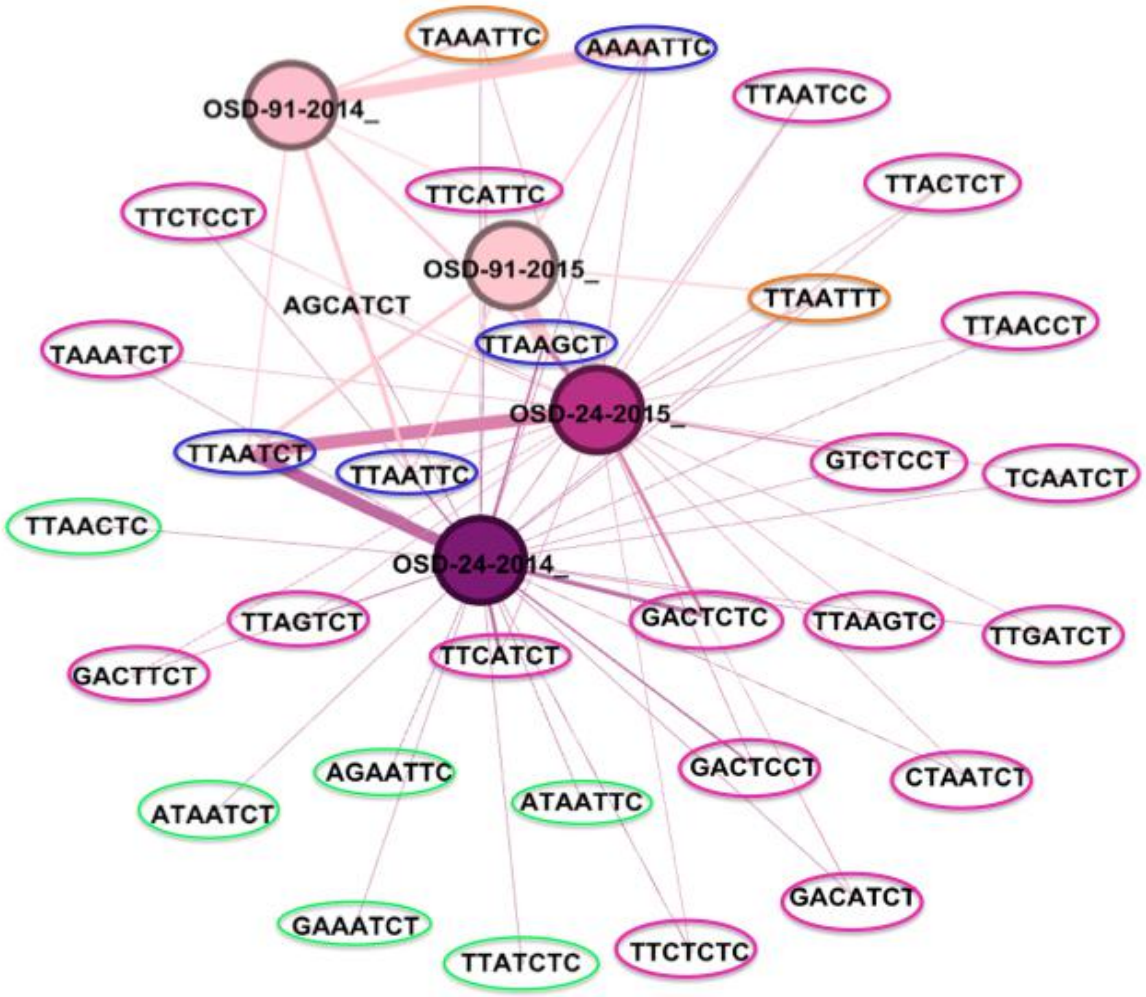


Figure 16. A network analysis of the oligotypes present at each site was performed using Gephi. Each dot indicates an oligotype present in at least one sample, and each edge on the network connects an oligotype to one or more samples. Colored circles represent oligotypes present in only one, two, three, or all samples.

6.4.Environmental variables influencing oligotype diversity

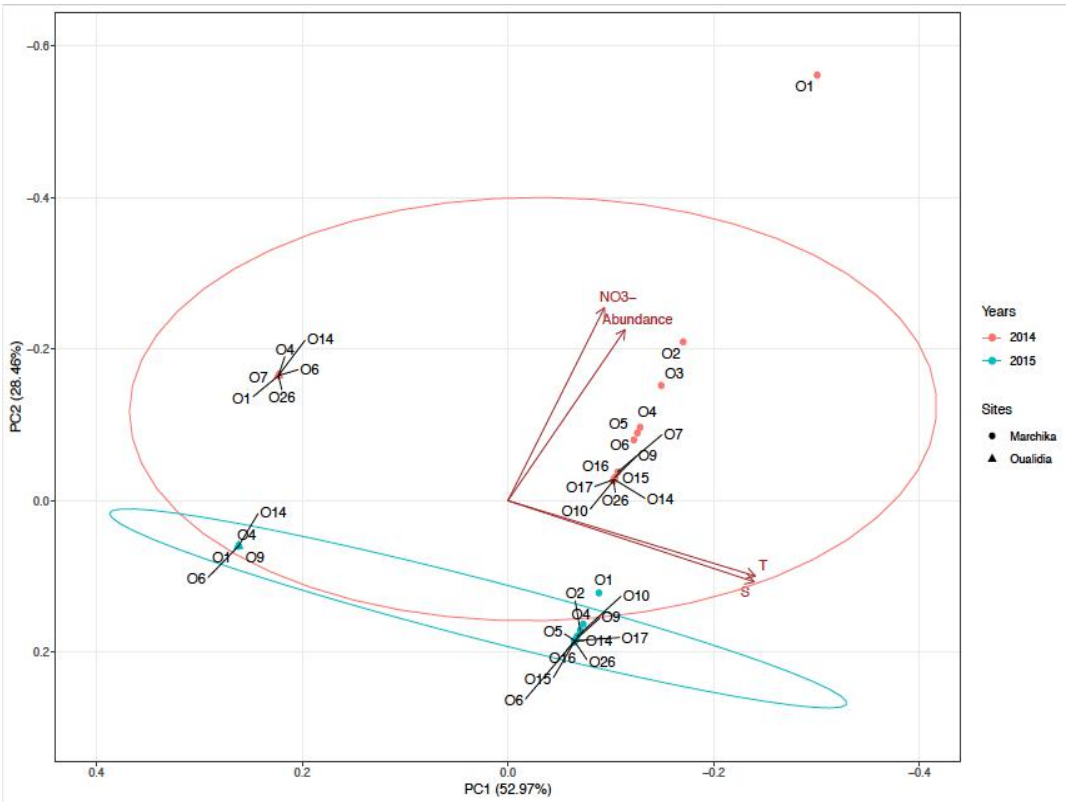


Figure 17. Principle component analysis of Synechococcus oligotype relative abundance. The plot is generated using the relative abundance of each oligotype, T: Temperature, S: Salinity, and NO3-: Nitrate. Each point represents an oligotype. Colors represent the year of sampling; red for 2014 and blue for 2015. The shape of point indicates the sampling site; rounded points refer to the Marchica lagoon, and triangles refer to Oualidia. Circles represent the normal distribution of oligotypes; the red circle refers to 2014, and the blue circle refers to 2015.

Following principal component analysis (PCA) of oligotype relative abundance (Figure 22), we observed that physicochemical factors in the lagoons correlated with oligotype cooccurrence patterns. The first principal component (PC1) captured 46% of the variance in oligotype relative abundance and discriminated oligotypes according to nitrate. The second principal component explained an additional 27% of the variation and discriminated oligotypes according to salinity and temperature.

The composition shift seen during the summer was supported by the statistical connections between dominating PCs and environmental variables. Notably, identified oligotypes in Marchica (O2, O3, O4, O5, O6, O7, O9, O10, O14, O15, O16, O17, O26) in 2014 and (O1, O2, O4, O5, O6, O9, O10, O14, O15, O16, O17, O26) in 2015 correlated with higher temperatures and salinities (26 Celcius, 27 Celcius, 35 ppt, 35 ppt, respectively), when oligotypes in Oualidia

(O1, O4, O6, O7, O14, O26) in 2014 and (O1, O4, O6, O9, O14) in 2015 correlated with lower ones (21 Celcius, 20 Celcius, 27 ppt, 29 ppt respectively). Both oligotypes found in Marchica and Oualidia in 2014 correlated with a higher nitrate concentration (12 mg/l, and 10 mg/l, respectively). However, those observed in 2015 correlated with a lower nitrate concentration (4 mg/l, 2 mg/l). Furthermore, Oligotype O1 in 2014 was spatially isolated, and not affiliated with the principal component.

Discussion

METAGENOMES AND PHYSICOCHEMICAL PROPERTIES OF MOROCCAN COASTAL LAGOONS

In this project, we intended to delineate the microbial diversity and abundance within the surface waters of two Moroccan coastal lagoons, Marchica and Oualidia, depict their functional capabilities, and establish their variation patterns across environmental, temporal, and spatial gradients. Through the use of shotgun and amplicon sequencing metagenomics technologies, we obtained a fair amount of read data from marine water-filtered extracted DNA, which allowed us to identify the taxonomic structure and functional profiles of the microbial communities in these lagoons.

The degradation of water quality in lagoons by both natural and anthropogenic activities has a direct impact on the biodiversity and ecological balance, which has triggered the need for suitable conservation and management strategies. The physicochemical parameter measurements of temperature, salinity, electrical conductivity, pH, dissolved oxygen, phosphate, turbidity, nitrates, and nitrites were set out to assess the ecological health status of the Marchica and Oualidia lagoons. The higher rate of salinity (35 ppt) observed in Marchica in comparison to the Mediterranean Sea and Oualidia makes it a hypersaline lagoon. During the dry season, salinity increases because of the temperature increase. Other factors also impact salinity, such as shallow depth, evaporation, and volume of freshwater influx. However, in Oualidia, the absence of freshwater inflow to the lagoon suggests that the lower recorded salinity (27 ppt) is influenced by underground water resurgence **(Natij *et al.* 2014)**. pH is measured to assess the health status of aquatic ecosystems; it is known to influence the development and spread of microbes and has a direct impact on the environmental conditions that contribute to microorganism survival and expansion **(Thompson *et al.* 2017)**. The pH of both lagoons was mildly alkaline, which is similar to results found by **(Matoir *et al.* 2015)**, where the pH of the Marchica lagoon oscillated between 8.1 and 8.7, corresponding to the wastewater treatment plant during the summer season **(Riouchi *et al.* 2021)**. In Oualidia, 8.03 was the pH value recorded for dry months, which implied an explicit marine influence **(Hassou *et al.* 2013)**. The concentrations of dissolved oxygen in Marchica were remarkably low (8.3 mg/L) compared to those reported at Mohandis station (13,76 mg/L) in June 2013 by **(Chagas *et al.* 2005)**, which is possibly related to the fair oxygenation observed at the surface after the installation of an artificial inlet in 2010 that helped with recycling the Marchica lagoon waters. In addition, algal death is another factor that

influences the decrease in DO in Nador Lagoon **(Chagas *et al.* 2005)**. In Oualidia Lagoon, samples were taken from the upstream region, which is known to be more oxygenated in spring and less oxygenated in summer. This is due to the growing rate of marine organism respiration and aerobic biodegradation of dissolved organic matter by heterotrophic bacteria **(Damsiri *et al.* 2014)**. In reality, three essential factors influence the good oxygenation of Oualidia lagoon waters: freshwater inflows, depth, and surface water agitation **(Damsiri *et al.* 2014)**. The recorded turbidity in both lagoons (3 NTU) fell under the normal value set by the U.S. Geological Survey (USGS) (1–50 NTU). This means that the water is clear, which increases the amount of sunlight to penetrate the water layers, thereby increasing the photosynthetic rate **(Aknaf *et al.* 2017)**. The low phosphate concentrations recorded at both Marchica (0.03 mg/L) and Oualidia (0 mg/L) might be explained by the high renewal rate of water **(Damsiri *et al.* 2014; Aknaf *et al.* 2015)**. Nitrate concentrations in 2014 in Marchica and Oualidia lagoons (12.5 mg/L and 10.79 mg/L, respectively) showed slightly higher values compared to the maximum contaminant level (MCL) set in the United States for the concentration of nitrates in water (10 mg L^{-1}) and to the World Health Organization (WHO) guidelines (11.3 mg L^{-1}). This could be explained in Marchica by terrigenous inputs from Selouane and Caballo Oueds (streams). Furthermore, Oued Caballo carries urban wastewater from the Zeghanghane neighborhood, while Oued Selouane transports urban and toxic waste toward the lagoon **(Aknaf *et al.* 2015)**. Farming in the Bou Areg adjacent plain is the main surface and groundwater contamination source of nitrates, nitrites, and phosphate in Marchica. On the other hand, the wastewater treatment plant unwittingly increases these nutrients in the lagoon system **(Aknaf *et al.* 2015)**. In Oualidia, the quantity of fertilizers used in this agricultural area is generally the main source of nitrates **(Damsiri *et al.* 2014)** and heavy metals **(Hassou *et al.* 2016)**. In 2015, we noticed a significant decrease in the nitrate level (from 12.5 mg/L to 4.8 mg/L in Marchica and from 10.79 mg/L to 2.68 mg/L in Oualidia), which could be due to remediation actions by local authorities in Marchica by recycling the wastewater and materials, collecting waste, and improving the renewal of the lagoon waters. Indeed, the human impact was still noticed in Oualidia but was much less compared to 2014. This could be explained by the newly installed pit, which aimed to promote water renewal, improve water quality, and reduce the rate of suspended matter **(Damsiri *et al.* 2014)**. Almost certainly, our results suggest that the opening of the new inlet and the installation of waste treatment plants had a positive impact on the physicochemical properties, which indicates an improvement in the lagoon's water quality.

