

**Metagenomics Assessment of Anthropogenic Impact on Coral
Reef-Associated Microorganisms on the Kenyan Indian Ocean**

by

SAMMY MUSEE WAMBUA

Submitted in fulfilment of the academic requirements of

Doctor of Philosophy

in Biological Sciences

School of Life Sciences

College of Agriculture, Engineering and Science

University of KwaZulu-Natal

Westville

South Africa

January 2020

PREFACE

The research contained in this thesis was completed by the candidate while based in the Discipline of Biological Sciences, School of Life Sciences of the College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Westville Campus, South Africa. The research was financially supported by the Swedish Research Council (grant number: 2015-03443_VR). The candidate was also supported by the Western Indian Ocean Marine Scientific Association (WIOMSA) through a conference scholarship, and a research grant (grant number: MARG II Contract 01/2019) to facilitate data analysis and writing. The candidate's postgraduate fees were remitted by the University of KwaZulu-Natal.

The contents of this work have not been submitted in any form to another university and, except where the work of others is acknowledged in the text, the results reported are due to investigations by the candidate.



Signed: Dr. Angus Macdonald

Date: 27th January 2020



Signed: Prof. Santie de Villiers

Date: 03.01.2020

DECLARATION 1: PLAGIARISM

I, Sammy Musee Wambua, declare that:

(i) the research reported in this dissertation, except where otherwise indicated or acknowledged, is my original work;

(ii) this dissertation has not been submitted in full or in part for any degree or examination to any other university;

(iii) this dissertation does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons;

(iv) this dissertation does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a) their words have been re-written, but the general information attributed to them has been referenced;

b) where their exact words have been used, their writing has been placed inside quotation marks, and referenced;

(v) where I have used material for which publications followed, I have indicated in detail my role in the work;

(vi) this dissertation is primarily a collection of material, prepared by myself, published as journal articles or presented as a poster and oral presentations at conferences. In some cases, additional material has been included;

(vii) this dissertation does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the dissertation and in the References sections.



Signed: Sammy Musee Wambua

Date: 3rd January 2020

DECLARATION 2: PUBLICATIONS AND PRESENTATIONS

My role in each paper and presentation is indicated. Papers that are still in progress and have a status other than accepted or published are indicated. The * indicates corresponding author.

Chapter 2

1. Wambua, S*, Macdonald, A., de Villiers, S. Marine Microbial Genomics: Unexplored Opportunities in Western Indian Ocean. [Under review]
2. Wambua, S*, de Villiers, S. and the Coral Metagenomic Study Group. Metagenomic Assessment of Anthropogenic Impact on Marine Microorganism: Applications for Sustainable Marine Resource Management. Global Sustainable Blue Economy Conference, University of Nairobi, Nairobi, Kenya held from 26 – 28 November 2018. Accepted as oral presentation. Presenter: S Wambua.
3. Wambua, S*, Gourelé, H., Curpen, K., Karlsson, O., Wambiji, N., Macdonald, A., Jaufeerally, Y., de Villiers, E., Bongcam-Rudloff, E., de Villiers, S. Modern DNA-based Assessments of Human Impacts on Kenya Corals. Global Sustainable Blue Economy Conference, University of Nairobi, Nairobi, Kenya held from 26 – 28 November 2018. Accepted as poster presentation. Presenter: S Wambua.
4. Wambua, S*. DNA-based Approaches for Marine Research. Address to the Kenya Wildlife Service (KWS) Management Staff given at the KWS Marine Centre, Mombasa, Kenya on 11th April 2019. Presenter: S Wambua.

The work in the above manuscript and presentations are a result of literature review for the methods that were to be employed in this project. The paper reviewed state-of-the-art genomic approaches in marine research and application to the WIO region. As for the presentations, I was invited to the Global Sustainable Blue Economy Conference in Nairobi to present my work and demonstrate the utility of molecular biology in the sustainable use of marine resources. Afterwards, Senior Warden John Wambua of the Mombasa National Marine Park and Reserve, having become aware of my presentation at the conference, requested if I could make a presentation to the KWS management staff at Mombasa to see which of the methods employed in my research they could incorporate or collaborate on in their survey and monitoring

strategies. I discussed non-invasive sampling of environmental DNA (eDNA), next generation sequencing (NGS) and bioinformatics analysis employed in my project and how they may be adapted in marine resource conservation and management. I conceived the layouts, conducted literature review and wrote the presentations.

Chapter 3

5. Kaimba, A., de Villiers, S., Wambua, S*. (2019). Does Protection of Marine Areas Safeguard Coral Reefs from Human-Source Pollution? *Frontiers in Environmental Science*, 7(89), 1–4. [Published]
6. Kaimba, A., de Villiers, S., Wambua, S*. *E. coli* and Nutrients Concentration Variations by Levels of Marine Protection on Coral Reefs of the Kenyan Western Indian Ocean. 11th WIOMSA Scientific Symposium, University of Mauritius, Réduit, Mauritius held from 1 – 6 July 2019. Accepted as poster presentation. Presenter: A Kaimba.

The above work sought to find potential surrogate markers for anthropogenic impact in marine environment. This resulted from review of marine protected areas (MPAs) and human activities in the coastal zones covered in chapter 3. I designed the experiments, performed sample collection and laboratory analysis, I performed data analysis, wrote sections of the manuscripts and edited it.

Chapter 4

7. Wambua, S*, Gourelé, H., de Villiers, E., Ngoi, J., Macdonald, A., Bongcam-Rudloff, E., de Villiers, S. A peek into western Indian Ocean microbial richness: a pilot for a coral microbiome study. *Frontiers in Marine Science*. [Under review]
8. Wambua, S*, Gourelé, H., de Villiers, E., Macdonald, A., Bongcam-Rudloff, E., de Villiers, S. Coral-reef and Sub-littoral Microbiome Metabarcoding by 16S rRNA Gene Ion Torrent Sequencing. 11th WIOMSA Scientific Symposium, University of Mauritius, Réduit, Mauritius held from 1 – 6 July 2019. Accepted as oral presentation. Presenter: S Wambua.

The data presented is from samples utilized for optimization of environmental DNA isolation methods for metagenomic work in chapter 4. I obtained a scholarship from WIOMSA to attend the 11th WIOMSA Scientific Symposium in Mauritius where I presented the data. I discussed

the microbial diversity richness of near-shore sublittoral and coral reef zones of western Indian Ocean and showcased the potential of eDNA metabarcoding approaching as bioprospecting and monitoring tool for marine resources. I designed the experiment, performed sample collections, did laboratory processing and data analyses on the sequence data for this work.

Chapter 5

9. Wambua, S*, Gourelé, H., de Villiers, Oskar Karlsson-Lindsjö, O., Wambiji, N., Macdonald, A., Bongcam-Rudloff, E., de Villiers, S. Molecular Ecology of Coral Reef Microorganisms in the Western Indian Ocean coast of Kenya. *Ecology and Evolution* [Under review]

This paper highlights the main patterns in the Kenyan coral reef microbial communities under gradient of human impacts. It combines the microbial taxonomic and functional diversity analysed in chapters 4 and 5. I designed the experiment, performed sample collections, did laboratory processing and data analyses on the sequence data for this work.



Signed: Sammy Musee Wambua

Date: 3rd January 2020

ABSTRACT

The Western Indian Ocean (WIO) is the world's second richest marine biodiversity hotspot. It is characterized by a diverse range of ecologically and nutritionally rich marine ecosystems that are increasingly under pressure from the impacts of human population growth and coastal development. A comprehensive understanding of variations of marine microbial community composition with environmental conditions is key to understanding and predicting responses to human and climate pressures because microorganisms are the main drivers of biogeographical processes, and they respond and adapt fast to climatic patterns. Also, because unique environments harbour unique microorganisms with unique properties, genomic exploration of marine microorganisms may lead to discovery of novel metabolic processes, and bioactive products with potential for novel biotechnological applications. This project aimed to assess the effect of local anthropogenic impacts on the community structure and the functional potential of microorganisms inhabiting the WIO coral reefs along the Kenyan coast by metagenomics. Reasons for low application of genomics specifically in marine research in the WIO region were examined through literature review and in-depth interviews with scientists in the region. Coral reef seawater and sediment samples were collected, for microbial assessments, along gradients of human impacts and protection regimes. Environmental variables were estimated, and microbial taxonomic and functional diversity analysed by metagenomic approaches. Comparisons were done between sites with differential human impacts, and with oceanic metagenomes from *Tara* Oceans expedition. Compartmentalised training of marine scientists and lack of collaboration with molecular scientists are key reasons highlighted for poor uptake of genomics in marine research. Significant differences in taxonomic and functional composition were observed between the coral reef and *Tara* Oceans datasets. Coral reef metagenomes had more diverse and even microbial taxa and gene groups. *Tara* Oceans metagenomes were enriched with groups of genes of functions in keeping with oligotrophic conditions, and reefs metagenomes with genes for functions related to adaptations to heterogenous environments. *E. coli* density decreased with increasing degree of protection, but physicochemical and nutrient variables did not differ across coral reefs in the protected coral reefs. Variations in relative abundances of copiotrophic bacteria and coliphages were observed in coral reefs corresponding to magnitude of the neighbouring human impacts. Malindi and Mombasa marine parks, the coral reef sites experiencing degradative human impacts, were significantly enriched with genes for functions suggestive of mitigation of environment perturbations e.g. capacity to reduce intracellular levels of environmental contaminants and

repair of DNA damage. This study establishes essential baselines for microbiome studies in the WIO region and provides insights to anthropogenic impact on microbial structure and functions, and potential indicators of health status coral reef ecosystems.

ACKNOWLEDGMENTS

I am forever indebted to a family of amazing souls who are the real power behind my coming this far:

First and most importantly, God my Almighty Father who supernaturally orchestrated such intricate details and all the events leading up to me being found by this PhD project, and working with the people I did. Words are not enough!

My supervisors Angus Macdonald (UKZN) and Santie de Villiers (PU), as well as Erik Bongcam-Rudloff (SLU), for your committed and sacrificial mentorship, and very kind generosity. Beyond academics you have reinforced my faith in humanity. If only the world had a few more people like you. I want to be to others what you have been to me.

Oskar Karlsson (SLU), for your support with procurement, shipment and sequencing logistics. Etienne de Villiers (KEMRI/WTRP) and Hadrien Gourelé (SLU), for your technical support with the bioinformatics work.

The Kenyan regulatory agencies for your cooperation. A special mention of Ms. Lilian Gorretti (NACOSTI) and Joseph Makau (NEMA) for your efficiency. KWS Senior Wardens Ms. Jane (Malindi Marine National Park & Reserve), Mr. John Wambua (Mombasa Marine National Park & Reserve), and Mr. Erick Aduda (Kisite-Mpunguti Marine National Park & Reserve), as well as Mr. John Haro (Kuruwitu Conservancy) and Nina Wambiji (KMFRI) for support and help with sampling logistics on site.

Swedish Research Council (Vetenskapsrådet) and the Western Indian Ocean Marine Association (WIOMSA), for funding.

The Swedish Embassy in Nairobi and my Sweden family: my sister Wandoye, daughters Zubeda and Manga, as well as mama Manga; brother Saria (TZ) and sister Winnie (UG). Ahsanteni for helping with travel and living logistics. Brother Prosper Hakizimana and your wife, sister Rebecka Risberg, thanks for your incessant prayers.

My late mom and dad, Mr. and Mrs. Wambua. I believe it is not too late to make you proud!

My girls, Tunu Muikiio and Neema Ndanu, for bearing with all the father-daughters time I deprived you of while I worked on the PhD. And for the concern, love and the encouragement you showered me with during the nights of writing. My main girl and loving wife Jedidah Kawuki, for playing dad when I wasn't available physically and emotionally, for understanding, supporting and stabilizing me emotionally, spiritually and physically. You girls are my chief motivation... and this PhD is dedicated to you.

TABLE OF CONTENTS

	<u>Page</u>
PREFACE	ii
DECLARATION 1: PLAGIARISM	iii
DECLARATION 2: PUBLICATIONS AND PRESENTATIONS	iv
Chapter 2	iv
Chapter 3	v
Chapter 4	v
Chapter 5	vi
ABSTRACT	vii
ACKNOWLEDGMENTS	ix
TABLE OF CONTENTS	xi
LIST OF TABLES	xiv
LIST OF FIGURES	xv
CHAPTER 1: INTRODUCTION	1
1.1 Rationale for the research (nature and scope)	1
1.2 Justification	2
1.3 Aims	4
1.4 Objectives	4
1.6 Outline of dissertation/thesis structure	4
CHAPTER 2: MARINE MICROBIAL GENOMICS: UNEXPLORED OPPORTUNITIES IN THE WESTERN INDIAN OCEAN	6
2.1 Abstract	6
2.2 Introduction	7
2.3 Ecological value of microorganisms	8
2.3.1 Human health	8
2.3.2 Environmental health	9
2.4 Coral microbiota	10
2.4.1 Coral reef-associated microorganisms	11
2.4.1.1 Coral reef pelagic microorganisms	11
2.4.1.2 Coral reef sediment microorganisms	12

2.5 Status of the WIO marine ecosystems.....	13
2.5.1 Mangrove forests	14
2.5.2 Seagrass beds	16
2.5.3 Corals reefs	19
2.6 Microbial genomics-related approaches appropriate for the WIO region.....	21
2.6.1 Metabarcoding	22
2.6.2 Metagenomics.....	24
2.6.4 PCR-based microbial markers	26
2.7 Strategies to enhance uptake of genomic methods in the WIO region	27
2.8 Conclusion.....	29
CHAPTER 3: WATER QUALITY IN KENYAN CORAL REEF ECOSYSTEMS ACROSS	
A COASTAL IMPACT GRADIENT	
3.1 Abstract	31
3.2 Introduction	32
3.3 Materials and methods	35
3.3.1 Study sites and human activities	35
3.3.2 Sampling and in situ assessments	36
3.3.3 Laboratory tests	38
3.3.3.1 Nutrient measurements.....	38
3.3.3.2 Microbial assessments.....	38
3.4 Results	40
3.4.1 Inter-site variations in “no-take” marine national parks	40
3.4.2 Variations across marine protection gradient	43
3.5 Discussion	44
3.6 Conclusion.....	48
CHAPTER 4: TAXONOMIC STRUCTURE OF THE KENYAN INDIAN OCEAN	
CORAL REEF MICROBIAL COMMUNITIES	
4.1 Abstract	49
4.2 Introduction	50
4.3 Materials and methods	51
4.3.1 Study sites and sampling	51
4.3.2 DNA isolation.....	51
4.3.3 Library preparation and DNA sequencing.....	52
4.3.4 Bioinformatics analyses.....	52

4.3.5 Statistical analyses	53
4.4 Results	53
4.5 Discussion	67
4.3 Conclusion.....	71
CHAPTER 5: FUNCTION POTENTIAL OF MICROBIAL COMMUNITIES ON KENYAN CORAL REEFS	72
5.1 Abstract	72
5.2 Introduction	73
5.3 Materials and methods	74
5.4 Results	75
5.5 Discussion	83
5.6 Conclusion.....	85
CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH.....	87
6.1 Introduction	87
6.2 Aims and objectives	89
6.3 Summary of major findings.....	89
6.4 Study challenges and limitations.....	91
6.5 Future possibilities	92
6.6 Final comments and summary conclusions.....	93
REFERENCES.....	94

LIST OF TABLES

<u>Table</u>	<u>Page</u>
Table 3.1 <i>Bacteroides-Prevotella</i> and universal prokaryote primers used in the study for identification of sources of faecal pollution in the coral reef sites.	39
Table 3.2 Measurements of physicochemical characteristics of the three coral reefs within the marine national parks at the time of sampling for metagenomic sequencing i.e. November -December 2016.	40
Table 3.3 Nutrient levels measurements of the three coral reefs within the marine national parks at the time of sampling for microbiome sequencing i.e. November -December 2016.	41
Table 3.4 Nutrient mean concentrations for in water samples collected at from the coral reef sites.	44
Table 4.1 A breakdown of the metagenomic sequences from the Kenyan coral reefs and <i>Tara</i> Oceans expedition (oceanic) samples processed on the MGX metagenomics analysis platform	53

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
Figure 3.1 Map showing location of the coral reef sites sampled for this study.....	37
Figure 3.2 Agarose gel (1% w/v) image showing amplification of <i>Bacteroides-Prevotella</i> marker genes on 1) Neat; 2) 10 ⁻¹ ; 3) 10 ⁻² ; and 4) 10 ⁻³ dilutions of ruminants, human and pig faecal control DNAs. L, GeneRuler™ 1kb DNA ladder 250-10000bp; and N, master mix with PCR water used as negative control.....	41
Figure 3.3 Agarose gel (1% w/v) image showing PCR results for control DNAs isolated from stool samples of human (H), cow (C), goat (G) and pig (P) amplified with faecal bacteria primers specific for: A) human, B) ruminants, C) all-source <i>Bacteroides-Prevotella</i> , and D) pig. L, GeneRuler™ 1kb DNA ladder 250-10000bp; and N, master mix with PCR water used as negative control	42
Figure 3.4 Agarose gel (1% w/v) image results for the universal bacteria 16S rRNA gene PCR. L, GeneRuler™ 1kb DNA ladder 250-10000bp; P, commercial bacterial control DNA from <i>E. coli</i> DNA; 1 to 3, eDNA from 1) Malindi, 2) Mombasa and 3) Kisite sites; 4, unrelated DNA that tested positive for bacteria in a previous experiment; N, master mix with PCR water used as negative control.....	42
Figure 3.5 Agarose gel (1% w/v) image showing replicate faecal bacteria PCR screening results for eDNA from 1) Malindi, 2) Mombasa and 3) Kisite sites, along with control DNAs isolated from stool samples of human (H), cow (C), goat (G) and pig (P); L, GeneRuler™ 1kb DNA ladder 250-10000bp; and N, master mix with PCR water used as negative control.....	43
Figure 3.6 <i>E. coli</i> concentrations (MPN/100 mL) in water sampled at three different coral reef sites. Error bars indicate 95% confidence intervals.	43
Figure 4.1 Rarefaction curves for community species richness on the sequences from <i>Tara</i> Oceans (oceanic) (A) and Kenya coral reefs (reef) (B) samples.....	54
Figure 4.2 Stacked charts of the distribution of community phyla with relative abundance > 1% sequenced from Kenya coral reef (Ken*), and the <i>Tara</i> Oceans expedition (Tar*) samples	55

Figure 4.3 Plots of alpha diversity metrics the observed and estimated bacterial, viral and archaeal species richness on the Oceanic and Reef sample sequences	56
Figure 4.4 Bacterial, viral and archaeal combined community composition analysed with nonmetric multidimensional scaling (nMDS) plots using Bray-Curtis dissimilarity for Kenya coral reefs (peach circles) and <i>Tara</i> Oceans (aqua circles) datasets. For this nMDS plot, stress = 0.04, $r^2 = 0.998$	56
Figure 4.5 Bacterial, viral and archaeal combined community composition Principal Coordinate Analysis (PCoA) plots of Bray–Curtis dissimilarity matrix for reefs and oceanic datasets	57
Figure 4.6 Bacterial community composition analysed with Principal Coordinate Analysis (PCoA) plots of Bray–Curtis dissimilarity matrix for reefs and oceanic datasets.	57
Figure 4.7 Distribution of oceanic (aqua) and reef (peach) sequences assigned to bacterial classes with relative abundance of 1% or higher in at least one of the datasets are plotted.	58
Figure 4.8 Distribution of oceanic (aqua) and reef (peach) sequences assigned to orders of the copiotrophic phyla Proteobacteria (A), and Bacteroidetes and Actinobacteria (B). Only orders with relative abundance of 1% or higher in at least one of the datasets are plotted.	59
Figure 4.9 Bacterial community composition analysed with nonmetric multidimensional scaling (nMDS) plots using Bray-Curtis dissimilarity for Kenya coral reefs at the national marine parks classified “exploited” (Olive circles; Malindi), “baseline” (Green circles; Kisite-Mpunguti), and “polluted” (Blue circles; Mombasa). For this nMDS plot, stress = 0.05, $r^2 = 0.998$	61
Figure 4.10 Distribution of the top ten most abundant orders of the copiotrophic phyla (Proteobacteria and Bacteroidetes) from sequences of samples from the Kenyan coral reefs at the national marine parks classified “baseline” (Sky; Kisite-Mpunguti), “exploited” (Olive; Malindi) and “polluted” (Bronze; Mombasa).....	62
Figure 4.11 Distribution of faecal indicator bacterial genera by study site i.e. Oceanic (blue), and the coral reefs at sites classified “baseline” (sky), “exploited” (olive) and “polluted” (bronze).	62

Figure 4.12 Venn diagrams showing the distribution of microbial species across A) <i>Tara</i> and reef, and B) Kenya coral reef metagenomes based on species with more than 2, 10, 50 reads.	63
Figure 4.13 Plots of alpha diversity metrics for the observed and estimated archaeal species richness on the Oceanic and Reef sample sequences	63
Figure 4.14 Archaeal community composition analysed with nonmetric multidimensional scaling (nMDS) plots using Bray-Curtis dissimilarity for Kenya coral reefs (peach circles) and <i>Tara</i> Oceans (aqua circles) datasets. For this nMDS plot, stress = 0.05, $r^2 = 0.998$	64
Figure 4.15 Relative abundance of archaeal phyla sequenced from Kenya coral reef (Ken*), and the <i>Tara</i> Oceans expedition (Tar*) samples.....	65
Figure 4.16 Distribution of archaeal orders with relative abundance >1% grouped by study site i.e. Oceanic (blue), and the coral reefs at sites classified “baseline” (sky), “exploited” (olive) and “polluted” (bronze)	66
Figure 4.17 Stacked charts of the distribution of viral phyla with relative abundance > 1% sequenced from Kenya coral reef (Ken*), and the <i>Tara</i> Oceans expedition (Tar*) samples.....	67
Figure 5.1 Relative abundances of COG functional categories. Asterisks denote categories with significant differences in abundance.	76
Figure 5.2 Distribution of COG categories in sample sequences from the Kenya coral reefs with varied human impacts i.e. “baseline” (sky blue), “exploited” (olive) and “polluted” (bronze).	77
Figure 5.3 Rarefaction curves for clusters of orthologous groups of proteins (COG) richness on the sequences from Kenya coral reefs (reef) (A) and <i>Tara</i> Oceans (oceanic) (B) samples.....	78
Figure 5.4 Composition of clusters of orthologous groups of proteins (COGs) analysed with Principal Coordinate Analysis (PCoA) plots of Bray–Curtis dissimilarity matrix for reefs and oceanic metagenomes.	79

Figure 5.5 Distribution of the 24 most abundant COGs (~10% of total reads) in oceanic and reef metagenomes. Asterisks denote COGs with significant differences in relative abundance between the datasets.	80
Figure 5.6 Comparison between reef-specific and oceanic-specific relative abundances of the 40 most represented COGs significantly enriched in the Kenya coral reef metagenomes	81
Figure 5.7 Plots of alpha diversity metrics for predicted COGs richness on metagenomes from the Kenya coral reef sites	82
Figure 5.8 Comparison between Kenya coral reef sites of relative abundances of the 25 most represented COGs that are significantly enriched in the “exploited” and “polluted” sites.	83

CHAPTER 1: INTRODUCTION

1.1 Rationale for the research (nature and scope)

The Western Indian Ocean (WIO) hosts the second hotspot of coral reef biodiversity globally. With over 60 million people (more than a quarter of the region's population) living within 100 km of the shoreline, the WIO coral reef ecosystems underpin the region's economies by providing food, recreation opportunities, livelihoods and jobs to local communities (Obura, 2017; Van der Elst et al., 2005). The untapped economic potential of coral reefs in the region has come to focus in recent debates around the blue economy. Kenya, for instance, has put in place institutional and legal frameworks to prioritize the blue economy as a strategy to achieving vision 2030 development agenda targets (Wairimu & Khainga, 2017). Inhabited by over 250 coral species (Obura, 2012; Tuda & Omar, 2012), the Kenyan coral reefs wield immense unexplored economic potential (Obura et al., 2017; Ransom & Mangi, 2010). However, this potential is under constantly increasing pressure from both global and local threats.

Since the 1980s global warming has caused three mass coral bleaching events (Hughes et al., 2017) resulting in over 90% coral mortality in some regions (Claar et al., 2018). In the WIO region Kenya and Seychelles were the hardest hit by the first event in 1998 (Obura et al., 2017). Significant recovery has since been noted especially in Kenya where corals have developed remarkable resilience against the 2016 event (Gardner et al., 2019; McClanahan, 2017; Obura et al., 2017). However, a variety of local stressors including overexploitation, nutrient pollution, and use of destructive fishing methods continually pose a threat to coral reefs. It is expected that these pressures will increase with time considering the ongoing population growth, increasing coastal development, and recent efforts to prioritize the blue economy, which emphasize utilization over stewardship. There is therefore an urgent need for innovative interventions to protect, conserve and restore coral reef ecosystems.

So far ecosystem-based management approaches, especially government administered marine protected areas (MPA) and locally managed marine areas (LMMAs) are the main strategies employed to protect marine biodiversity. In Kenya, MPAs and LMMAs have had great success in improving fish stocks (Ransom & Mangi, 2010; Samoilys & Obura, 2011). However, it is feared that MPAs and LMMAs may not be effective in mitigating against the indirect impacts

of human activities such as pollution, sedimentation and coastal development (Mora et al., 2006). Because of their fundamental contribution to the health and functioning of coral reefs, microorganisms can potentially serve as sensitive and rapid markers for ecosystem stress. Also, since microorganisms respond and adapt fast to change in their environments (Ainsworth et al., 2010), they may be used to predict response to stressors.

Modern high throughput DNA sequencing technologies have greatly improved the ability to study microbial diversity and ecology. For instance, metabarcoding – the use of environmental DNA (eDNA) to identify species by targeting a short sequences of a specific gene – has aided in the discovery of extant microbial diversity that had eluded detection by traditional cultivation approaches (Baharum et al., 2010; DeLong, 2005). On the other hand, metagenomics, which is the direct genetic analysis of genomes contained in an environmental sample, provides access to the full repertoire of genes of microbial communities allowing for both phylogenetic and functional description surveys (Thomas et al., 2012). These approaches provide a vast potential to survey and compare microbial communities from different environments, their evolution, as well as discovery of potentially novel compounds (Behzad et al., 2016; Thomas et al., 2012).

This project employed next generation sequencing (NGS) technology and bioinformatics analysis to profile population composition and the functional potential of microorganisms inhabiting the coral reefs of the Indian Ocean along the Kenyan coast. Sites managed with varying degrees of protection were sampled, creating a gradient of impacts of human activities. The sites were also characterized with respect to environmental variables to note any inter-site variations at the time of sampling. Comparison was also done with open ocean seawater (oceanic) metagenomes from the Tara Oceans expedition.

1.2 Justification

Rapid development of affordable and flexible high-throughput sequencing approaches and advances in bioinformatics have made studying microbial ecology easier and more efficient. The WIO, because of its dynamic tropical waters and being the second global hotspot of biodiversity, has immense unique ecological and economic potential which may be explored by molecular means. However scarce use of molecular methods is worse in the WIO region and is precipitated by inadequate technological capacity, shortage of skilled manpower, and lack of meaningful collaborations (H3Africa Consortium, 2014). Consequently, the molecular

microbial diversity of the WIO is arguably the least studied of all global oceans (Díez et al., 2016). To date, only five studies of WIO microbial populations have been reported. Of these, one characterized only viruses (Williamson et al., 2012), while two were large-scale expeditions on global oceans (Sunagawa et al., 2015; Zinger et al., 2011) which reported pooled findings with no specific details on the WIO. Such generalized findings have limited use for local application since it has been demonstrated that marine microorganisms associate in context-specific manner with regard to time, space, and host (Apprill et al., 2016; Gardner et al., 2019; Glasl et al., 2019; Walsh et al., 2017). The other two studies are, arguably, the only examples of the type needed in the region; they characterised microbial communities associated with coral disease in South Africa, Reunion and Mayotte (Séré et al., 2013) and with coral bleaching in the Seychelles (Gardner et al., 2019). It is worth noting also, that these studies included only limited sequencing of targeted genes for taxonomic diversity and did not do metagenomic sequencing to assess functional biodiversity.

Microorganisms are major determinants of marine biological processes (Glöckner & Joint, 2010). They also respond and adapt fast to environmental disturbance (Ainsworth et al., 2010), and are implicated in facilitating adaptation of corals to environmental changes (Gardner et al., 2019). Furthermore, microorganisms are believed to be critical for coral nutrition, homeostasis, health and protection against diseases (Glasl et al., 2019; Godoy-Vitorino et al., 2017; Krediet et al., 2013). It follows then that studying microbial ecology is necessary to understand and predict responses of the marine ecosystem to perturbations such as that resulting from anthropogenic and climate pressures. Indeed, the diagnostic value of microorganisms, especially those inhabiting coral reef, to infer the environmental health status of coral reef ecosystems has recently been demonstrated (Glasl et al., 2019).

Coral reef bacterial communities inhabit habitats such as the sediments, overlying water column, and benthic invertebrates such as corals and sponges (Bourne & Webster, 2013). Benthic-pelagic coupling occurs in shallow well-mixed tropical coral reefs where bacterial communities in benthic organisms, the sediments, and the overlying water column are strongly interlinked (Bourne & Webster, 2013; Vanwonderghem & Webster, 2020). These interactions within what has been termed the “coral ecosphere” are suspected to influence recruitment of coral-associated microorganisms (Weber et al., 2019). For instance, increases in the abundance of microbes in the reef water column has been correlated with an increase in coral disease and reduction in coral cover (Walsh et al., 2017). Furthermore, since anthropogenic activities most

likely impact coral health through the agency of the immediate pelagic and benthic surrounding, the water overlying corals and sediments are often the niches sampled to detect relevant signal in coral reef microorganisms (Bourne & Webster, 2013; Glasl et al., 2019; Kelly et al., 2014; McDole et al., 2012; Tout et al., 2014; Walsh et al., 2017).

Kenya marine ecosystems are reported to be fairly intact (Gudka et al., 2018; Obura, 2017; Obura et al., 2017) and contain corals that are more resilient to bleaching than those in other WIO states (Gardner et al., 2019; McClanahan, 2017). Studying the microbial populations in the coral reefs here has the potential to identify communities responsible for resilience, as well as establishing baseline profiles for future reference. Also, genomic exploration of the microorganisms adapted to the Kenyan marine environment will provide an opportunity for biodiscovery of unique metabolic processes with potential industrial or biotechnological applications.

1.3 Aims

The main aim of the project was to assess the effect of local anthropogenic impacts on the community structure and functional potential of microorganisms associated with coral reefs in the Kenyan Indian Ocean.

1.4 Objectives

The specific research objectives of the study were to:

- Characterise the physicochemical and biological variables of coral reefs neighbouring locations of known human activities
- Determine the taxonomic diversity of coral reef water and sediment micro-organisms by highthroughput sequencing
- Assess the genomic diversity and functional potential of coral reef water and sediment micro-organisms by metagenomic sequencing

1.6 Outline of dissertation/thesis structure

This thesis is structured as chapters written in manuscript format. A section of Chapter 3 has been published, while Chapter 2 and a section of Chapter 4 are under review in international journals. Chapters that have sections separately analysed for publication are organised in sub-studies so that the manuscripts are included in the format they were submitted for publication. The thesis chapters are summarized as follows:

Chapter 1 provides the general background of the project, introduces the research rationale and highlights the research gaps and the objectives of the study.

Chapter 2 is a review of the literature that is reiterated throughout the remaining chapters. It addresses the lacklustre adoption of molecular approaches in marine biology research within the WIO region and highlights reasons for the low uptake and possible mitigation. Specific molecular and genomic methods with potential application in the WIO are described along with the key marine ecosystems that could benefit from genomic approaches as well as their potential and status in the face of ongoing global and local pressures.

Chapter 3 describes the study sites sampled for the PhD with respect to their neighbouring human activities, the degree of protection, as well as their physicochemical and biological environmental characteristics. It addresses the possibility of MPAs protecting against the impact from indirect human activities and the potential to employ of such environmental characteristics as surrogates for human activities.

Chapter 4 profiles taxonomic diversity of microorganisms (bacteria, archaea and viruses) that inhabit Kenyan coral reef seawater and sediments. Annotation was done by metagenomic approach and comparisons done with WIO open seawater sequences from the Tara Oceans expedition, and between the different anthropogenic impacts at the coral reefs.

Chapter 5 annotates the potential functional diversity similarities and differences between microbial communities in the coral reef seawater and sediments from the study sites and the Tara Oceans expedition from chapter 4. It assesses the repertoire of gene groups differentially enriched in each of the sampling sites.

Chapter 6 integrates work from all the preceding chapters and makes conclusions about the value of the research. This chapter also outlines future perspectives and discusses the potential research opportunities that stem from this work.

CHAPTER 2: MARINE MICROBIAL GENOMICS: UNEXPLORED OPPORTUNITIES IN THE WESTERN INDIAN OCEAN

2.1 Abstract

The Western Indian Ocean (WIO) is characterized by a diverse range of ecologically and nutritionally rich ecosystems ranking it as the world's second richest marine biodiversity hotspot. The goods and services provided by these ecosystems are increasingly coming under pressure from the impacts of human population growth and coastal development. A comprehensive understanding of the biogeochemical processes is necessary to protect and sustain the WIO ecosystem. Marine microorganisms are the principal drivers of the biogeographical processes of the oceans and ultimately the atmosphere, and they also respond and adapt fast to climatic patterns. Therefore, studying microorganisms' biodiversity is key to understanding and predicting responses of the marine ecosystem to human and climate pressures. And because marine microorganisms have unique properties in terms of adaptation to the extreme environments, their genomic exploration may lead to discovery of metabolic processes, products and functions with potential for novel biotechnological applications. More than 99% of microorganisms cannot yet be cultured under standard laboratory conditions. Recent advances in genetics and genomic technologies offer efficient alternatives that have motivated intense research into the composition and activities of microorganisms from diverse habitats. However, the appreciation and application of these methods to address marine problems has been slower than in other disciplines of life sciences. Although molecular and genomic approaches are substantially employed in biomedical and agricultural research within some of the WIO countries, their application to address marine biology challenges is deficient due to the lack of collaborations between molecular and marine biologists and inadequate technological capacity. This chapter reviewed the state of marine ecosystems in the WIO region and examined the unique reasons for low use of genomics specifically in marine research in the WIO region. Compartmentalised subject-area-based, rather than holistic, training of marine scientists, and lack of collaboration with molecular scientists are some of the reasons highlighted. Solutions are suggested. Microbial genomic approaches that can be adopted and how they may augment marine resource conservation and management efforts in the region are discussed.

Keywords: *Marine microbiology, molecular ecology, metagenomics, microbial diversity*

2.2 Introduction

Since the 1990s DNA-based tests have increasingly gained credibility as powerful tools for accurate and definitive identification of organisms and genetic conditions, following the use of ‘DNA fingerprinting’ to convincingly solve a forensic case (Roewer, 2013). Advances in molecular biology technologies and their increased application has led to significant progress in many fields of life sciences. For instance, the emergence of high-throughput sequencing and computational tools have significantly advanced the field of microbial ecology by inspiring intense research focused on the composition and activities of microorganisms from diverse habitats (Heidelberg et al., 2010). Consequently, a vast diversity of previously unknown microbial life forms, especially in marine niches have been identified (Behzad et al., 2016). Despite this trend, few studies have applied molecular biology applications to address marine biology problems in the WIO region. Indeed, the Indian Ocean remains the least studied of the global oceans (Díez et al., 2016), especially in terms of microbial ecology.

Healthy ocean and coastal ecosystems play critical roles in supporting human well-being through the provision of ecosystem goods and services including food security, opportunities for livelihoods and recreation, to protection from coastal hazards such as storms (Ruckelshaus et al., 2013). The WIO features diverse marine ecosystems that include coral reefs, mangrove forests and seagrass meadows, which contribute substantially to the region’s economy (Obura, 2017). Their diversity and productivity attract humans to the coastal zones. It is estimated that 60 million people (more than a quarter of the region’s population) live within 100 kilometres of the WIO coast (Obura, 2017; Van der Elst et al., 2005), with human population and coastal urbanisation zones projected to continue growing (Neumann et al., 2015). This influx into the coastal zones generates unsustainable pressure on the coastal ecosystems and natural resources through overexploitation and pollution (Vikas & Dwarakish, 2015) leading to degradation. It is feared that the WIO marine ecosystems will suffer this fate as they already show signs of distress (Obura, 2017). There is therefore an urgent need to protect marine ecosystems from further destruction.

The importance of microbial communities to human and environmental health has frequently been demonstrated through genomic studies (Jansson & Baker, 2016). It has been shown that microbes are important for coral nutrition (Rosenberg et al., 2007), homeostasis and protection against diseases (Godoy-Vitorino et al., 2017). Therefore, comprehensive understanding of microbial population composition and ecological processes in marine ecosystems is necessary

to inform novel and augment existing conservation and management strategies. Genomic approaches that can be applied in marine research throughout the region and which are already in use in some WIO countries, are discussed. Specific challenges prevalent in the WIO region that can be addressed are highlighted alongside the contributions the techniques can make to study, protect and enhance marine resources.

2.3 Ecological value of microorganisms

Microorganisms are present nearly everywhere in nature; they live in all environments on Earth that are occupied by other organisms and are the sole life forms in some environments. They are the principal drivers of ecological processes, mainly through their influence on the biogeochemical cycling in ecosystems (Konopka, 2009). The key contributions of microorganisms to nature are detailed in “*The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet*” (National Research Council, 2007).

2.3.1 Human health

There are more microorganisms (formerly called ‘normal flora’) on the human body than human cells. Differences in microbial population structure between individuals and body sites have been noted (Huttenhower et al., 2012) indicating the possibility that microorganisms co-evolved with their human hosts (Turnbaugh et al., 2007). As such, their role in human health cannot be overstated. The human gut, particularly, is resident to highly specialized microorganisms that regulate the host immunity and metabolism; they offer defence against pathogens, detoxify potentially harmful chemicals if ingested, as well as help with metabolic functions that humans have not evolved, like extraction of calories from indigestible diets and the synthesis of some essential vitamins and amino acids (Petrof & Khoruts, 2014). Indeed, dysbiosis – microbial population imbalance and/or altered activity – have been associated with many diseases while deliberately regulating the microbiota has been shown to improve or prevent some pathologic conditions (Bäckhed et al., 2012; Cho & Blaser, 2012). The health conditions influenced by the gut-bacteria balance are diverse, covering both intestinal (including inflammatory bowel disease, irritable bowel syndrome, and coeliac disease) and extra-intestinal disorders such as blood pressure, anxiety, aging processes, allergies, asthma, cardiovascular disease, and obesity (Carding et al., 2015; Gareau et al., 2010). These observations have inspired a surge of research to understand how microbial communities at different sites of the human body affect health and disease with a view to formulate better diagnosis, prevention and treatment. Some radical interventions have resulted from the

appreciation of microbiota's importance to human immunity. For instance, in vaginal seeding a swab of a mother's vaginal fluid is applied to a new-born's mouth, nose, and eyes, soon after caesarean delivery, to restore 'friendly bacteria' to help the baby develop a healthy immune system and decreasing the likelihood of allergies and inflammation in the long term (Lokugamage & Pathberiya, 2019). In faecal transplantation, or bacteriotherapy, restoration of gut microbiota balance is achieved through transplanting faeces from a healthy donor into the gastrointestinal tract of a recipient. Other related interventions in use at the moment include probiotics, prebiotics and phage therapy (Bäckhed et al., 2012; Petrof & Khoruts, 2014).

2.3.2 Environmental health

The microbial communities found in the soil and the ocean play a central role in the health of these environments, and the productivity of their components. Some soil-dwelling bacteria, for example, fix atmospheric nitrogen converting it to ammonia, which is readily usable by plants and animals. When plants and animals die, bacteria aid in their decomposition thereby recycling nutrients back into the soil to be utilized in plant nutrition. Complex soil microbial communities are also suspected to protect against plant pathogens as indicated by the loss of soil's ability to suppress disease once it is sterilized (Mazzola, 2004).

Marine microorganisms, on the other hand, are thought to contribute about 50% of global primary productivity (Glöckner & Joint, 2010). Photosynthetic bacteria are a proportion of the organisms in the ocean that convert atmospheric carbon dioxide into sugars generating oxygen needed by aerobic forms of life. It is estimated that more than half of the oxygen we breathe is produced in the ocean (Obura, 2017). These microbial communities also help alleviate greenhouse gases released by human activities in the atmosphere by transforming inorganic carbon dioxide into organic carbon that is then sequestered in the ocean. Overall, the chemical transformations mediated by marine microorganisms are responsible for the ocean's chemistry and habitability of the earth. Marine microorganisms also serve as an essential source of nutrition to benthic organisms (Zobell & Eltham, 1938).

Increasing appreciation of the ecological function of marine microorganisms in the last two decades, is evident from the increasing number of studies seeking to catalogue microbial communities and functions in different ocean niches (Biller et al., 2018; Godoy-Vitorino et al., 2017; Sahoo & Dhal, 2009; Sogin et al., 2006; Sunagawa et al., 2015), and marine organisms (Grozdanov & Hentschel, 2007; Hernandez-Agreda et al., 2017; Krediet et al., 2013; Ribes et

al., 2012). The exact ecological functions played by the microbes in these niches and organisms are yet to be conclusively determined. However, it has been found that microbial communities associate with marine organisms in species-specific relations suggesting the microbes' relevance to the well-being of their hosts. There are marine sponge-specific microbiota, for instance, that are passed vertically between sponge generations which have not been found in seawater, sediments, other marine invertebrates nor in freshwater sponges (Grozdanov & Hentschel, 2007). This microbial by sponges species has been noted to impact the microbes at both the taxonomy and functional levels (Karimi et al., 2018). Furthermore, metabolic pathways involved in nutrient uptake were shown to be mediated in two sponge species by different, but host specific, microbial communities (Ribes et al., 2012).

Microorganisms' ability to rapidly sense and adapt to environmental changes is important to environmental health and has great potential for broad biotechnological applications. Adaptations involving rapid restructuring of the microbial population composition, for instance, are theorized to facilitate acclimatization of the coral holobiont to changes in the environment (Gardner et al., 2019). Also, some soil bacteria are known to adapt to contaminated environments so as to utilize compounds of the pollutant as food (Wardell, 1995). In this regard, subsurface soil microorganisms that evolve the ability to biodegrade gasoline constituents are credited for protecting potential leakages from underground gasoline storage into groundwater sources (Mazzola, 2004). The potential application of these abilities in bioremediation has been extensively explored for deployment in various environments (Miller, et al., 2009; Prince, 1993; Song et al., 1990; Wardell, 1995).

2.4 Coral microbiota

Coral-associated microorganisms are arguably the most studied of the marine invertebrate microbiota; coral reefs are the most diverse of marine ecosystems that have recently come under unprecedented degradation due to local and global pressures. Corals' optimum functioning is dependent on healthy co-existence of symbiotic associations with algae and microorganisms. Microbe-host specificity has been shown in great detail in corals as well as in sponges. Hosting of thousands of bacterial species in species-specific associations irrespective of distance and time scale has been consistently noted in a number of coral species (Hernandez-Agreda et al., 2017; Rohwer et al., 2002). Additionally, investigations of microbial consortia in coral mucus, tissue and skeleton have recorded unique population composition for each of the compartments which were also significantly different from microbiota in the surrounding seawater and

sediments (Apprill et al., 2016; Li et al., 2015). Host nutrition, homeostasis and protection against diseases are some of the specific functions that coral-associated microorganisms are believed to be responsible for (Godoy-Vitorino et al., 2017; Rosenberg et al., 2007). The role of coral-associated microorganisms in corals' health is further supported by the fact that shifts in population composition often correlate with the onset of diseases and bleaching (Krediet et al., 2013).

2.4.1 Coral reef-associated microorganisms

Coral reefs are some of the most productive ecosystems on Earth, with primary production rates likened to that of rain forests. The tropical coral reefs are a complex ecosystem consisting of a wide assortment of animals, plants, microorganisms, and viruses (Wegley et al, 2007). While most of the corals' energy reserves is supplied through a symbiotic relationship with the photosynthetic microalgae of the genus *Symbiodiniaceae* (zooxanthellae), other coral-associated microorganisms are known to provide important contributions to the holobiont functioning (Godoy-Vitorino et al., 2017; Rosenberg et al., 2007). Overall, reef productivity is significantly dependent on the capture and recycling of nutrients by reef- associated bacterial communities (Bourne & Webster, 2013).

Defining a coral reef and the associated microbial communities is complex because coral reefs comprise not only of the reef-building corals that form the complex substrata, but also of multiple species that use these structures for habitat and foraging (Bourne & Webster, 2013). As such, coral reef bacterial communities occupy a range of different habitats including the sediment, overlying water column, and benthic invertebrates. As a basis for understanding coral reef bacterial processes, studies have mainly focused on pelagic and sediment bacteria which are believed to drive water column and benthic processes of reef systems (Bourne & Webster, 2013; Glasl et al., 2019; Linda W. Kelly et al., 2014; McDole et al., 2012; Tout et al., 2014; Walsh et al., 2017).

2.4.1.1 Coral reef pelagic microorganisms

The support of functioning of a healthy coral reef depends on the facilitation of microorganisms to capture and recycle nutrients in oligotrophic waters (Garren & Azam, 2012). The column water overlaying coral reef hosts microbial taxa and genomic adaptations distinct to off-reef waters suggesting that unique microbial processes occur on coral reefs (Kelly et al., 2014; Weber et al., 2020). In the tropical coral reefs, this water comprises of a mixture of oligotrophic

offshore and reef waters with the organic carbon and nitrogen substrates exuded by the benthic community which define the composition, structure, and function of the microbial communities in the water column (Kelly et al., 2019). It has been demonstrated, for instance, that in contrast to open water and sandy substrate niches that are dominated by oligotrophs and metabolic potential for core house-keeping processes e.g. amino acid, carbohydrate and protein metabolism, coral surface seawater was dominated by copiotrophic bacteria with enhanced frequency of genes associated with dynamic processes including motility and chemotaxis, regulation and cell signalling (Tout et al., 2014).

Corals influence the growth and composition of reef microorganisms by forming momentum boundary layers surrounding individual colonies and with their metabolic products (Weber et al., 2019). Corals enrich the overlaying water by releasing with mucus large quantities of dissolved organic carbon and dissolved organic nitrogen which is degraded by- or serves as chemical cues for bacteria (Garren & Azam, 2012). This results to higher bacterial growth rates (50 times faster) and oxygen concentrations in the reef- than in the open ocean communities (Garren & Azam, 2012; Silveira et al., 2017). Some coral species also feed on bacterioplankton directly as a source of phosphorus, nitrogen and carbon (Garren & Azam, 2012). Human impacts, especially overfishing and pollution, may also cause reef trophic structure that favour microbial growth (Weber et al., 2020). Subsequent interactions within the reef may influence microbial symbiont acquisition and pathogen recruitment to the coral surface (Weber et al., 2019).

2.4.1.2 Coral reef sediment microorganisms

Marine sediments cover over two-thirds of the Earth's surface and are permeated with microbial life, as far as temperature gradients and porosity permit (Teske, 2013). Sediment matrices are believed to function as biocatalytic filter systems of efficient pelagic-benthic coupling as they retain, in the upper layer, dissolved and suspended matter metabolized by the associated bacteria which is transported between sediments and the overlaying water column (Schöttner et al., 2011).

Coral reef sediments have been shown to contain microbial assemblages distinct from the water column, nearshore and outer-shelf locations, and which has the highest diversity in the top 3 cm (Bourne & Webster, 2013; Hewson & Fuhrman, 2006). Due to their versatile metabolic capacities, coral reef sediment bacteria are involved in numerous pelagic and benthic processes.

Those inhabiting the reef frameworks and sandy sediments undertake key role in the remineralization of organic matter directly within the reef system (Rusch et al., 2009). This they do especially through nitrogen cycling and the metabolism of organic matter which are subsequently exported to the surrounding reef ecosystem, which is particularly crucial for coral reefs to maintain high levels of primary production and biomass in the oligotrophic seawater (Schöttner et al., 2011). Indeed, it has been estimated that over 25% of reef primary production is supported by algae within the coral rubble and 70% of nitrogen fixed within reef sediments is exported to the surrounding reef (Bourne & Webster, 2013).

2.5 Status of the WIO marine ecosystems

The WIO is one the regional seas identified by United Nations Environment Programme (UNEP). Its surface covers about 8 per cent of global oceans, and over 15,000 km of combined coastline. The region comprises 10 countries which include: five mainland states spanning Somali to South Africa, Madagascar, and four small island states beyond Madagascar. Because of the broad variety of biogenic habitats and oceanographic conditions, the region's waters encompasses a diverse range of ecologically and nutritionally rich ecosystems (Van der Elst et al., 2005). With its goods and services supporting over a quarter (more than 60 million people) of the region's population who live within 100 km of the shoreline (Obura, 2017; Van der Elst et al., 2005), WIO and coastal ecosystems is essential to the region's economy. A recent report estimated that the annual gross marine product of the WIO is more than US\$20 billion, and the total "ocean asset base" is more than US\$330 billion (Obura, 2017). These estimates are conservative and did not account for the invaluable indirect contributions of the ocean such as climate regulation, oxygen production, intrinsic value of biodiversity, subsistence fishing or its spiritual and cultural value. Therefore, WIO's real economic value is much more. Although the ecosystems driving its biodiversity and productivity have been relatively intact over time (Gudka et al., 2018; Obura, 2017; Obura et al., 2017), it is feared that with the rapidly increasing development and utilisation of coastal zones (Neumann et al., 2015), the WIO's ability is overwhelmed, overexploited and poorly managed, putting its wealth at risk. Indications of distress on the region's ecosystems are increasingly noted, over time, characterized by reducing fish catches, diminishing mangrove coverage, and declining coral reefs (Obura, 2017; Van Der Elst et al., 2005).

Individual members of different marine ecosystems are thought to be interconnected, partly through interactions with microorganisms, forming a complex that has been referred to as

“nested ecosystems” (Vanwonderghem & Webster, 2020). Interactions of constituent members can shape the overall functioning and fitness of holobionts nested within communities of neighbouring holobionts that, in turn, interact with and influence successively larger and more complex communities and marine environments (Pita et al., 2018). The concept of nested ecosystems implies that the actions of individual constituents of the coral microbiome can have effects far beyond the holobiont and may also be influenced by members of other distant ecosystems.

2.5.1 Mangrove forests

Mangrove forests occupy the largest areas of all coastal habitats in the WIO. They mostly occur in river deltas or estuaries and cover about 1 million ha, which is 5.2% of the world’s total mangrove forest cover, comprising of 10 of the 69 known true mangrove species worldwide (Samoilys et al., 2015). These are widely distributed in the region such that Mozambique has all of the 10 species, while nine occur in both Kenya and Tanzania. Mangrove trees are resistant to salinity, so they grow in brackish water and seawater from which they extract freshwater.

Mangrove forests are extremely productive ecosystems that provide numerous goods and services to both the marine environment and people. The economic value of the mangroves’ main goods and services, estimated from a 12-year plantation in Kenya, is about US\$ 2,900/ha/yr (Kairo et al., 2009). Firewood and building poles, coastal protection, ecotourism, research and education, carbon sequestration and fisheries were the major goods and services identified. A WIO-wide study conservatively estimated the overall asset value of mangroves in the region to be US\$ 42.7 billion, which was higher than coral reefs and seagrasses at US\$ 18.1 billion and US\$ 20.8 billion, respectively (Obura, 2017).

Mangrove forests are habitats to a wide variety of fish, crab, shrimp, and mollusc species which form an essential source of food for thousands of coastal communities. They are also nurseries for many fish species, including coral reef fish. Mangroves have been shown to enhance the biomass of some fish species on reefs close to mangroves up to 25 times higher compared to mangrove-scarce systems (Mumby et al., 2004) compared to where mangroves have been cut down. The forests also support a variety of shrubs and palms which are important to other animals. One such plant is the climbing legume *Derris trifoliata* fed on by elephants in Kenya from the mangrove forests (Samoilys et al., 2015).

Mangrove trees play critical role in stabilizing the coastline and preventing erosion from waves and storms with their dense root systems which trap sediments in river discharge that would otherwise enter the sea and potentially affect the productivity of the seagrass beds and coral reefs. Mangroves allocate proportionally more carbon belowground than terrestrial trees. Most mangrove carbon is stored as large pools in soil and dead roots. They are thus valued as carbon sinks that mitigate the effects of climate change through their high productivity and by trapping carbon in biomass and sediment. It is estimated that they account for 14% of the carbon sequestration by the global ocean (Alongi, 2012).

Mangrove wood is sturdy, resistant to rot and insects. This makes it the most valued wood among many coastal communities for construction and as fuel (firewood or charcoal). Medicinal plants are also collected by communities from the mangrove ecosystems and mangrove leaves are utilized as animal fodder.

Replanted and natural mangroves offer incalculable value as a tool for augmenting education and research to both researchers and students who use them as field laboratories. The interest to study varied aspects of mangroves and their associated ecosystems has steadily increased over the years and is projected to continue (Kairo et al., 2009).

Due to their associated diverse flora and fauna, mangrove ecosystems offer opportunities for tourists to enjoy bird watching and canoeing in mangrove areas. A conservative estimate of unit value of mangroves to ecotourism was US\$9.30/ha/year in Kenya in 2009 (Kairo et al., 2009). This was at the infancy of ecotourism and based on an area of 700 ha of natural and artificial mangrove plantations. With sustained structured promotion, and establishment of decent infrastructure to host and accommodate tourists, ecotourism has good potential to create substantial revenue for the region.

Worldwide, mangroves are under increasing pressure from anthropogenic disturbances. In 50 years, the mangrove forest cover of the tropical coasts declined from 75% to less than 50% worldwide (Bouillon et al., 2009). In the WIO region, an earlier estimate of cover loss for Mozambique, Tanzania and Kenya from 1980 to 2005 were 8% (Samoilys et al., 2015). However, more recent estimates showed increasing attrition rates with Kenya and Tanzania losing about 18% of their mangroves over 25 years, and Mozambique 27% over a shorter timeframe (Obura, 2017). This trend may be difficult to reverse considering that in the WIO

region mangroves are more commonly managed as forest reserves subject to regulations drawn up with a view to resource utilization, rather than biodiversity conservation (Samoilys et al., 2015).

Clearing of mangrove areas and conversion to other land uses is a major concern. On the Kenyan coast about 20% of the total mangrove forest was cleared to make way for salt manufacture, aquaculture, fuelwood and construction (Abuodha & Kairo, 2001). In addition, about 70% of the wood used for building and heating along the Kenyan coast is mangrove-sourced (Kairo et al., 2009) and significant mangrove deaths have also been observed on the Kenyan coast due to dredging and oil pollution.

Important potential threats to mangroves in the WIO region include oil and gas extraction, aquaculture and coastal development, particularly port construction. As oil and gas reserves are discovered in the region and mariculture expands to meet food and employment demands in the region, it is feared that mangrove forests may be damaged in the process if environmentally friendly approaches are not employed in their operations.

If not reversed, the decline of mangroves may lead to a cascade with dire consequences resulting from loss of their ecosystem services to the marine environment, fisheries and humans. There is therefore an urgent need for novel strategies to inform good management and maintain the health of mangrove forests.

2.5.2 *Seagrass beds*

Seagrasses are marine-adapted flowering plants covering 0.1–0.2% of the global ocean floor where they create some of the most productive aquatic ecosystems on earth, hosting and supporting a variety of marine organisms. The WIO region is one of the world's most species-rich areas for seagrass as it is home to 13 of the 60 species known worldwide (Gullström et al., 2002). They occur throughout the WIO coastline growing at a maximum depth of 40 meters due to the need for light for photosynthesis. In places other than the WIO, seagrasses have been known to grow at depths of up to 70 meters in water of exceptional clarity (Samoilys et al., 2015).

Compared to coral reefs and mangrove forests, less management and research efforts have been directed towards seagrass meadows despite their economic contribution in the WIO region and

the connectedness of the habitats. Although recently seagrasses have become one of the most discussed and noticeable ecological systems globally, there is still paucity of published research on the matter especially from the WIO region. The available studies from the WIO are localised, focussing on varied country-specific themes rather than the broad region. Nonetheless, the socio-economic significance of seagrass ecosystems in the region can be approximated from the state-limited observations as well as from what is known globally.

Seagrass beds are often used by both finfish and shellfish species for foraging and shelter. Notably, they are habitats for vulnerable dugongs and turtles. The significance of seagrass as habitats and food source has been established by several studies. For instance, the spatial patterns and variability of fish population composition has been demonstrated to be determined by seagrass structure and location (Gullström et al., 2008). Also, de la Torre-castro et al., (2008) found that the food contents of commercial fishes constituted seagrass-associated items, underscoring the importance of seagrass in food provisioning services. Consequently, seagrass ecosystems' economic contribution results primary from the essential role they serve in fisheries production. Seagrass fisheries have been shown to be the primary source of animal protein and the highest income earner for the local people in a rural tropical economy on the East Coast of Zanzibar (de la Torre-Castro & Rönnbäck, 2004). Moreover, seagrass habitats have been shown to be the most visited by small-scale fisheries, and provide more stable catches compared to coral reef and mangrove habitats (de la Torre-Castro et al., 2014). Up to 80% of prawn trawl catches in Tanga (Tanzania) have been known to consist of seagrass fish species (Samoilys et al., 2015).

Seagrass ecosystems further earn the WIO region foreign income from fish species and shrimps' exports products. They also serve as substrate for seaweed cultivation which is exported and used in cosmetic, pharmaceutical and food industries. Despite no regional data on the size and importance of seagrass fisheries, it is widely believed that this industry in the WIO region is substantial (Gullström et al., 2002). A recent report estimated the asset value of WIO seagrass meadows at US\$20.8 billion compared to US\$18.1 billion and US\$42.7 billion for mangroves and coral reefs (Obura, 2017).

Seagrasses provide important ecological functions by altering the ecosystems around them physically and chemically. Seagrass meadows mitigate climate change – they are considered key blue carbon sinks (Gullström et al., 2018) and are believed to store up about 10% of the

ocean's total carbon (Macreadie et al., 2014). They also help improve water quality by filtering contaminants and alleviate turbidity by trapping sediments with their root system. This stabilizing of sediments by seagrass root systems, also protect the shoreline from erosion. Following photosynthesis, seagrasses produce oxygen, which supports primary production necessary to marine ecosystems. These plants also have aesthetic and religious value and traditionally, some seagrass plants have been used for medicinal purposes. Seagrass beds are also used by fishermen and marine park tour-guides as navigation aids (Cullen-Unsworth et al., 2014; de la Torre-Castro & Rönnbäck, 2004)

Growing coastal populations and human disturbance are increasing pressure on seagrass beds in the region. As a result, the last few decades have been marked by widespread seagrass loss (Jahnke et al., 2019) due to varied direct human impacts, as well as from natural causes, such as cyclones and floods (Duarte, 2002). The greatest threat to WIO seagrass beds is posed by fishing activities. Trawling and seine netting are widespread along the coasts of WIO countries. These activities cause physical disturbance and destruction of seagrasses and may alter food webs in the habitat at the expense of the seagrasses. For instance, elimination of natural predators has caused seagrass overgrazing by sea urchins (Eklöf et al., 2009). Mechanical damage of seagrasses has also been noted as a result of dredging and anchoring, common in Kenya and Tanzania (Samoilys et al., 2015), as well as physical removal of the seagrasses in Mauritius (Daby, 2003) for what hoteliers considered as 'aesthetic' and 'safety' reasons. Farming activities have also been linked to reduced seagrass. Seagrass declines have been noted in estuarine and deltaic areas due to farming in the river basins of the Athi-Sabaki and Tana Rivers in Kenya and the Rufiji River in Tanzania (Samoilys et al., 2015). Moreover, seaweed farming in Zanzibar was shown to impact seagrass beds negatively (Eklöf et al., 2005). It is feared that coastal developments, such as ports construction in Lamu (Kenya) and Mtwara (Tanzania), will affect large tracts of seagrass which are important turtle and dugong feeding grounds.

Despite the recognition of the importance of WIO seagrass beds as some of the most productive aquatic ecosystems with an unclear future, not enough attention has been paid to their conservation and environmental management.

2.5.3 Corals reefs

The WIO is home to some 16% of the world's coral reefs which makes the region the second largest hotspot of coral reef biodiversity globally. Throughout the WIO region corals occur between the depths of 0 and 50 meters, where they meet their energetic requirements through endosymbiotic associations with photosynthetic dinoflagellates algae, which also enable them to contribute to high benthic autotrophic productivity (Medina, 2011; Obura, 2016). Four classes of coral reefs have been identified within the WIO region: fringing reefs, barrier reefs, atolls and coral banks (Obura, 2016). No comprehensive representative coral diversity study has yet been conducted in the region, but Obura (2012) identified 369 species of hard corals from 21 locations within the WIO. The exact potential pool of species in the region is estimated to be at least 450 (Obura, 2016; Samoilys et al., 2015).

Coral reefs are the most biologically diverse of shallow water marine ecosystems (Roberts et al., 2016). Covering only 0.1% of the ocean floor, coral reefs host at least 25% of the marine biodiversity (Hoegh-guldberg et al., 2017), providing ecosystem goods and services to a substantial number of people (Moberg & Folke, 1999). Coral reefs are custodians of significant ecological and socio-economic value (Obura et al., 2017) and are responsible, to a great extent, for the ocean-associated benefits relied upon by humans (Hughes et al., 2003). Coral reefs are habitats, spawning and nursery grounds for economically important fish species that form an essential source of food, protein and livelihood for coastal communities. The rough surfaces of coral reefs help dissipate wave energy from the ocean hence protecting coastlines from storms and erosion (Ruckelshaus et al., 2013). Indirectly, coral reefs provide jobs and income from fishing, recreation, and tourism (Halpern et al., 2012), and are potential sources of novel pharmaceuticals (Bruckner, 2002; Obura et al., 2017).

Coral reefs are considered some of the most threatened ecosystems on earth, facing increasing degradation due to the local and global impacts of human activities, the net effect of which has been rapid decline of coral reef habitats worldwide (Walsh et al., 2017). Population growth and coastal development that tends toward overexploitation, nutrient loading, and declining water quality are believed to be the main anthropogenic pressures responsible for localised degradation, as well as disease and outbreaks of coral predating crown-of-thorns starfish in some regions. At both local and global scales, climate change is the primary stressor (Glasl et al., 2019; Hughes et al., 2017; Mora, 2008; Neumann et al., 2015). Human-source nitrogen

enrichment has also been demonstrated, by a 30-years study, to play a role in coral bleaching and death, even in the absence of rising water temperatures (Lapointe et al., 2019). This observation underscores the contribution of anthropogenic activities, and the potential of local solutions in addressing coral reef degradation.

WIO coral reefs have been repeatedly impacted by heat waves in the last 20 years (Gardner et al., 2019). The unprecedented devastating bleaching event of 1998 led to decline of WIO coral reef cover by 25%, giving way to a 2.5-times increase in fleshy algae cover. A further 20% coral cover decline was experienced in the region's reefs during the 3rd global bleaching event of 2016. Following these episodes, up to two thirds of the affected corals recovered despite fears that increase in algal cover could potentially hamper coral recruitment and recovery (Gudka et al., 2018; Obura et al., 2017).

The bleaching episodes have also had an impact on the region's fish community which has shifted from the more balanced and diverse predator-dominated state to domination by smaller-bodied herbivores and detritivores (Obura et al., 2017). The shift may have been due to increase in algae, which selectively support herbivorous fish. Sustained loss of coral coverage and the associated biodiversity is likely to expose coastal dwellers to the risk of diminished fish catch and reduced protection from coastal hazards e.g. storms and rising seas.

The main strategy promoted by the United Nations Environment Programme (UNEP) for protecting coral reefs from the impacts of climate change and other human activities is integrated ecosystem-based management approaches guided by sound science. The most widely used of these approaches for coral reef conservation and management is marine protected areas (MPAs), including no-take fishing reserves (Bridge et al., 2013). The importance of limiting human impact on the marine ecosystems to save coral reefs was recently reiterated by a study that recommended prioritization for protection and management of reefs that survived the 2014-17 El Niño thermal stress (Darling et al., 2019). While establishment of MPAs has led to increased and improved fish populations (Ransom & Mangi, 2010; Samoilys & Obura, 2011), they have not been as effective in coral reefs protection due to poor management and enforcement, as well as vulnerability to indirect human activities such as sedimentation, pollution and coastal development (Mora et al., 2006).

2.6 Microbial genomics-related approaches appropriate for the WIO region

There are three key noticeable features common to the WIO ecosystems from the foregoing status review: i) they are habitats to immense biodiversity that provide invaluable goods and services, with potential for incalculable ecological and economic benefits to the region's economies and the marine ecosystem; ii) they are threatened by direct and indirect anthropogenic stressors, and they need efficient conservation strategies are put in place; and iii) they are disproportionately understudied, especially with regard to their microbial diversity. Microbial genomics studies, utilizing the latest nucleic acid sequencing technologies and bioinformatics analysis pipelines, offer a broad range of possibilities for generating information that could: increase our understanding of human activities' impact on the WIO and its ecosystems' responses, support creation of novel environmental health monitoring, conservation and management strategies, and help to point to new priority areas for research.

Marine ecosystems cover approximately 70% of the Earth's surface. About 90% of all marine biomass is microbial (Alvarez-Yela et al., 2019). Understanding microbial communities is therefore essential to understanding their natural habitats. Limited understanding of microorganisms, despite their ubiquitousness, fundamentally limits our understanding of the ocean (and by extension the Earth's biosphere) and its response to environmental perturbations, thereby jeopardizing efforts to create effective solutions. The need to study microorganisms to understand and address the ever-increasing effects of global climate change cannot be overstated. The urgency is captured in a consensus statement published by the alliance of world scientists as a 'warning to humanity' on the peril of failing to acknowledge the central role and global importance of microorganisms in climate change biology (Cavicchioli et al., 2019).

In the past, studying marine microorganisms was hampered by the fact that more than 99% of microorganisms are not cultivatable by standard laboratory methods. The past three decades have seen an explosion of advances in technologies for sequence-based methods that have revolutionized the efficiency of surveying marine microbial biodiversity. Limited appreciation of genetics-based studies in the WIO region has been noted before (Ridgway & Sampayo, 2005). Over time, the number of genetics studies in the region have increased and primarily focused on macroscopic organisms such as fish, molluscs, crabs, lobsters, corals, seagrasses and bats (Chan et al., 2011; Gagnaire et al., 2007; Gopal et al., 2006; Jahnke et al., 2019; Macdonald et al., 2011; Mzingirwa et al., 2019; O'Brien et al., 2009; Ragionieri et al., 2010; Visram et al., 2010), due to their economic or ecological importance. Of the sequence-based

published studies seen so far, only two were on microorganisms (Gardner et al., 2019; Séré et al., 2013). These studies do indicate that researchers in some WIO countries have access to modern sequence-based methods, but these have not been sufficiently utilized to study marine microorganisms. Accessible methods that are available to some of the region's scientists with potential applications in marine microbiology include metabarcoding and metagenomics. These and complementary methods are described below along with their potential applications in studying the WIO marine ecosystems.

2.6.1 Metabarcoding

Barcoding is a method of species identification using a short sequence of DNA, termed “barcode” or marker, from a specific gene. Suitable candidates for barcodes are conserved gene regions with less intraspecific variation than interspecific variation. In other words, these regions ought to be similar enough for a given taxon to be routinely identified (e.g. prokaryotes) yet varied enough to distinguish between members of that taxon (e.g. bacteria species). Selection of an informative DNA region therefore is essential for successful characterization of species based on DNA barcodes. The conserved sites may be used to design universal PCR primers for wide taxonomic application. Different gene regions are used to identify different organismal groups. Microorganisms, for instance, are widely detected using the 16S rRNA gene, for prokaryotes, and the 18S rRNA gene, for microbial eukaryotes.

The advent of next-generation sequencing (NGS) platforms endowed with massively parallel processing capabilities has made it possible to sequence and identify multiple samples simultaneously. The high-throughput sequencing of taxon identifiers has led to the emergence of DNA metabarcoding – rapid barcoding that allows simultaneous identification of many taxa within the same sample, often within the same organism group. The standard metabarcoding procedure involves community or environmental DNA (eDNA) isolation from samples e.g. soil, water, faeces, etc. The eDNA undergoes amplification by polymerase chain reaction (PCR) to enrich for the appropriate genetic marker or barcode that allows for downstream identification. Sequencing on a suitable NGS platform follows, and finally bioinformatics analysis to determine community composition of the sample used.

Metabarcoding of environmental samples is advantageous to conservation as sampling is non-invasive hence it does not damage the environment or the organisms (especially macroorganisms) as do many traditional ecological studies (Deiner et al., 2017). Environmental

DNA sampling is also a convenient and efficient approach to ecological studies as it can detect organisms, especially macroorganisms, without needing to capture them. For microbiological studies, metabarcoding transcends the limitations of culture-dependant methods that require development of suitable culture conditions to grow microorganisms before identifying them.

Determination of microbial species composition involved in ecological processes is the primary application for metabarcoding. This capability enables analysis of diverse elements of the microbial communities that would benefit microbial studies in the WIO region. The analysis of 16S rRNA gene sequences directly from environmental samples will facilitate estimations of microbial diversity, which is the central property in community ecology. Two aspects of diversity inventory are possible with this approach: taxon richness and the relative abundance of different taxa in the community (Konopka, 2009). Spatial-temporal establishment of these metrics for key habitats of the WIO would be valuable for potential novel conservation applications. Key to these applications is monitoring and assessing environmental ecosystem and biodiversity status, either for specific species groups, or for the entire ecosystem (Ruppert, Kline, & Rahman, 2019). Microbial communities are known to respond and adapt quickly to perturbations (Ainsworth et al., 2010). Therefore, microbial population structures determined in healthy marine environments may be used as an indicator for health while deviations from the baseline healthy community composition may be used to detect or predict impacts on the ecosystem. Some recent studies have demonstrated microbial community compositions correlate well with the impacts of human activities on the ocean (Glasl et al., 2019; Li et al., 2018), confirming the potential to use microbial population structures as a barometer for marine environment health by acting as surrogates for anthropogenic impacts. As discussed already, microorganisms tend to interact with a remarkable specificity with their habitats, both host organisms and physical environments. It has been shown that the taxa and functions of microorganisms in the water column are influenced by the closest dominant benthic macroorganisms forming what has been termed an “aura-biome” (Walsh et al., 2017). Therefore, microbial metabarcoding has considerable potential for use in the assessment of a broad range of ecosystem functionalities.

In cases where microbial identifications can be resolved to the species level, eDNA metabarcoding is useful to detect individual microorganisms with known functions. Numerous marine microorganisms have been identified, some with roles in coral disease (Krediet et al., 2013; Sawabe et al., 1998; Séré et al., 2013), in human disease (Abdelzaher et al., 2010), and

others with properties such as antibiotic, anticancer, antioxidant, antifungal, anti-HIV, antituberculosis, and antimalaria (Isnansetyo & Kamei, 2003, 2009; Synnes, 2007), among other discoveries. Metabarcoding can therefore be used as an efficient tool for surveillance of harmful microbes, for instance pathogenic ones, and for bioprospecting for the beneficial ones such as those with known biotechnological applications.

Metabarcoding studies also serve an essential role as precursor to more intensive genomic studies. Some approaches of direct genetic analysis such as metagenomics (discussed below), which uncovers complete microbial genomes from an environmental sample, are resource intensive. These methods also require sufficient quantities of high-quality nucleic acid to ensure confidence in results which is often difficult especially with marine ecosystem samples. The quality and quantity of nucleic acids is influenced by many factors including sample origin, state and transportation from sampling sites, to primers and library preparation biases (Deiner et al., 2017; Thomas et al., 2012). Metabarcoding assessments are less expensive and recommended prior to metagenomic studies (National Research Council, 2007) to provide insight into the microbial diversity expected and guidance on the extent and depth of metagenomic sequencing required.

2.6.2 Metagenomics

Although metabarcoding of an environmental sample elucidates the taxonomic richness and evolutionary history over time, it is unable to clarify the functional diversity of the microbial populations under study (Nogales et al., 2011). Often metagenomics, which entails sequencing of the collective genes of all organism communities in an environmental sample, is needed (Bourlat et al., 2013). Like in metabarcoding, the procedure involves isolation of eDNA from environmental samples which then undergoes deep whole metagenome sequencing (WMS) by NGS technologies followed by bioinformatics analyses.

Because metagenomics uncovers the complimentary genes of a community, the approach gives insight into population composition and the potential ecological functions of the community members (Bäckhed et al., 2012). Although this strategy is useful and could identify novel metabolic pathways and products, it does not explain the exact function of a microbial community; it catalogues total rather than expressed genes. It therefore can be complemented by such strategies as metatranscriptomics and metaproteomics which profiles the community expressed genes or proteins respectively (Titilade & Olalekan, 2015).

The collection of genes uncovered from microbial communities with metagenomics protocols allow assembly of new microbial genomes without the need for culture. Such assemblies are termed metagenomic assembled genomes (MAGs) and can be used to relate taxa with function; perform pangenome analysis of the genera from a sample; and discover new genes. One common challenge with taxonomic analysis of microbial communities is that microbial sequences from environmental samples lack taxonomically related sequences in existing reference databases (Mande et al., 2012). This is likely to be problematic for the initial sampling from the WIO assuming there will be new microbial populations that have not been characterized before. This problem will ultimately be addressed through subsequent posting of the novel MAGs to open access databases thereby helping reduce database bias.

Metagenomics is a powerful tool for generating novel hypotheses of microbial functions by identifying microbial composition structuring by habitats, and by referencing genes with already characterized functions. In the WIO region, seagrass species influence microbial diversity (Uku et al., 2007) - an association that is believed to help broaden the ecological role of seagrass (Hassenrück et al., 2015). Metagenomics analyses will confirm such associations and further suggest potential microbial functions. Also, it has been proposal to use growth-enhancing bacteria to enhance reforestation with mangrove seedlings (Bashan & Holguin, 2002). Metagenomics analysis of the productive microbial biodiversity that transforms dead mangrove vegetation into sources of nitrogen, phosphorous and other nutrients (Sahoo & Dhal, 2009) will help specifically identify and isolate the most efficient microorganisms suitable for use with WIO mangroves.

Sequence similarity with known genes is used to predict function in newly sequenced organisms. Bacteria-derived rhodopsin phototrophy (Beja et al., 2000) and ammonia-oxidizing archaea (Nicol & Schleper, 2006) were discovered in this way. Comparative metagenomics has further revealed that microbial diversity is a function of the local environment (Behzad et al., 2016); unique environments typically hosts novel microbial species with unique genes and metabolic pathways. Microorganisms living in extreme environments must have cellular biomolecules including proteins, nucleic acids, and lipids, that are adapted to function under such conditions. They, for instance, produce enzymes that catalyse specific biochemical reactions under conditions that would denature conventional proteins. These are of interest to the industry for applications like detergent production, sugar, lipid and oil chemistry, and food

processing. This potential has been recognized by the European Commission (EC) which, along with the European Science Foundation (ESF), consequently launched programmes to support and stimulate research on marine diversity and genomics (Glöckner & Joint, 2010). Biodiversity hotspots in the WIO region suggest that the area is rich in marine natural products. Only a small fraction of this biodiversity has been explored for its commercial potential (Wynberg, 2016). The large metagenomic microbial community datasets generated will offer considerable potential to champion novel gene and bioproducts biodiscovery from the unique niches of the WIO marine ecosystems.

Microorganisms' quick response and adaptability can be exploited by metagenomics approaches to not only understand and predict responses of marine ecosystem to disturbances, but also to discover microorganisms or processes that may be used to address environmental contaminants. Assessments of contaminated environments may lead to isolation of microorganisms or metabolic pathways capable of degrading the contaminant which may be used as a bioremediation or bioaugmentation agent for the contaminant. For instance, salt marsh grass (*Spartina alterniflora*) showed high resilience to the impacts of the 2010 Deepwater Horizon oil spill, because following the spill the microbial populations shifted to oil-degrading bacteria dominance (Kandalepas et al., 2015). This has led to oil-spill bioremediation trials with the identified bacteria.

Finally, since metagenomics assessments uncover whole genome sequences of microbial community members, this discipline offers the possibility to identify population compositions that help to describe and predict environmental changes, as well as for barcoding i.e. to design genetic markers that can subsequently be detected by more rapid, inexpensive and less rigorous approaches such as PCR.

2.6.4 PCR-based microbial markers

Genomics approaches are useful to study marine microbial structure and functional diversity and identify patterns that may be used to monitor environmental health and predict responses to calamities. Inexpensive tools for rapid and efficient evaluation are urgently needed, especially considering the intensity of human activities is expected to increase, to help discriminate stressors and prioritize conservation strategies. Genomics methods are not suitable for frequent use – as may be needed for monitoring – because they are neither cheap nor rapid. However, genomics studies provide the needed genetic information to enable the development

of PCR-based tests that are fast and cheap in comparison to sequence-based methods. Once an indicator microorganism or its gene is characterised through metabarcoding or metagenomics assessments, a signature sequence (or barcode) can be determined and targeted for PCR primers. Depending on the level of information required for meaningful interpretation to guide interventions, either endpoint or quantitative PCR may be designed. An example of such an emerging PCR-based test is microbial source tracking (MST) PCR, which identifies the source of pollution, especially in aquatic environments. The test employs primers to detect and quantify markers of members of *Bacteroidales*, which are abundant in the intestines and faeces of warm blooded animals (Odagiri et al., 2015). *Bacteroidales* populations adapt uniquely to different hosts. Therefore their hosts can be traced as the origin of faecal contamination using *Bacteroidales*-host-associated specific genetic markers (Bernhard & Field, 2000).

Following on from this example, if microbial associations with the ecosystem status of interest can be determined conclusively at the genomic level, it may be possible to identify associated genetic markers for which PCR primers would be designed to facilitate rapid and inexpensive testing. Some of the WIO marine ecosystem characteristics that can be targeted for this type of assessment include coral health – bleaching and resilience; declining fisheries catches; and the increasing incidence of turtle diseases e.g. fibropapillomatosis - the exact aetiology of which is not yet known, but is known to correlate strongly with anthropogenic activities that create high-nitrogen footprints in the surrounding environment (van Houtan et al., 2010).

2.7 Strategies to enhance uptake of genomic methods in the WIO region

Application of genetics and genomics methods in research is generally limited in Africa. The main reasons commonly advanced for this shortage include lack of technological capacity and shortage of local experts. In the last ten years, significant level of genomics capacity has become available - both infrastructure and experts - in a few of the WIO countries, particularly South Africa and Kenya. There are also spirited efforts to build genomics research capacity in African countries through training, funding for research as well as infrastructure, and establishment of networks (Dandara et al., 2014; Karikari et al, 2015; Mboowa & Sserwadda, 2019). These initiatives do not seem to have had as much impact in marine research as they have had in biomedical and agricultural research more than marine research, even though genomics applies to all life sciences. The Web of Science (2010 - 2020) records only 752 papers published in Africa employing genomics in marine research compared to 3,727 papers in biomedicine and 4,633 in agriculture. With specific respect to coral reef microbes, only four (4) papers were

published, in the last decade, that employed genomics in the WIO, compared to 17 in the Caribbean Sea and 114 in the Great Barrier Reef (GBR). Of the papers on the WIO reefs, only one (1) had the corresponding author with an African affiliation.

Perhaps lack of appreciation due to low literacy in genetics and genomics sciences is also an impediment to uptake by marine researchers. The WIO region's universities do offer undergraduate degree programs in marine-related programs. An informal survey was done to compare the level of genetics and microbiology training in marine-related degree programs (e.g. marine, aquaculture, aquatic, and fisheries) with other biological science programs (e.g. crop science, animal science, agriculture, and botany) that produce potential researchers in agriculture and biomedical (e.g. biomedical science, biology, and zoology) research, in the WIO universities. Course content details were obtained mainly from former students' transcripts or, where possible, from respective universities websites and university responses to questionnaire for 34 programs in 11 universities across 6 WIO states. Overall, genetics and microbiology training in marine-related programs is limited compared to the other biological science degree programs. Out of an average of 47 modules required for degree qualification, marine-related programs offered an average of 1 (range: 0-2) module for both genetics and microbiology, equivalent to 35-45 notional hours. In contrast, agriculture-related programs had an average of 3 (range: 2-4) genetics and 2 (range: 1-3) microbiology modules, while biomedical-related programs offered an average of 5 modules for both genetics (range: 2-6) and microbiology (range: 3-6). Furthermore, in the marine-related programs, the genetics and microbiology subjects were introductory in nature offered at the beginning of the programs compared to the agriculture- and biomedical-related programs where the said modules build up progressively during the programs' period. Also, genetics and microbiology courses are delivered with a bias towards biomedicine; because genomics and microbiology are mainly exploited in biomedical research and crop sciences, instructors of the courses are either drawn from biomedical research fraternity or were themselves instructed with biomedical bias. Graduates from the marine-related programs, therefore, are not adequately equipped to make use of genomics applications or to pursue further training in the related disciplines.

A long-term strategy to increase appreciation and applications in marine biology research is the revision of currently offered curricula to build competence and confidence in genomic sciences and microbiology. One way the WIO region universities may achieve this is to prioritize undergraduate programs that provide broad foundation in biology and allow specializations, at

that level, only fields with cross-cutting applications in life sciences such as molecular and genomic sciences. Training in less broad fields including marine sciences may then be offered at postgraduate level with reinforcement of the undergraduate cross-cutting knowledge and the option to utilize such knowledge, say in molecular sciences, for research projects. This might help promote genomic approaches as tools for studying life sciences and not a specialization with medical bias, as currently perceived. This may help mitigate shortage of local experts, albeit in the long run.