Looking at the viral signals, we believe our results nicely illustrate the fact that we started to have a good "baseline database" of oceanic viral genomes (and genome fragments), although TOV_43 (Tara oceans) is highly biased toward the open ocean, whereas OSD is biased toward coastal sites. We also think that the fact that we detected the same viral genome/contingent in several distant OSD samples confirms that ocean viruses have a wide distribution (Figures 6 and 7). This is consistent with what was observed in the TOV_43 study (https://www.ivirus.us/data_accessed on 15 December 2020), with the added information that the OSD samples are microbial metagenomes, so should include only "actively infecting" viruses, and were synchronized (whereas Tara sampling was spread across multiple years), so we have the actual "proof" that the same virus was active at the same time in multiple places. Finally, we think we may correlate this viral signal to putative hosts: we will not have strong statistical power, but for the "known" viruses, we should be able to check if the corresponding host was indeed abundant in the same sample, and, for the "unknowns", we should check which were the major microbial groups in the corresponding samples that would at least provide putative hosts.

By examining the community composition, samples from both studied locations appeared to be made up mostly of *Proteobacteria* members (α and γ), *Cyanobacteria* (*Synechococcus*), *Bacteroidetes* (*Flavobacteriales*), *Verrucomicrobia* (*Verrucomicrobiae*), and *Actinobacteria* (*Actinobacteria*). *Alphaproteobacteria* seemed to be the highest identified class across the different sampling locations, covering an average of 20% to 45% of the microbial communities in Oualidia and Marchica, respectively. The class consisted of *Rhodobacteraceae,* which are classified as denitrifiers capable of rapid growth and reducing nitrates in both oxic and anoxic conditions **(Morris *et al.* 2002)**. Their increasing abundance in lagoon coastal waters might be linked with the decrease in nitrogen content at both Marchica and Oualidia stations in 2015 (4 mg/L, 2 mg/L) compared to 2014 (14 mg/L, 2 mg/L). A sizeable proportion of reads belonged to *Candidatus Pelagibacter* Rappé (11% in Marchica; 7% in Oualidia) in 2014. Interestingly, in 2015, none were detected. This genus is known as an eminent member of SAR11 clusters, which are classified as the largest group of bacteria in the surface water of marine ecosystems **(Giovannoni *et al.* 2017)**. This is in line with findings from **(Ghai *et al.* 2012)** on Spanish Mar Menor Lagoon, where half of all 16S reads were assigned to *Candidatus Pelagibacter* (43%). In addition, SAR11 reads were encountered in all samples collected from the surface waters of different marine habitats either in the Pacific or Atlantic oceans, specifically coastal environments, as reported by Tara Oceans and the Global Ocean Sampling Expedition **(Venter**

***et al.* 2004)**. Different strains were within the *Rhodobacterales* group, particularly *Rickettsia conorii 7* Ogata, which formed the most abundant strain identified in both lagoons. It is generally present on the Mediterranean side of Europe and is known to cause boutonneuse fever **(MacConnachie *et al.* 2022)**. *Gammaproteobacteria*, in contrast, made up lower proportions compared to *Alphaproteobacteria*. This is typical of *Gammaproteobacteria* because they are found in large numbers in benthic and lower pelagic marine environments **(Zinger *et al.* 2011)**. This class was composed of *Alteromonadales,* which have been found to be related to sites affected by urbanization and eutrophication **(Juhmani *et al.* 2020)**. In addition, members of this group are resistant to metals **(Juhmani *et al.* 2020)**. Spatial and temporal variation in the distribution of *Alteromonadales* compared to *Rhodobacterales* might be assigned to competition for limiting nutrients **(Seo *et al.* 2014)**. The strain *Xylella fastidiosa* Temecula1 Sluys, classified among the top dangerous pathogenic bacteria infecting plants worldwide **(Schneider *et al.* 2020)**, was the dominant strain found in the *Gammaproteobacteria* group.

We found further evidence of the occurrence of almost exclusively the genus *Synechococcus* Nägeli as the most abundant *Cyanobacteria*, notably in Marchica, which was similarly observed in Mar Menor Lagoon, where *Synechococcus* formed a sizeable percentage and declined with increasing salinity in the Mediterranean Sea **(Ghai *et al.* 2012)**. Almost no *Prochlorococcus Penny* populations were found in Oualidia waters, which is similar to the results in the parent Atlantic water body from which Oualidia waters are derived (OSD 93). This is the usual behavior of *Prochlorococcus, which is* commonly not found in coastal marine environments **(Partensky *et al.* 1999)**. In addition, the *Synechococcus* growth rate increases with increasing temperature and nitrates, while increasing temperature and light levels might retard *Prochlorococcus* growth **(Kim *et al.* 2018)**. In fact, salinity and temperature are essential environmental factors for the composition and function of *Synechococcus* in the marine microbiome **(Kim *et al.* 2018)**. According to previous studies, *Cyanobacteria* grow in nutrient-enriched ecosystems, and their prevalence increases with eutrophication **(Lürling *et al.* 2018)**. However, when the macronutrient concentration is low in the summer season, *Synechococcus* shows fast growth **(Lürling *et al.* 2018)**. The highest number of *Synechococcus* read recruitment in Marchica in 2014 compared to 2015 could be explained by the lagoon's intensive environmental rehabilitation effort, which includes cleaning wastes and depollution. Based on strain identification in the metagenome, we were able to identify the cyanobacterium *Nostoc*

azollae 0708 *Ran,* which is famous for its nitrogen-fixation properties and is commonly used in farming **(Brouwer *et al.* 2017)**.

Most of the remaining sequences were mainly assigned to moderately smaller phyla, including *Bacteroidetes*, *Verrucomicrobia*, and *Actinobacteria*. We noticed the unique existence of *Bacteroides fragilis* 9343 Coyne (3%) in Venice Lagoon, and *Candidatus Amoebophilus asiaticus* 5a2 Schmitz-Esser (9%) comprised the underrepresented *Cytophagales* group in the Atlantic Ria Formosa Lagoon. The former strain could potentially be used as an organic pollution bioindicator **(Niestępski *et al.* 2020)**.

Archaeal assemblages made up a tiny percentage of the metagenome, forming less than 1% of the total sequences in both Marchica and Oualidia lagoons. These findings were in agreement with previous reports where barely detectable levels of archaeal sequences were found in surface marine environments **(Ghai *et al.* 2012; Rusch *et al.* 2007; Biers *et al.* 2009)**. In contrast, Archaea have a greater preference for deep waters, where they exist in significantly higher numbers **(Signori *et al.* 2014; Luria *et al.* 2014)**. *Euryarchaeota* were the dominant clade, which is normal as they are known to inhabit surface waters **(Brown *et al.* 2009)**.

Phytoplankton sequence data showed dissimilarities between the two lagoons, where Marchica was dominated by *dinoflagellates* and Oualidia by *Ochrophytes* and *Chlorophytes*. The red dinoflagellate *Gyrodinium* Hulbert belongs to the family *Dinophyceae*, which lacks armor protists responsible for red tides **(Gárate *et al.* 2013)**. The *Ochrophyta* genus *Pseudo-Nitzschia* Peragallo, which belongs to diatoms, is responsible for amnesic shellfish poisoning (ASP); blooms were observed globally in coastal waters and were linked to elevated nutrient concentrations in marine environments **(Brouwer *et al.* 2017)**. While the *Chlorophyta* genus *Tetraselmis* Stein is a green alga within the order *Chlorodendrales*, the species members' habitat preference depends on whether the water is deep or shallow because of their photosynthetic nature. Thus, they inhabit different water environments if sufficient light and nutrients are accessible **(Teo *et al.* 2014)**. In addition, many species belonging to the genus *Tetraselmis* have recently been investigated for use as biofuels because of their increased lipidic content **(Teo *et al.* 2014)**. Both *Gyrodinium* and *Nitzschia* were abundantly identified in the Mediterranean Spanish coastal lagoon Mar Menor using microscopic examination and enumeration. These diatom species are linked to high nutrient concentrations, especially phosphorus, and are hence associated with contaminated eutrophic waters **(Ghai *et al.* 2012)**. Overall, the taxonomy of

these groups remains to be elucidated **(Ghai *et al.* 2012)**. In Oualidia, most hits were assigned to the *Ostreococcus* Courties genus, a unicellular green alga that has a crucial role in the oceanic carbon cycle, and a model to study the adaptation of green algae in the marine environment, *Pelagodinium* HJ Spero and *Thalassiosira* Cleve. The latter is widely distributed throughout the world's oceans **(Chappell *et al.* 2013)**.

The functional diversity of microbial communities was explored in our study using a shotgun metagenomics approach. Across the different sampling sites, we observed similar patterns in the identified communities with regard to their functional features. The topmost functional genes were closely similar in Marchica and Oualidia, essentially comprising genes implicated in protein, carbohydrate, and amino acid metabolism. Previous reports supported these findings, as the surface microbial communities were found to be located in the Mediterranean Sea, Atlantic Ocean, and Pacific Ocean, where they share a close dissemination of gene-encoded functions **(Venter *et al.* 2007; Rusch *et al.* 2007)**. This would possibly suggest that the survival and adaptation of these microbial communities depends on the corset of genes they share in the surface water environments **(Hewson *et al.* 2009)**. Remarkably, *Methylacidiphilum infernorum* V4 Dunfield and *Rothia dentocariosa* 17931 Eisenberg exhibited the highest proportions of genes with metabolic functions in Marchica. However, these amounts were related to *Anaplasma marginale* Maries Brayton and *Nostoc azollae* 0708 Ran in Oualidia. Furthermore, the correlation between microbiome gene expression and water properties showed that the latter, namely, temperature, salinity, conductivity, oxygen, phosphate, nitrates, and nitrites, influenced the expression of genes with metabolic functions, as the correlation was positive. This was indeed reported by **(Scofield *et al.* 2015)**, who studied five shallow coastal lagoons in Brazil and concluded that temperature and the concentration of nutrients are responsible for bacterial metabolism regulation. However, the pH in our study showed a negative impact on metabolic processes. This result is in accordance with a study by **(Krause *et al.* 2012)**, which found that pH has no influence on bacterial abundance.