Collaborations between marine biology and genomics scientists is the most promising short-term strategy to ensure marine biology challenges are addressed through genomics applications. Collaborations would also ensure that resource-poor countries have access to advanced genomics infrastructure and expertise. Marine biologists are in a better position to initiate these collaborations since they are on the ground, appreciate the biology and are familiar with knowledge gaps and challenges that could be addressed by molecular methods. Marine biologists would particularly play a critical role in generating the environmental metadata needed to interpret genomics data. Collaborations of this nature could also be used as basis for mentoring future scientists in the field with a more holistic worldview of life sciences.

The costs associated with acquiring sequencing facilities and to archive sequenced data is often a prohibitory challenge in developing countries. Lately, it has become exceedingly easy and inexpensive to outsource sequencing services. Also, sequencing technologies evolve very fast and it is more logical to outsource to modern sequencing services rather than acquire a platform that becomes obsolete in a short period of time.

2.8 Conclusion

Mangrove forests, seagrasses and coral reefs are among the most biodiverse ecosystems of the WIO and play enormous ecological and economical roles in the region. The ocean has immense potential, owing to its highly dynamic tropical waters, currently under threat from rapid warming and increasing anthropogenic stressors with unforeseen consequences. Studying associated microbial diversity and genomics offer the opportunity to understand and predict the complex responses of the marine ecosystem to global changes. Microbial genomics also offers the possibilities of bioprospecting and biodiscovery of bioactive products of varied industrial and environmental interest, including potential applications in bioremediation and assisted evolution of endangered species. The developed world acknowledges the potential of marine

microbial genomics as evidenced through their proactive efforts in sponsoring initiatives that stimulate relevant research among their members. Despite the increased global interest in marine microorganisms for the pharmaceutical, cosmetic and food industries (Wynberg, 2016), the WIO remains understudied with regard to the distribution and occurrence of commercially interesting taxonomic groups (Díez et al., 2016). The main hinderance to the realisation of the WIO's potential include scant scientific and technical capacity as well as the lack of appreciation of the potential of molecular methods in marine biology research. These challenges may be surmounted by collaborations between marine and molecular scientists, as well as mentoring of well-rounded life scientists through delivery of holistic training curricula.

CHAPTER 3: WATER QUALITY IN KENYAN CORAL REEF ECOSYSTEMS ACROSS A COASTAL IMPACT GRADIENT

3.1 Abstract

The Kenyan coast hosts a rich diversity of coral reef ecosystem that provides essential source of food, protein and livelihood for coastal communities. As marine biodiversity is negatively impacted by climate change and human activities, the coral reef's untapped potential is increasingly under threat. Efficient strategies for detecting and safeguarding against negative impacts on coral reefs are needed to prevent or reverse decline in marine biodiversity. Marine protected areas (MPAs) have had success in protecting coral reefs and biodiversity by inhibiting overfishing. However, MPA's effectiveness against indirect impacts such as sedimentation, nutrient enrichment and pollution of coral reefs has not been specifically investigated. Measurements of physicochemical parameters and nutrient levels have been used as surrogates for human activities and have been shown to correlate well with proximity to agriculture and urban development activities. Also, faecal indicator bacteria (FIB) are used to detect human-source faecal pollution with molecular microbial source tracking approaches gaining popularity lately due to their ability to detect recent or ongoing faecal pollution and in identifying sources of the contamination. In this chapter, one time-point surveys of physicochemical and biological variables were conducted in the Kenyan coral reef water. To assess potential human influence in the coral reefs, nutrients and microbial indicators were determined in settings presenting a gradient of marine protection. Nutrient concentrations at the "no-take" coral reefs varied with the surrounding human activities, but between-site differences were not statistically significant (ANOVA $p = 0.068$ for nitrates, 0.552 for nitrites, 0.572 for ammonium, and 0.345 for phosphates). Overall, physicochemical characteristics were within the ranges considered suitable for coral growth in tropical oceans as well as previously determined measurements from Kenyan coral reefs. Faecal bacteria were not detected in the coral reef waters and *E. coli* density decreased with increasing marine protection. These findings suggest protective effect of MPAs against human-source pollution which need to be confirmed by future studies with robust sampling controlling for spatial-temporal variations.

Keywords: MPA, *E. coli*, nutrients, coral reefs, physicochemical, FIB, MST, PCR, pollution source identification

3.2 Introduction

Coral cover has been declining in recent decades due to environmental and climate change related stressors. Coral reefs are sensitive ecosystems – reef-building corals, particularly, have narrow tolerance limits with respect to environmental variables. Corals exist as holobionts, dependent on symbiotic associations with algae and prokaryotic microbes that contribute to host physiology (Medina, 2011). Stability of the physicochemical environment is key to corals resilience (Obura & Grimsditch, 2009) as variations may impact on the symbionts in turn undermining the corals physiology. Temperature, salinity, nutrients, and light are some of the environmental variables believed to be key to the fitness of tropical corals (Guan, Hohn, & Merico, 2015).

In equatorial oceans, coral reefs thrive in warm and oligotrophic conditions where they critically rely on sunlight (Roik et al., 2016). Since these waters have low concentrations of essential nutrients such as phosphate and nitrates, symbiotic algae benefit the coral through photosynthesis while the associated microorganisms help with nutrient recycling, homeostasis and immunity. High water temperatures disrupt the symbiotic relationships as thermal stress leads to the loss of symbionts often leading to coral bleaching. Indeed the mass bleaching of corals seen worldwide has mainly been attributed to global warming leading to increase in sea surface temperatures (Hughes et al., 2017).

Although rising sea surface temperature resulting from climate change has received the most attention, growing evidence show that other physicochemical factors are equally important. Salinity, nutrient concentrations, and aragonite saturation state, for instance, are believed to affect coral growth (Guan et al., 2015). Also, in a 30-years study, nitrogen enrichment from sewage, fertilizers and top soil, was shown to have caused coral bleaching and death before the effect of rising sea surface temperatures (Lapointe et al., 2019). The effects of human activities are apparent from observed gradients of some physicochemical properties associated with distances of reefs to the shore (Roik et al., 2016). Elevated input of nutrients, for instance, are often correlated with presence of humans close to coral reefs (D'Angelo & Wiedenmann, 2014). Anthropogenic nutrient enrichment leads to coral bleaching through phosphorus starvation of corals caused by elevated nitrogen levels, which reduces the corals' temperature threshold for bleaching. Determining locale-specific physicochemical properties of sites with thriving corals is therefore essential to better understand their biogeography. Such measurements can also help estimate the contribution of human activity to perturbations of marine ecosystems.

A potential strategy of defining drivers of coral reef stressors is identifying pollution sources. Indeed pollution source identification is the rationale behind the widespread testing for the presence of faecal indicator bacteria (FIB) e.g. faecal coliforms, *Escherichia coli*, and enterococci as surrogates for contamination with human faecal matter (Harwood et al., 2014). However, FIBs are no longer considered reliable as they have been shown to originate also from livestock, wildlife, urban runoff and even non-faecal sources (Boehm et al., 2013) – a limitation that renders their use inappropriate for pollution source identification. It has also been demonstrated that indicator bacteria include strains that naturally occur in soil, vegetation and water and that are capable of persisting and replicating in the environment (Ferguson & Signoretto, 2011). This further confounds FIBs’ usefulness to accurately resolve pollution sources, and detection does not indicate recent or ongoing pollution owing to their ability to persist outside the animal hosts. An alternative to testing for FIB is microbial source tracking (MST) where the source of faecal pollution is determined by exploiting the association of certain microorganisms with the gut and excreta of specific animal groups. Faecal bacteria of the order *Bacteroidales* are the most researched in this regard. They are considered particularly suitable candidates for MST applications because: they co-evolve with the host (Ahmed, Hughes, & Harwood, 2016) therefore ensuring specificity, they are highly abundant and prevalent in faeces (Okabe & Shimazu, 2007), their presence are better predictors of occurrence of waterborne pathogens than FIB (Schriewer et al., 2010), and detection methods for most markers are well established (Wuertz et al., 2011). Furthermore, being obligate anaerobes, members of the *Bacteroidales* are incapable of long-term survival or replication outside of their hosts. Their detection therefore usually indicates an ongoing or recent pollution. Qualitative and quantitative polymerase chain reaction (PCR) methods have been developed for MST markers meant to discriminate between human and non-human sources, with others designed to differentiate among animal sources (Boehm et al., 2013).

To prevent or reverse decline in marine biodiversity due to human disturbances, Marine Protected Areas (MPAs) are promoted globally and increasingly adopted as the principal conservation tool. In Kenya, MPAs are implemented by Kenya Wildlife Service (KWS) on behalf of the government to conserve marine biodiversity and promote tourism (Ngugi, 2002; Tuda & Omar, 2012). Owing MPAs’ success in restoring fish populations (Ransom & Mangi, 2010; Samoilys & Obura, 2011) locals also have adopted over 20 Locally Managed Marine Areas (LMMAs) responding to outcries over diminishing fish catches due to overfishing

(Obura, 2017). The protection strategy is designed to work by minimizing human activities in target marine areas. To date, Kenya has 6 national MPAs and 24 LMMAs which constitutes only 1.03% of her Exclusive Economic Zone (EEZ) (Obura, 2017). This is well below the Aichi Target of 10% MPA coverage required by 2020 to improve the status of biodiversity by safeguarding ecosystems, species and genetic diversity (SCBD, 2010). Furthermore, only few of these MPAs are effectively implemented.

Because MPAs are vulnerable to indirect human activities such as sedimentation, pollution and coastal development (Mora et al., 2006), there is a need to identify additional innovative strategies for protecting coral reefs from the impacts of human activities. A comprehensive understanding of the local causes of ecosystem decline is the foundation of conservation planning of coral reefs. However, marine ecosystems are usually under the influence of multiple anthropogenic stressors and the resulting responses by corals tend to be non-specific. There have been numerous efforts to prioritize drivers with the most impact on coral degradation. However, most research on human activities and their effects on coral reefs is selective limiting its applicability in local management (Freed & Granek, 2014). Furthermore, determining specific drivers of coral reef degradation has been challenging because one stressor can evoke different responses in different reef organisms (Mora, 2008) and findings from one site may not be reproducible or relevant to another site owing to the context-specificity of the drivers (Freed & Granek, 2014). Rather than determine the contribution of individual human activities, the next best option is to consider the cumulative impacts of activities from various sources over space and time (Halpern et al., 2009). Such an approach would entail mapping patterns of human activities along with the study of their associated effects on coral reef ecosystem over space and time.

Human activities cause degradation of coral reef ecosystem through physical destructions as well as impacting the chemical and biological quality of the surrounding reef waters (Balakrishnan et al., 2017). Physicochemical parameters such as temperature, pH and nutrients are reasonable indicators of the cumulative impacts of human activities on coral reefs. Characterization of baseline measurements in healthy coral reefs with limited human impacts is necessary for future referencing when deviations are suspected. But determining thresholds to protect coral reefs has been difficult because water quality is dynamic and shifts with many environmental factors and requires timeseries data for water quality and biological assemblages which is expensive (Houk et al., 2020). Instead many efforts have relied upon shorter-term

assessments along gradients that allow linking local changes in environmental conditions with coral reef ecosystem health (Cooper et al., 2009; Fabricius et al., 2012; Houk et al., 2020).

In this chapter, cross-sectional surveys of physicochemical characteristics of coral reef water in “no-take” marine national parks were conducted to assess between-site variations at the time of sampling for metagenomic sequencing for chapters 4 and 5. Indicators of human-source pollution were also determined across marine protection gradient to test human impact on the coral reefs.

3.3 Materials and methods

3.3.1 Study sites and human activities

This study sampled done in six coral reef sites with varying degrees of protection. Site selection was guided by the Kenya Marine and Fisheries Research Institute (KMFRI) ensuring inclusion of locations in proximity to the range of human activities typical in cosmopolitan urban and rural settlements along the coast of Kenyan. The sites are spread across three of the five counties on the Kenyan coast.

Three of the coral reefs were within marine national parks i.e. Malindi (3°15'35.1"S 40°08'40.0"E), Mombasa (3°59'45.7"S 39°44'50.1"E) and Kisite-Mpunguti (4°42'54.0"S 39°22'23.8"E). Here physicochemical characteristics and nutrient levels were estimated to assess inter-site variations at the time of sampling for microbial characterization i.e. in November and December 2016. The same reefs were sampled again for faecal pollution PCR tests in April 2019 during the long rain season. Located about 118 km north of Mombasa, Malindi Marine National Park and Reserve is the oldest MPA in Kenya having been gazetted in 1968 (McClanahan et al., 2010). This marine park is popular for glass bottom boat tours and snorkelling among other recreational touristic activities. It also experiences significant year-round discharge of freshwater and sediments from the Sabaki River which runs through a catchment area dominated by agricultural settlements (Katwijk et al., 1993; Munyao et al., 2003). KWS wardens administering the MPA advised that there is considerable illegal fishing in the park at night spilling over from the “reserve”, an area adjoining the park where artisanal fishing is allowed. This represented the “exploited” site. The tidal range at these sites is approximately 4 meters. Mombasa Marine National Park was considered the “polluted” site; owing to its proximity to the urbanised touristic city, the park experiences pollution from hotels,

hospitals, domestic and industrial waste disposal (Mwangi et al., 2001; Okuku et al., 2011). It was established in 1986 (Ngugi, 2002) with restrictions of protection being enforced commencing in 1991 (Tuda & Omar, 2012) and is the most visited of Kenya's marine parks by both local and international tourists (Owens, 1978). Because Mombasa is the primary port serving inland eastern and central African countries, the park is also impacted by marine traffic including oil pollution and dredge-spoil dumping (Mwangi et al., 2001). On the southern coast, 90 km from Mombasa, the Kisite-Mpunguti site served as the "baseline" site. This protected area comprises of Kisite Marine National Park and Mpunguti Reserve created in 1973 and gazetted in 1978. It is bordered by sparsely populated coral islands and experiences the least human activities because it is 11 km offshore (Emerton & Tessema, 2001). There is no river serving the site (Okuku et al., 2019) and the most common activity here is snorkelling.

Choice of the other three coral reef sites was designed to assess the effect a gradient of marine protection regimes (i.e. the "strictest", "moderate" and "least") on the densities of *E. coli* and nutrient levels as surrogates for human-source pollution. Sampling was done between March and April 2018 during the long rains. The coral reefs chosen were within the Kuruwitu Conservancy (3°48'27.7"S 39°49'60.0"E), Malindi Marine Reserve (3°17'41.3"S 40°07'07.3"E) and Kilifi creek (3°38'22.3"S 39°51'18.5"E).

Kuruwitu Conservancy is an LMMA administered as a "no-take" area, like the MPAs, since 2006 where no forms of consumptive utilization and human activities are allowed except for research and tourism. This site was therefore designated as guarded with the "strictest" protection. Malindi Marine Reserve was under "moderate" protection. The reserve acts as a buffer to the "no-take" Malindi Marine National Park (described above) adjoining it. Here, artisanal fishing as well as research and tourism is allowed, but commercial forms of utilization are prohibited (Tuda & Omar, 2012). Kilifi creek, on the other hand, was not protected hence it was designated the "least" protected. It is open to fishing, recreation and experiences effluent input from surrounding hotels and residential houses.

3.3.2 Sampling and in situ assessments

Sampling was done 200 - 500 m from the shore in the morning hours at a depth of 1-2 m in low tides and within 15-20 cm of a healthy-looking *Acropora* spp. colony – the dominant species on the Kenyan coral reefs. Seawater was collected in triplicate 50-mL samples (in the marine national parks) and duplicate 1L samples (in the other coral reefs) and transported on ice to the

laboratory for nutrient levels determination and microbial assessments. Physicochemical parameters including water temperature, pH, salinity, and dissolved oxygen were determined *in situ*, using portable multiprobe water quality meters per manufacturer's instructions (YSI Inc., Yellow Spring, OH). Sampling and measurements were completed within 5 days.

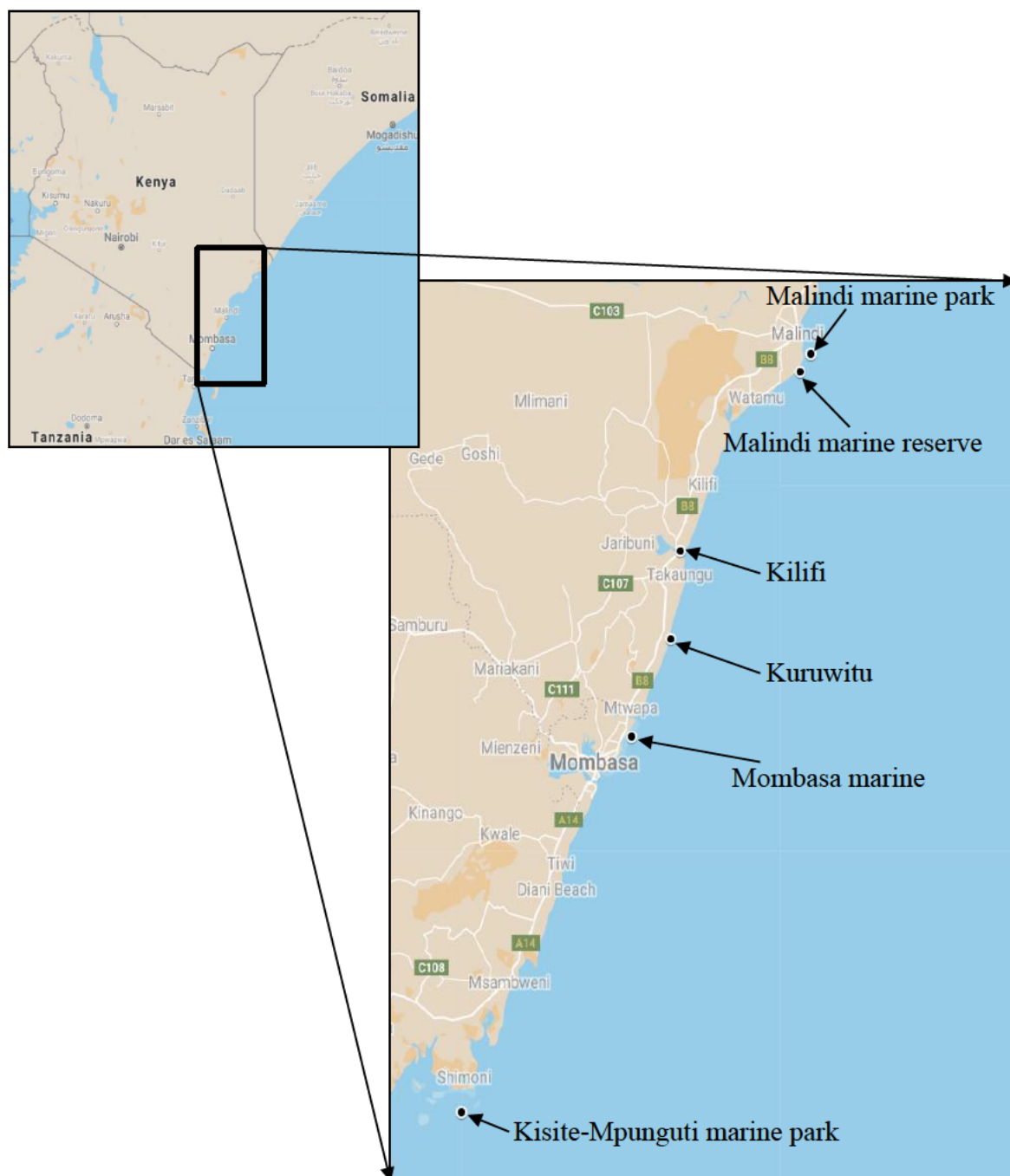


Figure 3.1 Map showing location of the coral reef sites sampled for this study.

Acknowledgment: Google maps.

3.3.3 Laboratory tests

3.3.3.1 Nutrient measurements

At the KMFRI nutrients laboratory, Mombasa, seawater samples collected in the marine national parks (Malindi, Mombasa and Kisite-Mpunguti) for assessment of inter-site variations were refrigerated for a week before processing. Concentrations of nitrate (NO_3^- -N), nitrite (NO_2^- -N), ammonium (NH_4^+ -N) and phosphate (PO_4^{3-} -P) nutrients were then determined by spectrophotometry as described in Ongore et al., 2013 (Ongore et al., 2013). Between-site comparison were tested by analysis of variance (ANOVA). For assessment of the effect of different marine protection regimes on the impact of human-source pollution, duplicate seawater from the other coral reefs was processed at Pwani University teaching laboratory 3 hours post-sampling. Here nitrate concentration was determined with ultraviolet spectrophotometry at 275 nm. Phosphate concentration was derived from orthophosphate colorimetric measurement at 690 nm using the Stannous Chloride method (APHA et al., 1999).

3.3.3.2 Microbial assessments

Laboratory processing for microbial assessments commenced within 3 hours post-sampling. PCR assays were optimized for *Bacteroides-Prevotella* primers that yield 541 – 694 bp product lengths (Table 3.1) for microbial-source identification. Faecal DNAs from human, cow, goat and pig were isolated from 0.20 g of freshly collected stool samples by DNeasy® PowerBiofilm® Kit (Qiagen) and used as controls. For the marine national park samples, DNA was isolated from 4 L of the coral reef seawater after vacuum-filtering (VWR, West Chester, PA, USA) through a 0.2-um pore size membrane (Pall Corporation, Port Washington, NY, USA) to capture microbial cells. Quality and quantity of DNA was confirmed by Nanodrop spectrophotometry and 1% agarose gel electrophoresis.

Two (2) μL of approximately 30 ng of faecal and seawater DNAs were amplified by PCR. PCR assays were performed on Veriti® Thermal Cycler (Thermo Fisher Scientific) in 25- μL reactions each containing 1x DreamTaq™ Green buffer with 2 mM MgCl_2 (Thermo Fisher Scientific), each primer at a concentration of 10 μM (Sigma), each deoxynucleoside triphosphate at a concentration of 200 μM (Thermo Fisher Scientific), 1.25 U of DreamTaq DNA Polymerase (Thermo Fisher Scientific) and 640 ng of bovine serum albumin (New England Biolabs) per μL . Reaction mixes were preheated at 94°C for 1 minute followed by 25 cycles of 94°C for 30 s, 61°C annealing for 30 s, 72°C for 1 min and a final 6-min extension at 72°C, for the detection of faecal bacteria. To increase the sensitivity of detection, 1 μL of each

PCR product was reamplified using the same conditions. Presence of prokaryotic DNA was confirmed separately with the commonly used primers 27F and 1492R, which amplify about 1500 bp of the bacterial 16S rRNA gene. For universal bacteria confirmation, denaturation at 94°C was done for 1 minute, annealing at 48°C for 30 seconds, and extension at 72°C for 2 minutes followed by a final extension at 72°C for 5 minutes. PCR products were visualized, against the GeneRuler™ 1kb DNA ladder 250-10000bp (Thermo Scientific), in a 1% agarose gel stained with GelRed.

Sensitivity of the microbial source tracking PCRs was evaluated by amplifying marker genes from serial dilutions of faecal control DNAs.

Table 3.1 *Bacteroides-Prevotella* and universal prokaryote primers used in the study for identification of sources of faecal pollution in the coral reef sites.

Primer	Sequence (5' – 3')	Target microbe	Reference
Bac708R	CAATCGGAGTTCTTCGTG	General <i>Bacteroidales</i>	Bernhard & Field, 2000
Bac32F	AACGCTAGCTACAGGCTT	General <i>Bacteroidales</i>	Bernhard & Field, 2000
CF128F	CCAACYTTCCCGWTACTC	Ruminant <i>Bacteroidales</i>	Bernhard & Field, 2000
HF183F	ATCATGAGTTCACATGTCCG	Human <i>Bacteroidales</i>	Bernhard & Field, 2000
PF163F	GCGGATTAATACCGTATGA	Pig <i>Bacteroidales</i>	Fremaux et al., 2009
27F	AGAGTTTGATCCTGGCTCAG	General bacteria	Frank et al., 2008
1492R	TACCTTGTTACGACTT	General bacteria	Frank et al., 2008

To estimate *E. coli* density across marine protection regimes gradient, triplicate 10, 1, and 0.1 mL seawater samples were inoculated in lauryl tryptose broth and incubated with inverted Durham's tubes at 37°C for 24 hrs for presumptive isolation of coliform bacteria. Negative samples were re-incubated up to 48 hrs. Loopful of culture from each tube showing growth and gas production were streaked on eosin-methylene blue agar and incubated at 37°C for 24 hrs for confirmation of *E. coli*. Concentration was determined from the standard three-tube most probable number statistical tables based on 95% confidence limits (Sutton, 2010).

3.4 Results

3.4.1 Inter-site variations in “no-take” marine national parks

Physicochemical parameters characteristics are presented in Table 3.2. Measurements did not vary much across the sites except for temperature and salinity recorded in Malindi that seemed at variance with the other sites: temperature recordings across the study sites ranged between 28.2°C (Kisite-Mpunguti) and 29.8°C (Malindi), while salinity was estimated as 27.6 ppt at Malindi compared to 30.1 ppt in both Mombasa and Kisite-Mpunguti. Dissolved oxygen and pH varied only slightly between the study sites.

Table 3.2 Measurements of physicochemical characteristics of the three coral reefs within the marine national parks at the time of sampling for metagenomic sequencing i.e. November - December 2016.

Variable	Kisite	Mombasa	Malindi
Temperature (°C)	28.2	28.6	29.8
pH (unit)	8.0	8.1	8.0
Dissolved Oxygen (ppm)	8.5	8.3	8.2
Salinity (ppt)	30.1	30.1	27.6

The nutrients assessed, except for nitrites, were highest at Malindi, the “exploited” site, and lowest at Kisite-Mpunguti, the “baseline” site (Table 3.3). Although observed, differences in nutrients concentrations between the study sites were not statistically significant (ANOVA).

PCR detection limit for *Bacteroides-Prevotella* marker gene was approximately the 10⁻² dilution, approximately 6 ng of faecal control DNA (figure 3.2). On plasmid DNA (pure target organism DNA) the PCR detection limit has been determined to be 10⁵ gene copies (Bernhard & Field, 2000) implying that 6 ng of total genomic faecal control DNA in this study contains at least 10⁵ gene copies of *Bacteroides-Prevotella*.

PCR results of DNAs freshly isolated from human, cow, goat and pig faeces tested with host-specific *Bacteroides-Prevotella* primers are shown in figure 3.3.

Table 3.3 Nutrient levels measurements of the three coral reefs within the marine national parks at the time of sampling for microbiome sequencing i.e. November -December 2016.

Nutrient	Kisite	Mombasa	Malindi	P
NITRATES (NO₃⁻ -N)				
(mg/L \pm SD)	0.002 \pm 0.003	0.046 \pm 0.029	0.138 \pm 0.096	0.068
NITRITES (NO₂⁻ -N)				
(mg/L \pm SD)	0.063 \pm 0.019	0.059 \pm 0.028	0.076 \pm 0.005	0.552
AMMONIUM (NH₄⁺ -N)				
(mg/L \pm SD)	0.035 \pm 0.006	0.040 \pm 0.006	0.040 \pm 0.07	0.572
PHOSPHATES (PO₄³⁻ -P)				
(mg/L \pm SD)	0.014 \pm 0.013	0.025 \pm 0.008	0.041 \pm 0.002	0.345

All primers worked optimally at an annealing temperature of 61°C. At this temperature, the in-house control DNAs were specifically amplified, each by the appropriate primers. None of the control DNAs were amplified by non-specific primers. Furthermore, control primers for all-source *Bacteroides-Prevotella* faecal bacteria amplified all the in-house control DNAs.

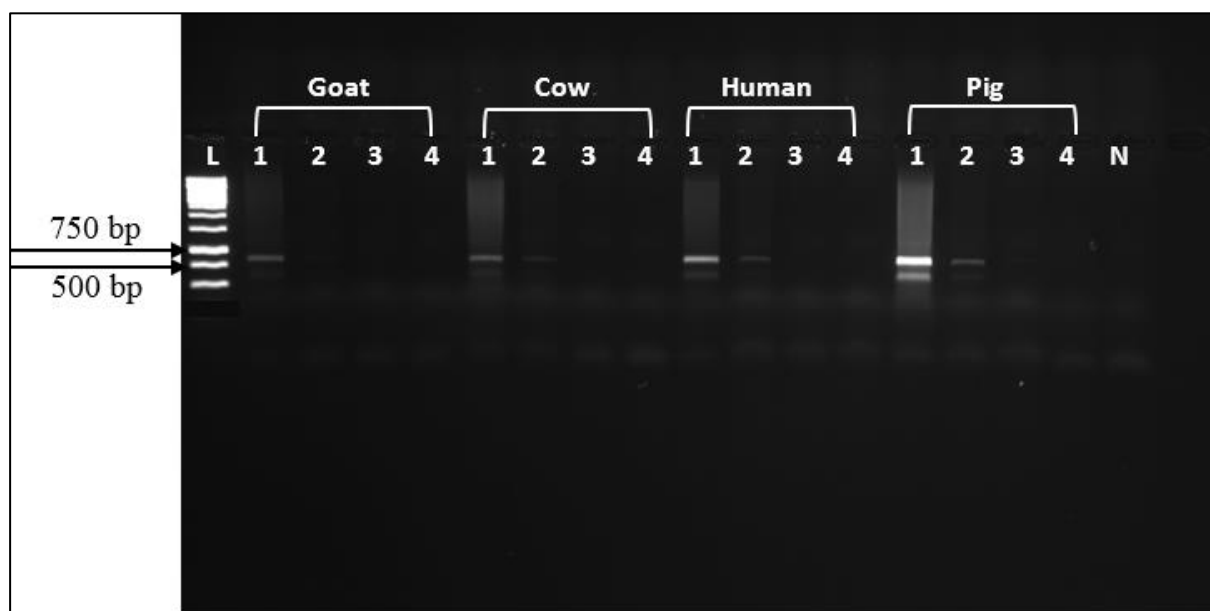


Figure 3.2 Agarose gel (1% w/v) image showing amplification of *Bacteroides-Prevotella* marker genes on 1) Neat; 2) 10⁻¹; 3) 10⁻²; and 4) 10⁻³ dilutions of ruminants, human and pig faecal control DNAs. L, GeneRuler™ 1kb DNA ladder 250-10000bp; and N, master mix with PCR water used as negative control

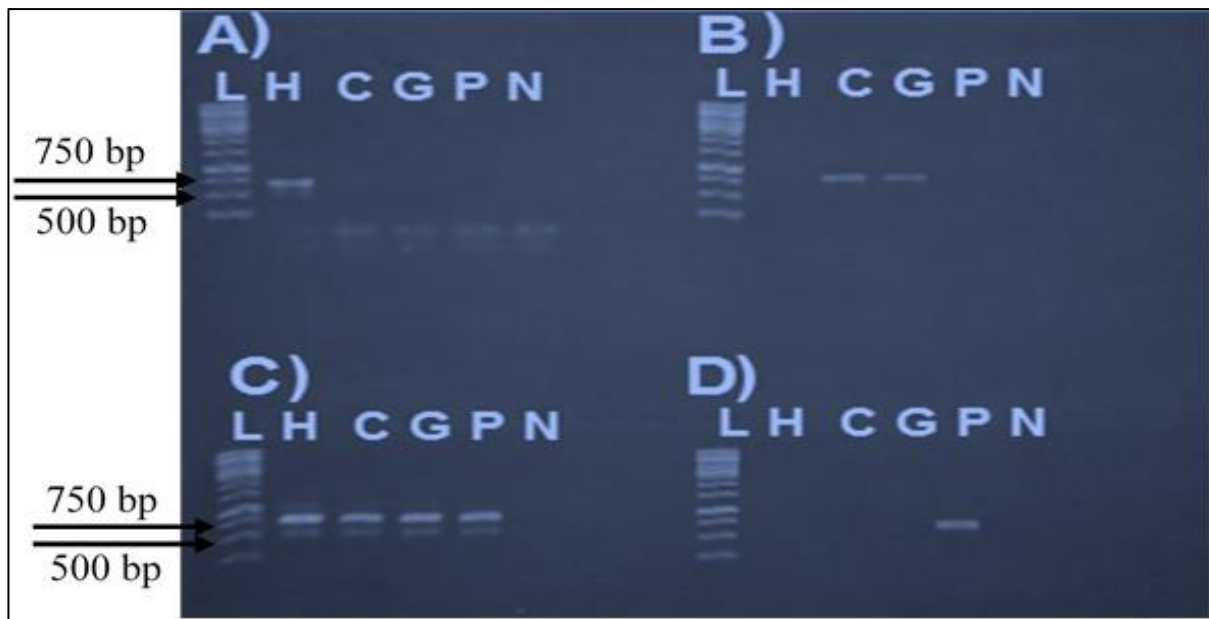


Figure 3.3 Agarose gel (1% w/v) image showing PCR results for control DNAs isolated from stool samples of human (H), cow (C), goat (G) and pig (P) amplified with faecal bacteria primers specific for: A) human, B) ruminants, C) all-source *Bacteroides-Prevotella*, and D) pig. L, GeneRuler™ 1kb DNA ladder 250-10000bp; and N, master mix with PCR water used as negative control

PCR with the universal prokaryote primers, 27F and 1492R, tested positive in all the seawater DNA samples from Malindi, Mombasa and Kisite-Mpunguti marine national parks (figure 3.4).

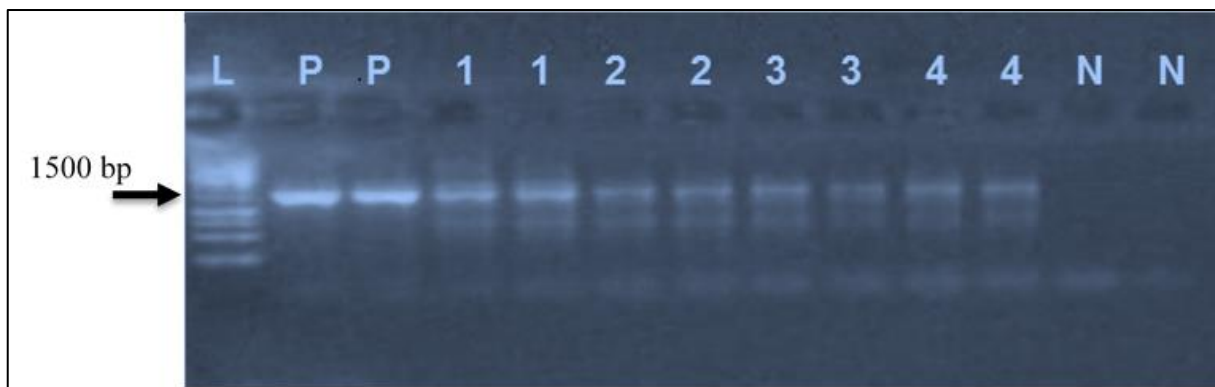


Figure 3.4 Agarose gel (1% w/v) image results for the universal bacteria 16S rRNA gene PCR. L, GeneRuler™ 1kb DNA ladder 250-10000bp; P, commercial bacterial control DNA from *E. coli* DNA; 1 to 3, eDNA from 1) Malindi, 2) Mombasa and 3) Kisite sites; 4, unrelated DNA that tested positive for bacteria in a previous experiment; N, master mix with PCR water used as negative control.

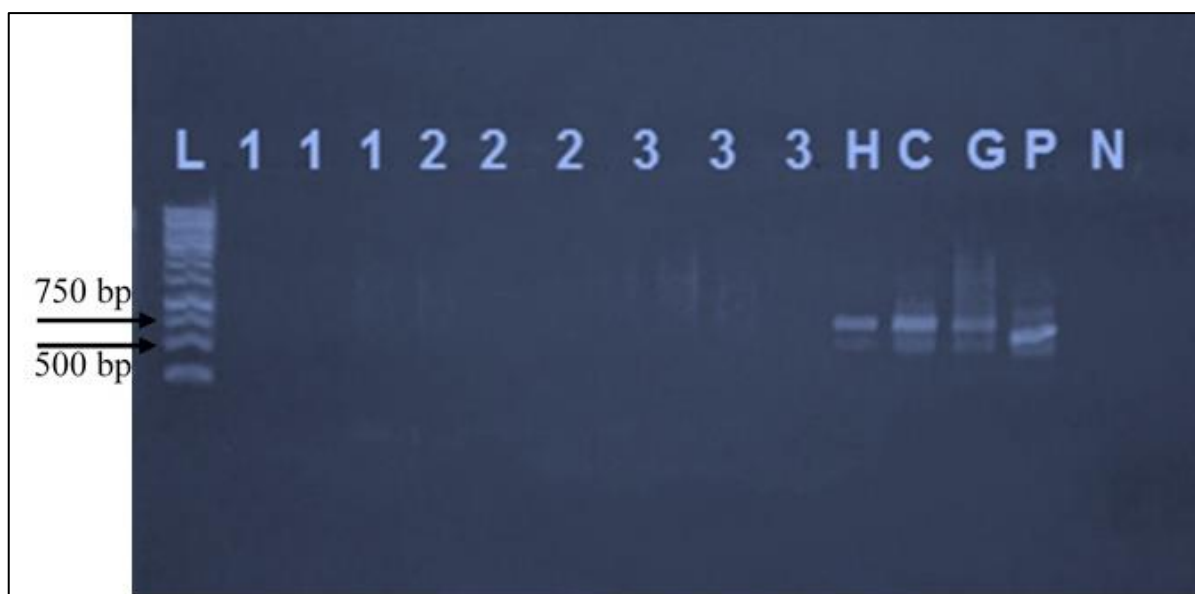


Figure 3.5 Agarose gel (1% w/v) image showing replicate faecal bacteria PCR screening results for eDNA from 1) Malindi, 2) Mombasa and 3) Kisite sites, along with control DNAs isolated from stool samples of human (H), cow (C), goat (G) and pig (P); L, GeneRuler™ 1kb DNA ladder 250-10000bp; and N, master mix with PCR water used as negative control

3.4.2 Variations across marine protection gradient

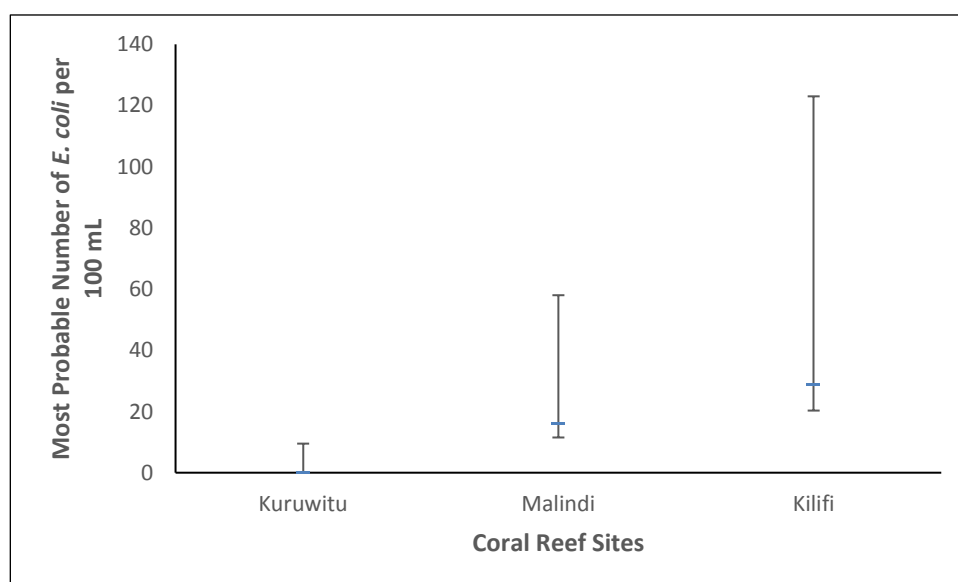


Figure 3.6 *E. coli* concentrations (MPN/100 mL) in water sampled at three different coral reef sites. Error bars indicate 95% confidence intervals.

However, DNA samples from the marine national parks had no detectable faecal bacteria when tested for all-source *Bacteroides-Prevotella* (figure 3.5).

E. coli concentrations decreased with increased levels of protection (figure 3.6). At Kuruwitu, the site with the strictest protection, *E. coli* was below detection limit (<3 MPN/100 mL), whilst Malindi (moderate protection) recorded 16 MPN/100 mL and Kilifi (least protection) 29 MPN/100 mL. Between-sites differences of *E. coli* density were not statistically significant.

Orthophosphate mean concentrations differed significantly among sites, increasing with the degree of marine protection (Table 3.4): the highest concentration, 0.524 mg/L, was observed at Kuruwitu followed by Malindi (0.422 mg/L) and Kilifi (0.326 mg/L) ($P < 0.01$). Nitrate mean concentrations also differed significantly among sites but did not correlate with the level of protection. Kuruwitu registered 0.566 mg/L, Kilifi 0.188, and Malindi, 1.402 ($P < 0.01$).

Table 3.4 Nutrient mean concentrations for in water samples collected at from the coral reef sites.

Nutrient	Kuruwitu	Malindi	Kilifi	<i>P</i>
ORTHOPHOSPHATE				
(mg/L \pm SD)	0.523 \pm 0.018	0.421 \pm 0.001	0.325 \pm 0.004	<0.01
NITRATE				
(mg/L \pm SD)	0.565 \pm 0.007	0.187 \pm 0.001	1.401 \pm 0.004	<0.01

3.5 Discussion

Poor water quality can impact coral reefs directly and indirectly diminishing their ability to withstand and recover from disturbance cycles. Characterizing natural environmental conditions supporting coral growth and health is necessary for understanding and potentially mapping the biogeography of corals in a region, especially in the context of the ongoing global coral reef ecosystems decline due to climate change and human activities. All environmental variables assessed at coral reefs in the “no-take” marine protected areas fell mostly with the ranges considered suitable for coral growth in tropical oceans (Guan et al., 2015; Henkel, 2010; Kannapiran et al., 2008; Roik et al., 2016), except for salinity which was low (i.e. 27.6 - 30.1 ppt) compared to previous estimates (32 - 33 ppt) (Obura, 2001). This may have been due

to high rainfalls during sampling and, especially for Malindi which had the lowest salinity, river Sabaki's discharge into the ocean. During heavy rainfalls, salinity has been known to decrease down to 12 ppt in shallow reef flats (Obura, 2001).

Except for nitrites, concentrations for all other nutrients assessed were higher at the "exploited" and "polluted" sites than at the "baseline" site suggesting influence by the surrounding human activities. These differences were not statistically significant though. It is not clear why nitrites were higher at the "baseline" site than the "exploited" and "polluted" sites, but it has previously been hypothesized that unexplained accumulation of nitrite might result from it being a common intermediate in numerous oxidative or reductive biochemical pathways in aquatic environments (Philips et al., 2002). Therefore, unlike nitrate levels, nitrite concentration is not a direct indicator of human impacts into the aquatic environments.

Although temporal variations were not investigated in this study, current estimates in the "no-take" marine protected areas are typical of previous measurements taken about two decades ago (McClanahan, 1988; Obura, 2001). Being cross-sectional surveys, the reported measurements represent quality of the coral reef water at the time of sampling, and not necessarily the stable state, which is essential for contextual understanding of microbiome data (presented in chapters 4 and 5) for which sampling was done during the survey. A recent study modelling the contribution of natural and human factors in predicting water quality identified human influences as the random components of variation associated with site-based differences (Houk et al., 2020). The estimates in this study suggest, therefore, that at the time of sampling human impacts to the coral reefs in the "no-take" marine protected areas were undetectable or indistinguishable by water quality measurements. Furthermore, no detection was made of faecal pollution by *Bacteroides-Prevotella* gene markers in any of the coral reef water sampled at the "no-take" MPAs. Tests for 16S rRNA gene which is universal for all prokaryotic taxa confirmed presence of bacteria in the coral reefs whose source was most likely not faecal.

Previous studies have reported substantial sewage disposal and nutrient enrichment at various locations on the Kenyan coast (Katwijk et al., 1993; Mwangi et al., 2001; Okuku et al., 2011). It is likely therefore that faecal pollution from both human and animals gets into the marine environment even at the time of this study. The fact that no detection was made may relate to one or a combination of three possible explanations. Firstly, the test method's detection limit which has previously been estimated to be 10^5 gene copies in approximately 10^{-12} g of plasmid

DNA (Bernhard & Field, 2000). The detection limit in faecal control DNAs was estimated to be about two orders of magnitude (i.e. 6×10^{-10} g) more than for the plasmid DNA. It is possible, therefore, that the approximately 6×10^{-9} of total genomic DNA from coral reef water used in this study did not contain adequate *Bacteroides-Prevotella* specific gene copies to be detected. Secondly, faecal bacteria of the order Bacteroidales are obligate anaerobes. As such, they are incapable of long-term survival or replication outside of their hosts. Therefore, *Bacteroides-Prevotella* from faecal pollution impacting coral reefs a few days before sampling would not still be viable to be detected. Lastly, faecal pollution discharged into the marine environment may not be getting to the coral reefs in the protected areas either because of control over human activities by the MPA enforcement or dilution of the pollutants due to the distance of the reefs from the shorelines, the portal of entry for the human-source pollution. In light of the findings from this chapter with distinct patterns of some water quality variables across sites, the protective benefits of MPAs are apparent, though further water quality studies are warranted.

At the coral reefs across marine protection gradient, *E. coli* concentrations decreased as the level of protection increased. Conversely, mean orthophosphate concentrations increased with increasing protection while mean nutrient concentration did not correlate with protection. These observations may be explained by the wide and differing range of human activities taking place around each study site, the level of marine protection and the proximity of each sampling site to a town.

In urban settings, wastewater and solid waste disposal is the principal cause of microbial pollution and nutrient loading (particularly nitrate) in the ocean (Okuku et al., 2011; Wakida & Lerner, 2005). The Kilifi site, which registered the highest *E. coli* and nitrate concentrations, is a creek dividing two business centers, Mnarani and Kilifi towns, which together constitute the capital of Kilifi County. The creek is surrounded by resorts and residential houses leaking and/or disposing wastewater into the ocean, and the beach is frequented by foreign tourists and residents, and holiday makers for recreation. In contrast, Malindi Marine Park and Kuruwitu Conservancy have limited, if any, exposure to urban pollution due to their distance from towns and levels of protection; the Malindi site (moderate protection) is about 10 km from the town, and Kuruwitu (strictest protection) about 23 km from Kilifi town.

The Kuruwitu site shares a beach with several affluent residential homes with swimming pools, lush lawns utilizing fertilizers, as well as commercial farms. The conservancy administration

reported seeking the intervention of National Environmental Management Agency (NEMA) to stop residents from discharging swimming pools wastewaters directly into the ocean during the initiation phase of the LMMA. This incidence highlights the need for diligence in enforcements of marine protection guidelines if MPAs are to be effective.

Agriculture-related activities are the main contributors of both orthophosphates and nitrate in rural settings (Coulter et al., 2004). Malindi (where the Sabaki river meets the ocean) was expected to contain higher mean nutrient concentrations than Kuruwitu since the Sabaki river runs through a catchment area with farming activities. However, the opposite was observed. Run-off from Kuruwitu lawns and commercial farms may explain the elevated orthophosphate and nitrate concentrations compared to Malindi where little or no farming occurs in the immediate vicinity of the sampling site. Furthermore, the Sabaki river has been shown to have the lowest discharge in the ocean during the north-eastern to south-eastern Monsoon transition period (March-April), which was when this study sampled. Most river sediment was deposited along the beach and intertidal area, and not in the neritic zone where this study sampled (Munyao et al., 2003). These observations support the results in this study, assuming that nutrients are deposited in a similar manner as sediments.

There are no universal *E. coli* and nutrients references for tropical coral reefs but in comparison to local standards elsewhere, the levels observed in this study are of public and environmental health concern. Although the *E. coli* concentration was lower than the threshold allowable for bathing beaches in parts of Europe (Baudart, et al., 2009), it may have been much higher at the shore close to the point where the pollutants entered the sea as bacterial numbers would decrease as they moved towards the open ocean due to tidal dilution and the bactericidal activity of seawater (Vaccaro et al., 1948). The nutrient concentrations assessed were also high, consistent with eutrophication (Okuku et al., 2011) or may have been the result of resuspended sediments (which accumulate nutrients at higher concentrations than seawater) following disturbances caused by rains and sampling activities. It is also possible that nutrient measurements in the three sites across marine protection gradient were systematically elevated through the test method. The nutrient measurements by ultraviolet spectrophotometry, the method used in this study, has been shown to be biasedly elevated in wastewaters due to interference by organic matter (Srivastava, Nahar, Brighu, & Gupta, 2019). Nonetheless, the elevated nutrients in these sites need to be confirmed by long-term follow-up study with supplementary data on possible impacts on reef biodiversity.

3.6 Conclusion

Marine pollution has been shown to be an important contributor to the decline of coral reef ecosystems. This is a major concern considering the growing number of people living within the coast, and especially in developing countries where 80 – 90% of the sewage is discharged in to the coastal zone (Okuku et al., 2011). Such disposal results in nutrient enrichment which causes corals' and reef organisms' death by anoxia or production of hydrogen sulphide (Dubinsky & Stambler, 1996). This also poses health risk to humans through consumption of fishery products and use of coastal zones for recreation (Kilinc & Besler, 2014). There is need therefore for regions endowed with rich diversity of coral reefs, such as WIO, to be able to detect and identify sources of pollution to design and implement appropriate interventions to safeguard coral reefs and their associated food webs.

This chapter aimed to survey variations in physicochemical and biological water quality parameters, between sites and across marine protection regimes gradient, that may be attributed to influence of human activities on Kenyan coral reefs. Physicochemical, nutrient levels and bacterial indicators were within the “healthy” range and they were comparable in all coral reefs despite differences in location and the surrounding human activities. Although the parameters evaluated could not detect human impacts within the MPAs, variation of *E. coli* density with degree of protection suggests that human activities contribute pollution that reaches Kenyan coral reefs in a manner that is inversely proportional to their restriction. Moreover, the unpredictable nutrient levels across protection gradient points to the possibility of multiple sources of marine pollution to which marine protection may have little or no effect.

Being a survey, these findings are limited to the state of coral reef water quality at the time of sampling and not necessarily representative of the long-term status. The study forms a needed basis for future spatiotemporal investigations for definitive determination of MPA's protective effect on pollution as well as specific links between human activities and environmental variables.

CHAPTER 4: TAXONOMIC STRUCTURE OF THE KENYAN INDIAN OCEAN CORAL REEF MICROBIAL COMMUNITIES

4.1 Abstract

Microorganisms contribute significantly to the functioning of ecosystems, as symbionts and pathogens, as well as essential components of food webs and nutrient cycles. Appreciation of this fact, and the emergence of advanced sequencing technologies that allow for improved understanding of microbial ecology have inspired research to establish indicators of coral reef stressors. Most phylogenetic profiling and piloting for metagenomic studies are done with 16S rRNA gene sequencing because it is relatively simple and cost effective. However, this approach has limitations, the main one being primer bias. This can be effectively mitigated through whole metagenomic sequencing. This study performed shotgun sequencing on environmental DNA from Kenyan coral reefs with neighbouring human activities characterised as either “exploited”, “polluted” or “baseline”. Microbial communities were compared between these sites (reefs) and with those classified from publicly available sequences of the *Tara* Oceans expedition (oceanic). Distribution was generally consistent with those published previously for marine ecosystems globally with the dominant bacteria being observed from Proteobacteria, Cyanobacteria and Bacteroidetes. Estimated species richness too was within comparable scale of previous estimate with oceanic sequences recording higher observed species while the reef sequences recorded higher diversity and evenness. Between the Kenyan coral reefs, there were variations in the relative abundances of ecologically significant taxa, especially copiotrophic bacteria and coliphages, that corresponded to the magnitude of human impacts in the respective sites. This debut metagenomic assessment of coral reef microbial diversity establishes a much-needed baseline for the WIO region and highlights a potential area for future research towards establishing indicators of environmental perturbations.

Keywords: *microbial diversity, coral reef microbiota, WIO biodiversity, metagenomic-profiling, whole metagenomic sequencing, ecosystem processes*

4.2 Introduction

There is growing interest in the distributions and ecology of microorganisms as potential monitoring tools for early signs of ecosystem degradation. The interest is inspired by the enhanced appreciation of the central role microorganisms play in ecosystems and the emergence of advanced technologies that allow for improved understanding of microbial ecology. Microorganisms contribute significantly to the functioning of ecosystems, as symbionts and pathogens, as well as essential components of food webs and nutrient cycles (Farrell et al., 2019). In coral reef ecosystems specifically, microorganisms have a central role in nutrient cycling (Hernandez-Agreda et al., 2017), coral nutrition (Krediet et al., 2013), response and acclimatization to stress (Gardner et al., 2019), and health and disease (Godoy-Vitorino et al., 2017; Rosenberg et al., 2007). This, and the fact that microorganisms respond and adapt fast to environmental disturbances (Ainsworth et al., 2010; Farrell et al., 2019) are the reasons it is believed that tracking of microbial communities and their response to environmental conditions across space and time may serve as indicators of ecosystem health in areas of conservation concern (McKew et al., 2012). This potential has recently been bolstered by a study that demonstrated free-living microbial communities, especially those in the coral reef seawater, have environmental sensitivity and predictability that can be used to infer shifts in the surrounding reef environment (Glasl et al., 2019).

The recent emergence of massively parallel sequencing approaches to environmental samples have greatly improved the efficiency of studying microbial ecology by allowing for rapid and culture-free characterization of bacterial communities (Farrell et al., 2019). Approaches such as metabarcoding, which utilizes phylogenetic markers to identify species in an environmental sample, have helped reveal a vast diversity of previously unknown microbial life forms in different marine niches. On the other hand, metagenomics – the direct sequencing and analysis of genomes in an environmental sample – have generated tremendous amounts of data enabling identification of microbial genes, their community interactions, and potential industrial applications (Behzad et al., 2016). Comparing metagenomes from different environments have led to the overarching hypothesis that microbial diversity is determined by the local environment i.e. unique environments harbour unique microbial species with unique genes and metabolic pathways.

The WIO has unique features that motivate studying its' microbial ecology: it is the warmest ocean with dynamic tropical waters (Díez et al., 2016), its ecosystems are under increasing

global and local pressure, yet they remain intact (Obura, 2017) with remarkably resilient corals (Gardner et al., 2019; McClanahan, 2017; Obura & Grimsditch, 2009; Obura et al., 2017). Despite this, the WIO remains understudied with respect to the baseline distributions of its microbial ecology (Díez et al., 2016). Tara Oceans expedition (Sunagawa et al., 2015) which assessed the microbiome of the main global oceanic regions, including WIO, sampled in the open ocean waters, not coral reefs. WIO-specific studies have mainly sampled corals with a specific focus on disease (Séré et al., 2013) and bleaching (Gardner et al., 2019).

This study assessed the microbial community compositions of Kenyan coral reef water and sediments by shotgun sequencing. Comparison were also done with metagenomic sequences from the Tara Oceans expedition sampled within the WIO oceanic zone for context.

4.3 Materials and methods

4.3.1 Study sites and sampling

Sampling was done in coral reefs at Malindi National Marine Park (3°15'35.1"S 40°08'40.0"E), and Mombasa National Marine Park (3°59'45.7"S 39°44'50.1"E) and Kisite-Mpunguti National Marine Park (4°42'54.0"S 39°22'23.8"E). The tidal range of these sites is approximately 4 meters. Detailed descriptions of the study sites and their neighbouring human activities and characteristic environmental variables are provided in Chapter three section 3.3.1.

Sampling was done within a window of 5 days in November and December 2016 in the coral reefs, 200 - 500 m from the shore, during low tides in the morning hours. Two samples were taken from each site: seawater collected in water bottle at 1-2 m depth within 15-20 cm of a healthy-looking *Acropora* spp. coral colony – the dominant species of coral, and sediment collected at the base of the same coral colony with a 10 mL syringe barrel. Samples were transported on ice to the laboratory for processing, typically within 3 hours of sampling.

4.3.2 DNA isolation

From each site, 4 L of seawater was vacuum-filtered (VWR, West Chester, PA, USA) through a 0.2-um pore size membrane (Pall Corporation, Port Washington, NY, USA) to capture microbial cells which were then added to a bead tube with lysis buffer. Microbial DNA from seawater samples was isolated with PowerWater®, and from sediment samples with

PowerSoil® DNA isolation kits according to the manufacturer's instructions (Mo Bio, Inc., Carlsbad, CA, USA). Quality and quantity of DNA was checked by Nanodrop spectrophotometry. Suitability for sequencing of DNA samples for metagenomics analysis was confirmed with 1% agarose gel electrophoresis following 16S rRNA PCR amplification (Rohwer et al., 2002).

4.3.3 Library preparation and DNA sequencing

Extracted DNA was subjected to whole-genome amplification by multiple displacement amplification (MDA) using the REPLI-g mini Kit according to the manufacturer's instructions (QiaGen, Hilden, Germany), except that incubations were held for 10 hours instead of the maximum recommended 16 hours. Sequencing libraries were prepared from 1 µg of DNA according to the manufacturers' preparation guide # 15036187 using the TruSeq DNA PCR-free library preparation kit (20015962/3, Illumina Inc.). Briefly, the DNA was fragmented using a Covaris E220 system, aiming at 350bp fragments. Resulting DNA fragments were end-repaired, and the 3' end adenylated to generate an overhang. Adapter sequences were ligated to the fragments via the A-overhang and the generated sequencing library was purified using AMPure XP beads (Beckman Coulter). The quality of the library was evaluated using the FragmentAnalyzer system and a DNF-910 kit. The adapter-ligated fragments were quantified by qPCR using the Library quantification kit for Illumina (KAPA Biosystems/Roche) on a CFX384Touch instrument (BioRad) before cluster generation and sequencing. A 200 pM solution of the individual sequencing libraries was subjected to cluster generation and paired-end sequencing with 150bp read length using an S2 flowcell on the NovaSeq system (Illumina Inc.) using the v1 chemistry according to the manufacturer's protocols. Base-calling was done on the instrument by RTA 3.3.3 and the resulting .bcl files were demultiplexed and converted to fastq format with tools provided by Illumina Inc., allowing for one mismatch in the index sequence. Sequencing was performed by the NGI SNP&SEQ Technology Platform in Uppsala, Sweden (www.sequencing.se).

4.3.4 Bioinformatics analyses

The raw Illumina reads were trimmed at Q5 threshold (MacManes, 2014), and the adapters were removed using fastp v0.19.5 (Chen et al., 2018). Trimmed sequences were deposited to the European Nucleotide Archive under the project accession # PRJEB30838. Publicly available metagenome sequences (project accession # PRJEB1787) from 14 oceanic samples

collected within the WIO during the 2009-2013 *Tara* Oceans expedition (Sunagawa et al., 2015), were also incorporated in the analysis.

Reads were uploaded into the MGX metagenomics analysis platform v20200508 (Jaenicke et al., 2018) where paired-end reads were merged and quality-filtered at Q35 for subsequent analysis. Taxonomic profiling was performed by applying the Lowest Common Ancestor (LCA) pipeline based on the Kraken tool, against NCBI nonredundant ('nr') database, enhanced by DIAMOND. Resulting taxonomy output files were exported to the R statistical environment (R Core Team, 2020) for analysis and visualization.

4.3.5 Statistical analyses

Taxa and metadata files were merged using *phyloseq* version 1.28.0 (McMurdie & Holmes, 2013) and used in subsequent microbial community analyses. Rarefaction curves were estimated with the *ranacapa* package (Kandlikar et al., 2018) and plotted using *ggplot2* (Wickham, 2016). Estimates of α -diversity were measured within sample categories using *estimate_richness()* function of the *phyloseq* package. Non-metric multidimensional scaling (nMDS) and principal coordinates analysis (PCoA) ordinations of Bray-Curtis dissimilarity were performed using taxonomy relative abundance matrix by the *vegan* 2.5-6 (Oksanen et al., 2019). Comparison of beta diversity between groups was assessed by permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2001) using *adonis* test based on Bray-Curtis distances with 999 permutations also with *vegan*.

4.4 Results

Table 4.1 A breakdown of the metagenomic sequences from the Kenyan coral reefs and *Tara* Oceans expedition (oceanic) samples processed on the MGX metagenomics analysis platform

	Coral reefs	Oceanic
Total reads processed	211,966,893	1,207,371,204
Assigned reads	19,038,026	589,613,571
Archaea	249,741	1,591,604
Bacteria	18,236,457	554,859,741
Eukaryota	331,859	2,098,101
Viruses	219,969	31,064,125

A breakdown of 1,419,338,097 total reads processed from the 20 samples is provided in table 4.1. Overall, 43% of the reads were assigned, the least percentage (~9%) being from the reef sequences. Taxonomic assignment resulted in four superkingdoms the majority (94.2% of the assigned reads) being Bacteria, then Viruses (5.1%), Eukaryota (0.4%) and Archaea (0.3%). Eukaryotic taxa included both unicellular (e.g. protists) and multicellular species (including *Acropora* spp and fish) whose gametes or fragmented cells may have been sampled – associated reads were excluded from analysis.

Rarefaction curves (figure 4.1) for most of the sequences from both datasets began to level off suggesting reasonable coverage of the microbial communities characterized with the processed reads.

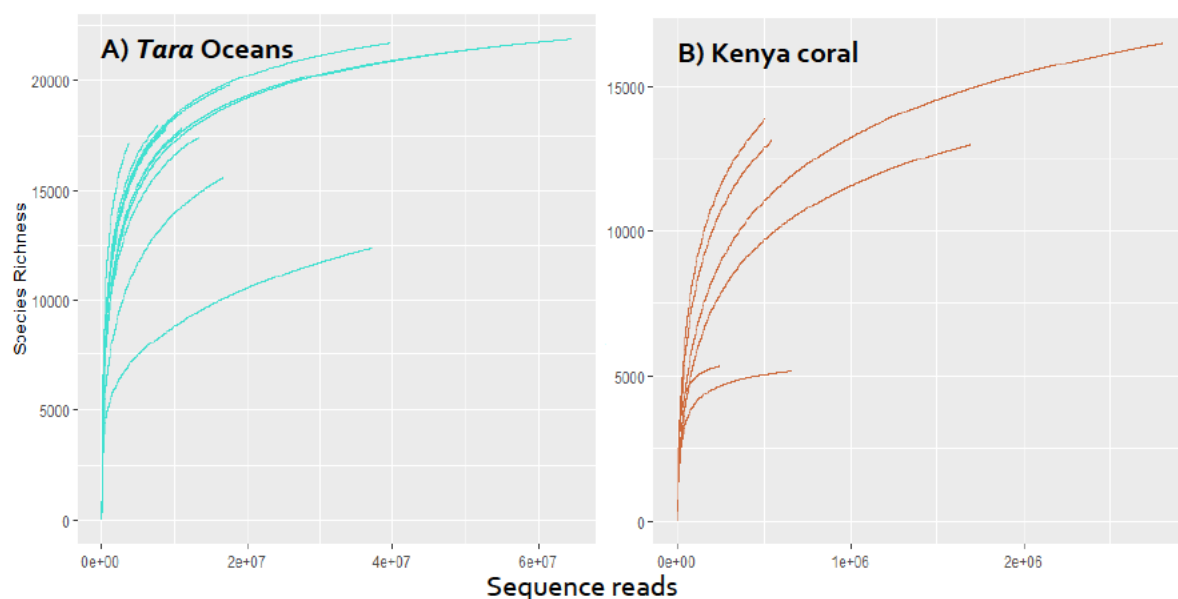


Figure 4.1 Rarefaction curves for community species richness on the sequences from *Tara* Oceans (oceanic) (A) and Kenya coral reefs (reef) (B) samples

In all, 25,460 species belonging to 76 phyla (bacterial, viral and archaeal combined) were assigned.

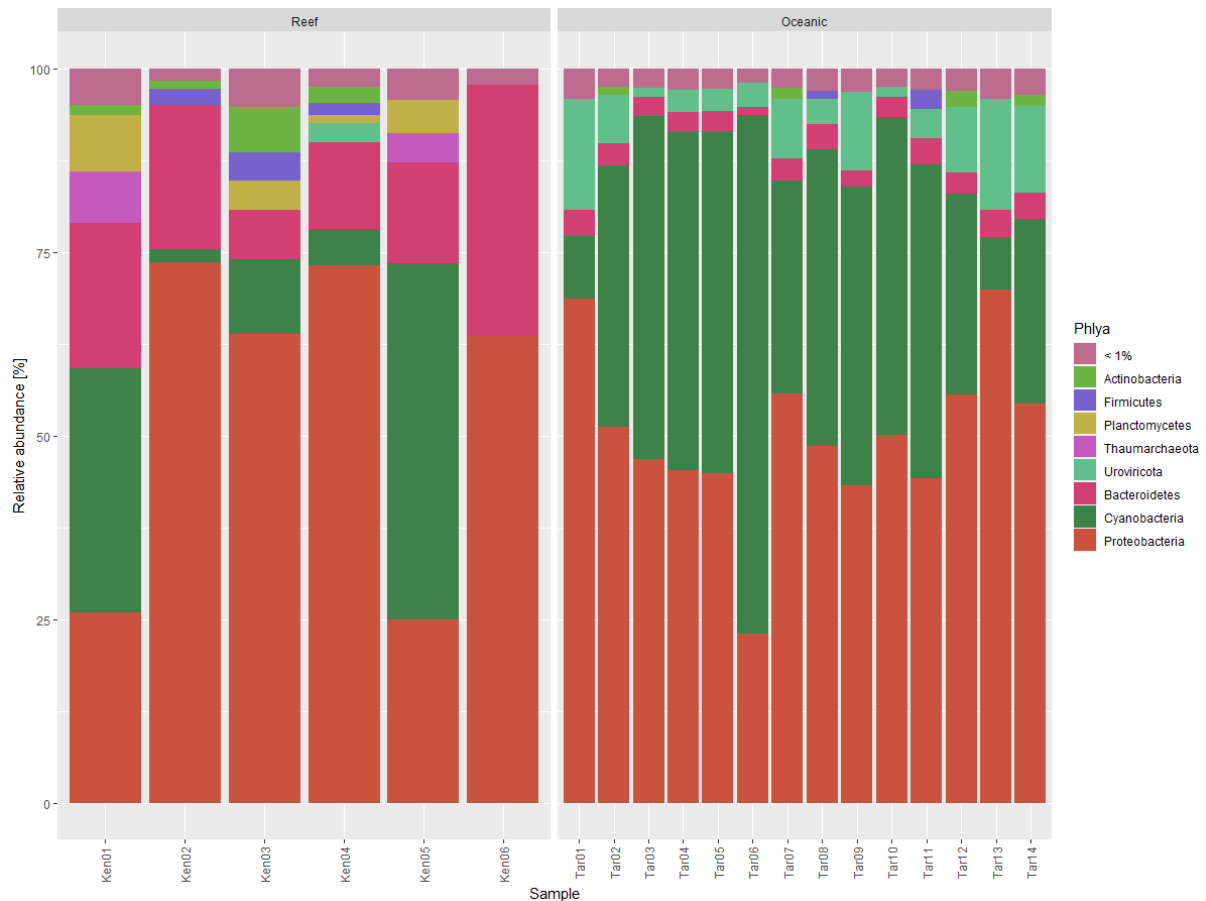


Figure 4.2 Stacked charts of the distribution of community phyla with relative abundance $\geq 1\%$ sequenced from Kenya coral reef (Ken*), and the *Tara* Oceans expedition (Tar*) samples

Proteobacteria was the most dominant phylum with over 50% overall relative abundance (figure 4.2). The second most abundant phylum was Cyanobacteria (39%) in oceanic samples (versus 11% in reef) and Bacteroides (20%) in the reef samples (versus 3% in oceanic samples). Two phyla, Candidatus Riflebacteria and Saleviricota, were found only in the oceanic dataset while the phylum Hofneiviricota was unique to the reefs.

Compared to reef samples, the oceanic sequences (from *Tara* Oceans dataset) had significantly higher observed ($p < 0.001$) and estimated ($p < 0.001$) species richness (figure 4.3). In contrast, reef samples showed higher diversity and evenness estimated by Shannon Index ($p < 0.001$). nMDS ordination (figure 4.4) based on the Bray-Curtis dissimilarities of the samples showed that reef samples had higher dispersion and were separated from the oceanic samples ($p = 0.001$). PCoA revealed that ecosystem (i.e. reef vs. oceanic) accounted for 48.1% of the variance (figure 4.5).

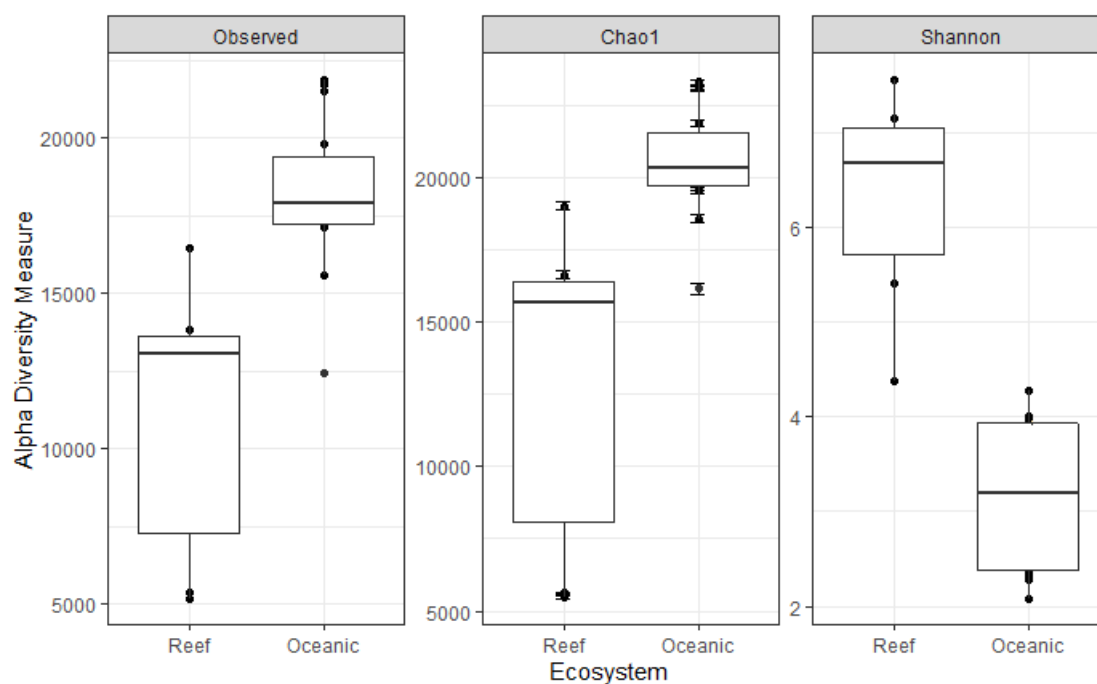


Figure 4.3 Plots of alpha diversity metrics the observed and estimated bacterial, viral and archaeal species richness on the Oceanic and Reef sample sequences

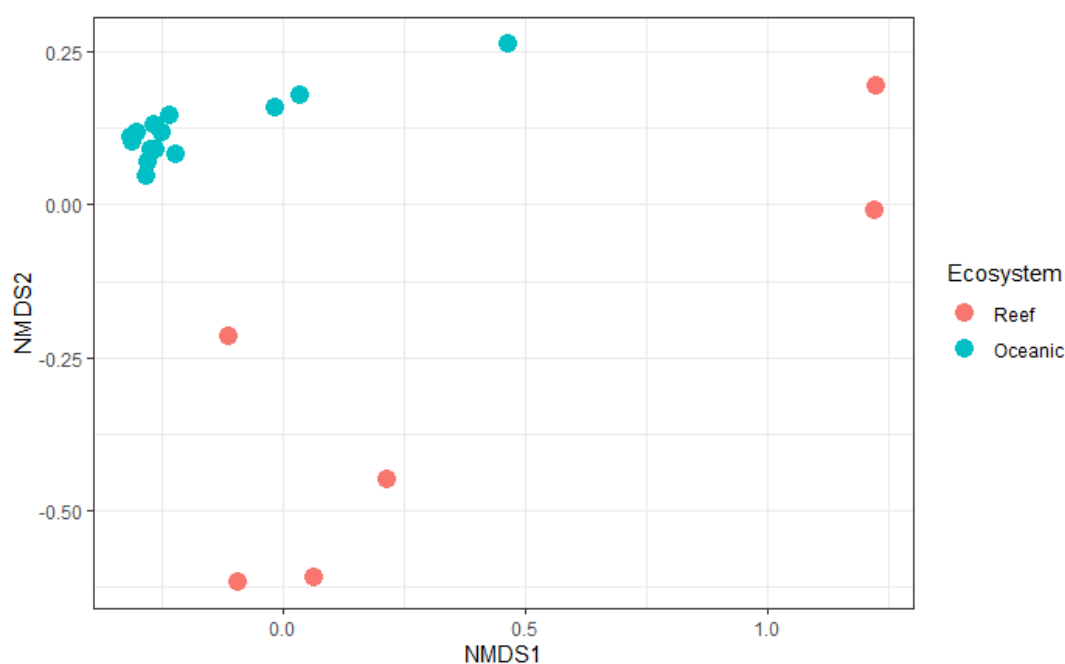


Figure 4.4 Bacterial, viral and archaeal combined community composition analysed with nonmetric multidimensional scaling (nMDS) plots using Bray-Curtis dissimilarity for Kenya coral reefs (peach circles) and *Tara* Oceans (aqua circles) datasets. For this nMDS plot, stress = 0.04, $r^2 = 0.998$.

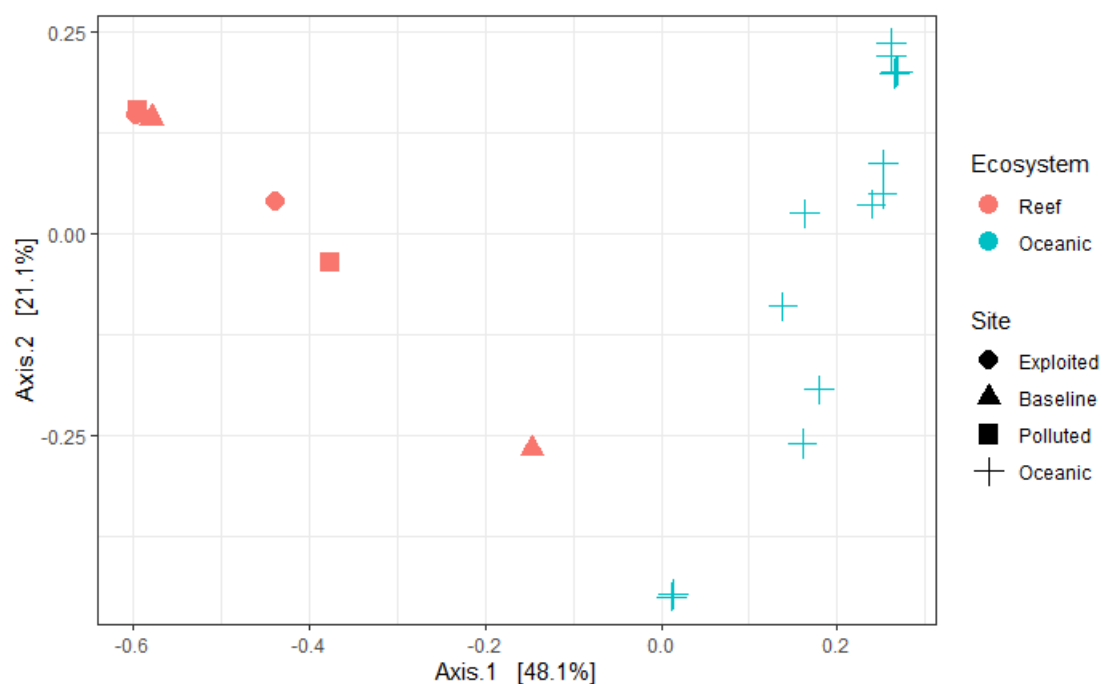


Figure 4.5 Bacterial, viral and archaeal combined community composition Principal Coordinate Analysis (PCoA) plots of Bray–Curtis dissimilarity matrix for reefs and oceanic datasets

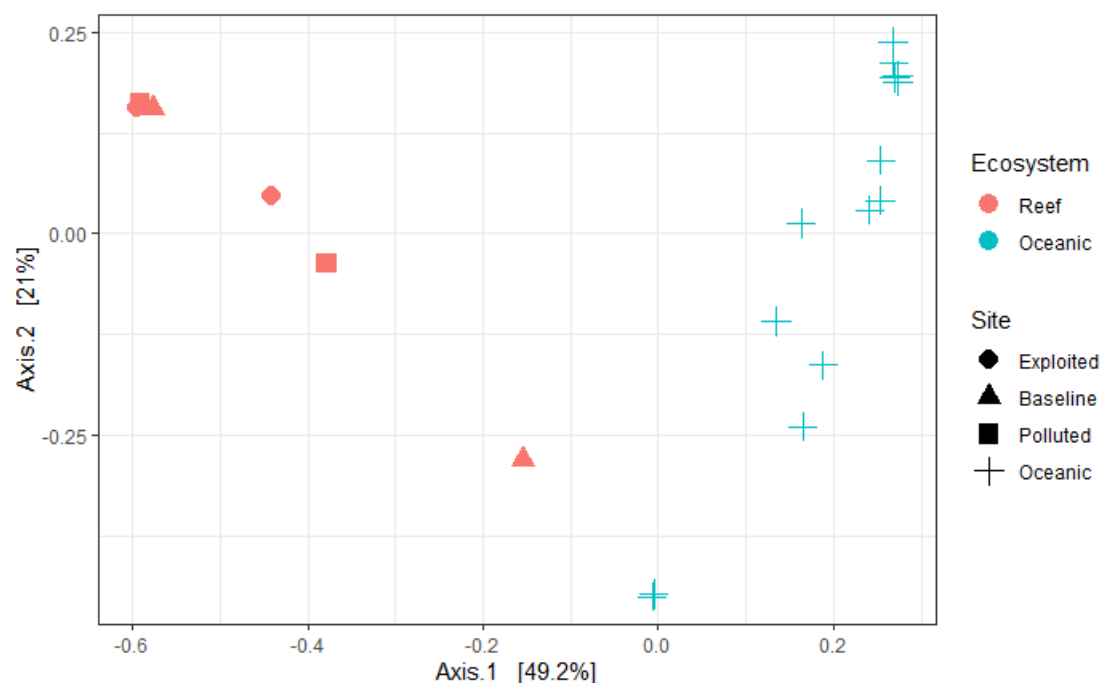


Figure 4.6 Bacterial community composition analysed with Principal Coordinate Analysis (PCoA) plots of Bray–Curtis dissimilarity matrix for reefs and oceanic datasets.

A total of 23,260 bacterial species assigned to 53 phyla were classified from both oceanic and reef sequences. Oceanic samples had significantly higher observed (23,003 versus 18,271) and estimated (23,822 versus 20,076, by Chao1) bacterial species richness ($p<0.001$), while the reefs had higher species diversity and evenness (6.15 versus 2.73, by Shannon index) ($p<0.001$). PCoA showed separation between oceanic and reef bacterial communities with over 49% of the variance being explained by ecosystem (figure 4.6).

Classes belonging to Proteobacteria and Bacteroidetes phyla had the highest relative abundances. Among the Proteobacteria classes, Alphaproteobacteria and Epsilonproteobacteria had higher relative abundance in the oceanic samples, while Gammaproteobacteria, Betaproteobacteria and Deltaproteobacteria were relatively higher in reefs (figure 4.7). Other classes that had a relative abundance higher than 1% belonged to the phyla Actinobacteria (Actinobacteria) and Firmicutes (Bacilli, Clostridia and Planctomycetes).

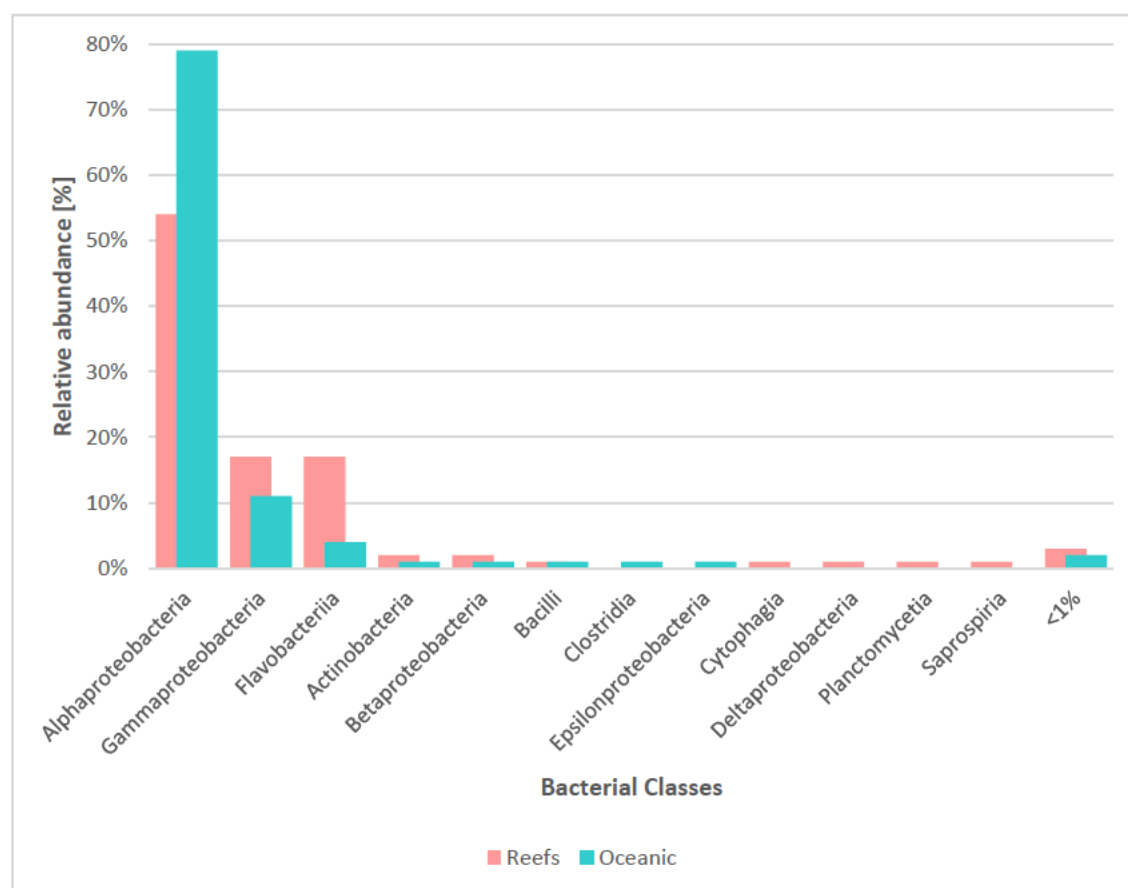


Figure 4.7 Distribution of oceanic (aqua) and reef (peach) sequences assigned to bacterial classes with relative abundance of 1% or higher in at least one of the datasets are plotted.