This metagenomics study enriched the Moroccan national catalog of marine microorganisms that can be used for bioindication, biodiversity protection, biotechnology valorization, biomonitoring, and lagoon health assessment. For instance, the presence of *Synechococcus* Nägeli and *Bacteroides fragilis* Coyne belonging to the *Cyanobacteria* and *Bacteroides* groups, respectively, could be associated with the eutrophic conditions of the lagoon Marchica. Indeed, the rate of nitrogen in the Marchica Lagoon is alarmingly high (12 mg/L). Hence, we may further suggest

using these marine bacteria as bioindicators of deteriorated water quality in coastal lagoons. Notably, *Cyanobacteria* species are good candidates for environmental sustainability and an excellent agricultural bioresources **(Singh *et al.* 2016)**. Take, for example, *Nostoc azollae* Ran, a biofertilizer used for either Taro or the cultivation of rice. It has also shown great bioremediation potential in reducing pollutants and improving waste management **(Niestępski *et al.* 2020)**. *Cyanobacteria* have been remarkably implemented in blue biotechnology, a newly emerging solution to the high demand for biomedicine, pharmaceutical and agricultural products, natural cosmetics, and sustainable energy sources by modern societies by exploiting marine biodiversity. They have shown their ability to generate molecules with antibacterial, antifungal, and antiviral traits. Today, they are being investigated for their anti-multiplicative effect on cancer cells. An average of 400 biomolecules secreted by *Cyanobacteria* strains have been listed **(Bajpai *et al.* 2018)**. Other identified marine bacteria in our study, although with low frequencies, expressed impressive prospects for biotechnology applications, for example, *Alcanivorax* Yakimov and *Acinetobacter* Brisou species. The former plays an important role in oil spill bioremediation **(Zadjelovic *et al.* 2020)** and when together they are of high interest for beauty technology since they have been shown to produce biosurfactants **(Antoniou *et al.* 2015)**. Proper management of lagoon health is another important aspect to take into consideration by implementing biomonitoring, which will help us eventually better understand exposures to environmental chemicals, shape public health, and prevent the development of specific diseases caused by particular microorganisms, such as *Xylella fastidiosa* Sluys, a pathogen infecting plants that in Brazil infected approximately 200 million citrus trees and caused damage estimated at USD 100 million in the grape industry in California **(Schneider *et al.* 2020)**. We were also able to detect *Mycobacterium leprae* Hansen in the waters of both lagoons, which could possibly lead to leprosy infection. However, further investigations are needed to support this claim.

In connection with spatial variation, the microbial community was dominated by similar phyla identified in 2014 and 2015. However, clear differences were manifested at the genus, species, and strain divisions among samples. Hence, the bacterioplankton were diverse but modestly varied across samples, as shown by the alpha diversity. The observed richness in Oualidia was 1144 in 2014 and 704 in 2015, while in Marchica, we found 462 in 2014 and 316 in 2015. This also elucidated the importance of performing more sampling as the number of species captured fluctuates and the taxonomic richness increases with a high number of reads. Each lagoon chosen for this work displayed differences in physicochemical parameters such as

temperature, salinity, DO, or nitrate concentrations in addition to reported anthropogenic pressures due to agricultural, industrial, and human impacts, which probably influenced microbial community distinctions at each site. The relationship between spatial variation and community structure needs more study for better understanding.

SPATIO-TEMPORAL PATTERNS OF *SYNECHOCOCCUS* OLIGOTYPES IN MOROCCAN LAGOONAL ENVIRONMENTS

In a previous study **(Chaouni *et al.* 2022)**, we used bioinformatics tools to analyze the metagenome and the amplicon 16S sequences to gain an insight into microbial diversity in Moroccan lagoons, namely Marchica and Oualidia. 16S rRNA gene classification revealed a high percentage of bacteria in both lagoons. On average, bacteria accounted for 90% of the total prokaryotes in Marchica and ~70% in Oualidia. The five phyla that were the most abundant in both lagoons, Marchica and Oualidia, respectively, were *Proteobacteria* (53.62%, 29.18%), *Bacteroidetes* (16.46%, 43.49%), *Cyanobacteria* (0.53%, 34.35%), *Verrucomicrobia* (1.75%, 15.82%), and *Actinobacteria* (7.42%, 13.98%). At the genus level, we found that the highest assigned hits were attributed to *Synechococcus*, which was highly abundant in Marchica (32%) compared to Oualidia (0.07%) in 2014. This amount dropped to 22% in Marchica and 0.04% in Oualidia in 2015. Hence, in this study we performed an analysis of the *Synechococcus* genus community using oligotyping to investigate their dynamics and understand their co-occurrence and covariation in space and time within fragile ecosystems such as lagoons.

We may divide our results into two emerging *Synechococcus* communities: one dominated in 2014 and the other was less present in 2015, each composed of different co-occurring *Synechococcus* oligotypes. The abundant *Synechococcus* community in Marchica in 2014 consisted of clades I, 5.3, III, IV, and VII. These clades are typically found in either warmer or more oligotrophic environments **(Zwirglmaier *et al.* 2008; Post *et al.* 2011)**. This result is in accordance with Marchica's environmental characteristics; it is an oligotrophic ecosystem with high primary production and warmer water in summer **(Aknaf *et al.* 2017)**. The community included clades CB5 and WPC1 when the number of *Synechococcus* reads was lower. Strains belonging to the CB5 clade lack phycourobilin (PUB), contain one motile strain **(Chen *et al.* 2004; Waterbury *et al.* 1986),** are present in temperate coastal waters and are prevalent in polar/subpolar waters (207, 208, 209). WPC1 strains are observed in open-ocean and near-shore waters **(Chen *et al.* 2006; Choi *et al.* 2009; Mazard *et al.* 2012)**. Clades IV and I usually co-occur and are more prevalent in cold coastal waters **(Zwirglmaier *et al.* 2008; Zwirglmaier *et al.***

2007; Tai *et al.* 2009; Mella-Flores *et al.* 2011). Interestingly, Clade III was prominent in Marchica. This clade is known to be motile and restricted to warm, oligotrophic water **(Zwirglmaier *et al.* 2008; Mella-Flores *et al.* 2011; Post *et al.* 2011).** Although at a smaller read number, clade III was also observed in Oualidia, where the temperature is cooler compared to Marchica. Furthermore, we found that clade III growth has been shown to be severely affected at low temperatures **(Varkey *et al.* 2015).** Moreover, representatives of both clades I and IV were present in Oualidia in both the summers of 2014 and 2015. Some *Synechococcus* strains, which are known to prefer cooler water temperatures and salinities, were in higher relative abundance in the waters of Marchica. This result agrees with a previous study showing that *Synechococcus* isolates of clades I and IV exhibited temperature preferences **(Varkey *et al.* 2015)**. Their growth rates were marginally lower at low temperatures in strains from clades I and IV, which were dominant in temperate regions.

Nitrate levels are typically low or undetectable in these lagoons, which allows the persistence of clades that would not typically thrive in coastal waters at other times of the year. In 2014, the nitrate concentration was higher than the average of 10 mg/l, which could be due to increased agricultural activities and wastewater treatment plant effluent. The decreasing nitrate concentrations in Marchica in 2015 could be explained by the newly installed inlet in 2010, which was designed to improve water exchange with the open sea and reduce the amount of suspended matter **(Aknaf *et al.* 2017)**. Temperature and salinity have a large effect on nitrate in marine ecosystems **(Ebrahimi *et al.* 2015)**; the highest nitrate degradation rates were observed at 35°C and at increasing salinity rates. Therefore, we expected to see correlations between salinity, temperature and nitrate concentrations. Interestingly, clades CB5 in Marchica and IV in Oualidia increased in relative abundance in summer 2015 compared to 2014 (Table 1), when the nitrate concentration decreased (Table 5). Moreover, *Synechococcus* microbial community diversity and density are variables that depend on variations in physical and chemical parameters. These parameters are strongly influenced by the marine waters passing through the artificial inlets, which have an impact on the internal hydrodynamics of both lagoons and hence the distribution and co-occurrence of *Synechococcus* strains. In addition, anthropogenic activities also have a great influence on *Synechococcales* population growth and interactions with their viruses **(Traving *et al.* 2014; Chaouni *et al.* 2022)**.

This study revealed some differences between Marchica and Oualidia in identified *Synechococcus* clades. The Marchica lagoon showed more heterogeneity (clades I, II, III, IV, VII,

VIII, 5.3, WPC1, CB5, and IX) than the Oualidia lagoon, where fewer clades were identified (I, III, IV, and VII). There was a clear variation in the pattern of correlation between oligotypes of the same or different clades for both the 2014 and 2015 samplings. Furthermore, we observed complex patterns of co-occurrence among oligotypes; in 2014 (clades I, III, IV, 5.3, VII) and in 2015, we found clades CB5 and WPC1. In Oualidia, values decreased in comparison to Marchica in both 2014 and 2015 summer samplings, following a pattern of co-occurrence, especially for clades I and IV in both sampling years. Many studies have shown that the relative proportions of cooccurring *Synechococcus* populations to each other at the clade and subclade levels vary in space and time based on environmental factors such as seasonal temperature fluctuations, nutrient availability and upwelling, circulation patterns, and abundance of other phytoplankton **(Mackey *et al.* 2017)**.

We presume that the greater variability in oligotype co-occurrence behavior observed in Marchica Lagoon, especially in the summer of 2014, could be due to the higher abundance and diversity of *Synechococcus* oligotypes, physico-chemical parameter fluctuations or rehabilitation of the lagoon.

Less abundant oligotypes could also be considered potential bioindicators of *Synechococcus* genetic diversity. Their seasonal occurrence might contribute to changing ecological and biogeochemical characteristics of the marine environment **(Gilbert *et al.* 2012)**. The *Synechococcus* relative abundance count revealed that the Marchica *Synechococcus* community included the least abundant oligotypes in 2015. For instance, O7 and O8 were detected in 2014 and were absent in 2015. It is unclear which factors served to constrain the relative abundances of these least present oligotypes, but temperature and salinity could have an impact on their distribution in Marchica and the opposite for Oualidia, which are cooler-temperature adapted ones. We noticed that the relative abundance of cooccurring *Synechococcus* was not constant. For instance, oligotype 4 belonging to Clade IV showed higher values in summer 2014 (974 reads) in Marchica compared to summer 2015 (319 reads), and the opposite was observed in Oualidia, with a lower abundance compared to Marchica. Increased values of cooccurring clade I oligotypes (69, 210, and 61) were detected in the summer of 2014 in both lagoons.

In comparing our results with a study from Little Sippewissett Marsh (LSM) **(Mackey *et al.* 2017)** that used oligotyping to investigate the distribution of the genus *Synechococcus* in

space and time sequencing the V4-V6 hypervariable region of the 16S rRNA gene, we found 31 unique oligotypes, while they were only able to identify 12. In both studies, the proportion of *Synechococcus* oligotypes increased in summer and in coastal waters compared to estuaries. In addition, Clades I and IV were more abundant in hypersaline conditions, such as Marchica Lagoon. However, these clades were found in greater relative abundances at cold temperatures, in contrast to our study, where they were identified in Marchica's warm waters. Moreover, clade CB5 tended to be prominent at relatively warm temperatures (17 - 20 °C) **(Xia *et al.* 2015)**. In our work, it was not prevalent either in cooler or warmer water. Notably, the relative abundance of rare oligotypes was higher in warm hypersaline estuary waters, while in our case study, they occurred in cooler moderately saline Oualidia waters.

While the rpoC1 gene is a higher resolution diversity marker, 16S amplicon data can be used to explore the entire bacterial assemblage*, including Synechococcus* clade designations by oligotyping. The dominance of a certain clade could have many different ecological ramifications, especially as the clades can be incredibly diverse in their growth, loss, nutrient utilization and other attributes. The dominant clade's growth and loss patterns will set the stage for population dynamics. For example, if the dominant clade only blooms in a given environmental factor such as temperature, light, or salinity, this will then affect the timing of blooms and follow-on effects of subsequent grazing, lysis or even biogeochemical cycling. Even if the population is diverse, the dynamics as a whole will be a composite response of each individual clade's ecophysiology, making it important to understand the composition and how it changes over space and time.