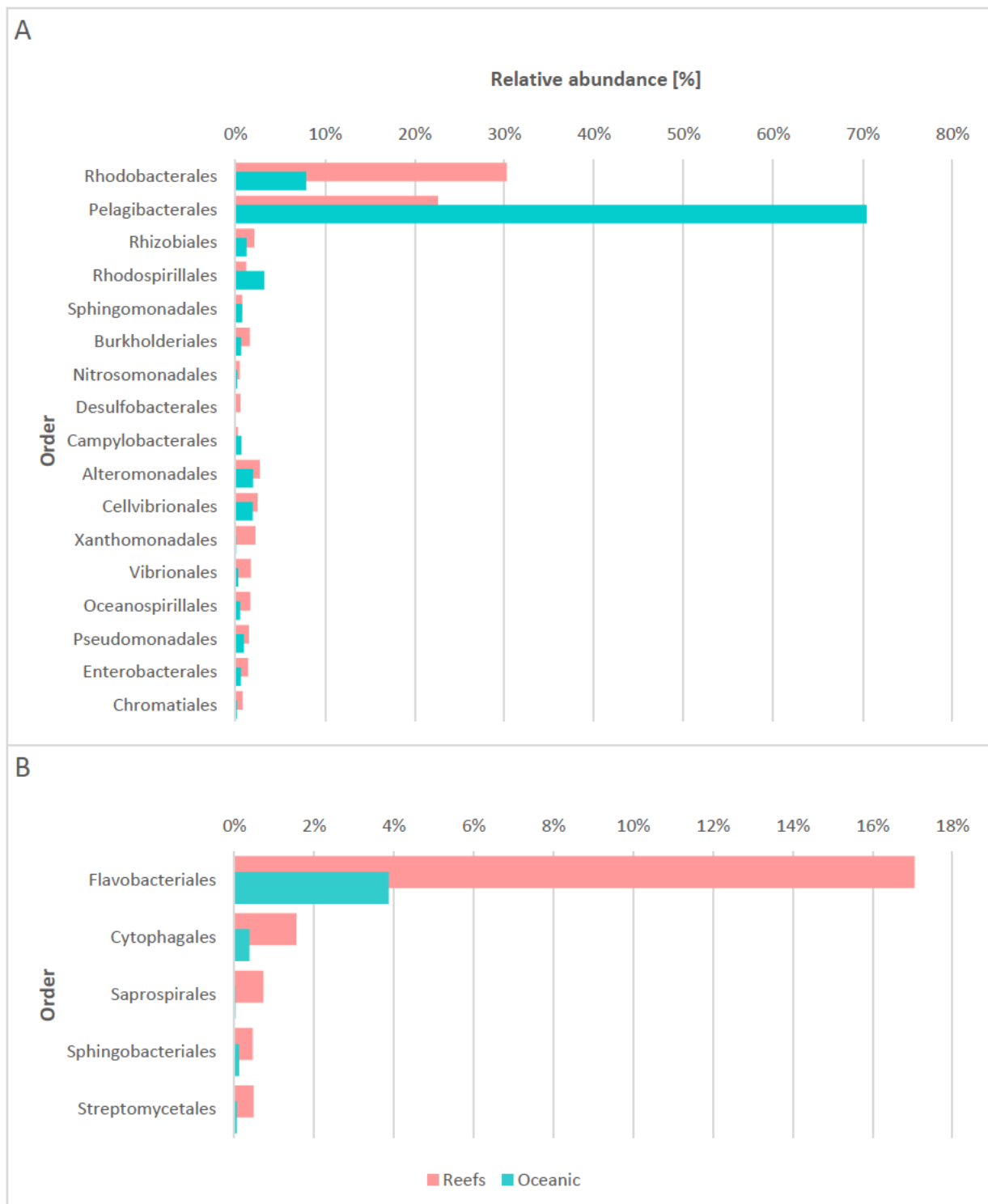


Figure 4.8 Distribution of oceanic (aqua) and reef (peach) sequences assigned to orders of the copiotrophic phyla Proteobacteria (A), and Bacteroidetes and Actinobacteria (B). Only orders with relative abundance of 1% or higher in at least one of the datasets are plotted.

Further inspection was made of the Proteobacteria phylum as well as the phyla Bacteroidetes and Actinobacteria which are considered copiotrophic i.e. found in nutrient-rich environments. The distribution of orders, from the copiotrophic phyla, with relative abundance $\geq 1\%$ are plotted in figure 4.8. Sequences for Pelagibacterales ($>20\%$) and Rhodobacterales ($>8\%$) had the highest relative abundance among the 17 Proteobacteria orders with a relative abundance $\geq 1\%$ (figure 4.8 A), while Flavobacteriales ($\geq 4\%$) had the dominant sequences of the Bacteroidetes orders i.e. Cytophagales, Saprospirales, Sphingobacteriales (figure 4.8 B). Of the Actinobacteria orders, only Streptomycetales had relative abundance of $\sim 1\%$. Reef sequences showed higher relative abundance in all copiotrophic orders, except Pelagibacterales and Rhodospirillales which were higher in oceanic sequences.

The Kenyan coral reef sequences assigned to 13,903 species harboured in copiotrophic phyla Proteobacteria, Bacteroidetes and Actinobacteria. The distribution of these copiotrophic sequences did not show clustering by site (figure 4.9) ($p = 0.798$).

The first ten most abundant orders of the copiotrophic phyla assigned to Kenyan coral reef sequences belonged to Proteobacteria and Bacteroidetes. By comparison, the relative abundance of each copiotrophic order was highest in either the exploited or polluted site except for the order Pelagibacterales whose relative abundance was highest in the baseline site (figure 4.10).

Relative abundance for sequences assigned to the traditional faecal indicator bacteria genera (*Escherichia*, *Salmonella* and *Campylobacter*) as well the alternative ones employed for microbial source tracking (i.e. *Bacteroides* and *Prevotella*) were less than 1%. No sequences were assigned to the *Enterococcus* genus. Of these, the oceanic sequences recorded lower relative abundance for all genera except *Campylobacter* which was more than two-fold higher than in coral reef sequences (figure 4.11).

Among the Kenya coral reef samples, the relative abundance of *Bacteroides* and *Prevotella* sequences from either “exploited” or “polluted” sites was more than three-fold higher than in the sequences from the “baseline” site. Relative abundances for *Campylobacter* and *Salmonella* sequences were comparable across the reef sites while that of *Escherichia* was more than two-fold higher in the sequences from the “baseline” site than in the sequences from either the “exploited” or “polluted” sites.

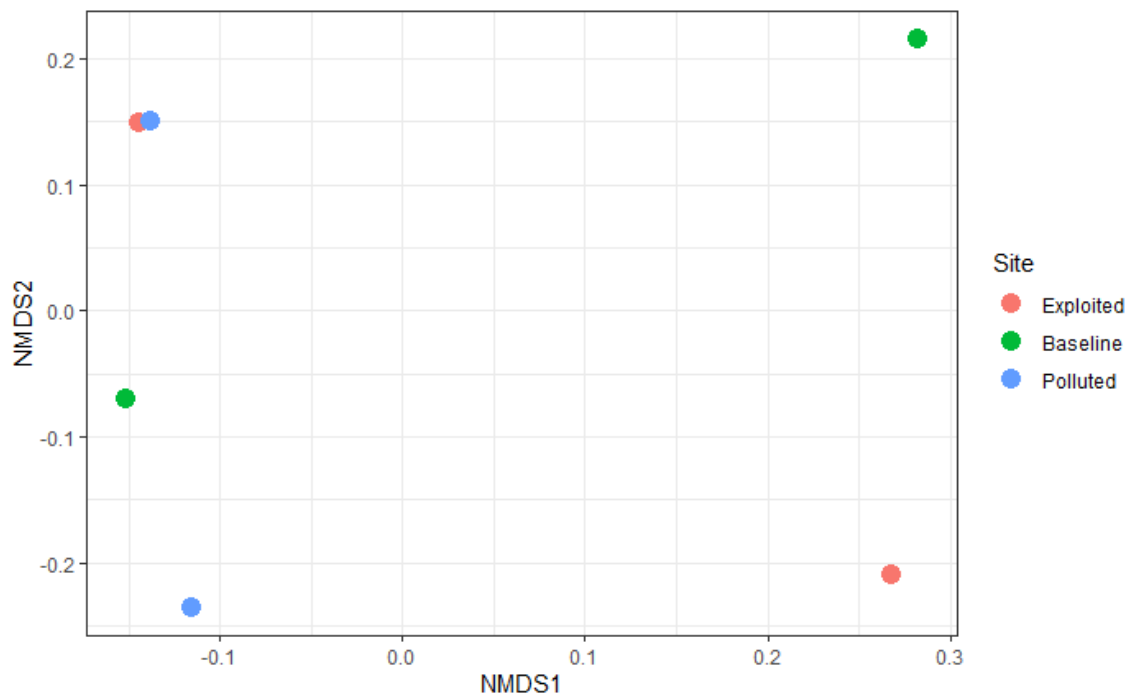


Figure 4.9 Bacterial community composition analysed with nonmetric multidimensional scaling (nMDS) plots using Bray-Curtis dissimilarity for Kenya coral reefs at the national marine parks classified “exploited” (Olive circles; Malindi), “baseline” (Green circles; Kisite-Mpunguti), and “polluted” (Blue circles; Mombasa). For this nMDS plot, stress = 0.05, $r^2 = 0.998$.

Of the species with more than 2 reads, 145 were specific for *Tara* Oceans while 186 were specific for Kenya coral reefs. The specificity of the *Tara* Oceans (i.e., the ratio of *Tara*-specific species to the total number species present in *Tara* Oceans and Kenyan coral reefs datasets) increased with the number of minimum sequences assigned to the species used for examination, at a higher rate than specificity for reefs (figure 4.12 A). Among the coral reef sites, the “baseline” site had the highest number of unique species and the “polluted” site the lowest (figure 4.12 B).

For the Archaea superkingdom, 679 species were assigned spread in 10 phyla. Sequences from the oceanic samples were assigned significantly higher observed (647 vs 558) and Chao1 estimated (Chao1 684 vs 618) species than sequences from Kenyan coral reefs, which in turn had had higher diversity and evenness (3.74 vs 2.45) ($p < 0.001$) (figure 4.13).

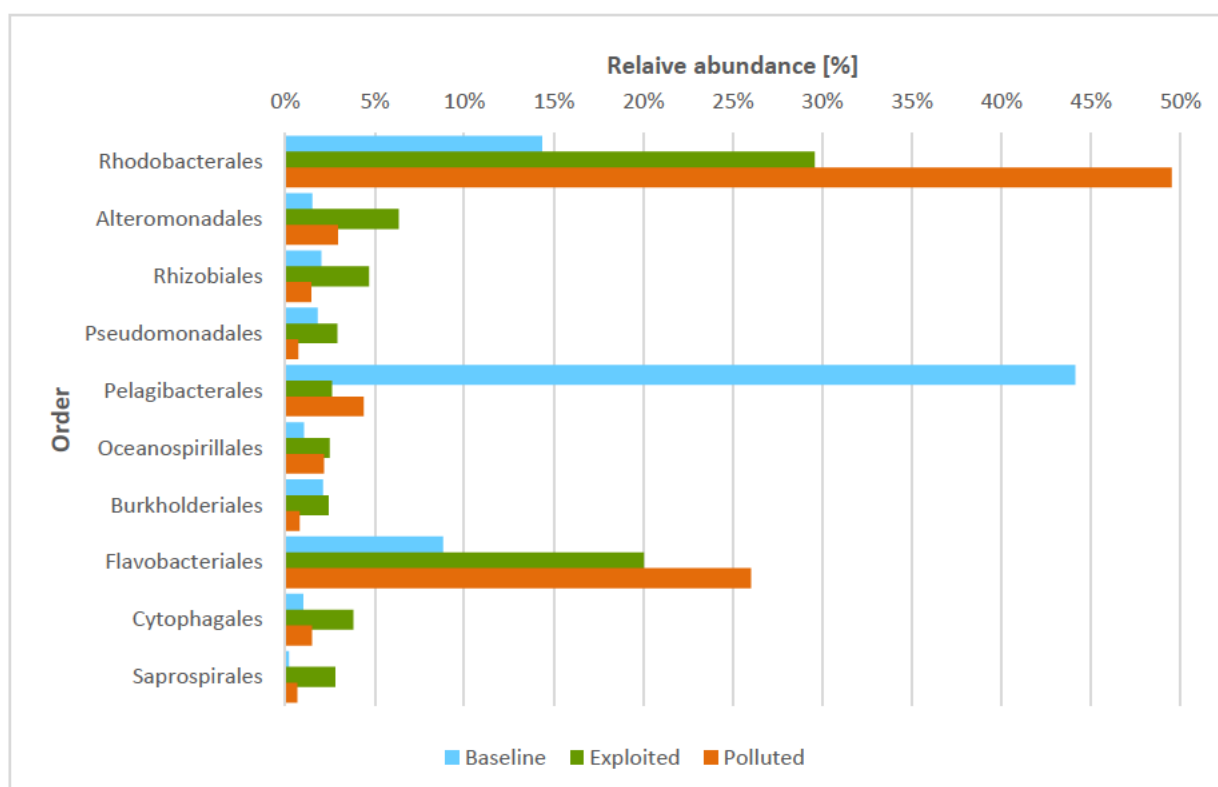


Figure 4.10 Distribution of the top ten most abundant orders of the copiotrophic phyla (Proteobacteria and Bacteroidetes) from sequences of samples from the Kenyan coral reefs at the national marine parks classified “baseline” (Sky; Kisite-Mpunguti), “exploited” (Olive; Malindi) and “polluted” (Bronze; Mombasa).

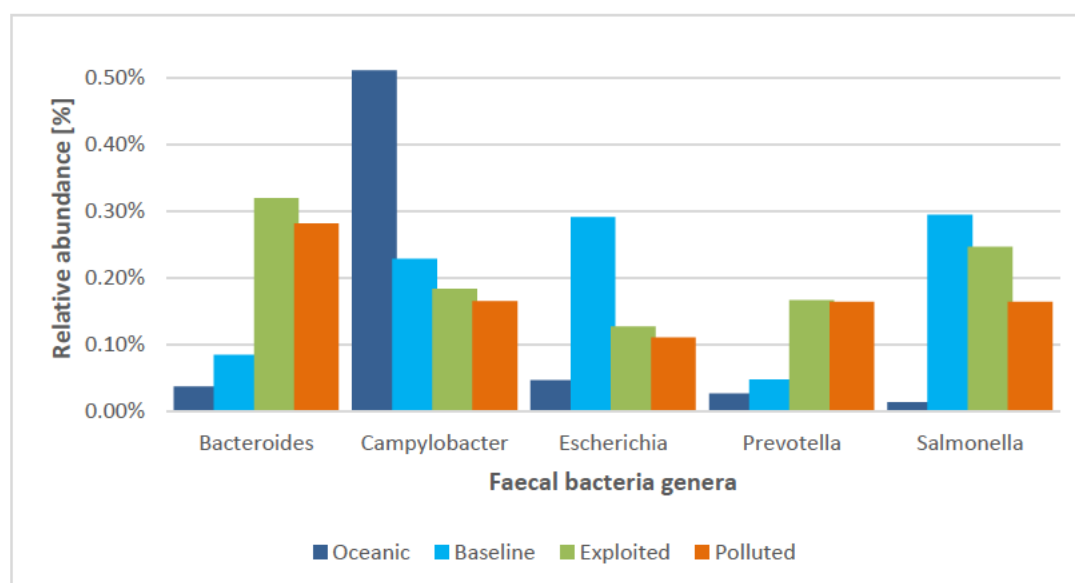


Figure 4.11 Distribution of faecal indicator bacterial genera by study site i.e. Oceanic (blue), and the coral reefs at sites classified “baseline” (sky), “exploited” (olive) and “polluted” (bronze).

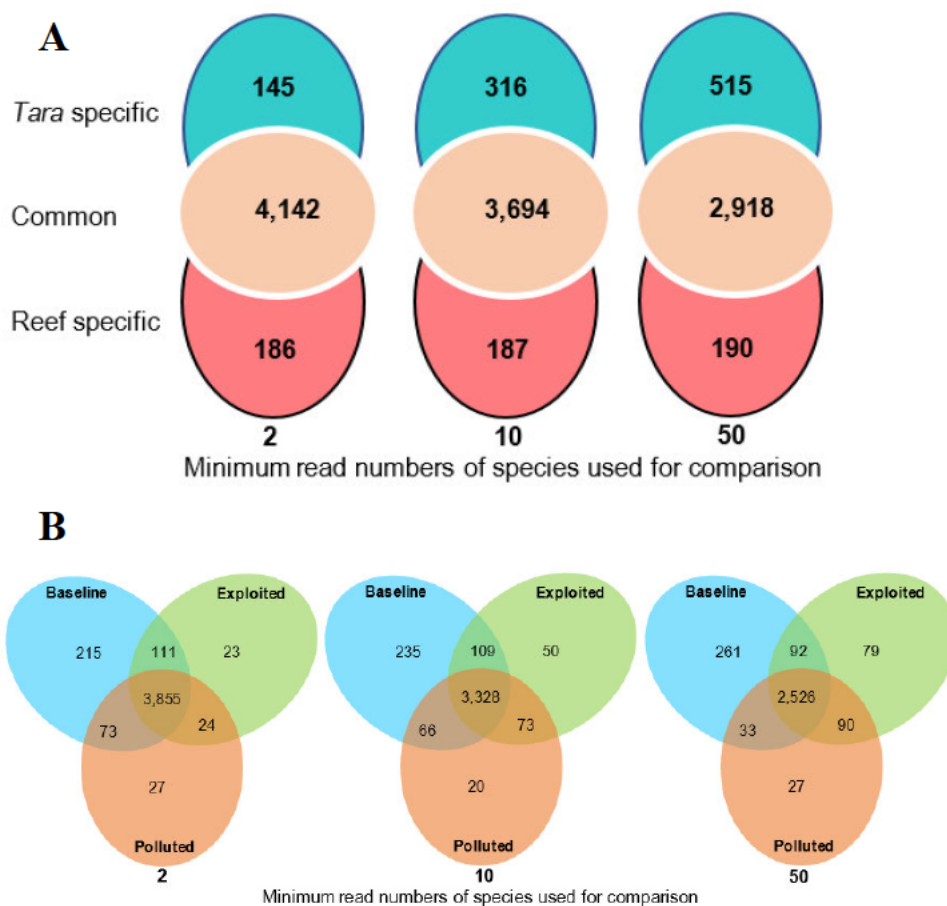


Figure 4.12 Venn diagrams showing the distribution of microbial species across A) *Tara* and reef, and B) Kenya coral reef metagenomes based on species with more than 2, 10, 50 reads.

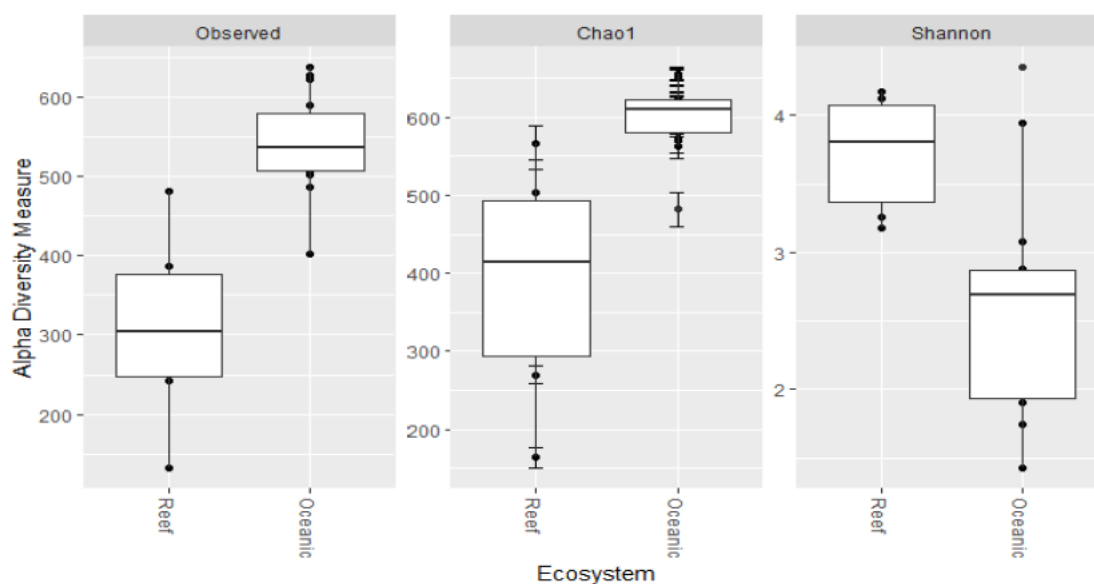


Figure 4.13 Plots of alpha diversity metrics for the observed and estimated archaeal species richness on the Oceanic and Reef sample sequences

Visualisation by nMDS ordination revealed clustering among the oceanic sequences, dispersion in the coral reef samples and no overlap between the two datasets ($P < 0.001$) (figure 4.14).

The phyla Euryarchaeota and Thaumarchaeota had the most sequence assignments in both datasets with overall average relative abundance of approximately 48% each. This relative abundance, for the two phyla, was comparable across all samples except in two sediment samples from the Kenya coral reefs (Ken01 and Ken05) where Thaumarchaeota was elevated to over 95% while Euryarchaeota reduced to below 4% (figure 4.15). The phylum Candidatus Riflebacteria was unique to the oceanic sequences.

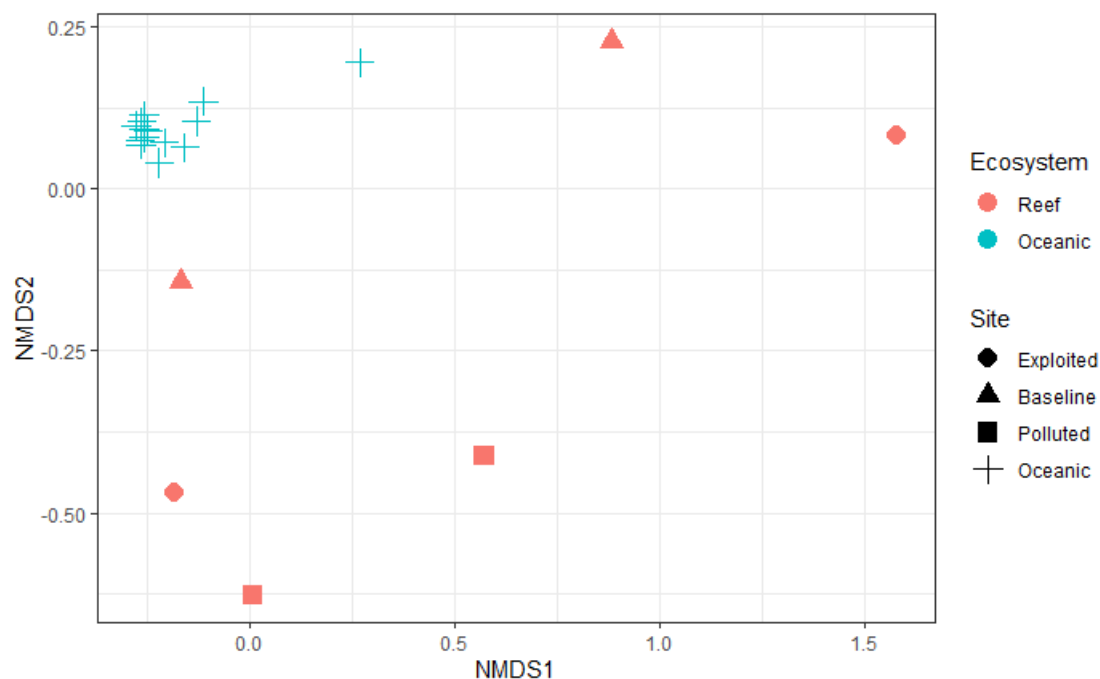


Figure 4.14 Archaeal community composition analysed with nonmetric multidimensional scaling (nMDS) plots using Bray-Curtis dissimilarity for Kenya coral reefs (peach circles) and *Tara* Oceans (aqua circles) datasets. For this nMDS plot, stress = 0.05, $r^2 = 0.998$.

Nitrosopumilales recorded the highest relative abundance particularly in the Kenyan coral reef sequences where it reached $> 90\%$ (figure 4.16). Between the Kenyan coral reefs, no site-specific structuring was noted ($p = 0.774$), but sequences from the “baseline” site had higher

relative abundance for all orders except Nitrosopumilales (31%) for which the “exploited” and “polluted” site sequences were three-fold higher. The oceanic sequences had higher relative abundance than the reef sequences for most orders except Nitrosopumilales, Nitrososphaerales, Haloferacales, Natrialbales and Methanococcales for which sequences from the “baseline” reef site had higher relative abundance.

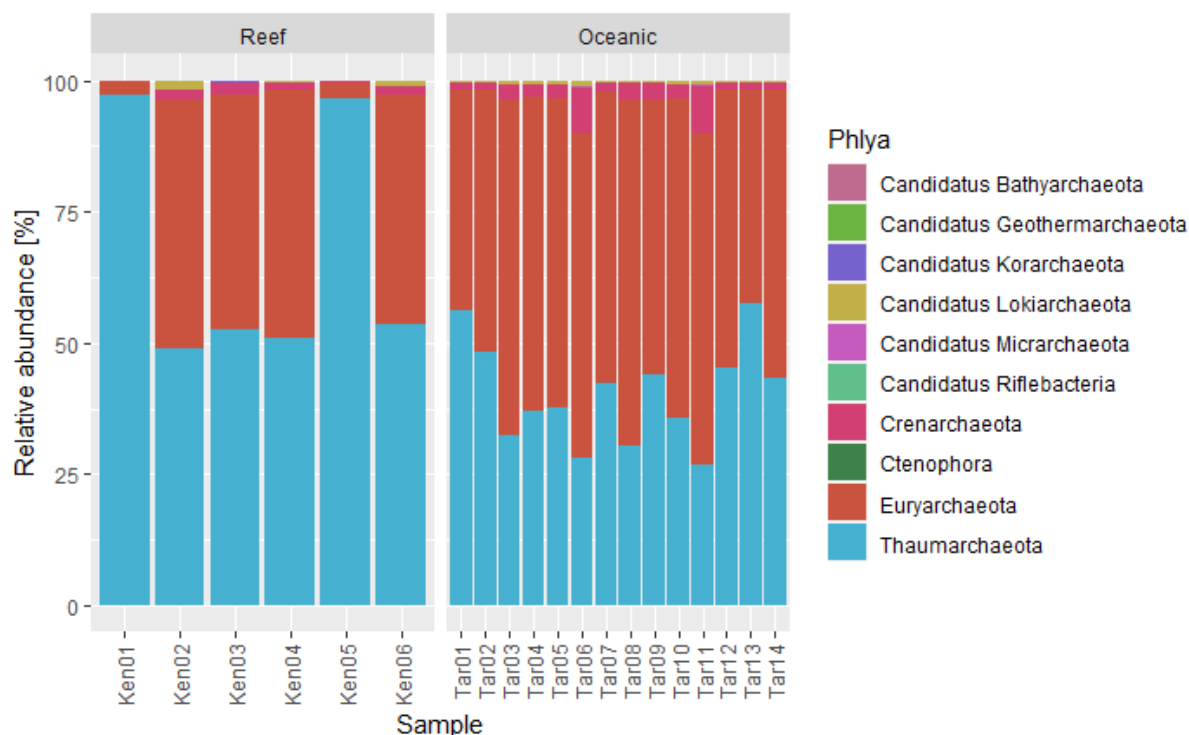


Figure 4.15 Relative abundance of archaeal phyla sequenced from Kenya coral reef (Ken*), and the *Tara* Oceans expedition (Tar*) samples

Furthermore 1,521 viral species, harboured in 15 phyla, were assigned to sequences from both oceanic and coral reef samples. Uroviricota dominated the sequences, particularly in the water samples (reef and oceanic) where its relative abundance was >90% (figure 4.17). In comparison, Uroviricota was relatively less abundant (31-65%) among reef sediment samples (Ken01, Ken03 and Ken05), where also the phyla Phixviricota and Cressdnviricota (which were less than 1% abundant in water samples) had relative abundance of 2-32% and 9-25% respectively.

Only two other phyla i.e. Nucleocytoviricota and Peploviricota had overall relative abundance of $\geq 1\%$.

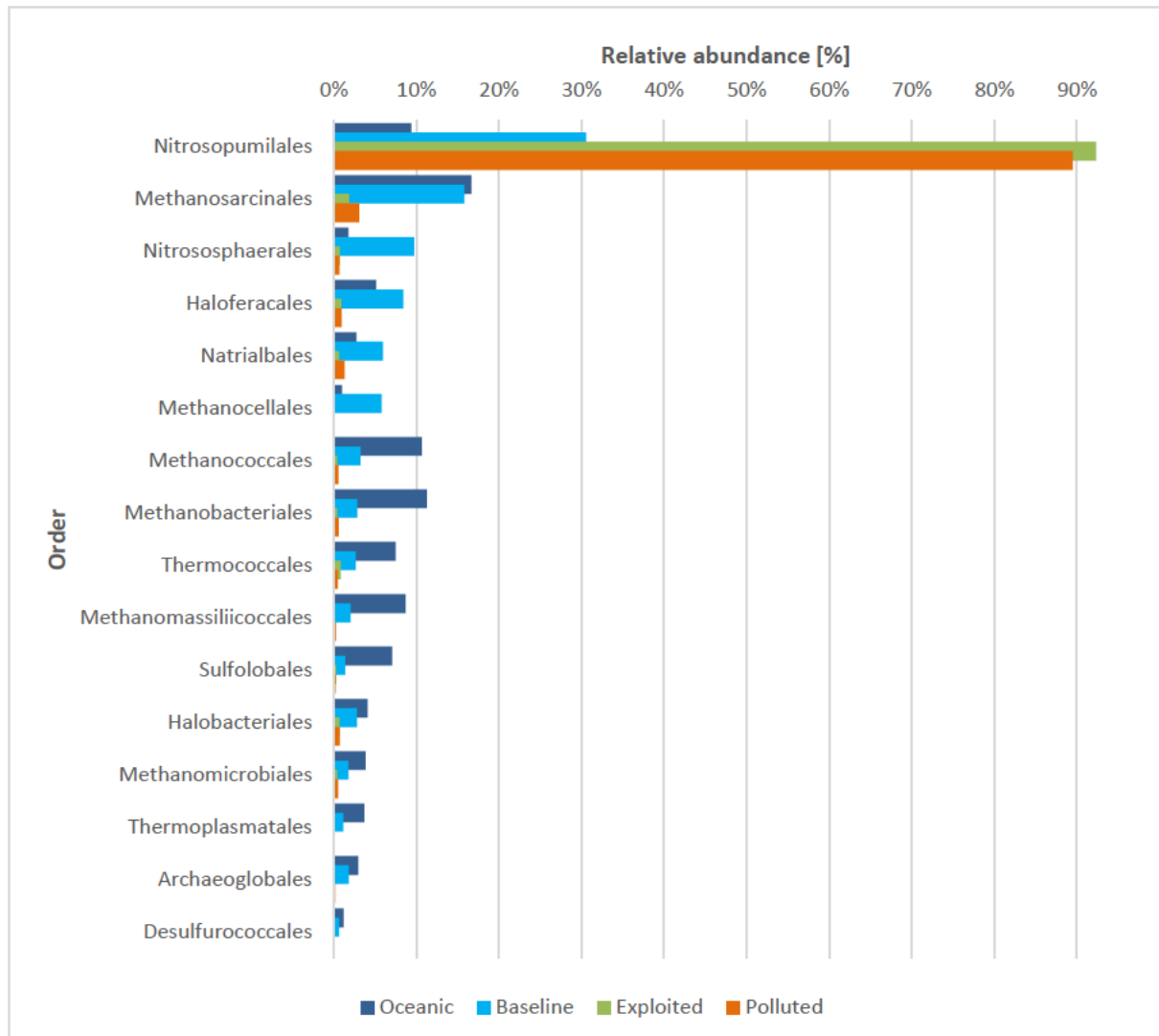


Figure 4.16 Distribution of archaeal orders with relative abundance $>1\%$ grouped by study site i.e. Oceanic (blue), and the coral reefs at sites classified “baseline” (sky), “exploited” (olive) and “polluted” (bronze)

Myoviridae was the most abundant (91%) virus family and had higher relative abundance in oceanic (91%) and “baseline” (82%) site sequences than in “exploited” (50%) and “polluted” (70%) site sequences. The other coliphage families, Podoviridae, Siphoviridae and

Microviridae, were also among the top ten most abundant families. In contrast to Myoviridae, the relative abundances of Podoviridae, Siphoviridae and Microviridae were generally higher in the “exploited” (2.4%, 2.7%, and 16.3%) and “polluted” (6.2%, 5.9%, and 5.5%) than in the “baseline” (2.0%, 1.6%, and <0.1%) and oceanic (0.8%, 0.6%, and <0.1%) site sequences.

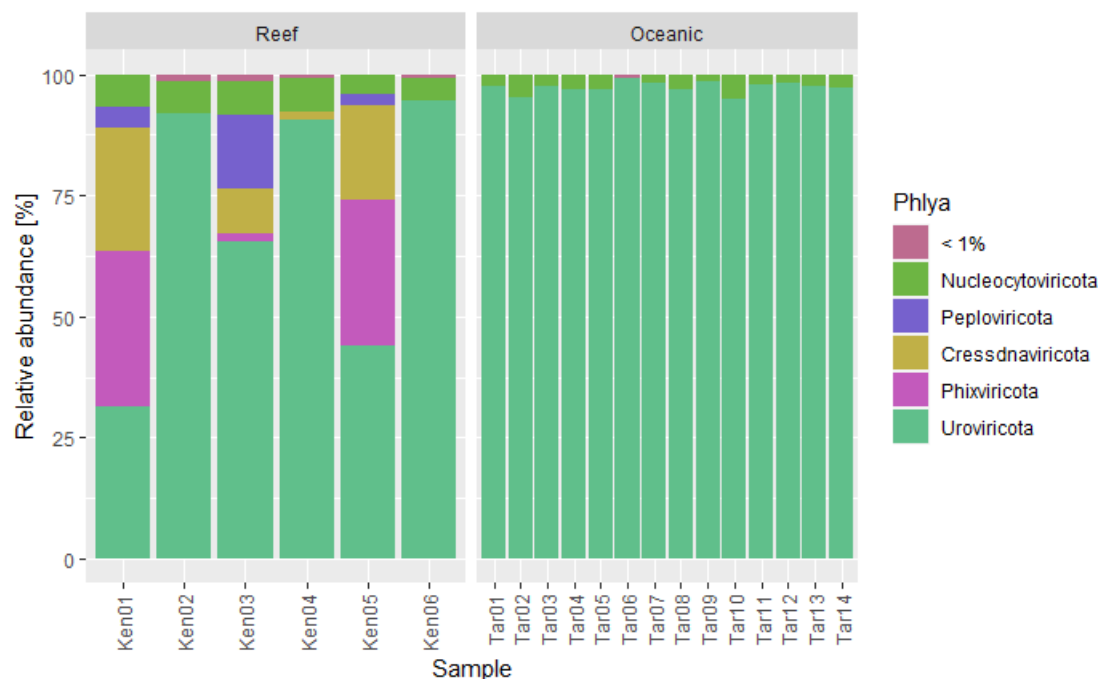


Figure 4.17 Stacked charts of the distribution of viral phyla with relative abundance > 1% sequenced from Kenya coral reef (Ken*), and the Tara Oceans expedition (Tar*) samples

Up to 96% of the relative abundance at the species level comprised of cyanophages particularly viruses for *Prochlorococcus* and *Synechococcus* whose relative abundances were consistently higher (up to 5-fold higher) in the oceanic and “baseline” site sequences than in either the “exploited” or “polluted” site sequences.

4.5 Discussion

Identification of patterns in community structure is increasingly recognised as an important initiative towards designing innovative environmental management strategies. Of the global oceans, the WIO and its marine ecosystems are the most poorly characterized with respect to

the associated microbial ecology. Although oceanic microbiome (Sunagawa et al., 2015) and virome (Williamson et al., 2012) have been explored, this is the first comparison with the WIO coral reef microbiome employing shotgun metagenomics approach.

Fewer reef than oceanic sequences were assigned most likely because of amplification of reef DNA samples with REPLI-g multiple displacement amplification (MDA) method which has been shown to produce fewer classified reads. Despite this limitation, REPLI-g MDA did not influence species abundance (Ahsanuddin et al., 2017), and considering that the level of saturation of rarefaction curves for both the oceanic and the reef samples were comparable, it is reasonable to make cautious comparisons of the resulting data.

Taxonomic composition, especially of the most abundant taxa, was largely consistent with other oceans ecosystems microbiota (Godoy-Vitorino et al., 2017; Santoro et al., 2019; Sunagawa et al., 2015; Williamson et al., 2012). The estimate for species richness was within the same order of magnitude as the previous metagenomic estimate for global oceans (Sunagawa et al., 2015). Estimated species diversity and evenness was higher in the reefs than in the oceanic samples. This is possibly due to the high primary productivity associated with coral reefs which has been shown to influence microbial diversity (Claire et al., 2003; Silveira et al., 2017). Also, coral reefs harbour up to 25% of the marine species each of which may influence their associated microbial species and also provide a range of habitat for the microbes (Bourne & Webster, 2013; Walsh et al., 2017). Although a substantial number of assignments were shared between the oceanic and the reef sequences, there were sequences unique to each ecosystem as well as significant differences in abundance and composition of all taxa assessed, reinforcing the thought that rather than every microbe being everywhere, they are selected or structure by the environment (Flaviani et al., 2018). Indeed, ecosystem or habitat (i.e. whether samples were collected from the oceanic or reefs) was estimated to account for over 40% of differences between microbial communities. The oceanic samples were dominated by typical bacterial and archaeal groups like Alphaproteobacteria, Epsilonproteobacteria, Cyanobacteria and Euryarchaeota that are known to thrive in oligotrophic ocean water columns (Overmann & Lepleux, 2016; Schöttner et al., 2011; Sunagawa et al., 2015; Wemheuer et al., 2019), along with the ubiquitous phages of the dominant bacterial groups (Biller et al., 2015; Vega et al., 2017; Williamson et al., 2012). On the other hand, samples collected within the coral reefs generally harboured microbial composition consistent with the typical taxa enriched by metabolic products of corals i.e. copiotrophic bacterial groups (including

Gammaproteobacteria, Betaproteobacteria and Bacteroidetes) as well as the ammonia oxidizing archaea (AOA) of the order Nitrosopumilales, and phages for the copiotrophic bacteria (coliphages) i.e. Podoviridae, Siphoviridae and Microviridae (Godoy-Vitorino et al., 2017; Kelly et al., 2014; Sunagawa et al., 2010; Vega et al., 2017; Weber et al., 2019).

There were also compositional differences between communities from the reef sequences which, though not statistically significant, may be subtle gradient signals of the human impacts in the coral reefs. For instance, most copiotrophic bacteria, especially the most abundant ones, had higher relative abundances in either the “exploited” or “polluted” than in the “baseline” site sequences. This is significant considering that copiotrophic bacteria thrive in high nutrient environments and include taxa utilized in water microbiology as faecal indicators (Buccheri et al., 2019). The genera *Bacteroides* and *Prevotella* especially, the alternative groups increasingly being exploited to replace the less reliable traditional faecal indicator bacteria, had higher relative abundance in both the “exploited” and “polluted” than in the “baseline” site sequences. Although coral reef seawaters are typically dominated by copiotrophic bacterial community due to increased organic matter from mucus and other nutrients released by corals (Bourne & Webster, 2013), between-sites differences observed in the bacterial relative abundances seemed to correspond to the varying magnitudes of human activities around the coral reef sites compared. As shown in chapter three, Malindi (the “exploited site”) and Mombasa (the “polluted site”) had higher nitrates, phosphates and ammonium concentrations, than Kisite-Mpunguti (the “baseline” site), though not statistically significant.