Oligotyping has a great advantage in answering unexplained diversity contained in taxa using 16S rRNA gene sequences. Nevertheless, it has some limitations, as it acts optimally only when performed on taxa that are closely related. Regarding distantly related taxa, the high number of increased-entropy locations makes the supervision steps difficult. In addition, although oligotyping does not rely on clustering conditions or the availability of existing reads within reference databases, it demands preliminary operational taxonomic unit clustering to find closely related species appropriate for the analysis. This method is under continuous improvement to better exploit the information within subtle variations in 16S rRNA gene sequences **(Eren *et al.* 2013)**.

Conclusion

The results gained from the first study provided the first comparison of microbial communities found in Moroccan coastal lagoon waters. They will serve as a valuable platform for further advanced studies to better understand the microbiome in lagoon ecosystems. More locations, various depths, and other lagoons should be sampled for more community descriptions within these waters. This will simultaneously strengthen existing findings of community patterns and confirm hypotheses that have been made on the relationship between microbial community structures, geography, and physical–chemical parameters. Seasonality studies can also be conducted to observe any changes in patterns by adding a broader time scale. The prokaryotic communities among a vertical profile can be analyzed and compared to those sampled across a horizontal gradient. It would also be interesting to further study microbial populations in surface lagoon waters and more deeply. A focus on minor and underrepresented taxa and a more profound and extensive analysis of the sequences obtained from these metagenomes can be carried out later to identify the taxonomic identity and functions of unknown reads and define more bioindicators/biomarkers of fragile marine ecosystems such as lagoons.

The second study explored the patterns of *Synechococcus* diversity in space and time using an oligotyping approach to examine these populations in lagoon waters of Mediterranean Marchica and Atlantic Oualidia, in Morocco. Patterns that have been observed at the clade and subclade levels, such as *Synechococcus*, relative abundance and the co-occurrence of groups from different clades, were shown to occur among oligotypes. The Marchica Lagoon showed a heterogeneous *Synechococcus* diversity compared to Oualidia in summer 2014. Thirty-one unique *Synechococcus* oligotypes were identified. Two distinct communities emerged in the 2014 and 2015 summer samplings, abundant and rare *Synechococcus* species, each comprising co-occurring *Synechococcus* oligotypes from different clades. Network analysis showed that six oligotypes were unique to Marchica Lagoon. The identified clades I, III, IV, VII, and 5.3 in Marchica were in accordance with its environmental characteristics. In addition, the relative abundance of some co-occurring *Synechococcus* strains was not constant over time and space (e.g., clades I and IV). Using gene oligotyping, we illustrated some of the challenges associated with the identification of novel *Synechococcus* strains or studied their co-occurrence in space and time. Oligotyping has been instrumental in discriminating closely related *Synechococcus* strains. However, this study leaves open questions about how samples differ by location and whether locations differ from year to year. Do co-occurring oligotypes interact with each other and to

what extent do they correlate with physicochemical parameters? What triggers the coexistence of clades I and IV with clade III in warm water or 5.3 with VII, which do not know much about. Finally, how do relative abundances change over seasons? Hence, future work needs to consider additional stations and seasons to provide better statistical support for our findings and to better understand their correlation with physical and chemical environmental parameters. Other factors were not considered in this study, such as nutrient availability, chlorophyll, irradiance, viral lysis, and greater sequencing depth, which could also influence the observed seasonal dynamics.

These studies all together confirmed the potential of metagenomics as a tool for microbial diversity characterization, uncultured bacteria exploration, and anthropogenic pressures delineation. It would be of great interest to investigate on a larger scale using whole shotgun sequencing the unknown groups of microorganisms, and metabolic pathways of typical genera and strains isolated from studied ecosystems for a deeper understanding of their interaction with the environment, the marine ecosystem, and impact on climate changes. As a result, we will be able to use these microbes for future remediation and restoration projects of lagoon and extreme environment waters. As a final note, generating a microbiome inventory will facilitate the building of a Moroccan lagoon microbiomes database reference, which will enable the exploration of specific microbial taxa as bioindicators of pollution, and enhancing monitoring strategies and bioremediation plans of similar aquatic ecosystems.

References

Cañedo-Argüelles, M.; Rieradevall, M.; Farrés-Corell, R.; Newton, A. Annual characterisation of four Mediterranean coastal lagoons subjected to intense human activity. Estuar. Coast. Shelf Sci. 2012, 114, 59–69. https://doi.org/1016/j.ecss.2011.07.017;

Ghai, R.; Hernandez, C.M.; Picazo, A.; Mizuno, C.M.; Ininbergs, K.; Díez, B.; Valas, R.; DuPont, C.L.; McMahon, K.D.; Camacho, A.; et al. Metagenomes of Mediterranean Coastal Lagoons. Sci. Rep. 2012, 2, 490. https://doi.org/10.1038/srep00490;

Soria, J; Pérez, R; Sòria-Pepinyà, X. Mediterranean Coastal Lagoons Review: Sites to Visit before Disappearance. Journal of Marine Science and Engineering. 2022; 10(3):347. https://doi.org/10.3390/jmse10030347;

Trombetta, T.; Bouget, F-Y.; Félix, C.; Mostajir, B.; and Vidussi, F. Microbial Diversity in a North Western Mediterranean Sea Shallow Coastal Lagoon Under Contrasting Water Temperature Conditions. Front. Mar. 2022; Sci.9:858744. doi: 10.3389/fmars.2022.858744;

Trombetta, T.; Vidussi, F.; Roques, C.; Mas, S.; Scotti, M.; Mostajir, B. Co-Occurrence Networks Reveal the Central Role of Temperature in Structuring the Plankton Community of the Thau Lagoon. Sci. Rep. 2021, 11, 17675. doi: 10.1038/s41598-021-97173-y;

Trombetta T., Vidussi F., Mas S., Parin D., Simier M., Mostajir B. Water Temperature Drives Phytoplankton Blooms in Coastal Waters. PloS One, 2019; 14, e0214933. doi: 10.1371/journal.pone.0214933;

Buchan, A.; LeCleir, G. R.; Gulvik, C. A.; González, J. M. Master Recyclers: Features and Functions of Bacteria Associated With Phytoplankton Blooms. Nat. Rev. Microbiol. 2014; 12, 686–698. doi: 10.1038/nrmicro3326;

Dolan J. R. Microbial Ecology of the Oceans. J. Plankton Res. 2018; 40, 500–502. doi: 10.1093/plankt/fby022;

Yunfeng, S.; Shuai, W.; Hui, W.; Zhaoyang, L.; Jiali, C.; Qiuying H.; and Muqiu, Z. "Diversity and succession of microbial communities on typical microplastics in Xincun Bay, a long-term mariculture tropical lagoon" Oceanological and Hydrobiological Studies, vol.51, no.1, 2022, pp.10-22. https://doi.org/10.26881/oahs.2022.1.02;

Needham D, M.; Fuhrman, J. A. Pronounced Daily Succession of Phytoplankton, Archaea and Bacteria Following a Spring Bloom. Nat. Microbiol. 2016; 1, 16005. doi: 10.1038/nmicrobiol.2016.5;

Chaouni, B.; Idrissi Azami, A.; Essayeh, S.; Arrafiqui, E.H.; Bailal, A.; Raoui, S.; Amzazi, S.; Twaddle, A.; El Hamouti, C.; Boukhatem, N.; Timinouni, M.; El Otmani, F.; Chahboune, R.; Barrijal, S.; El Homani, A.; Nejjari, C.; Zaid, E.H.; Hamamouch, N.; Bakkali, F.; Amaral-Zettler, L.; Ghazal, H. Moroccan Lagoon Microbiomes. Water 2022, 14, 1715. https://doi.org/10.3390/w14111715;

Walter, JM.; de Oliveira, LS.; Tschoeke, DA.; Meirelles, PM.; Neves, MHCB.; Batista, D.; Carvalho, AP.; Dos SantosCosta, R.; Dobretsov, S.; Coutinho, R.; Swings, J.; Thompson, CC.; and Thompson, FL. Metagenomic Insights Into Ecosystem Function in the Microbial Mats of a Large Hypersaline Coastal Lagoon System. Front. Mar. 2021; Sci. 8:715335. doi: 10.3389/fmars.2021.715335;

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

Águila, B.; Yanez-Montalvo, A.; Mercado-Juárez, RA.; Montejano, GA.; Becerra-Absalón, I.; Falcón, LI. Microbialites show a distinct cyanobacterial phylogenetic structure and functional redundancy in Bacalar lagoon and Cenote Azul sinkhole, Yucatan Peninsula, Mexico. FEMS Microbiol Ecol. 2022 May 14;98(5):fiac039. doi: 10.1093/femsec/fiac039;

Venter, J. C. et al. Environmental genome shotgun sequencing of the Sargasso Sea. Science. 2004; 304, 66–74.doi: 10.1126/science.1093857

Rusch, D. B. et al. The Sorcerer II Global Ocean Sampling expedition: northwest Atlantic through eastern tropical Pacific. PLoS Biol. 2007; 5, e77–431.https://doi.org/10.1371/journal.pbio.0050077;

Pesant, S. et al. Open science resources for the discovery and analysis of Tara Oceans data. Scientific Data. 2015; 2, 150023–16.https://doi.org/10.1038/sdata.2015.23;

Mackey, KRM.; Hunter-Cevera, K.; Britten, GL.; Murphy, LG.; Sogin, ML.; and Huber, JA. Seasonal Succession and Spatial Patterns of Synechococcus Microdiversity in a Salt Marsh

Estuary Revealed through 16S rRNA Gene Oligotyping. Front. Microbiol. 2017; 8:1496. doi: 10.3389/fmicb.2017.01496.

Falkowski, P. G. The ocean's invisible forest. Sci Am. 2022; 287, 54–61. doi: 10.1038/scientificamerican0802-54;

Eggleston, EM.; and Hewson, I. Abundance of Two Pelagibacter ubique Bacteriophage Genotypes along a Latitudinal Transect in the North and South Atlantic Oceans. Front. Microbiol. 2016; 7:1534. doi: 10.3389/fmicb.2016.01534;

Li, Y.; Sun, LL.; Sun, ML.; et al. Vertical and horizontal biogeographic patterns and major factors affecting bacterial communities in the open South China Sea. Sci Rep. 2018; 8, 8800. https://doi.org/10.1038/s41598-018-27191-w;

Tan, J.; Zuniga, C.; Zengler, K. Unraveling interactions in microbial communities - from co-cultures to microbiomes. J Microbiol. 2015 May;53(5):295-305. doi: 10.1007/s12275-015-5060-1;

Shade, A.; Handelsman, J. Beyond the Venn diagram: the hunt for a core microbiome. Environ Microbiol. 2012 Jan;14(1):4-12. doi: 10.1111/j.1462-2920.2011.02585.x;

Kopf, A.; Schnetzer, J.; and Oliver Glöckner, F. Understanding Marine Microbes, the Driving Engines of the Ocean. Front. Young Minds. 2016; 4:1. doi: 10.3389/frym.2016.00001;

Doney, S. C.; Ruckelshaus, M.; Duffy, J. E.; Barry, J. P.; Chan, F.; English, C. A.; et al. Climate change impacts on marine ecosystems. Mar. Sci. 2012; 4, 11–37. doi: 10.1146/annurev-marine-041911-111611.

Ruiz, F.; Abad, M.; Olías, M.; Galán, E.; González, I.; Aguilá, E.; Hamoumi, N.; Pulido, I.; Cantano, M. The Present Environmental Scenario of the Nador Lagoon (Morocco). Environ. Res. 2006, 102, 215–229. https://doi.org/10.1016/j.envres.2006.03.001.

El Madani, F.; Chiaar, A.; Chafi, A. Phytoplankton Composition and Abundance Assessment in the Nador Lagoon (Mediterranean Coast of Morocco). Acta Bot. Croat. 2011, 70, 269–288. https://doi.org/10.2478/v10184-010-0016-3.

Natij, L.; Damsiri, Z.; Khalil, K.; Loudiki, M.; Ettahiri, O.; Elkalay, K. Phytoplankton Abundance and Diversity in the Coastal Waters of Oualidia Lagoon, South Moroccan Atlantic in Relation to Environmental Variables. Int. J. Adv. Res. 2014, 2, 1022–1032.