There were exceptions in the order Pelagibacterales as well the traditional faecal indicator bacteria genera *Escherichia*, *Salmonella* and *Campylobacter* whose relative abundance was highest in the “baseline” than in the “exploited” and “polluted” site sequences. Pelagibacterales belongs to the class Alphaproteobacteria whose ubiquitous members, although heterotrophs, are known to thrive at low nutrient concentrations typical of open ocean conditions (Kelly et al., 2014; Zhao et al., 2017). The higher relative abundance of the group in “baseline” than in the “exploited” and “polluted” site sequences may have been, therefore, indicative of differences in nutrient levels at the sites due to the surrounding human activities. It is not clear why relative abundances for *Escherichia*, *Salmonella* and *Campylobacter*, did not parallel human impacts gradient at the coral reef sites but again these genera are generally unreliable as faecal indicators because their presence does not correlate with counts of faecal coliforms (Alonso & Alonso, 1993; Fremaux et al., 2009; Schriewer et al., 2010).

Compared with bacterial indicators, coliphages i.e. viruses that infect coliform bacteria, have been reported to be better predictor of faecal pollution in seawater (Burbano-Rosero et al., 2011) because, they outlive their bacterial hosts, they can thrive in marine environments unlike some of their anaerobic hosts, they are more resistant to disinfection and diffuse further distances from pollution sources (Bari & Yeasmin, 2014; Ebdon et al., 2012). In this study, relative abundances of all coliphage families assessed, except one, were higher in the “exploited” and “polluted” site sequences than in the “baseline” site sequences. The coliphage family, Myoviridae, had higher relative abundance in the “baseline” site sequences than in the “exploited” and “polluted” site sequences. The cause for this exception is not clear but it may have been influenced by the relative abundance of *Escherichia* which was higher in the “baseline” site sequences and which members of the Myoviridae family are known to preferentially infect (Tao & Rao, 2019). Cyanophages too recorded relative abundances commensurate to the bacteria they infect, cyanobacteria, which in turn correlates inversely with nutrient levels; phages for *Prochlorococcus* and *Synechococcus* were up to 5-fold higher in the “baseline” site sequences than in either the “exploited” or “polluted” site sequences. *Synechococcus* and *Prochlorococcus* are considered the most important primary producers in the tropical oceans, responsible for a large percentage of the photosynthetic production of oxygen (Biller et al., 2015; Kim et al., 2018; Waterbury et al., 1979) and it is thought that their phages mediate their population sizes and evolutionary trajectories (Sullivan et al., 2005). Being autotrophs, members of these genera are usually found in great abundance in ocean zones low in nutrients (Dinsdale et al., 2008; Kelly et al., 2014). These findings call for further investigations to establish the stability of these trends over time and space for the potential utilization of these virus groups, especially the coliphages, as surrogates for human impacts in coral reefs.

Among the archaea groups, the AOA order Nitrosopumilales had the most obvious variation with human activities gradient; in comparison to the “baseline” site sequences, the order was three times higher in the “exploited” and “polluted” site sequences. Members of this order are key players in nitrification processes (Alvarez-Yela et al., 2019) that have been shown to seek nutrient-rich environments where they utilize urea and ammonia as substrate (Bayer et al., 2016). All the other orders with over 1% relative abundance were higher in the “baseline” than in the “exploited” and “polluted” site sequences.

Whether the compositional variations observed in this study are environmentally significant will need to be confirmed in more robust studies controlling for seasonal variations and spatial stability, but they point to the possibility of using microbial ecology as indicators of perturbations.

4.3 Conclusion

This chapter comprises the first coral reef microbial diversity assessment by metagenomic approach in the WIO. Significantly, the results established much-needed microbial baselines for the Kenyan Indian Ocean and potentially for the WIO region. Relative abundance variations of microbial communities between study sites suggestive of human impacts influence were observed. The sensitivity to environmental variations by coral reef water microorganisms, and therefore their potential use as disturbance indicators, is particularly underscored. Areas for future investigations are pointed to, including determining the environmental significance of the compositional variations observed here, broader context characterization of environmental conditions, and associations of these with the neighbouring human activities.

CHAPTER 5: FUNCTION POTENTIAL OF MICROBIAL COMMUNITIES ON KENYAN CORAL REEFS

5.1 Abstract

Variations in the microbial community composition and diversity may be a sensitive and accurate indicator for marine ecosystems health and predictors of responses to pressures. The search for microbial indicators of environmental stressors has traditionally been through a taxonomic approach. Functional diversity is increasingly recognized as the essential link between biodiversity patterns and ecosystem functioning and determining microorganisms' responses to environmental changes. Comparative genomics can reveal genotypes and other adaptive responses linked to ecosystem-specific selective pressures. Here a metagenomic study of seawater and sediment samples collected from Kenyan coral reefs (reef) experiencing varying human impacts was performed, together with metagenomes from the *Tara* Oceans expedition (oceanic). Functional annotation was done against the clusters of orthologous groups of proteins (COG) database and comparisons performed between reef and oceanic metagenomes as well as between the different human impacts on the coral reefs. More than 4,500 protein groups were annotated for potential functions which showed clear compositional separation between the oceanic and reefs. The Kenyan coral reef datasets had higher functional richness, diversity and evenness. *Tara* Oceans metagenomes were enriched with COGs normally overrepresented under oligotrophic conditions, and reefs metagenomes with COGs for functions related to adaptations to heterogenous environments. Malindi and Mombasa marine parks, the coral reef sites closest to densely populated settlements with human impacts, were especially significantly enriched with genes for functions suggestive of mitigation of environment perturbations including capacity to reduce intracellular levels of environmental contaminants and repair of DNA damage. This study establishes essential baselines for microbiome studies in the WIO region and provides insights to anthropogenic impact on microbial functions, and potential indicators of health status for marine ecosystems especially coral reefs.

Keywords: *microbiome, clusters of orthologous groups of proteins (COGs), molecular traits, human impacts, environmental stressors*

5.2 Introduction

Coastal ecosystems are some of the most dynamic and vulnerable environments under various pressures from anthropogenic activities and climate change. Coral reef ecosystems, particularly, are of interest because of their importance for biodiversity, their productivity, and their worrisome decline globally. It is widely accepted now that shifts in the composition and diversity of microbial communities may be a good indicator for marine ecosystems health and for predicting responses to stresses (Won et al., 2017). The search for indicators of coral reef ecosystem stressors has commonly been through taxonomic studies of biogeographic variation of marine microorganisms.

Microbial communities function in a highly interdependent manner (Gianoulis et al., 2009). To survive in the marine habitat, there are core (housekeeping) genes which are shared among numerous taxa. There are also additional non-taxa-specific genes (the pan-genome) associated with specialized pathways that contribute to fitness under particular local conditions (Kelly et al., 2014). These genes may be found in multiple taxa adapted to similar environmental conditions. Therefore, in the search for indicators of environmental disturbance, functional profiling rather than taxa classification might be more informative because similar community metabolism may comprise members of phylogenetically variable groups while communities comprised of similar taxa may differ in metabolic capabilities. Indeed, functional diversity has repeatedly been observed to predict ecosystem processes and properties better than taxonomic or phylogenetic diversity (Johnson & Pomati, 2020). For this reason, functional evaluations are becoming a mainstay of monitoring and management approaches (Bellwood et al., 2019).

Traditionally, functional assessments have been achieved through homology-based approaches in which reads are annotated by mapping them to orthologous genes of similar function, usually available in public reference databases. Technological advances in environmental genomics offer unprecedented alternatives of studying trait-based biogeography (Green et al., 2008). Studying covarying traits in an environmental context is essential to understand the ecological mechanisms that structure community composition (Barberán et al., 2012; Sunagawa et al., 2015). Metagenomics allows for community-level functional analysis which circumvents the confounding effects of horizontal gene transfer common in approaches analysing species or populations (Barberán et al., 2012). Metagenomic studies also allow for comparison of community genomes from multiple environments (comparative genomics) enabling the revelation of genotypes and other adaptive responses linked to ecosystem-specific selective

pressures. This approach has been used to determine variations in gene functions, and metabolic potential with ocean depth, and elevated phosphorus uptake genes in response to low phosphorus levels (Thompson et al., 2013). Furthermore, patterns of trait variation help provide explanations and generate hypotheses to fundamental biological phenomena. For instance, Red Sea microbiota have been hypothesised to have acquired specific environmental adaptation because they were observed to harbour increased numbers of genes involved in DNA repair, high-intensity light responses, and osmoregulation (Behzad et al., 2016).

Comparative metagenomics studies have led to the overarching hypothesis that microbial diversity is determined by the local environment i.e. unique environments harbour unique microbial species with unique genes and metabolic pathways. The WIO is an oceanic region in the warmest ocean (Indian Ocean) that is currently under pressure from global warming and increasing anthropogenic stressors, with unpredictable consequences, yet it is the least studied of all global oceans (Díez et al., 2016).

This study annotated the *Tara* Oceans and Kenyan coral reefs metagenomes analysed in chapter 4 and compared their functional diversity, between the two datasets as well as between the coral reef sites experiencing varying magnitudes of anthropogenic impacts.

5.3 Materials and methods

Functional analyses in this chapter used the reef and oceanic sequences used for Chapter 4 analyses. The reef sequences were generated from microbial communities inhabiting the coral reef waters and sediments at Malindi (“exploited”), Mombasa (“polluted”) and Kisite-Mpunguti (“baseline”) marine parks on the Kenyan Indian Ocean. The study sites are described in chapter three (section 3.3.1) while the specifics of sampling, DNA extraction, sequencing and quality control are detailed in chapter four (section 4.3.1 - 4.3.4). The oceanic sequences were publicly available metagenome sequences from samples collected by the *Tara* Oceans expedition (Sunagawa et al., 2015) in the open ocean within the WIO.

To characterize the gene content, quality-filtered merged reads were functionally annotated using the clusters of orthologous groups (COG) of proteins database (Tatusov et al., 2000; Tatusov et al., 2001) with a BLASTX search of reads vs the COG database applying MGX pipeline defaults (E-value cutoff $1e^{-5}$) (Jaenicke et al., 2018). Resulting COG output files were exported to the R statistical environment (R Core Team, 2020) for analysis and visualization.

The abundance of each COG was counted as the sum of reads mapping to it (Tatusov et al., 2000), which was then normalized by the size of the dataset. COGs were assigned into functional categories. Functional diversity was estimated by Shannon index based on COG richness and evenness (Johnson & Pomati, 2020). Bray-Curtis distance based on the relative abundances were calculated to represent functional compositional variation among the samples (Bray & Curtis, 1957), and principal coordinate analysis (PCoA) was used to visualize the relative differences. Normality of distribution of the relative abundances were assessed using the Shapiro-Wilk normality test. Differences in functional relative abundances were tested for statistical significance using Wilcoxon rank-sum test, for COG categories, and chi-squared test, for COG groups of proteins (p value < 0.05 as significance cut-off). False Discovery Rate (FDR) was corrected for using the Benjamini-Hochberg's method (Benjamini & Hochberg, 1995). Comparison of beta diversity between groups was assessed by permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2001) using *adonis* test based on Bray-Curtis distances with 999 permutations.

5.4 Results

A total of 4,656 clusters of orthologous groups of proteins (COGs) belonging to 24 COG categories were classified. Overall, the categories of amino acid metabolism and transport (category E), general functional prediction only (category R), energy production and conversion (category C), and translation (category J) had the most relative abundance together constituting ~42% of all sequences. In contrast, categories of nuclear structure (category Y, <0.01%), RNA processing and modification (category A, 0.02%), cytoskeleton (category Z, 0.05%), and chromatin structure and dynamics (category B, 0.08%) had the least relative abundances. Relative abundances were significantly different between oceanic and reef sequences in 14 of the 24 categories (marked with asterisks in figure 5.1). Of the 14 categories, sequences from the Kenyan reef samples had higher relative abundances, than oceanic sequences, in the categories of general functional prediction only (category R), inorganic ion transport and metabolism (category P), signal transduction (category T), secondary structure (category Q), intracellular trafficking and secretion (category V), cell motility (category N), and chromatin structure and dynamics (category B).

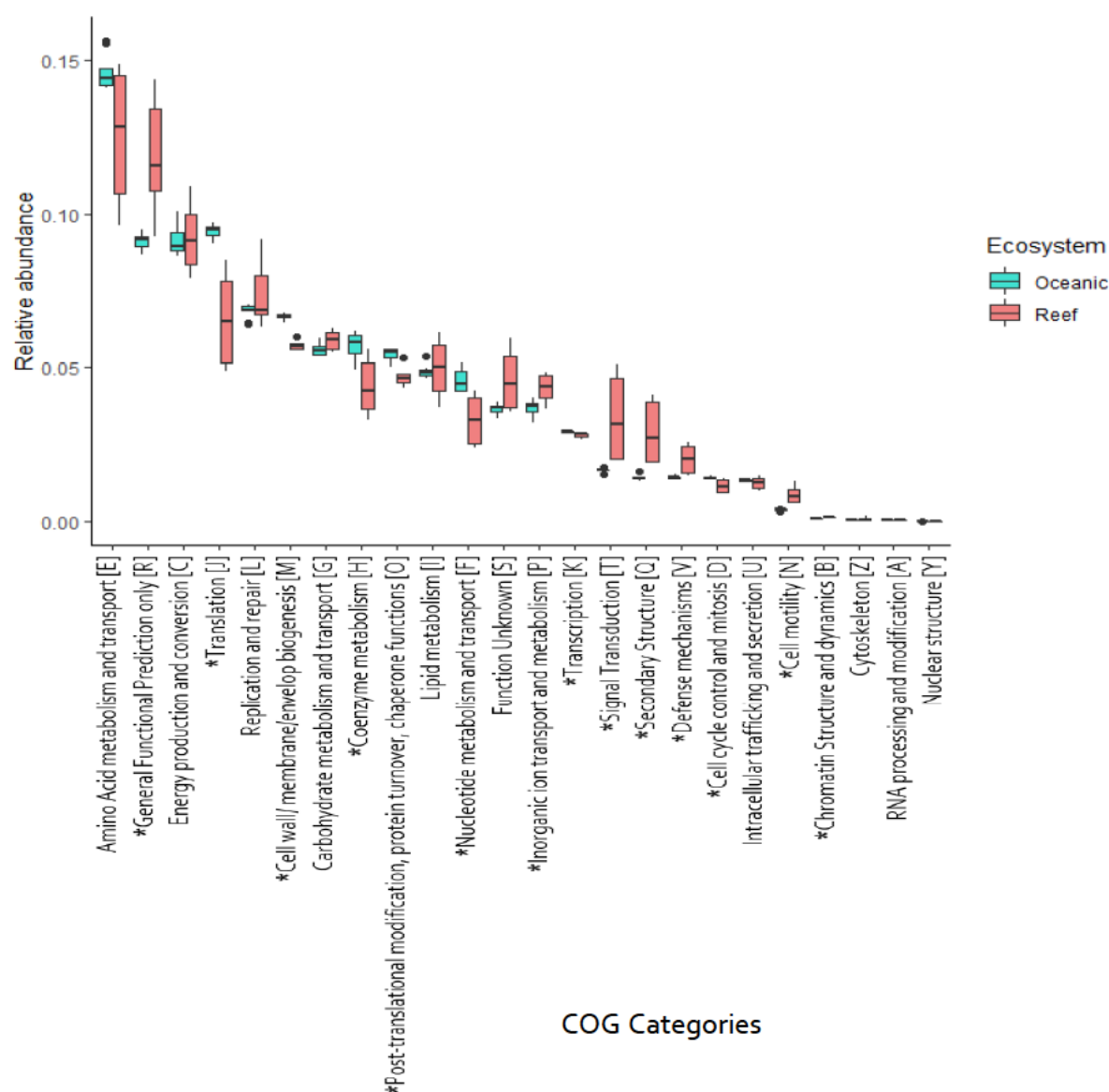


Figure 5.1 Relative abundances of COG functional categories. Asterisks denote categories with significant differences in abundance.

Between sequences from Kenyan coral reef samples (figure 5.2), no statistically significant differences in relative abundances were observed, except for subtle variations in a number of COG categories including energy production and conversion (category C), translation (category J), and signal transduction (category T) where the “baseline” site sequences seemed to have higher relative abundance compared to the “exploited” and “polluted” site sequences, and categories of general functional prediction only (category R), replication and repair (category L), lipid metabolism (category I), coenzyme metabolism (category H), and secondary structure (category Q) where both “exploited” and “polluted” site sequences had higher relative

abundances than the “baseline” site sequences.

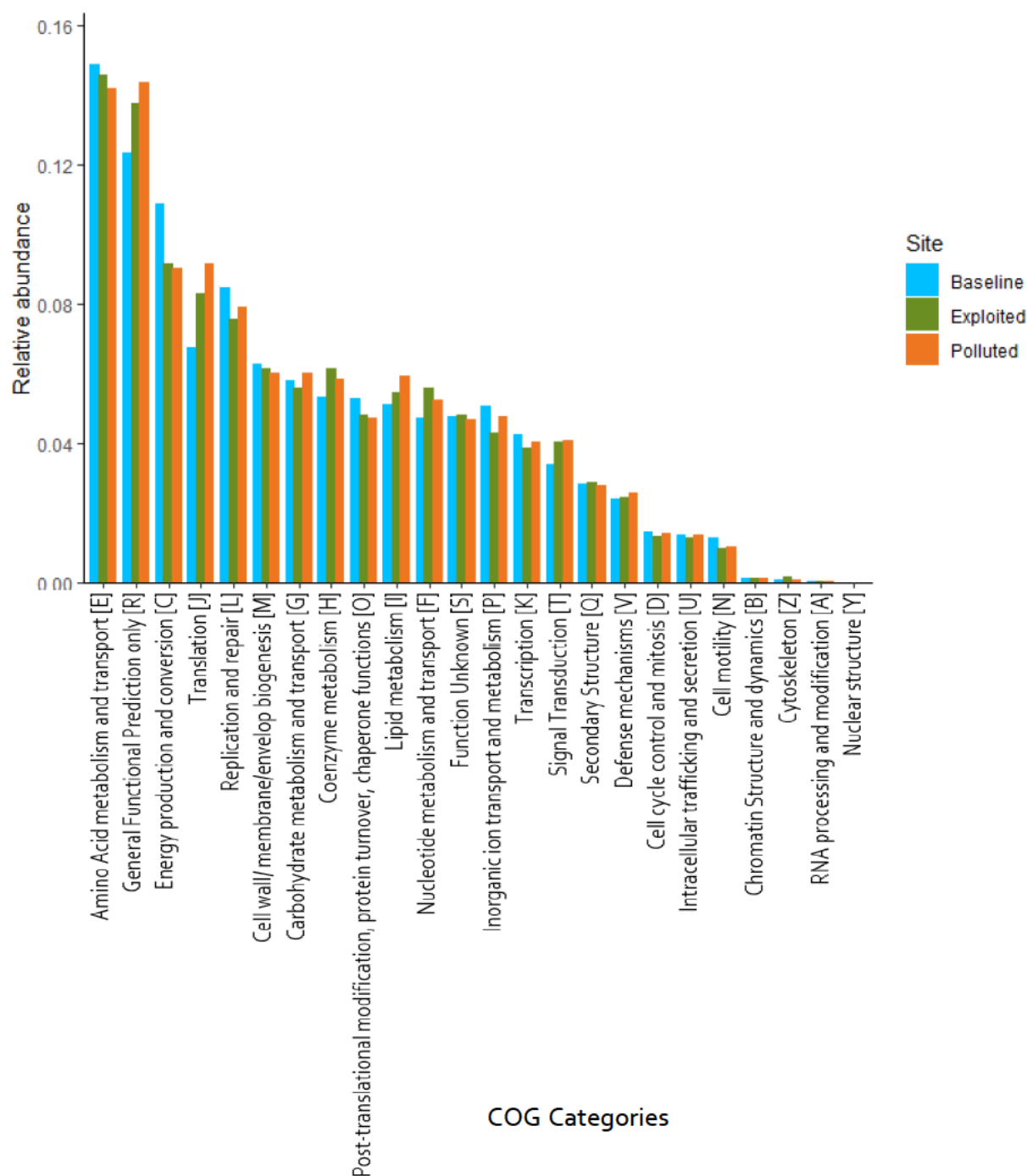


Figure 5.2 Distribution of COG categories in sample sequences from the Kenya coral reefs with varied human impacts i.e. “baseline” (sky blue), “exploited” (olive) and “polluted” (bronze).

The rarefaction curves (figure 5.3) shows the number of prevalent COGs identified with increasing sample number. The extent of sequencing showed that most prevalent COGs were recovered.

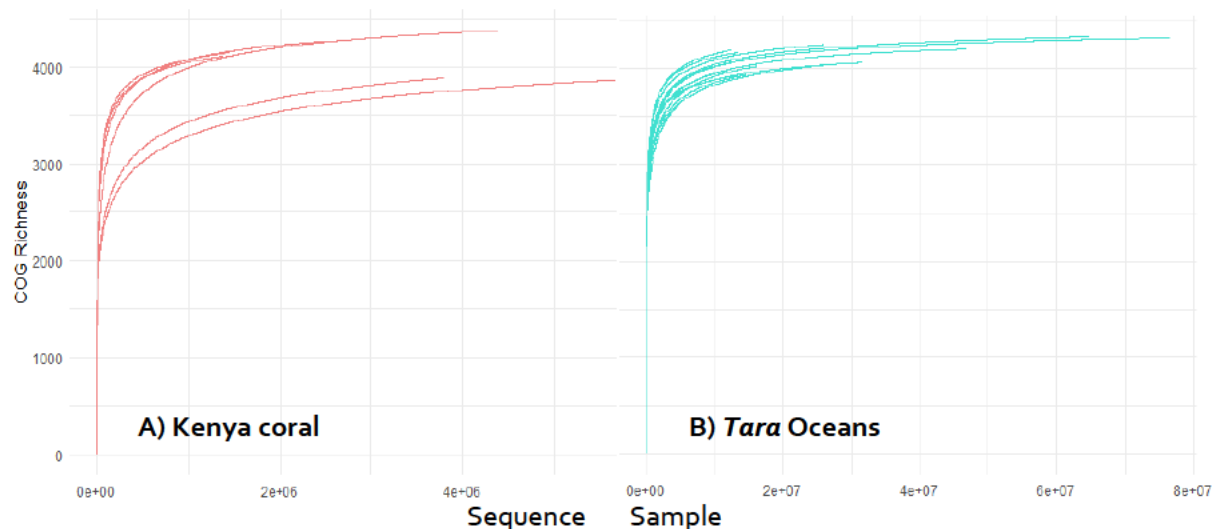


Figure 5.3 Rarefaction curves for clusters of orthologous groups of proteins (COG) richness on the sequences from Kenya coral reefs (reef) (A) and *Tara* Oceans (oceanic) (B) samples

Sequences from the reef samples had significantly higher number of predicted functional protein groups, 4,584 COGs versus 4,449 ($p < 0.001$), which were more diverse and even, Shannon index = 7.18 versus 7.03 ($p < 0.001$) compared to the oceanic sequences. Significant compositional differences of COGs relative abundances between the oceanic and reef sequences ($p < 0.001$) were observed, with PCoA attributing over 66% of the variation to the ecosystem (Figure 5.4). Oceanic sequences were clustered and did not overlap with the reef site sequences which, in turn, were more dispersed.

Average per sample COG richness was also higher in the reef samples $4,111 \pm 205$ than in the oceanic samples $4,074 \pm 182$ ($p < 0.001$). The more highly represented COGs in all metagenomes corresponded to dehydrogenases with different specificities (COG1028), NAD-dependent aldehyde dehydrogenases (COG1012), and nucleoside-diphosphate-sugar epimerases (COG0451). Significant relative abundance differences between oceanic and reef

metagenomes were noted in nine of the 24 most abundant predicted protein groups (marked with asterisks in figure 5.5). The predicted protein groups with significant relative abundance differences were higher in oceanic than reef metagenomes, except for signal transduction histidine kinase (COG642) whose relative abundance was more than twice higher in reef sequences compared to the oceanic sequences.

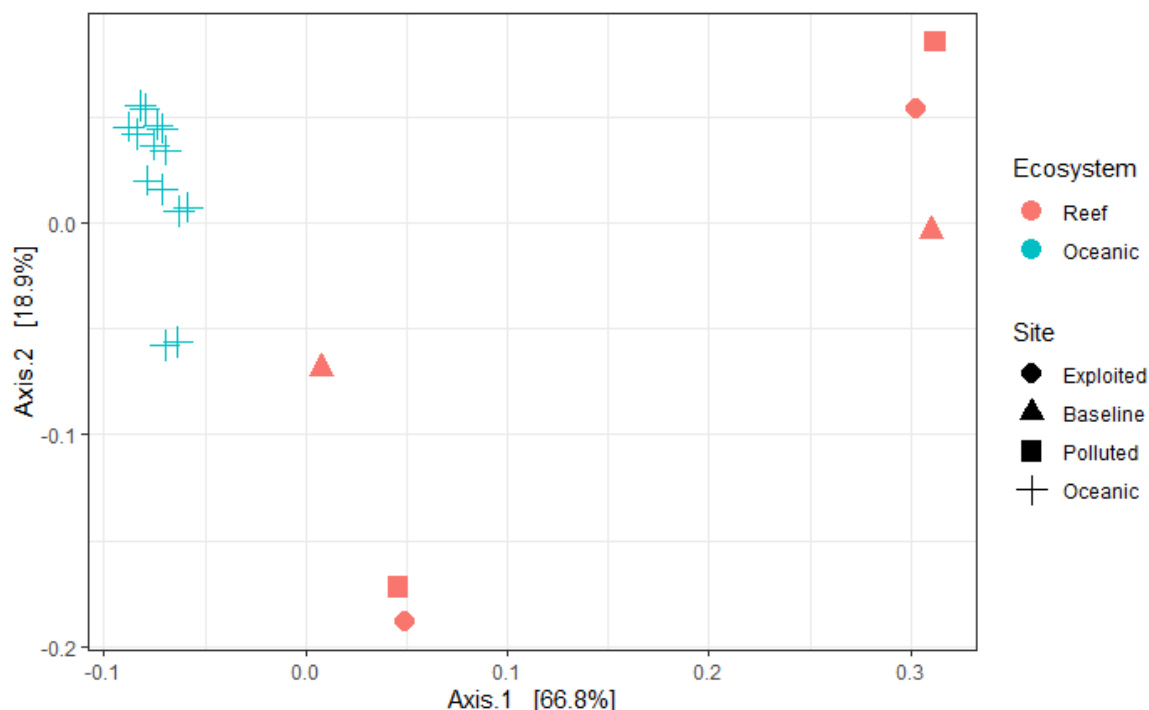


Figure 5.4 Composition of clusters of orthologous groups of proteins (COGs) analysed with Principal Coordinate Analysis (PCoA) plots of Bray–Curtis dissimilarity matrix for reefs and oceanic metagenomes.

Statistical analysis based on the COG relative abundances revealed that 1,078 COGs were significantly enriched in oceanic metagenomes the most abundant of which included predicted proteins for amino acid transport and metabolism (COG1008, COG0069, COG1003, COG0458, COG0028 and COG0147) and transcription (COG0085, COG0086 and COG0568). In comparison, 2,812 COGs were significantly enriched in reefs. The 40 most abundant COGs significantly enriched in the reefs are shown in figure 5.6 which were dominated by predicted proteins for replication, recombination and repair (COG0675, COG1943, COG2801,

COG3293, COG3316, COG3335, COG3344, COG3385, COG5421, COG5433, COG5655 and COG5659), and energy production and conversion (COG1018, COG1151, COG1882 and COG2414).

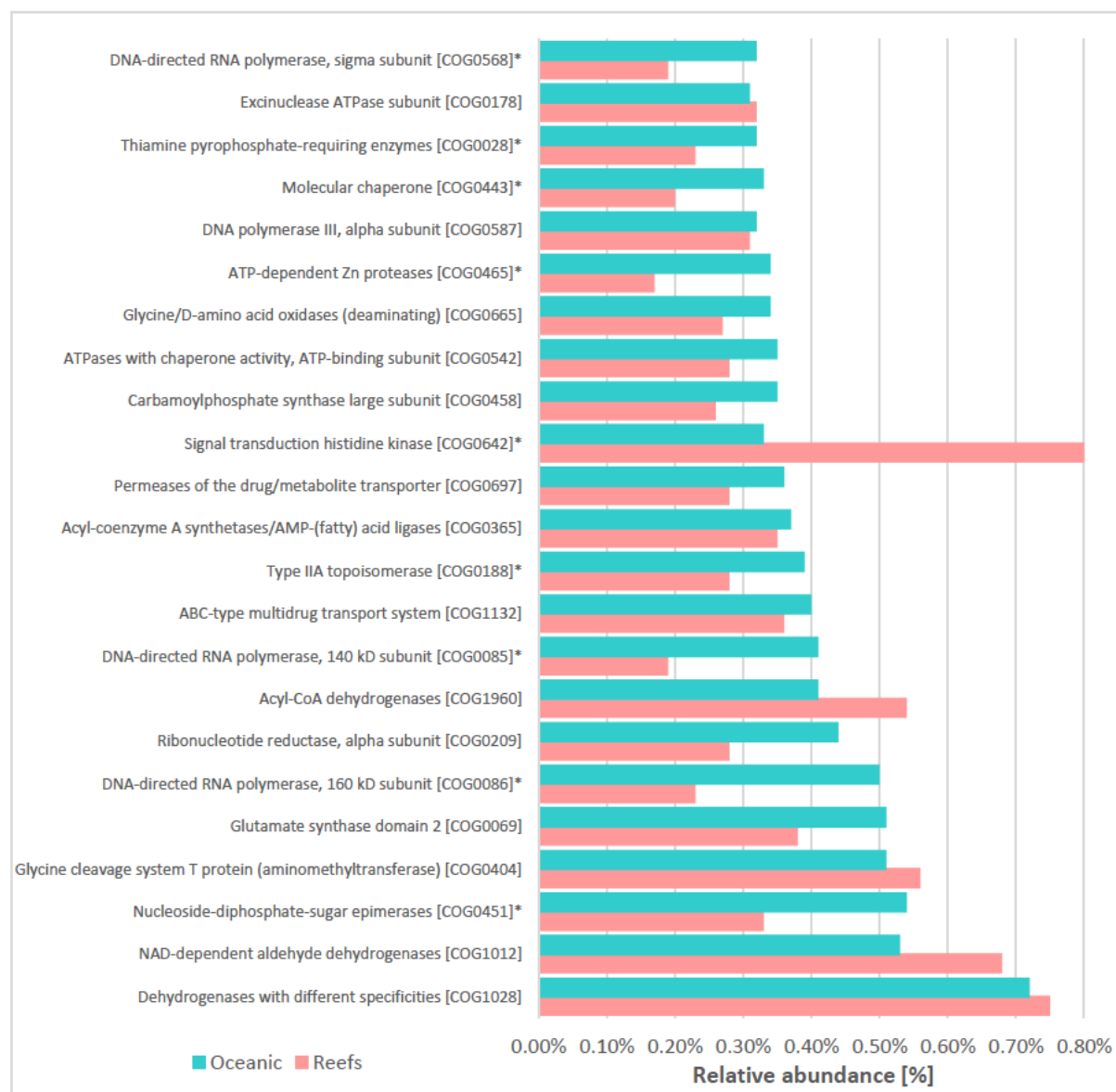


Figure 5.5 Distribution of the 24 most abundant COGs (~10% of total reads) in oceanic and reef metagenomes. Asterisks denote COGs with significant differences in relative abundance between the datasets.

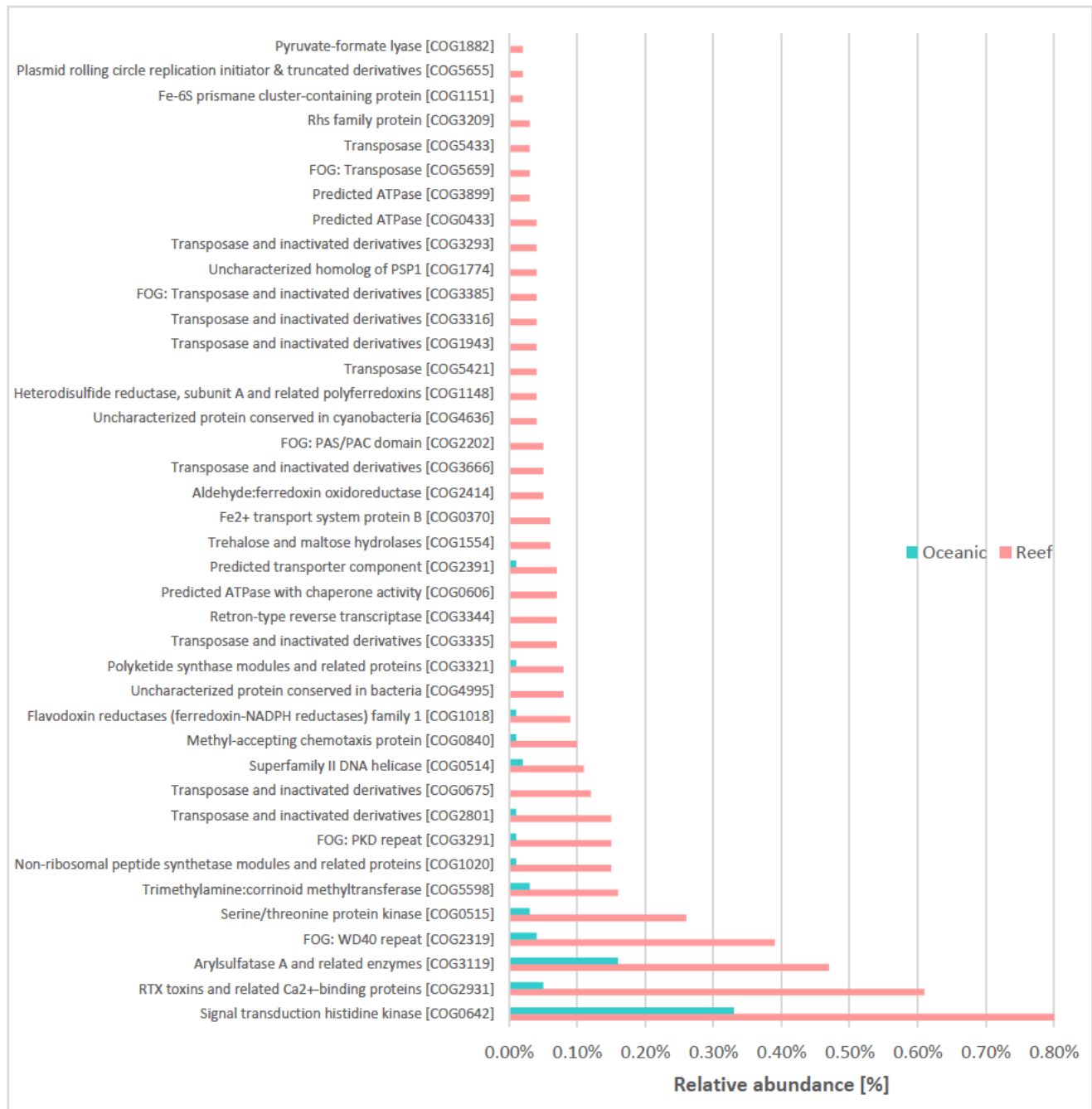


Figure 5.6 Comparison between reef-specific and oceanic-specific relative abundances of the 40 most represented COGs significantly enriched in the Kenya coral reef metagenomes

Among the Kenyan coral reef metagenomes, the “baseline” site sequences recorded higher number of observed COGs, 4,517 than the “exploited”, 4,283, and the “polluted”, 4,289 ($p = 0.207$), site sequences, which were also more diverse and even, Shannon index = 7.19, compared to 7.13 and 7.15 ($p = 0.822$), respectively (figure 5.7).

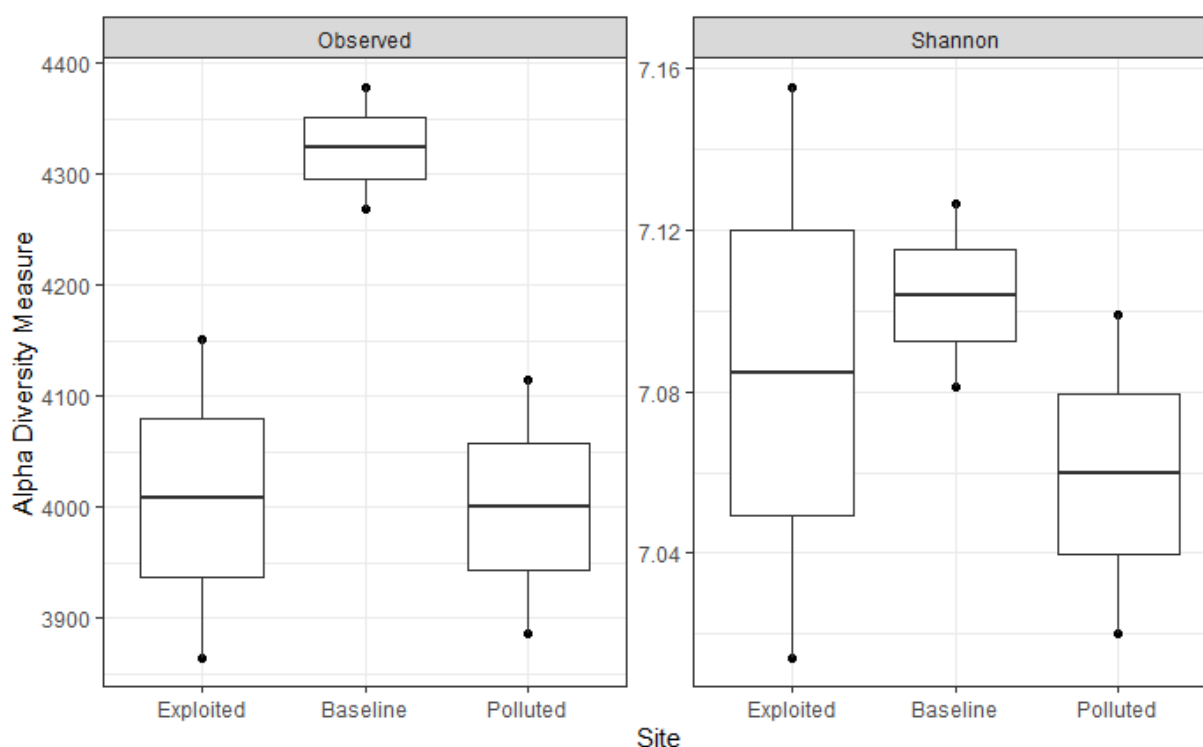


Figure 5.7 Plots of alpha diversity metrics for predicted COGs richness on metagenomes from the Kenya coral reef sites

The “baseline” sequences had 2,153 COGs that were significantly enriched, the most abundant of which were functions for energy production and conversion including COG2414, COG0674, COG1013, COG1145, COG1148 and COG3808. In contrast, 565 COGs were significantly enriched in the “exploited” and “polluted” metagenomes. In these sites, the most abundant COGs (figure 5.8) were dominated by predicted proteins related to defence mechanisms and removal or sequestration of unwanted compounds (e.g. COG1132, COG0534, COG3491 and COG0488), replication, recombination and repair (e.g. COG0514, COG1193, COG1793 and COG1201), and inorganic ion transport and metabolism (e.g. COG0659, COG4771 and COG4772).

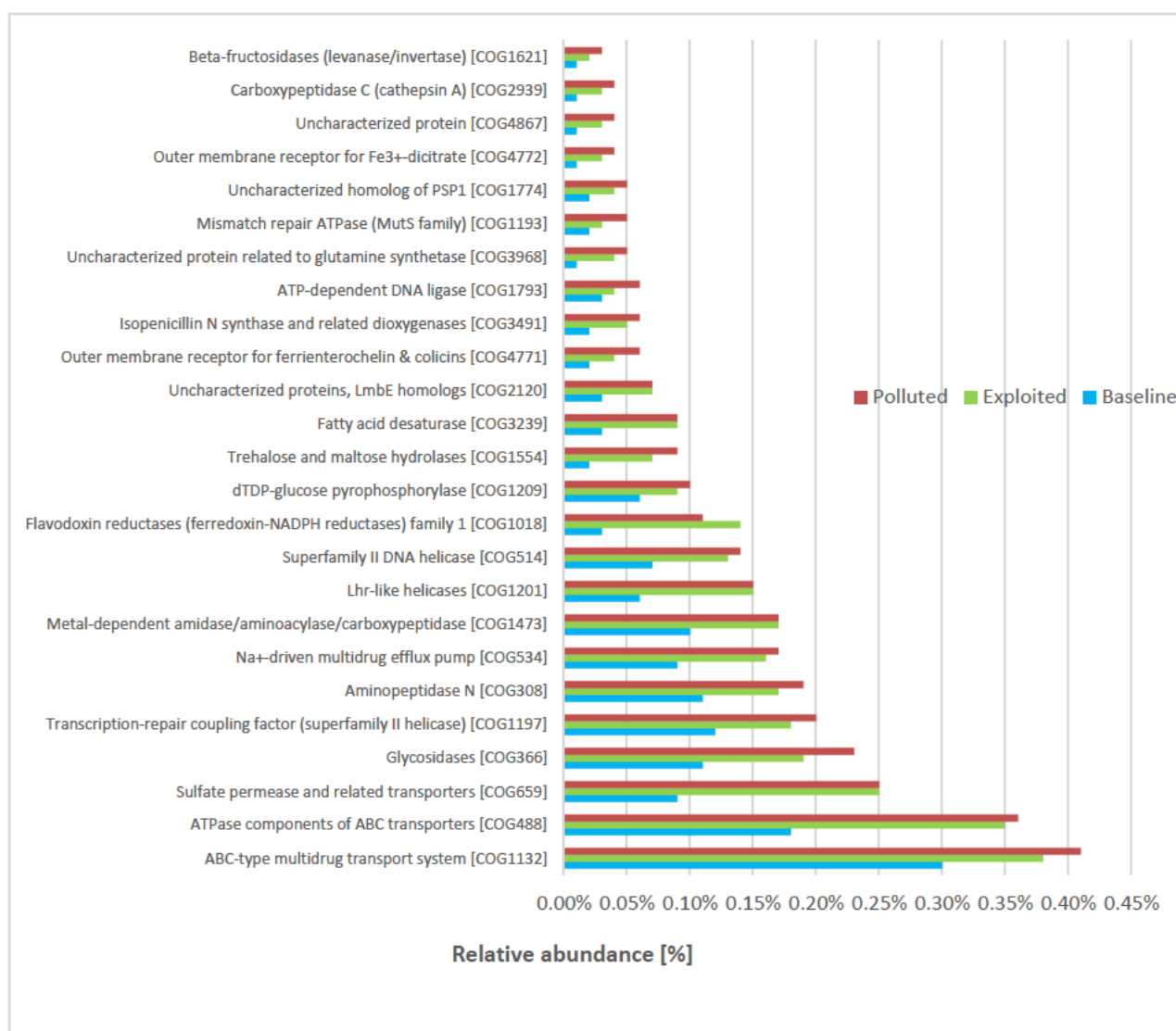


Figure 5.8 Comparison between Kenya coral reef sites of relative abundances of the 25 most represented COGs that are significantly enriched in the “exploited” and “polluted” sites.

5.5 Discussion

Characterization of coral reef microorganisms is essential in helping to clarify their ecological roles and the environmental conditions they respond to. Metagenomics provides the opportunity to access microbial community taxonomic and genomic content, as well as their functional potential in an ecosystem. The functional structure of an assemblage is considered a key indicator of the status of an ecosystem (Bellwood et al., 2019). For instance functional category profile by COGs, has been adopted by the Genome Standards Consortium as a useful

comparative genomic feature essential for newly sequenced genomes (Galperin & Kolker, 2006; Galperin et al., 2019). In this project, metagenomes from Kenyan coral reefs and from the open WIO collected by the *Tara* Oceans expedition, were classified into COGs and analysed to compare the functional potential of microbes between the two ecosystems, as well as between the coral reefs exposed to different human activities.

The COG richness observed in this study was slightly higher but comparable to the others previously reported by studies of global oceans and coral ecosystems (Carlos et al., 2016; Varasteh et al., 2020; Zhang et al., 2020). Significant compositional differences were observed between the reef and oceanic datasets. Compared to the *Tara* Oceans metagenomes, the Kenyan coral reef datasets had higher functional richness, diversity and evenness. Similar comparison was observed in the taxonomic profiles (Chapter 4), in-keeping with experiments that have demonstrated correlations between taxonomic and functional diversity (Hooper et al., 2002). These similarities suggests that the influence of coral reefs' primary productivity on taxonomic richness and diversity (Claire et al., 2003; Silveira et al., 2017) may extend to the microorganisms' functions. Differences in relative abundances may have been due to environmental structuring as principal coordinates analysis (PCoA) revealed clear separation of the *Tara* Oceans from the Kenyan coral reef COGs with PCo1 explaining 66.8% of the variability.

As for differential COG enrichment, predicted proteins for amino acid transport and metabolism were significantly overrepresented in the *Tara* Oceans metagenomes. This is in agreement with the most abundant marine microbial communities, especially members of class Alphaproteobacteria, that express disproportionate abundance of high-affinity transporters under oligotrophic conditions (Herlemann et al., 2019). Signal transduction histidine kinase which enables microbes to sense, respond, and adapt to a wide range of environments, stressors, and growth conditions (Skerker et al., 2005) were significantly enriched in the reef metagenomes. Furthermore, the coral reef sequences had overrepresentation of predicted proteins related to replication, recombination and repair especially transposases that have been implicated in microbial adaptive evolution to local conditions in diverse environments (Vigil-Stenman et al., 2017), as well as proteins for energy production and conversion consistent with adaptation for high nutrient uptake and synthesis needed in low nutrient environments (Gudhka et al., 2015). Although most of these predicted proteins are otherwise common in microbial communities in varied environments, their differential representation especially across the

human impacts' gradient in the Kenyan coral reefs suggests heterogeneity, possibly influenced by human activities.

Among the coral reef sites sequences from the “baseline” site, Kisite-Mpunguti marine park, were enriched with essential protein groups for functions related to uptake and synthesis of nutrients and transcription. Conversely, metagenomes from Malindi (“exploited”) and Mombasa (“polluted”) marine parks had overrepresentation of COGs suggestive of mitigation of environment stressors. Here the enriched COGs included, for instance, proteins related to defence mechanisms such as transporters that prevent intracellular accumulation of toxic compounds (Wilkens, 2015), which serves as the major defence mechanism against antimicrobial compounds (Lubelski et al., 2007). A recent study assessing antibiotic resistance genes along a pollution gradient also found highest abundance of transporters in the most polluted site (Chen et al., 2019). COGs for DNA replication, recombination and repair were also significantly overrepresented suggesting exposure to agents of DNA damage (Buckley et al., 2020; Zhang et al., 2020).

It is potentially significant that the genes that ward off environmental insults are significantly enriched in the metagenomes from Malindi and Mombasa marine parks, the sites with known degradative human activities ashore, and not in Kisite-Mpunguti marine park which is situated farther from human settlements thereby experiencing limited impacts. These observations provide insights into the functional shifts contributed by human impacts, and potential COGs that may be utilized as indicators of marine health generally, and coral reef specifically. However, being DNA-based observations, these findings need validation by transcriptomics and proteomics approaches to confirm the actual proteins that are overexpressed in the respective sites.

5.6 Conclusion

Metagenomics provide efficient strategy to predict microbial function in an ecosystem and identification of novel metabolic pathways. This study has generated invaluable novel data, from the functional annotation of oceanic and coral reef microbiomes of the WIO, which will serve as reference of future studies.

Differences in COG relative abundances across the sampling sites suggests an environmental role in structuring microbial functions which gives support to the usefulness of metagenomics

in identifying microbial indicators of ecosystem status. The preferential overrepresentation, at the “exploited” and “polluted” sites (and not the “baseline” site), of genes related to the capacity to reduce intracellular levels of environmental contaminants and to DNA repair is especially significant as it provides potential COGs that may be exploited as indicators for rapid monitoring of coral reef health.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER RESEARCH

6.1 Introduction

Worldwide, the WIO is the second most diverse hotspot of coral reef biodiversity, hosting 16% of the world's corals. WIO coral reef ecosystems form a critical component of the region's economies; providing habitat, spawning and nursery grounds for economically important fish species, coral reefs are essential sources of protein, food, and livelihood for coastal communities (Obura, 2017; Van der Elst et al., 2005). Directly, coral reefs also provide jobs and income from fishing, recreation, and tourism (Halpern et al., 2012), as well as protection of coastlines from hazards such as storms and erosion through dissipation of ocean wave energy by the rough surfaces of corals (Ruckelshaus et al., 2013).

Kenyan coral reefs, which consists of more than 250 coral species (Obura, 2012; Tuda & Omar, 2012), have enormous unexplored economic potential (Obura et al., 2017; Ransom & Mangi, 2010). The economic potential has particularly come to attention recently through government institute frameworks to maximize the blue economy as a strategy to achieve its vision 2030 development agenda (Wairimu & Khainga, 2017). This potential is increasingly at risk of being lost due to worldwide coral reefs degradation as a result of threats of anthropogenic origin at global and local scales. Stressors include climate change and ocean acidification globally, as well as overexploitation, nutrient pollution, and destructive fishing methods, locally (Emerton & Tessema, 2001; Fleming et al., 2006; Tuda & Omar, 2012). As populations grow and coastal development increases, these pressures are expected to get worse (Mora, 2008).

Government administered marine protected areas (MPA) and locally managed marine areas (LMMAs) are employed in Kenya, and the WIO region generally, as the main strategy for protecting marine biodiversity including coral reef ecosystems (Tuda & Omar, 2012). This approach has had great success in improving fish stocks (Ransom & Mangi, 2010; Samoilys & Obura, 2011), but it is feared that MPAs and LMMAs may not be effective in mitigating against the indirect impacts of human activities such as pollution, sedimentation and coastal development (Mora et al., 2006). There is therefore an urgent need for innovative interventions to protect, conserve and restore coral reef ecosystems.

Microorganisms wield considerable potential as a monitoring tool for coral reef ecosystem health for several reasons. They are ubiquitous in the ocean environment and play fundamental roles in the health, protection and functioning of corals and reef ecosystem homeostasis (Glasl et al., 2019; Godoy-Vitorino et al., 2017). Moreover, microorganisms respond and adapt quickly to shifts in environmental conditions (Ainsworth et al., 2010) through modification of their communities to exploit available resources. Microorganisms can, therefore, be sensitive and rapid markers for ecosystem stress because variations of their community compositions with environmental variables may be predictive of responses to stressors impacting coral reefs.

Modern advances in community DNA sequencing and computational analysis offer effective means of assessing marine microbial structural patterns, diversity, and functional potential (Sunagawa et al., 2015). Metabarcoding, the rapid simultaneous identification of species by sequencing a marker gene from an environmental sample, is a commonly used approach for determination of taxon richness and relative abundance of different taxa in a microbial community (Konopka, 2009), due to its cost effectiveness. On the other hand, metagenomics – where whole genomes in an environmental sample are sequenced – allow for the assessment of both taxonomic and genomic content of microbial communities (Sleator et al., 2008). Comparative genomics i.e. the comparison of community genomes from multiple environments, can reveal microbial adaptive responses to ecosystem-specific stressors. This is an innovative functional analysis alternative to the traditional homology-based approach.

Other molecular methods exist which complement these microbial genomic approaches. For instance, the potential functional repertoires generated by metagenomic assessments can be confirmed by a metatranscriptomics approach, which reveals expressed genes, or metaproteomics, which catalogues synthesized proteins from sequences of environmental samples. Furthermore, once microorganisms or genes of interests are identified through metabarcoding or metagenomics assessments, marker sequences (barcode) can be determined and used to design cost effective PCR-based methods for rapid detection. One such application is the microbial source tracking (MST) PCR for identifying pollution sources in aquatic environments by primers that detect and quantify markers of host specific bacteria members of the *Bacteroidales* which are abundant in the intestines and faeces of warm blooded animals (Bernhard & Field, 2000; Odagiri et al., 2015).

Application of genomic approaches to marine research has generally been slow in the WIO, the least studied of the global oceans with regard to its microbial ecology (Díez et al., 2016). The widely cited reasons for this are the lack of infrastructure, funding and trained personnel (Karikari et al., 2015). However, the WIO region, and Africa generally, has a reasonable level of uptake of genomic and molecular methods employed mainly for biomedical and agricultural research. Low genomics literacy and lack of meaningful collaborations are equally important explanations why the WIO region has not adopted microbial genomics studies. The WIO region therefore lacks coral reef microbial baselines which hinders the ability to study links between microbial composition variations and ecosystem pressures. This project employed genomic and molecular approaches to assess impacts of human activities on the taxonomic and functional diversity of microbial communities inhabiting the Kenyan coral reefs and the WIO open ocean. This chapter summarises the key findings of the study, reviews the implications of the research and highlights opportunities for future research.

6.2 Aims and objectives

The aim of the project was to use a metagenomic strategy to assess the microbiomes of the Kenyan coral reefs. Compositional and functional shifts of coral-associated microbial communities vary along gradients of anthropogenic impact and with changes in water quality (Angly et al., 2016; Kelly et al., 2014), indicating that links between microbes, human activities and environmental variables are correlated. Between-site differences in microbial community structure and molecular traits would be identified and matched with the neighbouring human activities as well as the local environmental variables to determine whether such variations could be used to predict coral reef health and response to perturbations.

To achieve the above aim, the study set specific objectives to: i) identify coral reefs at locations neighbouring settlements with varied human activities, and determine their characteristics with respect to physicochemical and biological variables, ii) determine the taxonomic diversity of coral reef seawater and sediment micro-organisms by massively parallel sequencing strategies, and iii) assess the functional potential of coral reef water and sediment micro-organisms by metagenomic sequencing.

6.3 Summary of major findings

Literature review, in chapter 2, revealed the WIO mangrove forests, seagrasses and coral reefs to wield vast yet threatened ecological and economical potential whose management and

sustainable utilization could benefit from application of genetics and genomic sciences. Compared to biomedical and agricultural research, marine research showed the lowest application of genomics in the region judging from publication outputs, with inadequate training of marine scientists and lack of meaningful collaborations being highlighted as some of the key possible explanation for this.

Three coral reef sites were identified which were each managed with varying degrees of protection creating a gradient of impact of human activities categorized as “strictest” (Kuruwitu), “moderate” (Malindi) and “least” (Kilifi). These sites were chosen due to their proximity to the laboratory and were used for methods optimization, and for assessing the effect of human activities on *E. coli* density and nutrient concentrations. *E. coli* density decreased with increasing degree of marine protection while nutrients showed no obvious trend. Physicochemical parameters and nutrients concentrations were also determined for three more coral reefs, which were identified with the guidance of KMFRI. These were situated within three of the six Kenyan MPAs adjacent to settlements with varied human activities categorized as “exploited” (Malindi), “polluted” (Mombasa) and “baseline” (Kisite-Mpunguti). Microbial source tracking by PCR assays were optimized and employed on these sites to detect and identify source of faecal pollution. Physicochemical and nutrient variables were found to be within the normal range for tropical coral reefs and did not differ across study sites. Furthermore, faecal bacteria were not detected in any of the tested coral reef samples. These results are detailed in Chapter 3.

Chapter 4 focussed on the metagenomic assessment of microbial taxonomic diversity of the Kenyan coral reefs in Malindi, Mombasa and Kisite-Mpunguti marine parks. Metagenomes from Tara Oceans were also incorporated into the analysis for comparison. At the phylum level Proteobacteria, Bacteroidetes and Cyanobacteria were the most dominant taxa, consistent with previously observed distribution in other global oceans. Compared to Tara Oceans the reef samples recorded higher taxonomic diversity and evenness. Variations in the relative abundances of copiotrophic bacteria and coliphages, were observed across the coral reef sites corresponding to human impacts gradient.

For Chapter 5, functional annotation was done for the datasets analysed in chapter 4, sequences from coral reefs experiencing varying human impacts and metagenomes from the Tara Oceans expedition, and comparisons done. Over 4,500 gene groups were assigned which showed clear

compositional separation between the oceanic (*Tara*) and reef metagenomes. The Kenyan coral reef datasets had higher functional richness, diversity and evenness. *Tara* Oceans metagenomes were enriched with group of genes for functions related to adaptation to oligotrophic conditions, and reefs for heterogenous environments. Among the Kenya coral reef sites, group of genes for functions suggestive of mitigation of environment perturbations such as capacity to reduce intracellular levels of environmental contaminants and repair of DNA damage were significantly enriched only in Malindi and Mombasa marine parks, the coral reef sites closest to human settlements with degradative activities.

The need for resource-poor countries to harness marine resources cannot be overstated. The African Union (AU) for instance, see this as the 'new frontier of the African renaissance', and has designed strategies in partnership with the European Union (EU), which include fostering scientific cooperation, to restore ocean health, and strengthen growth of blue economy sectors (Pichon, 2019). Marine research, as well as conservation and bioprospecting initiatives can benefit from genomics approaches. However, such modern research tools needed to understand and sustainably utilize the marine resources are inaccessible in most of the Africa partly due to insufficient investment capacity and lack of technical competence. This project highlights the extent of this shortfall and proposes practical short-term and long-term solutions that African marine scientists may adopt to benefit from application of genomics. Findings from this project provide insight into the state of marine ecosystems, especially coral reefs, useful for environmental management, as well as a foundation for future genomics work in the WIO. The fact that standard water quality parameters did not vary with coral reef sites, as did microbial taxa and functions, may be indicative of the superiority and sensitivity of microbes-based approaches in early prediction of ecosystem stress. This is welcome news considering the urgent need for sensitive and rapid markers for ecosystem perturbations to underpin effective management and restoration strategies (Glasl et al., 2019) especially in the face of the global coral reef crisis. This study provides pointers to potential microbial taxa and genes that may be targeted as indicators of coral reef perturbation. However, these will need to be confirmed first by robust studies controlling for temporal variations.

6.4 Study challenges and limitations

The implementation of this project was fraught with several challenges that impacted the analysis and extrapolation of its findings. The main challenge was the cost of sequencing. The depth of sequencing needed for this study necessitated use of Illumina NovaSeq 6000

sequencing platform which was expensive and unavailable locally. Ultimately, the per-sample cost of sequencing meant that only some of the samples from the initial plan could be sequenced. The computational time and storage space needed for efficient analysis was another challenge encountered. The size of the raw data generated could not be shared remotely; I had to travel to do initial analyses from Sweden where the sequencing was done. Reanalysis, which incorporated bigger dataset from Tara Oceans expedition, was constrained by a shortage of efficient web-accessible analysis servers compatible with the slow network I had access to; access to robust applications was not possible due to the costs involved and movement cessation necessitated by the corona pandemic.

6.5 Future possibilities

This study has laid foundation for future microbial studies in the WIO region by a cross-sectional description of coral reef physicochemical and biological characteristics, and by pointing out areas for further investigations. For future studies, comprehensive temporal and spatial descriptions of the habitats of the microorganism under study as well as mapping of human activities with the associated environmental variables is necessary for rigorous comparisons and precise contextual interpretation. Also, future investigations should sample coral compartments that are habitat to microorganisms i.e. skeleton, tissue and mucus, for holistic understanding of microbial interactions and their ecological function in coral reef ecosystems. Complementary approaches such as metatranscriptomics and metaproteomics should also be employed to confirm the proteins and functions predicted *in silico* through metagenomics.

The current study did not detect faecal pollution within the coral reefs. One possible reason for this is the fact that the markers used targeted members of *Bacteroides* and *Prevotella* genera which are anaerobic and do not live long outside their hosts. This, coupled with the dilution effect of the seawater, makes detection difficult especially when the contamination is not recent or ongoing. However, relative abundances of copiotrophic bacteria and coliphages seemed to correspond to anthropogenic impacts. Future work could focus on alternative taxa, for instance, e.g. viruses for the detection and source-identification of faecal contamination. Furthermore, phages infecting *Bacteroides* are host-specific, they outlive their hosts and can thrive in marine environments (Ebdon et al., 2012).

Finally, this study underscores the potential use of microbial variables as indicators of environmental disturbance. The findings of this study will be presented to the requisite environmental managers at the sampled marine national parks with a view to recommend the integration of microbial assessments in their coral reef monitoring initiatives.

6.6 Final comments and summary conclusions

This project evaluated physicochemical and microbial characteristics of Kenyan coral reefs experiencing a range of known human impacts. Taxonomic and functional diversity were compared with those annotated from *Tara* Oceans metagenomes. Water quality parameters were well within the expected ranges for tropical coral reefs and did not vary across the coral reefs. Species and gene richness were comparable to previous observations from global marine ecosystems. The reefs had more diverse and even species and functional repertoire. Significant variations in relative abundances of taxa and functions of ecological significance were observed both between *Tara* Oceans and reef datasets, and especially, between the coral reef sites, in-keeping with the human impacts surrounding the respective coral reefs.

This study - the first metagenomic study of the WIO coral reefs - has established a useful baseline for future microbial studies in the region and indicated areas for follow up research. The potential for microbial genomics to identify adaptive responses to ecosystem-specific selective pressures, and to generate new research hypotheses was demonstrated.

Overall, this study reiterates the need for marine microbial genomics studies in the WIO region and begins to lay a needed foundation.

REFERENCES

- Abdelzaher, A. M., Wright, M. E., Ortega, C., Solo-Gabriele, H. M., Miller, G., Elmir, S., ... Fleming, L. E. (2010). Presence of pathogens and indicator microbes at a non-point source subtropical recreational marine beach. *Applied and Environmental Microbiology*, 76(3), 724–732. <https://doi.org/10.1128/AEM.02127-09>
- Abuodha, P. A. W., & Kairo, J. G. (2001). Human-induced stresses on mangrove swamps along the Kenyan coast. *Hydrobiologia*, 458, 255–265. <https://doi.org/10.1023/A:1013130916811>
- Ahmed, W., Hughes, B., & Harwood, V. J. (2016). Current status of marker genes of bacteroides and related taxa for identifying sewage pollution in environmental waters. *Water (Switzerland)*, 8(6). <https://doi.org/10.3390/w8060231>
- Ahsanuddin, S., Afshinnkoo, E., Gandara, J., Hakyemezoglu, M., Bezdan, D., Minot, S., ... Mason, C. E. (2017). Assessment of REPLI-g multiple displacement whole genome amplification (WGA) techniques for metagenomic applications. *Journal of Biomolecular Techniques*, 28(1), 46–55. <https://doi.org/10.7171/jbt.17-2801-008>
- Ainsworth, T. D., Thurber, R. V., & Gates, R. D. (2010). The future of coral reefs: a microbial perspective. *Trends in Ecology and Evolution*, 25(4), 233–240. <https://doi.org/10.1016/j.tree.2009.11.001>
- Alongi, D. M. (2012, June). Carbon sequestration in mangrove forests. *Carbon Management*. <https://doi.org/10.4155/cmt.12.20>
- Alonso, J. L., & Alonso, M. A. (1993). Presence of *Campylobacter* in marine waters of Valencia, Spain. *Water Research*, 27(10), 1559–1562. [https://doi.org/10.1016/0043-1354\(93\)90100-V](https://doi.org/10.1016/0043-1354(93)90100-V)
- Alvarez-Yela, A. C., Mosquera-Rendón, J., Noreña-P, A., Cristancho, M., & López-Alvarez, D. (2019). Microbial Diversity Exploration of Marine Hosts at Serrana Bank, a Coral Atoll of the Seaflower Biosphere Reserve. *Frontiers in Marine Science*, 6(JUN), 338. <https://doi.org/10.3389/fmars.2019.00338>
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, 26(1), 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
- Angly, F. E., Heath, C., Morgan, T. C., Tonin, H., Rich, V., Schaffelke, B., ... Tyson, G. W. (2016). Marine microbial communities of the Great Barrier Reef lagoon are influenced by riverine floodwaters and seasonal weather events. *PeerJ*, 2016(1), e1511. <https://doi.org/10.7717/peerj.1511>

- APHA, AWWA, & AEF. (1999). *Standard Methods for the Examination of Water and Wastewater* (20th Edn). Washington DC: American Public Health Association.
- Apprill, A., Weber, L. G., & Santoro, A. E. (2016). Distinguishing between Microbial Habitats Unravels Ecological Complexity in Coral Microbiomes. *MSystems*, 1(5), e00143-16. <https://doi.org/10.1128/mSystems.00143-16>
- Bäckhed, F., Fraser, C. M., Ringel, Y., Sanders, M. E., Sartor, R. B., Sherman, P. M., ... Finlay, B. B. (2012). Defining a healthy human gut microbiome: Current concepts, future directions, and clinical applications. *Cell Host and Microbe*, 12(5), 611–622. <https://doi.org/10.1016/j.chom.2012.10.012>
- Baharum, S., Beng, E., & Mokhtar, M. (2010). Marine microorganisms: Potential application and challenges. *Journal of Biological Sciences*, 10(6), 555–564.
- Balakrishnan, S., Chelladurai, G., Mohanraj, J., & Poongodi, J. (2017). Seasonal variations in physico-chemical characteristics of Tuticorin coastal waters, southeast coast of India. *Applied Water Science*, 7(4), 1881–1886. <https://doi.org/10.1007/s13201-015-0363-2>
- Barberán, A., Fernández-Guerra, A., Bohannan, B. J. M., & Casamayor, E. O. (2012). Exploration of community traits as ecological markers in microbial metagenomes. *Molecular Ecology*, 21(8), 1909–1917. <https://doi.org/10.1111/j.1365-294X.2011.05383.x>
- Bari, M. L., & Yeasmin, S. (2014). Water Quality Assessment: Modern Microbiological Techniques. In *Encyclopedia of Food Microbiology: Second Edition* (pp. 755–765). Elsevier Inc. <https://doi.org/10.1016/B978-0-12-384730-0.00353-0>
- Bashan, Y., & Holguin, G. (2002). Plant growth-promoting bacteria: A potential tool for arid mangrove reforestation. *Trees - Structure and Function*, 16(2–3), 159–166. <https://doi.org/10.1007/s00468-001-0152-4>
- Baudart, J., Servais, P., Paoli, H. De, Henry, A., & Lebaron, P. (2009). Rapid enumeration of *Escherichia coli* in marine bathing waters : potential interference of nontarget bacteria. *Journal of Applied Microbiology*, 107, 2054–2062. <https://doi.org/10.1111/j.1365-2672.2009.04392.x>
- Bayer, B., Vojvoda, J., Offre, P., Alves, R. J. E., Elisabeth, N. H., Garcia, J. A. L., ... Herndl, G. J. (2016). Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME Journal*, 10(5), 1051–1063. <https://doi.org/10.1038/ismej.2015.200>
- Behzad, H., Ibarra, M. A., Mineta, K., & Gojobori, T. (2016). Metagenomic studies of the Red Sea. *Gene*, 576(2), 717–723. <https://doi.org/10.1016/j.gene.2015.10.034>

- Beja, O., Aravind, L., Koonin, E. V., Suzuki, M. T., Hadd, A., Nguyen, L. P., ... DeLong, E. F. (2000). Bacterial rhodopsin: Evidence for a new type of phototrophy in the sea. *Science*, 289(5486), 1902–1906. <https://doi.org/10.1126/science.289.5486.1902>
- Bellwood, D. R., Streit, R. P., Brandl, S. J., & Tebbett, S. B. (2019, June 1). The meaning of the term ‘function’ in ecology: A coral reef perspective. *Functional Ecology*. Blackwell Publishing Ltd. <https://doi.org/10.1111/1365-2435.13265>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
- Bernhard, A. E., & Field, K. G. (2000a). A PCR assay to discriminate human and ruminant feces on the basis of host differences in *Bacteroides-Prevotella* genes encoding 16S rRNA. *Applied and Environmental Microbiology*, 66(10), 4571–4574. <https://doi.org/10.1128/AEM.66.10.4571-4574.2000>
- Bernhard, A. E., & Field, K. G. (2000b). Identification of nonpoint sources of fecal pollution in coastal waters by using host-specific 16S ribosomal DNA genetic markers from fecal anaerobes. *Applied and Environmental Microbiology*, 66(4), 1587–1594. <https://doi.org/10.1128/AEM.66.4.1587-1594.2000>
- Biller, S. J., Berube, P. M., Dooley, K., Williams, M., Satinsky, B. M., Hackl, T., ... Chisholm, S. W. (2018). Marine microbial metagenomes sampled across space and time. *Scientific Data*, 5, 1–7. <https://doi.org/10.1038/sdata.2018.176>
- Biller, S. J., Berube, P. M., Lindell, D., & Chisholm, S. W. (2015). *Prochlorococcus*: the structure and function of collective diversity. *Nature Reviews Microbiology*, 13(1), 13–27. <https://doi.org/10.1038/nrmicro3378>
- Boehm, A. B., Van De Werfhorst, L. C., Griffith, J. F., Holden, P. A., Jay, J. A., Shanks, O. C., ... Weisberg, S. B. (2013). Performance of forty-one microbial source tracking methods: A twenty-seven lab evaluation study. *Water Research*, 47(18), 6812–6828. <https://doi.org/10.1016/j.watres.2012.12.046>
- Bouillon, S., Rivera-Monroy, V. H., Twilley, R. R., & Kairo, J. G. (2009). Mangroves. In *The management of natural coastal carbon sinks*. IUCN, Gland, Switzerland (p. 53 pp.).
- Bourlat, S. J., Borja, A., Gilbert, J., Taylor, M. I., Davies, N., Weisberg, S. B., ... Obst, M. (2013). Genomics in marine monitoring: New opportunities for assessing marine health status. *Marine Pollution Bulletin*, 74(1), 19–31. <https://doi.org/10.1016/j.marpolbul.2013.05.042>
- Bourne, D. G., & Webster, N. S. (2013). Coral reef bacterial communities. In E Rosenberg, E.

- . DeLong, S. Lory, E. Stackebrandt, & F. L. Thompson (Eds.), *The Prokaryotes - Prokaryotic Communities and Ecophysiology* (pp. 163–187). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-30123-0>
- Bray, J. R., & Curtis, J. T. (1957). An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs*, 27(4), 325–349. <https://doi.org/10.2307/1942268>
- Bridge, T. C. L., Hughes, T. P., Guinotte, J. M., & Bongaerts, P. (2013). Call to protect all coral reefs. *Nature Climate Change*, 3(6), 528–530. <https://doi.org/10.1038/nclimate1879>
- Bruckner, A. W. (2002). Life-Saving Products from Coral Reefs. *Issues in Science and Technology*, 18(3), 39–44.
- Buccheri, M. A., Salvo, E., Coci, M., Quero, G. M., Zoccarato, L., Privitera, V., & Rappazzo, G. (2019). Investigating microbial indicators of anthropogenic marine pollution by 16S and 18S High-Throughput Sequencing (HTS) library analysis. *FEMS Microbiology Letters*, 366(14). <https://doi.org/10.1093/femsle/fnz179>
- Buckley, R. J., Kramm, K., Cooper, C. D. O., Grohmann, D., & Bolt, E. L. (2020). Mechanistic insights into Lhr helicase function in DNA repair. *Biochemical Journal*, 477(16), 2935–2947. <https://doi.org/10.1042/BCJ20200379>
- Burbano-Rosero, E. M., Ueda-Ito, M., Kisielius, J. J., Nagasse-Sugahara, T. K., Almeida, B. C., Souza, C. P., ... Rivera, I. N. G. (2011). Diversity of Somatic coliphages in coastal regions with different levels of anthropogenic activity in São Paulo state, Brazil. *Applied and Environmental Microbiology*, 77(12), 4208–4216. <https://doi.org/10.1128/AEM.02780-10>
- Carding, S., Verbeke, K., Vipond, D. T., Corfe, B. M., & Owen, L. J. (2015). Dysbiosis of the gut microbiota in disease. *Microbial Ecology in Health & Disease*, 26(0). <https://doi.org/10.3402/mehd.v26.26191>
- Carlos, C., Pereira, L. B., & Ottoboni, L. M. M. (2016). Comparative genomics of *Paracoccus* sp. SM22M-07 isolated from coral mucus: insights into bacteria-host interactions. *Current Genetics*, 63(3), 509–518. <https://doi.org/10.1007/s00294-016-0658-3>
- Cavicchioli, R., Ripple, W. J., Timmis, K. N., Azam, F., Bakken, L. R., Baylis, M., ... Webster, N. S. (2019). Scientists' warning to humanity: microorganisms and climate change. *Nature Reviews Microbiology*, 17(September). <https://doi.org/10.1038/s41579-019-0222-5>
- Chan, L. M., Goodman, S. M., Nowak, M. D., Weisrock, D. W., & Yoder, A. D. (2011). Increased population sampling confirms low genetic divergence among *Pteropus* (Chiroptera: Pteropodidae) fruit bats of Madagascar and other western Indian Ocean islands. *PLoS Currents*. <https://doi.org/10.1371/currents.RRN1226>

- Chen, J., McIlroy, S. E., Archana, A., Baker, D. M., & Panagiotou, G. (2019). A pollution gradient contributes to the taxonomic, functional, and resistome diversity of microbial communities in marine sediments. *Microbiome*, 7(1), 104. <https://doi.org/10.1186/s40168-019-0714-6>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/bioinformatics/bty560>
- Cho, I., & Blaser, M. J. (2012). The human microbiome: At the interface of health and disease. *Nature Reviews Genetics*, 13(4), 260–270. <https://doi.org/10.1038/nrg3182>
- Claar, D. C., Szostek, L., McDevitt-Irwin, J. M., Schanze, J. J., & Baum, J. K. (2018). Global patterns and impacts of El Niño events on coral reefs: A meta-analysis. *PLoS ONE*, 13(2), 1–22. <https://doi.org/10.1371/journal.pone.0190957>
- Claire Horner-Devine, M., Leibold, M. A., Smith, V. H., & Bohannan, B. J. M. (2003). Bacterial diversity patterns along a gradient of primary productivity. *Ecology Letters*, 6(7), 613–622. <https://doi.org/10.1046/j.1461-0248.2003.00472.x>
- Consortium, H. (2014). Enabling the genomic revolution in Africa. *Science*, 344(June), 1346–1348.
- Cooper, T. F., Gilmour, J. P., & Fabricius, K. E. (2009, June 2). Bioindicators of changes in water quality on coral reefs: Review and recommendations for monitoring programmes. *Coral Reefs*. Springer. <https://doi.org/10.1007/s00338-009-0512-x>
- Coulter, C. B., Kolka, R. K., & Thompson, J. a. (2004). Water quality in agricultural, urban, and mixed land use watersheds. *Journal of the American Water Resources Association*, 40(6), 1593–1601. <https://doi.org/10.1111/j.1752-1688.2004.tb01608.x>
- Cullen-Unsworth, L. C., Nordlund, L. M., Paddock, J., Baker, S., McKenzie, L. J., & Unsworth, R. K. F. (2014). Seagrass meadows globally as a coupled social-ecological system: Implications for human wellbeing. *Marine Pollution Bulletin*, 83(2), 387–397. <https://doi.org/10.1016/j.marpolbul.2013.06.001>
- D’Angelo, C., & Wiedenmann, J. (2014, April). Impacts of nutrient enrichment on coral reefs: New perspectives and implications for coastal management and reef survival. *Current Opinion in Environmental Sustainability*. <https://doi.org/10.1016/j.cosust.2013.11.029>
- Daby, D. (2003). Effects of seagrass bed removal for tourism purposes in a Mauritian bay. *Environmental Pollution (Barking, Essex: 1987)*, 125(3), 313–324. [https://doi.org/10.1016/s0269-7491\(03\)00125-8](https://doi.org/10.1016/s0269-7491(03)00125-8)
- Dandara, C., Huzair, F., Borda-Rodriguez, A., Chirikure, S., Okpechi, I., Warnich, L., &

- Masimirembwa, C. (2014, December 1). H3Africa and the African life sciences ecosystem: Building sustainable innovation. *OMICS A Journal of Integrative Biology*. Mary Ann Liebert Inc. <https://doi.org/10.1089/omi.2014.0145>
- Darling, E. S., McClanahan, T. R., Maina, J., Gurney, G. G., Graham, N. A. J., Januchowski-Hartley, F., ... Mouillot, D. (2019). Social–environmental drivers inform strategic management of coral reefs in the Anthropocene. *Nature Ecology & Evolution*, 3(9), 1341–1350. <https://doi.org/10.1038/s41559-019-0953-8>
- de la Torre-castro, M., Eklöf, J. S., Rönnbäck, P., & Björk, M. (2008). Seagrass Importance in Food Provisioning Services : Fish Stomach Content as a Link between Seagrass Meadows and Local Fisheries. *Western Indian Ocean Journal of Marine Science*, 7(1), 95–110.
- de la Torre-Castro, M., & Rönnbäck, P. (2004). Links between humans and seagrasses—an example from tropical East Africa. *Ocean & Coastal Management*, 47(7–8), 361–387. <https://doi.org/10.1016/j.ocecoaman.2004.07.005>
- De la Torre-Castro, Maricela, Di Carlo, G., & Jiddawi, N. S. (2014). Seagrass importance for a small-scale fishery in the tropics: The need for seascape management. *Marine Pollution Bulletin*, 83(2), 398–407. <https://doi.org/10.1016/j.marpolbul.2014.03.034>
- Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., ... Bernatchez, L. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, 26(21), 5872–5895. <https://doi.org/10.1111/mec.14350>
- DeLong, E. F. (2005). Microbial community genomics in the ocean. *Nature Reviews Microbiology*, 3(6), 459–469. <https://doi.org/10.1038/nrmicro1158>
- Díez, B., Nylander, J. A. A., Ininbergs, K., Dupont, C. L., Allen, A. E., Yooseph, S., ... Bergman, B. (2016). Metagenomic Analysis of the Indian Ocean Picocyanobacterial Community: Structure, Potential Function and Evolution. *Plos One*, 11(5), e0155757. <https://doi.org/10.1371/journal.pone.0155757>
- Dinsdale, E. A., Pantos, O., Smriga, S., Edwards, R. A., Angly, F., Wegley, L., ... Rohwer, F. (2008). Microbial Ecology of Four Coral Atolls in the Northern Line Islands. *PLoS ONE*, 3(2), e1584. <https://doi.org/10.1371/journal.pone.0001584>
- Duarte, C. M. (2002). The future of seagrass meadows. *Environmental Conservation*, 29(2), 192–206. <https://doi.org/10.1017/S0376892902000127>
- Dubinsky, Z., & Stambler, N. (1996). Marine pollution and coral reefs. *Global Change Biology*, 2(6), 511–526. <https://doi.org/10.1111/j.1365-2486.1996.tb00064.x>
- Ebdon, J. E., Sellwood, J., Shore, J., & Taylor, H. D. (2012). Phages of bacteroides (GB-124):

- A novel tool for viral waterborne disease control? *Environmental Science and Technology*, 46(2), 1163–1169. <https://doi.org/10.1021/es202874p>
- Eklöf, J.S., de la Torre Castro, M., Adelsköld, L., Jiddawi, N. S., & Kautsky, N. (2005). Differences in macrofaunal and seagrass assemblages in seagrass beds with and without seaweed farms. *Estuarine, Coastal and Shelf Science*, 63(3), 385–396. <https://doi.org/10.1016/j.ecss.2004.11.014>
- Eklöf, Johan S., Fröcklin, S., Lindvall, A., Stadlinger, N., Kimathi, A., Uku, J. N., & McClanahan, T. R. (2009). How effective are MPAs? Predation control and “spill-in effects” in seagrass-coral reef lagoons under contrasting fishery management. *Marine Ecology Progress Series*, 384, 83–96. <https://doi.org/10.3354/meps08024>
- Emerton, L., & Tessema, Y. (2001). Marine Protected Areas: the Case of Kisite Marine National Park and Mpunguti Marine National Reserve, Kenya. *IUCN Eastern Africa Regional Office, Nairobi, Kenya*, (May), 32.
- Fabricius, K. E., Cooper, T. F., Humphrey, C., Uthicke, S., De’ath, G., Davidson, J., ... Schaffelke, B. (2012). A bioindicator system for water quality on inshore coral reefs of the Great Barrier Reef. *Marine Pollution Bulletin*, 65(4–9), 320–332. <https://doi.org/10.1016/j.marpolbul.2011.09.004>
- Farrell, M. J., Govender, D., Hajibabaei, M., Van Der Bank, M., & Davies, T. J. (2019). Bacterial diversity in the waterholes of the Kruger National Park: An eDNA metabarcoding approach. *Genome*, 62(3), 229–242. <https://doi.org/10.1139/gen-2018-0064>
- Ferguson, D., & Signoretto, C. (2011). Environmental Persistence and Naturalization of Fecal Indicator Organisms. In C. et al. Hagedorn (Ed.), *Microbial Source Tracking: Methods, Applications, and Case Studies* (pp. 379–397). Springer Science+Business Media, LLC 2011. <https://doi.org/10.1007/978-1-4419-9386-1>
- Flaviani, F., Schroeder, D. C., Lebre, K., Balestreri, C., Highfield, A. C., Schroeder, J. L., ... Rybicki, E. P. (2018). Distinct oceanic microbiomes from viruses to protists located near the Antarctic Circumpolar current. *Frontiers in Microbiology*, 9(JUL), 1474. <https://doi.org/10.3389/fmicb.2018.01474>
- Fleming, L. E., Broad, K., Clement, A., Dewailly, E., Elmir, S., Knap, A., ... Walsh, P. (2006). Oceans and human health: Emerging public health risks in the marine environment. *Marine Pollution Bulletin*, 53(10–12), 545–560. <https://doi.org/10.1016/j.marpolbul.2006.08.012>
- Frank, J. A., Reich, C. I., Sharma, S., Weisbaum, J. S., Wilson, B. A., & Olsen, G. J. (2008).