Charette, MA & Smith, WHF. 'The Volume of the Earth's Ocean', Oceanography, 2010, vol. 23, no. 2, pp. 112-114.

Das, S, Lyla, PS & Ajmal Khan, S, 'Marine microbial diversity and ecology: importance and future perspectives', Current Science, 2006, vol. 90, no. 10, pp. 1325-1335.

Gray, JS. 'Marine biodiversity: patterns, threats and conservation needs', Biodiversity and Conservation, 1997, vol. 6, pp. 153-175.

Falkowski, PG. 'The ocean's invisible forest', Scientific American, 2002, vol. 287, pp. 54-61.

Karl, DM. 'Nutrient dynamics in the deep blue sea', TRENDS in Microbiology, 2002, vol. 10, no. 9, pp. 410-418.

Glöckner F.O., Stal L.J., Sandaa R.-A., Gasol J.M., O'Gara F., Hernandez F., Labrenz M., Stoica E., Varela M.M., Bordalo A., Pitta P. Marine Microbial Diversity and its role in Ecosystem Functioning and Environmental Change. Marine Board Position Paper 17. Calewaert, J.B. and McDonough N. (Eds.). 2012, Marine Board-ESF, Ostend, Belgium.

Azam, F & Malfatti, F. 'Microbial structuring of marine ecosystems', Nature Reviews. Microbiology, 2007, vol. 5, pp. 782-791.

Karner, MB, DeLong, EF & Karl, DM. 'Archaeal dominance in the mesopelagic zone of the Pacific Ocean', Nature, 2001, vol. 409, pp. 507-510.

Viklund, J, Ettema, TJ & Andersson, SG. 'Independent genome reduction and phylogenetic reclassification of the oceanic SAR11 clade', Molecular biology and evolution, 2011, p. msr203.

Malmstrom, RR, Cottrell, MT, Elifantz, H & Kirchman, DL. 'Biomass Production and Assimilation of Dissolved Organic Matter by SAR11 Bacteria in the Northwest Atlantic Ocean', Applied Environmental Microbiology, 2005, vol. 71, no. 6, pp. 2979-2986.

Zubkov, MV. 'Photoheterotrophy in marine prokaryotes', Journal of Plankton Research, 2009, vol. 31, no. 9, pp. 933-938.

Signori, CN, Thomas, F, Prast, AE, Pollery, RCG & Sievert SM, ‘Microbial diversity and community structure across environmental gradients in Bransfield Strait, Western Antartic Peninsula’, Frontiers in Microbiology, 2014, vol. 5, pp. 647.

Luria, CM, Ducklow, HW & Amaral-Zettler, LA. ‘Marine bacterial, archaeal and eukaryotic diversity and community structure on the continental shelf of the western Antarctic Peninsula’, Aquatic Microbial Ecology, 2014,vol. 73, pp. 107-121.

Brown, MV, Philip, GK, Bunge, JA, Smith, MC, Bissett, A, Lauro, FM, Fuhrman, JA & Donachie, SP. 'Microbial community structure in the North Pacific ocean', ISME J, 2009,vol. 3, no. 12, Dec, pp. 1374-86.

DeLong, EF & Karl, DM. ‘Genomic perspectives in microbial oceanography’, Nature, 2005,vol. 437, pp. 336-342.

Moran, MA, Buchan, A, González, JM, Heidelberg, JF, Whitman, WB, Kiene, RP, Henriksen, JR, King, GM, Belas, R & Fuqua, C. 'Genome sequence of Silicibacter pomeroyi reveals adaptations to the marine environment', Nature, 2004, vol. 432, no. 7019, pp. 910-913.

DeLong, EF & Karl, DM. ‘Genomic perspectives in microbial oceanography’, Nature, 2005, vol. 437, pp. 336-342.

Strom, SL. ‘Microbial Ecology of Ocean Biogeochemistry: A Community Perspective’, Science, 2008, vol. 320, pp. 1043-1045.

Reisch, CR, Moran, MA & Whitman, WB. ‘Bacterial catabolism of dimethylsulfoniopropionate (DMSP)’, Frontiers in Microbiology, 2011,vol. 2, pp. 172.

Vallina, SM, & Simo, R. ‘Strong Relationship Between DMS and the Solar Radiation Dose over the Global Surface Ocean’, Science, 2007, vol. 315, pp. 506-508.

Welsh, DT. ‘Ecological significance of compatible solute accumulation by microorganisms: from single cells to global climate’, FEMS Microbiology Reviews, 2000, vol. 24, pp. 263-290.

Vila-Costa, M, Rinta-Kanto, JM, Sun, S, Sharma, S, Poretsky, R & Moran, MA. ‘Transcriptomic analysis of a marine bacterial community enriched with dimethylsulfoniopropionate’, The ISME Journal, 2010, vol. 4, pp. 1410-1420.

González, JM & Moran, MA. 'Numerical dominance of a group of marine bacteria in the alpha-subclass of the class Proteobacteria in coastal seawater', Applied and Environmental Microbiology, 1997, vol. 63, no. 11, pp. 4237-4242.

Fuhrman, JA, McCallum, K & Davis, AA. 'Phylogenetic Diversity of Subsurface and Marine Microbial Communities from the Atlantic and Pacific Oceans', Applied and Environmental Microbiology, 1993,vol. 59, no. 5, pp. 1294-1302.

Mazard S, Ostrowski M, Partensky F, Scanlan DJ. Multilocus sequence analysis, taxonomic resolution and biogeography of marine Synechococcus. Environ Microbiol. 2012, 14(2):372-386. doi:10.1111/j.1462-2920.2011.02514.x.

Scanlan DJ, and West NJ. Molecular ecology of the marine cyanobacterial genera Prochlorococcus and Synechococcus. FEMS Microbiol. Ecol. 2002, 40, 1–12.

Scanlan, DJ, Ostrowski M, Mazard S, Dufresne A, Garczarek L, Hess WR, et al. Ecological genomics of marine picocyanobacteria. Microbiol Mol Biol Rev, 2009, 73: 249–299. doi: 10.1128/MMBR.00035-08.

Ahlgren NA, and Rocap G.Diversity and distribution of marine Synechococcus: multiple gene phylogenies for consensus classification and development of qPCR assays for sensitive measurement of clades in the ocean. Front. Microbiol. 2012, 3:213. doi: 10.3389/fmicb.2012.00213.

Eren AM, Maignien L, Sul WJ, Murphy LG, Grim SL, Morrison HG, et al. Oligotyping: differentiating between closely related microbial taxa using 16S rRNA gene data. Methods Ecol. Evol. 2013, 4, 1111–1119. doi: 10.1111/2041-210X.12114.

Xia X, Vidyarathna NK, Palenik B, Lee P, and Liu H. Comparison of the seasonal variations of Synechococcus assemblage structures in estuarine waters and coastal waters of Hong Kong. Appl. Environ. Microbiol. 2015, 81, 7644–7655. doi: 10.1128/AEM.01895-15.

Hunter-Cevera KR, Post AF, Peacock EE, and Sosik HM. Diversity of Synechococcus at the Martha's Vineyard Coastal observatory: insights from culture isolations, clone libraries, and flow cytometry. Microb. Ecol. 2016, 71, 276–289. doi: 10.1007/s00248-015-0644-1.

Mackey KRM, Hunter-Cevera K, Britten GL, Murphy LG, Sogin ML and Huber JA. Seasonal Succession and Spatial Patterns of Synechococcus Microdiversity in a Salt Marsh Estuary

Revealed through 16S rRNA Gene Oligotyping. Front. Microbiol. 2017, 8:1496. doi: 10.3389/fmicb.2017.01496.

Nagarkar M, Wang M, Valencia B, & Palenik B. Spatial and temporal variations in Synechococcus microdiversity in the Southern California coastal ecosystem. Environmental Microbiology, 2021. 23 (1). https://doi.org/10.1111/1462-2920.15307.

Robidart JC, Preston CM, Paerl RW, Turk KA, Mosier AC, Francis CA, et al. Seasonal Synechococcus and Thaumarchaeal population dynamics examined with high resolution with remote in situ instrumentation. ISME J. 2012.6, 513–523. 10.1038/ismej.2011.127.

Rajaneesh K, Mitbavkar S. Factors controlling the temporal and spatial variations in Synechococcus abundance in a monsoonal estuary. Mar. Environ. Res. 2013, 92 133–143. 10.1016/j.marenvres.2013.09.010.

Flombaum P, Gallegos JL, Gordillo RA, Rincon J, Zabala LL, Jiao N, et al. Present and future global distributions of the marine cyanobacteria Prochlorococcus and Synechococcus. Proc. Natl. Acad. Sci. 2013, U.S.A. 110, 9824–9829. 10.1073/pnas.1307701110.

Paulsen ML, DoréH, GarczarekL, SeutheL, MüllerO, Sandaa RA, et al. Synechococcus in the Atlantic Gateway to the Arctic Ocean. Front. Mar. Sci. 2016, 3:191. doi: 10.3389/fmars.2016.00191.

Pedrós-Alió C. Marine microbial diversity: can it be determined? Trends Microbiol. 2006. 14, 257–263. doi: 10.1016/j.tim.2006.04.007.

Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM., Neal PR, et al.Microbial diversity in the deep sea and theunderexplored “rare biosphere.” Proc. Natl. Acad. Sci. 2006, U.S.A. 103, 12115–12120. doi: 10.1073/pnas.0605127103.

Fuhrman JA. Microbial community structure and its functional implications. Nature, 2009, 459, 193–199. doi: 10.1038/nature 08058.

González-Regalado M.L. Ruiz F. Abad M. Hamoumi N. Boumaggard E.H. Bouamterhane I. Guddari F. Toumi A. Dassy K. Ben Ahmed R. (2005). Distribución de los foraminíferos bentónicos en climas semiáridos: las lagunas de Nador (Marruecos) y El Meleh (Túnez): Geogaceta, 37, 211-214.

Maanan M. Saddik M., Maanan M., Chaibi M., Assobhei O., Zourarah B. 2014. Environmental and ecological risk assessment of heavy metals in sediments of Nador lagoon, Morocco. Ecological Indicators; 48.

Elmadani F. Chiaar A., Chafi A. 2011. Phytoplankton composition and abundance assessment in the Nador lagoon (Mediterranean coast of Morocco). Acta Bot. Croat. 70 (2), 269–288.

Carruesco, C. Génèse et évolution à l'holocène de trois lagunes de la façade Atlantique: Moulay Bousselham, Oualidia (Maroc) et Arcachon (France), thèse d'état, université de Bodeaux I, Bordeaux, France, 1989.

Hilmi, K. et al. Oualidia lagoon, Morocco: An estuary without a river. African Journal of Aquatic Science, 2005 30(1):1-10.

Bidet, J. C. Carruessco, C. Etude sédimentologique de la lagune de Oualidia (Maroc). Oceanologica Acta, 1982, 29-37.

Beaubrun PC. Les huîtres au Maroc et l'ostréiculture dans la lagune de Oualidia. Bulletin Institut des Pêches Maritimes 22:13–143. 41 Carruesco, C.Génèse et évolution à l'holocène de trois lagunes de la façade Atlantique: Moulay Bousselham, Oualidia (Maroc) et Arcachon (France), thèse d'état, université de Bodeaux I, Bordeaux, France, 1989.

Bennouna A. et al. Study of phytoplankton populations of Oualidia Lagoon (Morocco); potentially harmful dinoflagellates. Mar. Life 2000 – Vol. 10 (1-2): 3-18.

Petrosino, JF, Highlander, S, Luna, RA, Gibbs, RA & Versalovic. 'Metagenomic pyrosequencing and microbial identification', Clin Chem, 2009,vol. 55, no. 5, May, pp. 856-66.

Jorgensen, BB. 'Bacteria and Marine Biogeochemistry', In: Schulz, HD. and Zabel, M., ed. Marine Geochemistry. Springer. Ch. 5, 2006.

Kennedy, J, Flemer, B, Jackson, SA, Lejon, DP, Morrissey, JP, O'Gara, F & Dobson, AD. 'Marine metagenomics: new tools for the study and exploitation of marine microbial metabolism', Mar Drugs, 2010,vol. 8, no. 3, pp. 608-28.

Caporaso, JG, Lauber, CL, Walters, WA, Berg-Lyons, D, Lozupone, CA, Turnbaugh, PJ, Fierer, N & Knight, R. 'Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample', Proc Natl Acad Sci U S A, 2011,vol. 108 Suppl 1, Mar 15, pp. 4516-22.

Tringe, SG, Von Mering, C, Kobayashi, A, Salamov, AA, Chen, K, Chang, HW, Podar, M, Short, JM, Mathur, EJ & Detter, JC. 'Comparative metagenomics of microbial communities', Science, 2005,vol. 308, no. 5721, pp. 554-557. Tripp, HJ 2007, Genomic-assisted determination of the natural nutrient requirements of the cosmopolitan marine bacterium'Candidatus Pelagibacter ubique', thesis.

Morozova, O & Marra, MA. 'Applications of next-generation sequencing technologies in functional genomics', Genomics, 2008,vol. 92, no. 5, Nov, pp. 255-64.

Reis-Filho, JS. 'Next-generation sequencing', Breast Cancer Res, 2009,vol. 11 Suppl 3, p. S12.

Liu, C, Kroeze, C, Hoekstra, AY & Gerbens-Leenes, W. 'Past and future trends in grey water footprints of anthropogenic nitrogen and phosphorus inputs to major world rivers', Ecological Indicators, 2012,vol. 18, pp. 42-49.

Bentley et al. 'Accurate whole human genome sequencing using reversible terminator chemistry', Nature, 2008,vol. 456, no. 7218, Nov 6, pp. 53-9.Brown,

Quail, MA, Smith, M, Coupland, P, Otto, TD, Harris, SR, Connor, TR, Bertoni, A, Swerdlow, HP & Gu, Y. 'A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers', BMC Genomics, 2012,vol. 13, p. 341.

Caporaso, J. G., J. Kuczynski, J. Stombaugh, K. Bittinger, F. D. Bushman, E. K. Costello, N. Fierer, A. G. Pena, J. K. Goodrich, J. I. et al. QIIME allows analysis of high-throughput community sequencing data, Nat Methods, 2010, 7: 335-6.

Edgar, R. C. 'Search and clustering orders of magnitude faster than BLAST', Bioinformatics, 26: 2460-1. 2013. 'UPARSE: highly accurate OTU sequences from microbial amplicon reads', Nat Methods, 10: 996-8.

Uchiyama, T., and K. Miyazaki.'Functional metagenomics for enzyme discovery: challenges to efficient screening', Curr Opin Biotechnol, 2009, 20: 616-22.

Langille, M. G., J. Zaneveld, J. G. Caporaso, D. McDonald, D. Knights, J. A. Reyes, J. C. Clemente, D. E. Burkepile, R. L. Vega Thurber, R. Knight, R. G. Beiko, and C. Huttenhower. 'Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences', Nat Biotechnol, 2013, 31: 814-21.

Huson, D. H., A. F. Auch, J. Qi, and S. C. Schuster. 'MEGAN analysis of metagenomic data', Genome Res, 2007, 17: 377-86.

Meyer, F., D. Paarmann, M. D'Souza, R. Olson, E. M. Glass, M. Kubal, T. Paczian, A. Rodriguez, R. Stevens, A. Wilke, J. Wilkening, and R. A. Edwards. 'The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes', BMC Bioinformatics, 2008, 9: 386.

Abubucker, S., N. Segata, J. Goll, A. M. Schubert, J. Izard, B. L. Cantarel, B. Rodriguez-Mueller, J. Zucker, M. Thiagarajan, B. Henrissat, O. White, S. T. Kelley, B. Methe, P. D. Schloss, D. Gevers, M. Mitreva, and C. Huttenhower.'Metabolic reconstruction for metagenomic data and its application to the human microbiome', PLoS Comput Biol, 2012, 8: e1002358.

Kanehisa, M., S. Goto, S. Kawashima, Y. Okuno, and M. Hattori. 'The KEGG resource for deciphering the genome', Nucleic Acids Res, 2004, 32: D277-80.

Punta, M., P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. L. Sonnhammer, S. R. Eddy, A. Bateman, and R. D. Finn. 'The Pfam protein families database', Nucleic Acids Res, 2012, 40: D290-301.

Selengut, J. D., D. H. Haft, T. Davidsen, A. Ganapathy, M. Gwinn-Giglio, W. C. Nelson, A. R. Richter, and O. White. 'TIGRFAMs and Genome Properties: tools for the assignment of molecular function and biological process in prokaryotic genomes', Nucleic Acids Res, 2007, 35: D260-4.

Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov, A. N. Nikolskaya, B. S. Rao, S. Smirnov, A. V. Sverdlov, S. Vasudevan, Y. I. Wolf, J. J. Yin, and D. A. Natale. 'The COG database: an updated version includes eukaryotes', BMC Bioinformatics, 2003, 4: 41.

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. 2016, 18, 1403–1414. https://doi.org/10.1111/1462-2920.13023.

Huse, S.M.; Mark Welch, D.B.; Voorhis, A.; Shipunova, A.; Morrison, H.G.; Eren, A.M.; Sogin, M.L. VAMPS: A Website for Visualization and Analysis of Microbial Population Structures. BMC Bioinform. 2014, 15, 41. https://doi.org/10.1186/1471-2105-15-41.

Huse, S.M.; Dethlefsen, L.; Huber, J.A.; Welch, D.M.; Relman, D.A.; Sogin, M.L. Exploring Microbial Diversity and Taxonomy Using SSU RRNA Hypervariable Tag Sequencing. PLoS Genet. 2008, 4, e1000255. https://doi.org/10.1371/journal.pgen.1000255.

Quast, C.; Pruesse, E.; Yilmaz, P.; Gerken, J.; Schweer, T.; Yarza, P.; Peplies, J.; Glöckner, F.O. The SILVA Ribosomal RNA Gene Database Project: Improved Data Processing and Web-Based Tools. Nucleic Acids Res. 2013, 41, D590–D596. https://doi.org/10.1093/nar/gks1219.

Callahan, B.J.; McMurdie, P.J.; Rosen, M.J.; Han, A.W.; Johnson, A.J.A.; Holmes, S.P. DADA2: High Resolution Sample Inference from Illumina Amplicon Data. Nat. Methods, 2016, 13, 581–583. https://doi.org/10.1038/nmeth.3869.

Nurk, S.; Meleshko, D.; Korobeynikov, A.; Pevzner, P.A. MetaSPAdes: A New Versatile Metagenomic Assembler. Genome Res. 2017, 27, 824–834. https://doi.org/10.1101/gr.213959.116.

Mikheenko, A.; Saveliev, V.; Gurevich, A. MetaQUAST: Evaluation of Metagenome Assemblies. Bioinformatics 2016, 32, 1088–1090. https://doi.org/10.1093/bioinformatics/btv697.

Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. BMC Bioinform. 2010, 11, 119. https://doi.org/10.1186/1471-2105-11-119.

Truong, D.T.; Franzosa, E.A.; Tickle, T.L.; Scholz, M.; Weingart, G.; Pasolli, E.; Tett, A.; Huttenhower, C.; Segata, N. MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling. Nat. Methods 2015, 12, 902–903. https://doi.org/10.1038/nmeth.3589.

Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. Bioinformatics 2014, 30, 2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

Brum, J.R.; Ignacio-Espinoza, J.C.; Roux, S.; Doulcier, G.; Acinas, S.G.; Alberti, A.; Chaffron, S.; Cruaud, C.; de Vargas, C.; Gasol, J.M.; et al. Patterns and Ecological Drivers of Ocean Viral Communities. Science 2015, 348, 1261498. https://doi.org/10.1126/science.1261498.

Langmead, B.; Salzberg, S.L. Fast Gapped-Read Alignment with Bowtie 2. Nat. Methods 2012, 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Aknaf, A.; Akodad, M.; Layachi, M.; El Madani, F.; Jaddar, A.; Mesfioui, A.; Baghour, M. Study of the Spatial and Temporal Variation of Physical-Chemical Parameters Characterizing the Quality of Surface Waters of the Lagoon Marchica–North-East Morocco. J. Mater. Environ. Sci. 2017, 8, 3216–3225.

Padan, E.; Bibi, E.; Ito, M.; Krulwich, T.A. Alkaline PH Homeostasis in Bacteria: New Insights. Biochim. Biophys. Acta 2005, 1717, 67–88. https://doi.org/10.1016/j.bbamem.2005.09.010.

Kort, A.R.; Magurran, A.E. Increases in Local Richness (α-Diversity) Following Invasion Are Offset by Biotic Homogenization in a Biodiversity Hotspot. Biol. Lett. 2019, 15, 20190133. https://doi.org/10.1098/rsbl.2019.0133.

Magurran, A.E. Measuring Biological Diversity; John Wiley & Sons: Hoboken, NJ, USA, 2003.

Simpson, E.H. Measurement of Diversity. Nature 1949, 163, 688–688. https://doi.org/10.1038/163688a0.

Kim, B.-R.; Shin, J.; Guevarra, R.; Lee, J.H.; Kim, D.W.; Seol, K.-H.; Lee, J.-H.; Kim, H.B.; Isaacson, R. Deciphering Diversity Indices for a Better Understanding of Microbial Communities. J. Microbiol. Biotechnol. 2017, 27, 2089–2093. https://doi.org/10.4014/jmb.1709.09027.

Yilmaz, P.; Yarza, P.; Rapp, J.Z.; Glöckner, F.O. Expanding the World of Marine Bacterial and Archaeal Clades. Front. Microbiol. 2016, 6, 1524.

Gupta, R.S. The Phylogeny of Proteobacteria: Relationships to Other Eubacterial Phyla and Eukaryotes. FEMS Microbiol. Rev. 2000, 24, 367–402. https://doi.org/10.1111/j.1574-6976.2000.tb00547.x.

Dang, H.; Lovell, C.R. Bacterial Primary Colonization and Early Succession on Surfaces in Marine Waters as Determined by Amplified RRNA Gene Restriction Analysis and Sequence Analysis of 16S RRNA Genes. Appl. Environ. Microbiol. 2000, 66, 467–475. https://doi.org/10.1128/AEM.66.2.467-475.2000.

Yu, Z.; Yang, J.; Liu, L.; Zhang, W.; Amalfitano, S. Bacterioplankton Community Shifts Associated with Epipelagic and Mesopelagic Waters in the Southern Ocean. Sci. Rep. 2015, 5, 12897. https://doi.org/10.1038/srep12897.

Gilbert, J.A.; Steele, J.A.; Caporaso, J.G.; Steinbrück, L.; Reeder, J.; Temperton, B.; Huse, S.; McHardy, A.C.; Knight, R.; Joint, I.; et al. Defining Seasonal Marine Microbial Community Dynamics. ISME J. 2012, 6, 298–308. https://doi.org/10.1038/ismej.2011.107.

Hartsock, A.; Shapleigh, J.P. Physiological Roles for Two Periplasmic Nitrate Reductases in Rhodobacter Sphaeroides 2.4.3 (ATCC 17025). J. Bacteriol. 2011, 193, 6483–6489. https://doi.org/10.1128/JB.05324-11.

Vincent, P.; Pignet, P.; Talmont, F.; Bozzi, L.; Fournet, B.; Guezennec, J.; Jeanthon, C.; Prieur, D. Production and Characterization of an Exopolysaccharide Excreted by a Deep-Sea Hydrothermal Vent Bacterium Isolated from the Polychaete Annelid Alvinella Pompejana. Appl. Environ. Microbiol. 1994, 60, 4134–4141. https://doi.org/10.1128/aem.60.11.4134-4141.1994.

Morris, R.M.; Rappé, M.S.; Connon, S.A.; Vergin, K.L.; Siebold, W.A.; Carlson, C.A.; Giovannoni, S.J. SAR11 Clade Dominates Ocean Surface Bacterioplankton Communities. Nature 2002, 420, 806–810. https://doi.org/10.1038/nature01240.

Solano-Gallego, L.; Kidd, L.; Trotta, M.; Marco, M.D.; Caldin, M.; Furlanello, T.; Breitschwerdt, E.B. Febrile Illness Associated with Rickettsia Conorii Infection in Dogs from Sicily. Emerg. Infect. Dis. 2006, 12, 1985–1988. https://doi.org/10.3201/eid1212.060326.

Bullerjahn, G.S.; Post, A.F. Physiology and Molecular Biology of Aquatic Cyanobacteria. Front. Microbiol. 2014, 5, 359.

Cooke, G.D.; Welch, E.B.; Peterson, S.; Nichols, S.A. Restoration and Management of Lakes and Reservoirs, 3rd ed.; CRC Press: Boca Raton, FL, USA, 2005; ISBN 978-0-429-18923-4.

Fernández-Gómez, B.; Richter, M.; Schüler, M.; Pinhassi, J.; Acinas, S.G.; González, J.M.; Pedrós-Alió, C. Ecology of Marine Bacteroidetes: A Comparative Genomics Approach. ISME J. 2013, 7, 1026–1037. https://doi.org/10.1038/ismej.2012.169.

Gómez-Pereira, P.R.; Fuchs, B.M.; Alonso, C.; Oliver, M.J.; van Beusekom, J.E.E.; Amann, R. Distinct Flavobacterial Communities in Contrasting Water Masses of the North Atlantic Ocean. ISME J. 2010, 4, 472–487. https://doi.org/10.1038/ismej.2009.142.

Mann, A.J.; Hahnke, R.L.; Huang, S.; Werner, J.; Xing, P.; Barbeyron, T.; Huettel, B.; Stüber, K.; Reinhardt, R.; Harder, J.; et al. The Genome of the Alga-Associated Marine Flavobacterium Formosa Agariphila KMM 3901T Reveals a Broad Potential for Degradation of Algal

Polysaccharides. Appl. Environ. Microbiol. 2013, 79, 6813–6822. https://doi.org/10.1128/AEM.01937-13.

Cardman, Z.; Arnosti, C.; Durbin, A.; Ziervogel, K.; Cox, C.; Steen, A.D.; Teske, A. Verrucomicrobia Are Candidates for Polysaccharide-Degrading Bacterioplankton in an Arctic Fjord of Svalbard. Appl. Environ. Microbiol. 2014, 80, 3749–3756. https://doi.org/10.1128/AEM.00899-14.

Bünger, W.; Jiang, X.; Müller, J.; Hurek, T.; Reinhold-Hurek, B. Novel Cultivated Endophytic Verrucomicrobia Reveal Deep-Rooting Traits of Bacteria to Associate with Plants. Sci. Rep. 2020, 10, 8692. https://doi.org/10.1038/s41598-020-65277-6.

Jenkins, C.; Samudrala, R.; Anderson, I.; Hedlund, B.P.; Petroni, G.; Michailova, N.; Pinel, N.; Overbeek, R.; Rosati, G.; Staley, J.T. Genes for the Cytoskeletal Protein Tubulin in the Bacterial Genus Prosthecobacter. Proc. Natl. Acad. Sci. USA 2002, 99, 17049–17054. https://doi.org/10.1073/pnas.012516899.

Dunfield, P.F.; Yuryev, A.; Senin, P.; Smirnova, A.V.; Stott, M.B.; Hou, S.; Ly, B.; Saw, J.H.; Zhou, Z.; Ren, Y.; et al. Methane Oxidation by an Extremely Acidophilic Bacterium of the Phylum Verrucomicrobia. Nature 2007, 450, 879–882. https://doi.org/10.1038/nature06411.

Hou, S.; Makarova, K.S.; Saw, J.H.W.; Senin, P.; Ly, B.V.; Zhou, Z.; Ren, Y.; Wang, J.; Galperin, M.Y.; Omelchenko, M.V.; et al. Complete Genome Sequence of the Extremely Acidophilic Methanotroph Isolate V4, Methylacidiphilum Infernorum, a Representative of the Bacterial Phylum Verrucomicrobia. Biol. Direct 2008, 3, 26. https://doi.org/10.1186/1745-6150-3-26.

Ventura, M.; Canchaya, C.; Tauch, A.; Chandra, G.; Fitzgerald, G.F.; Chater, K.F.; van Sinderen, D. Genomics of Actinobacteria: Tracing the Evolutionary History of an Ancient Phylum. Microbiol. Mol. Biol. Rev. 2007, 71, 495–548. https://doi.org/10.1128/MMBR.00005-07.

Shivlata, L.; Satyanarayana, T. Thermophilic and Alkaliphilic Actinobacteria: Biology and Potential Applications. Front. Microbiol. 2015, 6, 1014. https://doi.org/10.3389/fmicb.2015.01014.

Pinto, I.; Calisto, R.; Serra, C.R.; Lage, O.M.; Antunes, S.C. Bacterioplankton Community as a Biological Element for Reservoirs Water Quality Assessment. Water 2021, 13, 2836. https://doi.org/10.3390/w13202836.

Pinheiro, R.O.; de Souza Salles, J.; Sarno, E.N.; Sampaio, E.P. Mycobacterium Leprae-Host-Cell Interactions and Genetic Determinants in Leprosy: An Overview. Future Microbiol. 2011, 6, 217–230. https://doi.org/10.2217/fmb.10.173.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997 Sep 1;25(17):3389-402. doi: 10.1093/nar/25.17.3389. PMID: 9254694; PMCID: PMC146917.

Hewson, I.; Paerl, R.W.; Tripp, H.J.; Zehr, J.P.; Karl, D.M. Metagenomic Potential of Microbial Assemblages in the Surface Waters of the Central Pacific Ocean Tracks Variability in Oceanic Habitat. Limnol. Oceanogr. 2009, 54, 1981–1994.

Schloerke, B.; Cook, D.; Larmarange, J.; Briatte, F.; Marbach, M.; Thoen, E.; Elberg, A.; Crowley, J. Extension to "ggplot2" [R Package GGally Version 2.0.0]. 2020.

Castellano-Hinojosa, A., Correa-Galeote, D., Carrillo, P., Bedmar, E.J., Medina-Sánchez, J.M.. Denitrification and biodiversity of denitrifiers in a highmountain Mediterranean lake. Frontiers in Microbiology, 2017, 8, 1911.

Edgar R.C. MUSCLE: Multiple sequence align-ment with high accuracy and high throughput. Nucleic Acids Research, 2004, 32, 1792-1797. doi:10.1093/nar/gkh340.

Thompson, L.R.; Sanders, J.G.; McDonald, D.; Amir, A.; Ladau, J.; Locey, K.J.; Prill, R.J.; Tripathi, A.; Gibbons, S.M.; Ackermann, G.; et al. A Communal Catalogue Reveals Earth's Multiscale Microbial Diversity. Nature 2017, 551, 457–463. https://doi.org/10.1038/nature24621.

Matoir, M.; Belabed, A.; Najih, M.; Kada, O.; Rezzoume, N.; Surrounding influence on the Ecological state of the lagoon of Marchica. J. Mater. Environ. Sci.. 2015, 1260–1265.

Riouchi, O.; Madani, F.E.; Abadie, E.; Skalli, A.; Baghour, M. The Spatio-Temporal Evolution of the Genus Nitzschia Longissima at the Level of the Lagoon in Nador, Morocco. E3S Web Conf. 2021, 234, 00081. https://doi.org/10.1051/e3sconf/202123400081.

Hassou, N.; Boujaber, N.; Oumaskour, K.; Lakhdar, F.; Benba, J.; Benhaddou, Z.; Assoobhei, O.; Etahiri, S. Nutrients Variation in the Coastal Waters of Oualidia Lagoon, Morocco. Relation to the Contribution of Watershed and Hydro-Chemical Characteristics. Int. J. Adv. Res. 2016, 4, 1697–1706.

Chagas, G.G.; Suzuki, M.S. Seasonal Hydrochemical Variation in a Tropical Coastal Lagoon (Açu Lagoon, Brazil). Braz. J. Biol. 2005, 65, 597–607. https://doi.org/10.1590/s1519-69842005000400006.

Damsiri, Z.; Natij, L.; Khalil, K.; Loudiki, M.; Rabouille, C.; Ettahiri, O.; Bougadir, B.; Elkalay, K. Spatio-Temporal Nutrients Variability in the Oualidia Lagoon (Atlantic Moroccan Coast). Int. J. Curr. Adv. Res. 2014, 2, 609–618.

Aknaf, A.; Akodad, M.; Moumen, A.; Chekroun, K.B.; Elhamouti, C.; Bailal, A.; Baghour, M. Impact of the New Pass on the Eutrophication of the Lagoon Marchica: Study of the Two Sites Bou Areg and Mohandis. J. Mater. Environ. Sci. 2015, 6, 2939–2943.

Giovannoni, S.J. SAR11 Bacteria: The Most Abundant Plankton in the Oceans. Annu. Rev. Mar. Sci. 2017, 9, 231–255. https://doi.org/10.1146/annurev-marine-010814-015934.

MacConnachie, K.; Tishkowski, K. Boutonneuse Fever. In StatPearls; StatPearls Publishing: Treasure Island, FL, USA, 2022.

Zinger, L.; Amaral-Zettler, L.A.; Fuhrman, J.A.; Horner-Devine, M.C.; Huse, S.M.; Welch, D.B.M.; Martiny, J.B.H.; Sogin, M.; Boetius, A.; Ramette, A. Global Patterns of Bacterial Beta-Diversity in Seafloor and Seawater Ecosystems. PLoS ONE 2011, 6, e24570. https://doi.org/10.1371/journal.pone.0024570.

Juhmani, A.-S.; Vezzi, A.; Wahsha, M.; Buosi, A.; Pascale, F.D.; Schiavon, R.; Sfriso, A. Diversity and Dynamics of Seaweed Associated Microbial Communities Inhabiting the Lagoon of Venice. Microorganisms 2020, 8, 1657. https://doi.org/10.3390/microorganisms8111657.

Seo, J.-H.; Kang, I.; Yang, S.-J.; Cho, J.-C. Characterization of Spatial Distribution of the Bacterial Community in the South Sea of Korea. PLoS ONE 2017, 12, e0174159. https://doi.org/10.1371/journal.pone.0174159.

Schneider, K.; van der Werf, W.; Cendoya, M.; Mourits, M.; Navas-Cortés, J.A.; Vicent, A.; Oude Lansink, A. Impact of Xylella Fastidiosa Subspecies Pauca in European Olives. Proc. Natl. Acad. Sci. USA 2020, 117, 9250–9259. https://doi.org/10.1073/pnas.1912206117.

Partensky, F.; Blanchot, J.; Vaulot, D. DifferentiaI Distribution and Ecology of Prochlorococcus and Synechococcus in Oceanic Waters: A Review. Bull. L'institut Océanographique Monaco 1999, 19, 457–475.

Kim, Y.; Jeon, J.; Kwak, M.S.; Kim, G.H.; Koh, I.; Rho, M. Photosynthetic Functions of Synechococcus in the Ocean Microbiomes of Diverse Salinity and Seasons. PLoS ONE 2018, 13, e0190266. https://doi.org/10.1371/journal.pone.0190266.

Lürling, M.; Mello, M.M.E.; van Oosterhout, F.; de Senerpont Domis, L.; Marinho, M.M. Response of Natural Cyanobacteria and Algae Assemblages to a Nutrient Pulse and Elevated Temperature. Front. Microbiol. 2018, 9, 1851. https://doi.org/10.3389/fmicb.2018.01851.

Brouwer, P.; Bräutigam, A.; Buijs, V.A.; Tazelaar, A.O.E.; van der Werf, A.; Schlüter, U.; Reichart, G.-J.; Bolger, A.; Usadel, B.; Weber, A.P.M.; et al. Metabolic Adaptation, a Specialized Leaf Organ Structure and Vascular Responses to Diurnal N2 Fixation by Nostoc Azollae Sustain the Astonishing Productivity of Azolla Ferns without Nitrogen Fertilizer. Front. Plant Sci. 2017, 8, 442.

Niestępski, S.; Harnisz, M.; Korzeniewska, E.; Osińska, A. Markers Specific to Bacteroides Fragilis Group Bacteria as Indicators of Anthropogenic Pollution of Surface Waters. Int. J. Environ. Res. Public Health 2020, 17, 7137. https://doi.org/10.3390/ijerph17197137.

Rusch, D.B.; Halpern, A.L.; Sutton, G.; Heidelberg, K.B.; Williamson, S.; Yooseph, S.; Wu, D.; Eisen, J.A.; Hoffman, J.M.; Remington, K.; et al. The Sorcerer II Global Ocean Sampling Expedition: Northwest Atlantic through Eastern Tropical Pacific. PLoS Biol. 2007, 5, e77. https://doi.org/10.1371/journal.pbio.0050077.

Biers, E.J.; Sun, S.; Howard, E.C. Prokaryotic Genomes and Diversity in Surface Ocean Waters: Interrogating the Global Ocean Sampling Metagenome. Appl. Environ. Microbiol. 2009, 75, 2221–2229. https://doi.org/10.1128/AEM.02118-08.

Signori, C.N.; Thomas, F.; Enrich-Prast, A.; Pollery, R.C.G.; Sievert, S.M. Microbial Diversity and Community Structure across Environmental Gradients in Bransfield Strait, Western Antarctic Peninsula. Front. Microbiol. 2014, 5, 647.

Luria, C.M.; Ducklow, H.W.; Amaral-Zettler, L.A. Marine Bacterial, Archaeal and Eukaryotic Diversity and Community Structure on the Continental Shelf of the Western Antarctic Peninsula. Aquat. Microb. Ecol. 2014, 73, 107–121. https://doi.org/10.3354/ame01703.

Brown, M.V.; Philip, G.K.; Bunge, J.A.; Smith, M.C.; Bissett, A.; Lauro, F.M.; Fuhrman, J.A.; Donachie, S.P. Microbial Community Structure in the North Pacific Ocean. ISME J. 2009, 3, 1374–1386. https://doi.org/10.1038/ismej.2009.86.

Gárate Lizárraga, I.; Sevilla Torres, G.; Álvarez Añorve, M.; Aguirre Bahena, F.; Violante González, J.; Rojas Herrera, A. FIRST RECORD OF A RED TIDE CAUSED BY Gyrodinium Instriatum (DINOPHYCEAE: GYMNODINIALES) IN BAHÍA DE ACAPULCO, GUERRERO. CICIMAR Oceánides 2013, 28, 43–47. https://doi.org/10.37543/oceanides.v28i1.120.

Teo, C.L.; Jamaluddin, H.; Zain, N.A.M.; Idris, A. Biodiesel Production via Lipase Catalysed Transesterification of Microalgae Lipids from Tetraselmis Sp. Renew. Energy 2014, 68, 1–5. https://doi.org/10.1016/j.renene.2014.01.027.

Chappell, P.; Whitney, L.; Haddock, T.; Menden-Deuer, S.; Roy, E.; Wells, M.; Jenkins, B. Thalassiosira Spp. Community Composition Shifts in Response to Chemical and Physical Forcing in the Northeast Pacific Ocean. Front. Microbiol. 2013, 4, 273.

Scofield, V.; Jacques, S.M.S.; Guimarães, J.R.D.; Farjalla, V.F. Potential Changes in Bacterial Metabolism Associated with Increased Water Temperature and Nutrient Inputs in Tropical Humic Lagoons. Front. Microbiol. 2015, 6, 310. https://doi.org/10.3389/fmicb.2015.00310.

Krause, E.; Wichels, A.; Giménez, L.; Lunau, M.; Schilhabel, M.B.; Gerdts, G. Small Changes in PH Have Direct Effects on Marine Bacterial Community Composition: A Microcosm Approach. PLoS ONE 2012, 7, e47035. https://doi.org/10.1371/journal.pone.0047035.

Singh, J.S.; Kumar, A.; Rai, A.N.; Singh, D.P. Cyanobacteria: A Precious Bio-Resource in Agriculture, Ecosystem, and Environmental Sustainability. Front. Microbiol. 2016, 7, 529. https://doi.org/10.3389/fmicb.2016.00529.

Bajpai, V.K.; Shukla, S.; Kang, S.-M.; Hwang, S.K.; Song, X.; Huh, Y.S.; Han, Y.-K. Developments of Cyanobacteria for Nano-Marine Drugs: Relevance of Nanoformulations in Cancer Therapies. Mar. Drugs 2018, 16, E179. https://doi.org/10.3390/md16060179.

Zadjelovic, V.; Chhun, A.; Quareshy, M.; Silvano, E.; Hernandez-Fernaud, J.R.; Aguilo-Ferretjans, M.M.; Bosch, R.; Dorador, C.; Gibson, M.I.; Christie-Oleza, J.A. Beyond Oil Degradation: Enzymatic Potential of Alcanivorax to Degrade Natural and Synthetic Polyesters. Environ. Microbiol. 2020, 22, 1356–1369. https://doi.org/10.1111/1462-2920.14947.

Antoniou, E.; Fodelianakis, S.; Korkakaki, E.; Kalogerakis, N. Biosurfactant Production from Marine Hydrocarbon-Degrading Consortia and Pure Bacterial Strains Using Crude Oil as Carbon Source. Front. Microbiol. 2015, 6, 274.

Zwirglmaier K, Jardillier L, Ostrowski M, Mazard S, Garczarek L, Vaulot D, Not F, Massana R, Ulloa O, and Scanlan DJ. Global phylogeography of marine Synechococcus and Prochlorococcus reveals a distinct partitioning of lineages among oceanic biomes. Environmental microbiology, 2008, 10(1), 147-161. https://doi.org/10.1111/j.1462-2920.2007.01440.x.

Post AF, Penno S, Zandbank K, Paytan A, Huse SM, Welch DM. Long term seasonal dynamics of Synechococcus population structure in the Gulf of Aqaba, Northern Red Sea. Front Microbiol, 2011, 2(131). doi: https://doi.org/10.3389/fmicb.2011.00131.

Chen F, Wang K, Kan J, Bachoon D, Lu J, Lau S, and Campbell L. Phylogenetic diversityof Synechococcus in the Chesapeake Bayrevealed byribulose-1,5 bisphosphate carboxylase-oxygenase (RuBisCO) large subunit gene (rbcL) sequences. Aquat. Microb.Ecol, 2004, 36, 153–164. doi: 10.3354/ame036153.

Waterbury JB, Watson SW, Valois FW, and Franks DG. Biological and ecological characterization of the marine unicellular cyanobacteria Synechococcus. Can. J.Fish. Aquat. Sci, 1986. 214,71–120.

Chen F, Wang K, Kan J, Suzuki M, and Wommack K. Diverse and uniquepicocyanobacterian ChesapeakeBay, revealedby16S-23S rRNAinternaltranscribedspacer sequences. Appl.Environ.Microbiol, 2006, 72, 2239–2243. doi: 10.1128/AEM.72.3.2239-2243.2006.

Cai H, Wang K, Huang S, Jiao N, and Chen F. Distinct patterns of picocyanobacterial communities in winter and summer in the Chesapeake Bay. Appl.Environ. Microbiol. 2010, 76, 2955–2960. doi: 10.1128/AEM.02868-09.

Huang S, Wilhelm SW, Harvey HR, Taylor K, Jiao N, and Chen F. Novel lineages of Prochlorococcus and Synechococcus in theglobaloceans. ISME J., 2012, 6, 285–297.

Choi D, and Noh J. Phylogenetic diversity of Synechococcus strainsisolatedfromtheEastChina Sea andtheEast Sea. FEMS Microbiol. Ecol., 2009, 69, 439–448. doi: 10.1111/j.1574-6941.2009.00729.x.

Zwirglmaier K, Heywood JL, Chamberlain K, Woodward EM, Zubkov MV, and Scanlan DJ. Basin-scale distribution patterns of picocyanobacterial lineages in the Atlantic Ocean. Environ.Microbiol. 2007, 9, 1278–1290.

Tai V, and Palenik B. Temporal variation of Synechococcus clades at a coastalPacific Ocean monitoring site. ISME J. 2009, 3, 903–915. doi: https://doi.org/10.1038/ismej.2009.35.

Mella-Flores D, Mazard S, Humily F, Partensky F, Mahe F, Bariat L, Courties C, Marie D, Ras J, Mauriac, R, Jeanthon C, Bendif EM, Ostrowski M, Scanlan DJ, and Garczarek L. Is the distribution of Prochlorococcus and Synechococcus ecotypes in the Mediterranean Sea affected by global warming. Biogeosciences, 2011, 8(9), 2785-2804. doi: https://doi.org/10.5194/bg-8-2785-2011.

Varkey D, Mazard S, Ostrowski M. et al. Effects of low temperature on tropical and temperate isolates of marine Synechococcus. ISME J. 2016, 10, 1252–1263. https://doi.org/10.1038/ismej.2015.179.

Ebrahimi S, Thi Hau Nguyen TH, Roberts DJ. Effect of Temperature & Salt Concentration on Salt Tolerant Nitrate-Perchlorate Reducing Bacteria: Nitrate Degradation Kinetics. Water Res, 2015, 83:345-53. doi: 10.1016/j.watres.2015.07.006.

Traving SJ, Clokie MR, Middelboe M. Increased acidification has a profound effect on the interactions between the cyanobacterium Synechococcus sp. WH7803 and its viruses. FEMS Microbiol Ecol. 2014, 87(1):133-141. doi:10.1111/1574-6941.12199.

Résumé

Les lagunes sont des écosystèmes marins fragiles affectés par les polluants anthropiques. Nous avons réalisé une caractérisation spatio–temporelle du microbiome de deux lagunes marocaines, Marchica et Oualidia, classées comme sites Ramsar, la première sur la côte méditerranéenne et la seconde sur la côte atlantique. Nous avons étudié leur diversité et leur abondance microbiennes à l'aide d'approches métagénomiques durant les étés 2014 et 2015. Le microbiome bactérien était composé principalement de *Proteobacteries* (25—53%, 29—29%), *Cyanobacteries* (34—12%, 11—0.53%), *Bacteroidetes* (24—16%, 23—43%), *Actinobacteria* (7—11%, 13—7%), and *Verrucomicrobia* (4—1%, 15—14%) à Marchica et Oualidia en 2014 et 2015, respectivement. Fait intéressant, 48 souches ont été nouvellement signalées dans les écosystèmes lagunaires, tandis que huit virus inconnus ont été détectés uniquement dans la Marchica méditerranéenne. L'analyse statistique a montré une diversité microbienne plus élevée dans la lagune atlantique que dans la lagune méditerranéene et une relation robuste entre la diversité alpha et les emplacements géographiques d'échantillonnage. Cette première étude métagénomique sur les écosystèmes aquatiques marocains a enrichi le catalogue national des microorganismes marins.

Mots–clefs (5) : Métagénomique, microbiome marin, lagunes marocaines, biodiversité, bioindicateurs.

Abstract

Lagoons are fragile marine ecosystems that are considerably affected by anthropogenic pollutants. We performed a spatiotemporal characterization of the microbiome of two Moroccan lagoons, Marchica and Oualidia, both classifified as Ramsar sites, the former on the Mediterranean coast and the latter on the Atlantic coast. We investigated their microbial diversity and abundance using 16S rRNA amplicon– and shotgun–based metagenomics approaches during the summers of 2014 and 2015. The bacterial microbiome was composed primarily of Proteobacteria (25—53%, 29—29%), Cyanobacteria (34—12%, 11—0.53%), Bacteroidetes (24—16%, 23—43%), Actinobacteria (7—11%, 13—7%), and Verrucomicrobia (4—1%, 15—14%) in Marchica and Oualidia in 2014 and 2015, respectively. Interestingly, 48 strains were newly reported in lagoon ecosystems, while eight unknown viruses were detected in Mediterranean Marchica only. Statistical analysis showed higher microbial diversity in the Atlantic lagoon than in the Mediterranean lagoon and a robust relationship between alpha diversity and geographic sampling locations. This fifirst–ever metagenomics study on Moroccan aquatic ecosystems enriched the national catalog of marine microorganisms.

Key Words (5) : Metagenomics; marine microbiome; moroccan lagoons; biodiversity; bioindicators.