- Critical evaluation of two primers commonly used for amplification of bacterial 16S rRNA genes. *Applied and Environmental Microbiology*, 74(8), 2461–2470. <https://doi.org/10.1128/AEM.02272-07>
- Freed, S., & Granek, E. F. (2014). Effects of Human Activities on the World's Most Vulnerable Coral Reefs: Comoros Case Study. *Coastal Management*, 42(3), 280–296. <https://doi.org/10.1080/08920753.2014.904261>
- Fremaux, B., Gritzfeld, J., Boa, T., & Yost, C. K. (2009). Evaluation of host-specific Bacteroidales 16S rRNA gene markers as a complementary tool for detecting fecal pollution in a prairie watershed. *Water Research*, 43(19), 4838–4849. <https://doi.org/10.1016/j.watres.2009.06.045>
- Functional, A. N. R. C. (US) C. on M. C. and. (2007). The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet. In *Washington (DC): National Academies Press (US)*. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK54011/>
- Gagnaire, P. A., Tsukamoto, K., Aoyama, J., Minegishi, Y., Valade, P., & Berrebi, P. (2007). RFLP and semi-multiplex PCR-based identification of four eel species from the southwestern Indian Ocean region. *Journal of Fish Biology*, 71(SUPPL. B), 279–287. <https://doi.org/10.1111/j.1095-8649.2007.01605.x>
- Galperin, M. Y., & Kolker, E. (2006). New metrics for comparative genomics. *Current Opinion in Biotechnology*, 17(5), 440–447. <https://doi.org/10.1016/j.copbio.2006.08.007>
- Galperin, M. Y., Kristensen, D. M., Makarova, K. S., Wolf, Y. I., & Koonin, E. V. (2019). Microbial genome analysis: The COG approach. *Briefings in Bioinformatics*, 20(4), 1063–1070. <https://doi.org/10.1093/bib/bbx117>
- Gardner, S. G., Camp, E. F., Smith, D. J., Kahlke, T., Osman, E. O., Gendron, G., ... Suggett, D. J. (2019). Coral microbiome diversity reflects mass coral bleaching susceptibility during the 2016 El Niño heat wave. *Ecology and Evolution*, 9(3), 938–956. <https://doi.org/10.1002/ece3.4662>
- Gareau, M. G., Sherman, P. M., & Walker, W. A. (2010). probiotics and the gut microbiota in intestinal health and disease. *Nature Publishing Group*, 7(9), 503–514. <https://doi.org/10.1038/nrgastro.2010.117>
- Garren, M., & Azam, F. (2012). New directions in coral reef microbial ecology. *Environmental Microbiology*, 14(4), 833–844. <https://doi.org/10.1111/j.1462-2920.2011.02597.x>
- Gianoulis, T. A., Raes, J., Patel, P. V., Bjornson, R., Korbel, J. O., Letunic, I., ... Gerstein, M. B. (2009). Quantifying environmental adaptation of metabolic pathways in metagenomics. *Proceedings of the National Academy of Sciences of the United States of America*, 106(5),

- 1374–1379. <https://doi.org/10.1073/pnas.0808022106>
- Glasl, B., Bourne, D. G., Frade, P. R., Thomas, T., Schaffelke, B., & Webster, N. S. (2019). Microbial indicators of environmental perturbations in coral reef ecosystems. *Microbiome*, 7(1), 94. <https://doi.org/10.1186/s40168-019-0705-7>
- Glöckner, F. O., & Joint, I. (2010). Marine microbial genomics in Europe: Current status and perspectives: Minireview. *Microbial Biotechnology*, 3(5), 523–530. <https://doi.org/10.1111/j.1751-7915.2010.00169.x>
- Godoy-Vitorino, F., Ruiz-Diaz, C. P., Rivera-Seda, A., Ramírez-Lugo, J. S., & Toledo-Hernández, C. (2017). The microbial biosphere of the coral *Acropora cervicornis* in Northeastern Puerto Rico. *PeerJ*, 5, e3717. <https://doi.org/10.7717/peerj.3717>
- Gopal, K., Tolley, K. A., Groeneveld, J. C., & Matthee, C. A. (2006). Mitochondrial DNA variation in spiny lobster *Palinurus delagoae* suggests genetically structured populations in the southwestern Indian Ocean. *Marine Ecology Progress Series*, 319, 191–198. <https://doi.org/10.3354/meps319191>
- Green, J. L., Bohannan, B. J. M., & Whitaker, R. J. (2008). Microbial biogeography: From taxonomy to traits. *Science*, 320(5879), 1039–1043. <https://doi.org/10.1126/science.1153475>
- Grozdanov, L., & Hentschel, U. (2007). An environmental genomics perspective on the diversity and function of marine sponge-associated microbiota. *Current Opinion in Microbiology*, 10(3), 215–220. <https://doi.org/10.1016/j.mib.2007.05.012>
- Guan, Y., Hohn, S., & Merico, A. (2015). Suitable Environmental Ranges for Potential Coral Reef Habitats in the Tropical Ocean. *PLOS ONE*, 10(6), e0128831. <https://doi.org/10.1371/journal.pone.0128831>
- Gudhka, R. K., Neilan, B. A., & Burns, B. P. (2015). Adaptation, ecology, and evolution of the halophilic stromatolite archaeon *Halococcus hamelinensis* inferred through genome analyses. *Archaea*, 2015. <https://doi.org/10.1155/2015/241608>
- Gudka, M., Obura, D., Mwaura, J., Porter, S., Yahya, S., & Mabwa, R. (2018). *Impact of the 3rd global coral bleaching event on the Western Indian Ocean Impact of the 3rd Global Coral Bleaching Event on the Western Indian Ocean in 2016. Global Coral Reef Monitoring Network (GCRMN)/Indian Ocean Commission*. <https://doi.org/10.13140/RG.2.2.32306.71365>
- Gullström, M., Bodin, M., Nilsson, P. G., & Öhman, M. C. (2008). Seagrass structural complexity and landscape configuration as determinants of tropical fish assemblage composition. *Marine Ecology Progress Series*, 363, 241–255.

<https://doi.org/10.3354/meps07427>

- Gullström, M., De La Torre Castro, M., Bandeira, S. O., Björk, M., Dahlberg, M., Kautsky, N., ... Öhman, M. C. (2002). Seagrass ecosystems in the Western Indian Ocean. *AMBIO: A Journal of the Human Environment*, 31(7), 588–596. <https://doi.org/10.1579/0044-7447-31.7.588>
- Gullström, M., Lyimo, L. D., Dahl, M., Samuelsson, G. S., Eggertsen, M., Anderberg, E., ... Björk, M. (2018). Blue Carbon Storage in Tropical Seagrass Meadows Relates to Carbonate Stock Dynamics, Plant–Sediment Processes, and Landscape Context: Insights from the Western Indian Ocean. *Ecosystems*, 21(3), 551–566. <https://doi.org/10.1007/s10021-017-0170-8>
- Halpern, B. S., Kappel, C. V., Selkoe, K. A., Micheli, F., Ebert, C. M., Kontgis, C., ... Teck, S. J. (2009). Mapping cumulative human impacts to California Current marine ecosystems. *Conservation Letters*, 2(3), 138–148. <https://doi.org/10.1111/j.1755-263X.2009.00058.x>
- Halpern, B. S., Longo, C., Hardy, D., McLeod, K. L., Samhouri, J. F., Katona, S. K., ... Zeller, D. (2012). An index to assess the health and benefits of the global ocean. *Nature*, 488(7413), 615–620. <https://doi.org/10.1038/nature11397>
- Harwood, V. J., Staley, C., Badgley, B. D., Borges, K., & Korajkic, A. (2014). Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. *FEMS Microbiology Reviews*, 38(1), 1–40. <https://doi.org/10.1111/1574-6976.12031>
- Hassenrück, C., Hofmann, L. C., Bischof, K., & Ramette, A. (2015). Seagrass biofilm communities at a naturally CO₂-rich vent. *Environmental Microbiology Reports*, 7(3), 516–525. <https://doi.org/10.1111/1758-2229.12282>
- Heidelberg, K. B., Gilbert, J. A., & Joint, I. (2010). Marine genomics: at the interface of marine microbial ecology and biodiscovery: Minireview. *Microbial Biotechnology*, 3(5), 531–543. <https://doi.org/10.1111/j.1751-7915.2010.00193.x>
- Henkel, T. P. (2010). Coral Reefs. Retrieved November 6, 2019, from <https://www.nature.com/scitable/knowledge/library/coral-reefs-15786954/>
- Herlemann, D. P. R., Markert, S., Meeske, C., Andersson, A. F., de Bruijn, I., Hentschker, C., ... Schweder, T. (2019). Individual physiological adaptations enable selected bacterial taxa to prevail during long-term incubations. *Applied and Environmental Microbiology*, 85(15), e00825-19. <https://doi.org/10.1128/AEM.00825-19>
- Hernandez-Agreda, A., Gates, R. D., & Ainsworth, T. D. (2017). Defining the Core Microbiome in Corals' Microbial Soup. *Trends in Microbiology*, 25(2), 125–140.

<https://doi.org/10.1016/j.tim.2016.11.003>

- Hewson, I., & Fuhrman, J. A. (2006). Spatial and vertical biogeography of coral reef sediment bacterial and diazotroph communities. *Marine Ecology Progress Series*, 306(Capone 1996), 79–86. <https://doi.org/10.3354/meps306079>
- Hoegh-guldberg, O., Poloczanska, E. S., Skirving, W., & Dove, S. (2017). Coral Reef Ecosystems under Climate Change and Ocean Acidification. *Frontiers in Marine Science*, 4(May). <https://doi.org/10.3389/fmars.2017.00158>
- Hooper, D., Solan, M., Symstad, A., Diaz, S., Gessner, M., Buchmann, N., ... van Peer, L. (2002). Species diversity, functional diversity, and ecosystem functioning. In M. Loreau, S. Naeem, & P. Inchausti (Eds.), *Biodiversity and Ecosystem Functioning: Synthesis and Perspectives* (pp. 195–202). Oxford University Press. <https://doi.org/10.1126/science.312.5775.846a>
- Houk, P., Comerros-Raynal, M., Sudek, M., Vaeoso, M., McGuire, K., & Regis, J. (2020). Nutrient thresholds to protect water quality and coral reefs. *Marine Pollution Bulletin*, 159(111451). <https://doi.org/10.1016/j.marpolbul.2020.111451>
- Hughes, T. P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K. D., Baird, A. H., ... Wilson, S. K. (2017). Global warming and recurrent mass bleaching of corals. *Nature*, 543(7645), 373–377. <https://doi.org/10.1038/nature21707>
- Hughes T.P., Baird A.H., Bellwood D.R., Card M., Connolly S.R., Folke C., ... Roughgarden J. (2003). Climate change human impacts and the resilience of coral reefs. *Science*, 301(August), 929–934. Retrieved from file:///d:/climate change-Converted.Data/climate change-Converted.Data/PDFHughes_et_al_2003-3109958400/Hughes_et_al_2003.pdf
- Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., ... White, O. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207–214. <https://doi.org/10.1038/nature11234>
- Isnansetyo, A., & Kamei, Y. (2003). MC21-A , a Bactericidal Antibiotic Produced by a New Marine against Methicillin-Resistant Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, 47(2), 480–488. <https://doi.org/10.1128/AAC.47.2.480>
- Isnansetyo, A., & Kamei, Y. (2009). Anti-methicillin-resistant Staphylococcus aureus (MRSA) activity of MC21-B, an antibacterial compound produced by the marine bacterium Pseudoalteromonas phenolica O-BC30T. *International Journal of Antimicrobial Agents*, 34(2), 131–135. <https://doi.org/10.1016/j.ijantimicag.2009.02.009>
- Jaenicke, S., Albaum, S. P., Blumenkamp, P., Linke, B., Stoye, J., & Goesmann, A. (2018). Flexible metagenome analysis using the MGX framework. *Microbiome*, 6(1), 76.

<https://doi.org/10.1186/s40168-018-0460-1>

- Jahnke, M., Gullström, M., Larsson, J., Asplund, M. E., Mgeleka, S., Silas, M. O., ... Nordlund, L. M. (2019). Population genetic structure and connectivity of the seagrass *Thalassia hemprichii* in the Western Indian Ocean is influenced by predominant ocean currents. *Ecology and Evolution*, 9(16), 8953–8964. <https://doi.org/10.1002/ece3.5420>
- Jansson, J. K., & Baker, E. S. (2016). A multi-omic future for microbiome studies. *Nature Microbiology*, 1(5), 1–3. <https://doi.org/10.1038/nmicrobiol.2016.49>
- Johnson, D. R., & Pomati, F. (2020). A brief guide for the measurement and interpretation of microbial functional diversity. *Environmental Microbiology*, 22(8), 3039–3048. <https://doi.org/10.1111/1462-2920.15147>
- Kairo, J. G., Wanjiru, C., & Ochiewo, J. (2009). Net Pay: Economic Analysis of a Replanted Mangrove Plantation in Kenya. *Journal of Sustainable Forestry*, 28(3–5), 395–414. <https://doi.org/10.1080/10549810902791523>
- Kandalepas, D., Blum, M. J., & Van Bael, S. A. (2015). Shifts in symbiotic endophyte communities of a foundational salt marsh grass following oil exposure from the deepwater horizon oil spill. *PLoS ONE*, 10(4). <https://doi.org/10.1371/journal.pone.0122378>
- Kandlikar, G. S., Gold, Z. J., Cowen, M. C., Meyer, R. S., Freise, A. C., Kraft, N. J. B., ... Curd, E. E. (2018). Ranacapa: An R package and shiny web app to explore environmental DNA data with exploratory statistics and interactive visualizations [version 1; referees: 1 approved, 2 approved with reservations]. *F1000Research*, 7. <https://doi.org/10.12688/f1000research.16680.1>
- Kannapiran, E., Kannan, L., Purushothaman, A., & Thangarajdou, T. (2008). Physico-chemical and microbial characteristics of the coral reef environment of the Gulf of Mannar marine biosphere reserve, India. *Journal of Environmental Biology*, 29(2), 215–222.
- Karikari, T. K., Quansah, E., & Mohamed, W. M. Y. (2015). Developing expertise in bioinformatics for biomedical research in Africa. *Applied and Translational Genomics*, 6, 31–34. <https://doi.org/10.1016/j.atg.2015.10.002>
- Karimi, E., Slaby, B. M., Soares, A. R., Blom, J., Hentschel, U., & Costa, R. (2018). Metagenomic binning reveals versatile nutrient cycling and distinct adaptive features in alphaproteobacterial symbionts of marine sponges. *FEMS Microbiology Ecology*, 94(6), 1–18. <https://doi.org/10.1093/femsec/fiy074>
- Katwijk, M. M. Van, Meier, N. E., Loon, R. Van, Hove, E. M. Van, Giesen, W. B. J. T., Velde, G. Van Der, & Hartog, C. Den. (1993). Sabaki River sediment load and coral stress: correlation between sediments and condition of the Malindi-Watamu reefs in Kenya

- (Indian Ocean). *Marine Biology*, 683, 675–676.
- Kelly, Linda W., Williams, G. J., Barott, K. L., Carlson, C. A., Dinsdale, E. A., Edwards, R. A., ... Rohwer, F. (2014). Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proceedings of the National Academy of Sciences*, 111(28), 10227–10232. <https://doi.org/10.1073/pnas.1403319111>
- Kelly, Linda Wegley, Nelson, C. E., Haas, A. F., Naliboff, D. S., Calhoun, S., Carlson, C. A., ... Rohwer, F. (2019). Diel population and functional synchrony of microbial communities on coral reefs. *Nature Communications*, 10(1), 1–9. <https://doi.org/10.1038/s41467-019-09419-z>
- Kilinc, B., & Besler, A. (2014). The occurrence of enteric bacteria in marine environment and pollution ., 3, 39–43.
- Kim, Y., Jeon, J., Kwak, M. S., Kim, G. H., Koh, I., & Rho, M. (2018). Photosynthetic functions of *Synechococcus* in the ocean microbiomes of diverse salinity and seasons. *PLOS ONE*, 13(1), e0190266. <https://doi.org/10.1371/journal.pone.0190266>
- Konopka, A. (2009). What is microbial community ecology. *ISME Journal*, 3(11), 1223–1230. <https://doi.org/10.1038/ismej.2009.88>
- Krediet, C. J., Ritchie, K. B., Paul, V. J., & Teplitski, M. (2013). Coral-associated micro-organisms and their roles in promoting coral health and thwarting diseases. *Proceedings of the Royal Society B: Biological Sciences*, 280(1755), 20122328.
- Lapointe, B. E., Brewton, R. A., Herren, L. W., Porter, J. W., & Hu, C. (2019). Nitrogen enrichment, altered stoichiometry, and coral reef decline at Looe Key, Florida Keys, USA: a 3-decade study. *Marine Biology*, 166(8), 108. <https://doi.org/10.1007/s00227-019-3538-9>
- Li, F., Peng, Y., Fang, W., Altermatt, F., Xie, Y., Yang, J., & Zhang, X. (2018). Application of Environmental DNA Metabarcoding for Predicting Anthropogenic Pollution in Rivers. *Environmental Science and Technology*, 52(20), 11708–11719. <https://doi.org/10.1021/acs.est.8b03869>
- Li, J., Chen, Q., Long, L.-J., Dong, J.-D., Yang, J., & Zhang, S. (2015). Bacterial dynamics within the mucus, tissue and skeleton of the coral *Porites lutea* during different seasons. *Scientific Reports*, 4(1), 7320. <https://doi.org/10.1038/srep07320>
- Lokugamage, A. U., & Pathberiya, S. D. C. (2019, June 28). The microbiome seeding debate - Let's frame it around women-centred care. *Reproductive Health*. BioMed Central Ltd. <https://doi.org/10.1186/s12978-019-0747-0>
- Lubelski, J., Konings, W. N., & Driessen, A. J. M. (2007). Distribution and Physiology of ABC-

- Type Transporters Contributing to Multidrug Resistance in Bacteria. *Microbiology and Molecular Biology Reviews*, 71(3), 463–476. <https://doi.org/10.1128/mmbr.00001-07>
- Macdonald, A. H. H., Schleyer, M. H., & Lamb, J. M. (2011). Acropora austera connectivity in the south-western Indian Ocean assessed using nuclear intron sequence data. *Marine Biology*, 158(3), 613–621. <https://doi.org/10.1007/s00227-010-1585-3>
- MacManes, M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in Genetics*, 5(JAN), 13. <https://doi.org/10.3389/fgene.2014.00013>
- Macreadie, P. I., Baird, M. E., Trevathan-Tackett, S. M., Larkum, A. W. D., & Ralph, P. J. (2014). Quantifying and modelling the carbon sequestration capacity of seagrass meadows - A critical assessment. *Marine Pollution Bulletin*, 83(2), 430–439. <https://doi.org/10.1016/j.marpolbul.2013.07.038>
- Mande, S. S., Mohammed, M. H., & Ghosh, T. S. (2012). Classification of metagenomic sequences: Methods and challenges. *Briefings in Bioinformatics*, 13(6), 669–681. <https://doi.org/10.1093/bib/bbs054>
- Mazzola, M. (2004). Assessment and management of soil microbial community structure for disease suppression. *Annual Review of Phytopathology*, 42(1), 35–59. <https://doi.org/10.1146/annurev.phyto.42.040803.140408>
- Mboowa, G., & Sserwadda, I. (2019, July 1). Role of genomics literacy in reducing the burden of common genetic diseases in Africa. *Molecular Genetics and Genomic Medicine*. Wiley-Blackwell. <https://doi.org/10.1002/mgg3.776>
- McClanahan, T. (1988). Seasonality in East Africa's coastal waters. *Marine Ecology Progress Series*, 44, 191–199. <https://doi.org/10.3354/meps044191>
- McClanahan, T. R. (2017). Changes in coral sensitivity to thermal anomalies. *Marine Ecology Progress Series*, 570, 71–85. <https://doi.org/10.3354/meps12150>
- McClanahan, Tim R, Graham, N. A. J., Calnan, J. M., & Macneil, M. A. (2010). Toward pristine biomass: reef fish recovery in coral reef marine protected areas in Kenya. *Ecological Applications*, 17(4), 1055–1067.
- McDole, T., Nulton, J., Barott, K. L., Felts, B., Hand, C., Hatay, M., ... Rohwer, F. (2012). Assessing Coral Reefs on a Pacific-Wide Scale Using the Microbialization Score. *PLoS ONE*, 7(9), e43233. <https://doi.org/10.1371/journal.pone.0043233>
- McKew, B. A., Dumbrell, A. J., Daud, S. D., Hepburn, L., Thorpe, E., Mogensen, L., & Whitby, C. (2012). Characterization of geographically distinct bacterial communities associated with coral mucus produced by Acropora spp. and Porites spp. *Applied and Environmental Microbiology*, 78(15), 5229–5237. <https://doi.org/10.1128/AEM.07764-11>

- McMurdie, P. J., & Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, 8(4), e61217. <https://doi.org/10.1371/journal.pone.0061217>
- Medina, M. (2011). Analyzing Coral Reefs and Their Microbial Assemblages. *Microbe*, 6(5), 226–232.
- Miller, L. D., Russell, M. H., & Alexandre, G. (2009). Chapter 3 Diversity in Bacterial Chemotactic Responses and Niche Adaptation. In *Advances in Applied Microbiology* (Vol. 66, pp. 53–75). [https://doi.org/10.1016/S0065-2164\(08\)00803-4](https://doi.org/10.1016/S0065-2164(08)00803-4)
- Moberg, F., & Folke, C. (1999). Ecological goods and services of coral reef ecosystems. *Ecological Economics*, 29(2), 215–233. [https://doi.org/10.1016/S0921-8009\(99\)00009-9](https://doi.org/10.1016/S0921-8009(99)00009-9)
- Mora, C. (2008). A clear human footprint in the coral reefs of the Caribbean. *Proceedings of the Royal Society B: Biological Sciences*, 275(1636), 767–773. <https://doi.org/10.1098/rspb.2007.1472>
- Mora, C., Costello, M. J., Kranenburg, C., Rollo, A., Veron, J., Gaston, K. J., & Myers, R. A. (2006). Coral Reefs and the Global Network of Marine Protected Areas. *Science*, 312(June), 1750–1751.
- Mumby, P. J., Edwards, A. J., Ernesto Arias-González, J., Lindeman, K. C., Blackwell, P. G., Gall, A., ... Llewellyn, G. (2004). Mangroves enhance the biomass of coral reef fish communities in the Caribbean. *Nature*, 427(6974), 533–536. <https://doi.org/10.1038/nature02286>
- Munyao, T., Tole, M. P., & Jungerius, P. D. (2003). Sabaki River transport and deposition in the Indian Ocean. *Recent Advances in Coastal Ecology, Studies from Kenya*, 119–132.
- Mwangi, S., Dzaha, T., & Kimathi, A. (2001). State of Marine Pollution in Mombasa Marine ational Park Marine Reserve and Mtwapa Creek. *Unpublished Technical Report*, 11 p.
- Mzingirwa, F. A., Stomeo, F., Kaunda-Arara, B., Nyunja, J., & Mujibi, F. D. N. (2019). Genetic Connectivity of the Sky Emperor, *Lethrinus mahsena* Populations Across a Gradient of Exploitation Rates in Coastal Kenya. *Frontiers in Genetics*, 10. <https://doi.org/10.3389/fgene.2019.01003>
- Neumann, B., Vafeidis, A. T., Zimmermann, J., & Nicholls, R. J. (2015). Future coastal population growth and exposure to sea-level rise and coastal flooding - A global assessment. *PLoS ONE*, 10(3). <https://doi.org/10.1371/journal.pone.0118571>
- Ngugi, I. (2002). Economic Impacts of Marine Protected Areas : a Case Study of the Mombasa Marine Park (Kenya). *Journal of the Social Sciences Graduate Student Association*, 1(1), 1–11.

- Nicol, G. W., & Schleper, C. (2006). Ammonia-oxidising Crenarchaeota: important players in the nitrogen cycle? *Trends in Microbiology*, 14(5), 207–212. <https://doi.org/10.1016/j.tim.2006.03.004>
- Nogales, B., Lanfranconi, M. P., Piña-Villalonga, J. M., & Bosch, R. (2011). Anthropogenic perturbations in marine microbial communities. *FEMS Microbiology Reviews*, 35(2), 275–298. <https://doi.org/10.1111/j.1574-6976.2010.00248.x>
- O'Brien, J., Mariani, C., Olson, L., Russell, A. L., Say, L., Yoder, A. D., & Hayden, T. J. (2009). Multiple colonisations of the western Indian Ocean by Pteropus fruit bats (Megachiroptera: Pteropodidae): The furthest islands were colonised first. *Molecular Phylogenetics and Evolution*, 51(2), 294–303. <https://doi.org/10.1016/j.ympev.2009.02.010>
- Obura, D.O., & Grimsditch, G. (2009). *Resilience assessment of coral reefs - Assessment protocol for coral reefs, focusing on coral bleaching and thermal stress. Rapport, IUCN, Gland, Switzerland. IUCN working group on Climate Change and Coral Reefs. IUCN, Gland, Switzerland.* Retrieved from http://cmsdata.iucn.org/downloads/resilience_assessment_final.pdf
- Obura, D. (2017). *Reviving the Western Indian Ocean Economy: Actions for a Sustainable Future. WWF International, Gland, Switzerland.*
- Obura, David. (2012). The Diversity and Biogeography of Western Indian Ocean Reef-Building Corals. *PLoS ONE*, 7(9). <https://doi.org/10.1371/journal.pone.0045013>
- Obura, David. (2016). *Coral and Biogenic Reef Habitats.*
- Obura, David, Gudka, M., Bijoux, J., Freed, S., Gian, S. B., Maharavo, J., ... Ahamada, S. (2017). Coral reef status report for the Western Indian Ocean (2017) Global Coral Reef Monitoring Network.
- Obura, David O. (2001). Kenya. *Marine Pollution Bulletin*, 42(12), 1264–1278. [https://doi.org/10.1016/S0025-326X\(01\)00241-7](https://doi.org/10.1016/S0025-326X(01)00241-7)
- Ochieng Okuku, E., Linet Imbayi, K., Gilbert Omondi, O., Veronica Ogolla Wayayi, W., Catherine Sezi, M., Mokeira Maureen, K., ... Oduor, N. (2019). Decadal Pollution Assessment and Monitoring along the Kenya Coast. In *Monitoring of Marine Pollution* (pp. 1–15). IntechOpen. <https://doi.org/10.5772/intechopen.82606>
- Odagiri, M., Schriewer, A., Hanley, K., Wuertz, S., Misra, P. R., Panigrahi, P., & Jenkins, M. W. (2015). Validation of Bacteroidales quantitative PCR assays targeting human and animal fecal contamination in the public and domestic domains in India. *Science of the Total Environment*, 502, 462–470. <https://doi.org/10.1016/j.scitotenv.2014.09.040>

- Okabe, S., & Shimazu, Y. (2007). Persistence of host-specific *Bacteroides-Prevotella* 16S rRNA genetic markers in environmental waters: Effects of temperature and salinity. *Applied Microbiology and Biotechnology*, 76(4), 935–944. <https://doi.org/10.1007/s00253-007-1048-z>
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2019, February 4). vegan: Community Ecology Package. *R Package Version 2.5-6*. <https://doi.org/10.5281/ZENODO.2556745>
- Okuku, E. O., Ohowa, B., Mwangi, S. N., Munga, D., Kiteresi, L. I., Wanjeri, V. O., ... Kilonzo, J. (2011). Sewage pollution in the coastal waters of Mombasa City, Kenya: A norm rather than an exception. *International Journal of Environmental Research*, 5(4), 865–874.
- Ongore, C. O., Okuku, E. O., Mwangi, S. N., Kiteresi, L. I., Ohowa, B. O., Wanjeri, V. O., ... Kilonzi, J. (2013). Characterization of nutrients enrichment in the estuaries and related systems in Kenya coast. *Journal of Environmental Science and Water Resources*, 2(6), 181–190.
- Overmann, J., & Lepleux, C. (2016). Marine Bacteria and Archaea: Diversity, Adaptations, and Culturability. In L. Stal & M. Cretoiu (Eds.), *The Marine Microbiome* (pp. 21–55). Springer, Cham. <https://doi.org/10.1007/978-3-319-33000-6>
- Owens, J. D. (1978). Coliform and escherichia coli bacteria in seawater around penang island, Malaysia. *Water Research*, 12(6), 365–370. [https://doi.org/10.1016/0043-1354\(78\)90101-X](https://doi.org/10.1016/0043-1354(78)90101-X)
- Petrof, E. O., & Khoruts, A. (2014). From stool transplants to next-generation microbiota therapeutics. *Gastroenterology*, 146(6), 1573–1582. <https://doi.org/10.1053/j.gastro.2014.01.004>
- Philips, S., Laanbroek, H. J., & Verstraete, W. (2002). Origin , causes and effects of increased nitrite concentrations in aquatic environments, (2), 115–141.
- Pichon, E. (2019). The African Union’s blue strategy. *European Parliamentary Research Service*, (March). Retrieved from <http://www.europarl.europa.eu/thinktank>
- Pita, L., Rix, L., Slaby, B. M., Franke, A., & Hentschel, U. (2018). The sponge holobiont in a changing ocean: from microbes to ecosystems. *Microbiome*, 6, 46. <https://doi.org/10.1186/s40168-018-0428-1>
- Prince, R. C. (1993). Petroleum spill bioremediation in marine environments. *Critical Reviews in Microbiology*, 19(4), 217–242. <https://doi.org/10.3109/10408419309113530>
- Ragionieri, L., Cannicci, S., Schubart, C. D., & Fratini, S. (2010). Gene flow and demographic history of the mangrove crab *Neosarmatium meinerti*: A case study from the western

- Indian Ocean. *Estuarine, Coastal and Shelf Science*, 86(2), 179–188.
<https://doi.org/10.1016/j.ecss.2009.11.002>
- Ransom, K. P., & Mangi, S. C. (2010). Valuing recreational benefits of coral reefs: The case of Mombasa Marine National Park and Reserve, Kenya. *Environmental Management*, 45(1), 145–154. <https://doi.org/10.1007/s00267-009-9402-9>
- Ribes, M., Jiménez, E., Yahel, G., López-Sendino, P., Diez, B., Massana, R., ... Coma, R. (2012). Functional convergence of microbes associated with temperate marine sponges. *Environmental Microbiology*, 14(5), 1224–1239. <https://doi.org/10.1111/j.1462-2920.2012.02701.x>
- Ridgway, T., & Sampayo, E. M. (2005). Population Genetic Status of the Western Indian Ocean: What do we Know? *Western Indian Ocean Journal of Marine Science*, 4(1), 1–9. <https://doi.org/10.4314/wiojms.v4i1.28468>
- Roberts, C. M., Mcclean, C. J., Veron, J. E. N., Hawkins, J. P., Allen, G. R., Mcallister, D. E., ... Werner, T. B. (2016). Marine Biodiversity Hotspots and Conservation Priorities for Tropical Reefs Published by : American Association for the Advancement of Science Linked references are available on JSTOR for this article : Marine Biodiversity Hotspots and Conservation Priorities, 295(5558), 1280–1284.
- Roewer, L. (2013). DNA fingerprinting in forensics: Past, present, future. *Investigative Genetics*, 4(1), 1. <https://doi.org/10.1186/2041-2223-4-22>
- Rohwer, F., Seguritan, V., Azam, F., & Knowlton, N. (2002). Diversity and distribution of coral-associated bacteria. *Marine Ecology Progress Series*, 243, 1–10. <https://doi.org/10.3354/meps243001>
- Roik, A., Röthig, T., Roder, C., Ziegler, M., Kremb, S. G., & Voolstra, C. R. (2016). Year-Long Monitoring of Physico-Chemical and Biological Variables Provide a Comparative Baseline of Coral Reef Functioning in the Central Red Sea. *PLOS ONE*, 11(11), e0163939. <https://doi.org/10.1371/journal.pone.0163939>
- Rosenberg, Eugene, Koren, O., Reshef, L., & Efrony, R. (2007). The role of microorganisms in coral health , disease and evolution, 5(May), 355–362. <https://doi.org/10.1038/nrmicro1635>
- Ruckelshaus, M., Doney, S. C., Galindo, H. M., Barry, J. P., Chan, F., Duffy, J. E., ... Talley, L. D. (2013). Securing ocean benefits for society in the face of climate change. *Marine Policy*, 40(1), 154–159. <https://doi.org/10.1016/j.marpol.2013.01.009>
- Ruppert, K. M., Kline, R. J., & Rahman, M. S. (2019, January 1). Past, present, and future perspectives of environmental DNA (eDNA) metabarcoding: A systematic review in

- methods, monitoring, and applications of global eDNA. *Global Ecology and Conservation*. Elsevier B.V. <https://doi.org/10.1016/j.gecco.2019.e00547>
- Rusch, A., Hannides, A. K., & Gaidos, E. (2009). Diverse communities of active Bacteria and Archaea along oxygen gradients in coral reef sediments. *Coral Reefs*, 28(1), 15–26. <https://doi.org/10.1007/s00338-008-0427-y>
- Sahoo, K., & Dhal, N. K. (2009). Potential microbial diversity in mangrove ecosystems: A review. *Indian Journal of Marine Sciences*, 38(2), 249–256.
- Samoilys, M. A., & Obura, D. O. (2011). Marine conservation successes in Eastern Africa. *Cordio East Africa*, (December), 1–28.
- Samoilys, M., Pabari, M., Andrew, T., Maina, G. W., Church, J., Momanyi, A., ... Mutta, D. (2015). Resilience of Coastal Systems and Their Human Partners. *Nairobi, Kenya: IUCN ESARO, WIOMSA, CORDIO and UNEP Nairobi Convention*, (June), x + 74pp. Retrieved from www.iucn.org
- Santoro, A. E., Richter, R. A., & Dupont, C. L. (2019, January 3). Planktonic marine archaea. *Annual Review of Marine Science*. Annual Reviews Inc. <https://doi.org/10.1146/annurev-marine-121916-063141>
- Sawabe, T., Makino, H., Tatsumi, M., Nakano, K., Tajima, K., Iqbal, M. M., ... Christen, R. (1998). *Pseudoalteromonas bacteriolytica* sp. nov., a marine bacterium that is the causative agent of red spot disease of *Laminaria japonica*. *International Journal of Systematic Bacteriology*, 48(3), 769–774.
- SCBD. (2010). Decision adopted by the Conference of the Parties to the Convention on Biological Diversity at its tenth meeting. In *X/2. The Strategic Plan for Biodiversity 2011–2020 and the Aichi Biodiversity Targets*. https://doi.org/10.1007/978-90-481-9659-3_119
- Schöttner, S., Pfitzner, B., Grünke, S., Rasheed, M., Wild, C., & Ramette, A. (2011). Drivers of bacterial diversity dynamics in permeable carbonate and silicate coral reef sands from the Red Sea. *Environmental Microbiology*, 13(7), 1815–1826. <https://doi.org/10.1111/j.1462-2920.2011.02494.x>
- Schriewer, A., Miller, W. A., Byrne, B. A., Miller, M. A., Oates, S., Conrad, P. A., ... Wuertz, S. (2010). Presence of Bacteroidales as a predictor of pathogens in surface waters of the central California coast. *Applied and Environmental Microbiology*, 76(17), 5802–5814. <https://doi.org/10.1128/AEM.00635-10>
- Séré, M. G., Tortosa, P., Chabanet, P., Turquet, J., Quod, J. P., & Schleyer, M. H. (2013). Bacterial communities associated with Porites white patch syndrome (PWPS) on three western Indian Ocean (WIO) coral reefs. *PLoS ONE*, 8(12).

- <https://doi.org/10.1371/journal.pone.0083746>
- Silveira, C. B., Cavalcanti, G. S., Walter, J. M., Silva-Lima, A. W., Dinsdale, E. A., Bourne, D. G., ... Thompson, F. L. (2017). Microbial processes driving coral reef organic carbon flow. *FEMS Microbiology Reviews*, 41(4), 575–595. <https://doi.org/10.1093/FEMSRE>
- Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G., & Laub, M. T. (2005). Two-Component Signal Transduction Pathways Regulating Growth and Cell Cycle Progression in a Bacterium: A System-Level Analysis. *PLoS Biology*, 3(10), e334. <https://doi.org/10.1371/journal.pbio.0030334>
- Sleator, R. D., Shortall, C. and, & Hill, C. (2008). Metagenomics. *Letters in Applied Microbiology*, 47, 361–366. <https://doi.org/10.1016/B978-0-12-374984-0.00933-5>
- Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., ... Herndl, G. J. (2006). Microbial diversity in the deep sea and the underexplored “rare biosphere.” *Proceedings of the National Academy of Sciences*, 103(32), 12115–12120. <https://doi.org/10.4278/ajhp.28.4.tahp>
- Song, H. G., Wang, X., & Bartha, R. (1990). Bioremediation potential of terrestrial fuel spills. *Applied and Environmental Microbiology*, 56(3), 652–656. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16348139>
- Srivastava, S., Nahar, A. S., Brighu, U., & Gupta, A. B. (2019). Comparative study of three methods for the analysis of nitrate nitrogen in synthetic water and wastewater samples. *International Journal of Environmental Analytical Chemistry*, 99(12), 1164–1185. <https://doi.org/10.1080/03067319.2019.1618456>
- Sullivan, M. B., Coleman, M. L., Weigle, P., Rohwer, F., & Chisholm, S. W. (2005). Three Prochlorococcus Cyanophage Genomes: Signature Features and Ecological Interpretations. *PLoS Biology*, 3(5), e144. <https://doi.org/10.1371/journal.pbio.0030144>
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., ... Velayoudon, D. (2015). Structure and function of the global ocean microbiome. *Science*, 348(6237), 1261359–1261359. <https://doi.org/10.1126/science.1261359>
- Sunagawa, S., Woodley, C. M., & Medina, M. (2010). Threatened Corals Provide Underexplored Microbial Habitats. *PLoS ONE*, 5(3), e9554. <https://doi.org/10.1371/journal.pone.0009554>
- Sutton, S. (2010). The Most Probable Number Method and Its Uses in Enumeration, Qualification, and Validation. *Journal of Validation Technology*, 16(3), 35–38. Retrieved from www.microbiol.org
- Synnes, M. (2007). Bioprospecting of organisms from the deep sea: Scientific and

- environmental aspects. *Clean Technologies and Environmental Policy*, 9(1), 53–59. <https://doi.org/10.1007/s10098-006-0062-7>
- Tao, P., & Rao, V. B. (2019). Bacteriophage Vaccines. In *Reference Module in Life Sciences*. Elsevier. <https://doi.org/10.1016/b978-0-12-809633-8.20953-9>
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., & Koonin, E. V. (2000, January 1). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research*. Oxford University Press. <https://doi.org/10.1093/nar/28.1.33>
- Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., ... Koonin, E. V. (2001). The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Research*, 29(1), 22–28. <https://doi.org/10.1093/nar/29.1.22>
- Team, R. C. (2020). R: A Language and Environment for Statistical Computing. <https://www.R-project.org/>. Retrieved from <https://www.r-project.org/>
- Teske, A. (2013). Marine Deep Sediment Microbial Communities. In E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, & F. L. Thompson (Eds.), *The Prokaryotes - Prokaryotic Communities and Ecophysiology* (pp. 123–138). Springer Berlin Heidelberg. <https://doi.org/10.1007/978-3-642-30123-0>
- Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics - a guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1), 3. <https://doi.org/10.1186/2042-5783-2-3>
- Thompson, L. R., Field, C., Romanuk, T., Kamanda Ngugi, D., Siam, R., El Dorry, H., & Stingl, U. (2013). Patterns of ecological specialization among microbial populations in the Red Sea and diverse oligotrophic marine environments. *Ecology and Evolution*, 3(6), 1780–1797. <https://doi.org/10.1002/ece3.593>
- Titilade, P. R., & Olalekan, E. I. (2015). The Importance of Marine Genomics to Life. *Journal of Ocean Research*, 3(1), 1–13. <https://doi.org/10.12691/jor-3-1-1>
- Tout, J., Jeffries, T. C., Webster, N. S., Stocker, R., Ralph, P. J., & Seymour, J. R. (2014). Variability in microbial community composition and function between different niches within a coral reef. *Microbial Ecology*, 67(3), 540–552. <https://doi.org/10.1007/s00248-013-0362-5>
- Tuda, A., & Omar, M. (2012). Protection of Marine Areas in Kenya. *The George Wright Forum*, 29(1), 43–50. Retrieved from <http://www.georgewright.org/291tuda.pdf>
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The Human Microbiome Project. *Nature*, 449(7164), 804–810.

<https://doi.org/10.1038/nature06244>

- Uku, J., Björk, M., Bergman, B., & Díez, B. (2007). Characterization and comparison of prokaryotic epiphytes associated with three East African seagrasses. *Journal of Phycology*, 43(4), 768–779. <https://doi.org/10.1111/j.1529-8817.2007.00371.x>
- Vaccaro, R. F., Briggs, M. P., Carey, L., & Ketchum, B. H. (1950). Viability of *Escherichia coli* in Sea Water. *American Journal of Public Health*, 40(10), 1257–1266.
- Van Der Elst, R., Everett, B., Jiddawi, N., Mwatha, G., Afonso, P. S., & Boulle, D. (2005). Fish, fishers and fisheries of the Western Indian Ocean: Their diversity and status. A preliminary assessment. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 363(1826), 263–284. <https://doi.org/10.1098/rsta.2004.1492>
- van Houtan, K. S., Hargrove, S. K., & Balazs, G. H. (2010). Land Use, Macroalgae, and a tumor-forming disease in Marine Turtles. *PLoS ONE*, 5(9). <https://doi.org/10.1371/journal.pone.0012900>
- Vanwonderghem, I., & Webster, N. S. (2020, April 24). Coral Reef Microorganisms in a Changing Climate. *IScience*. Elsevier Inc. <https://doi.org/10.1016/j.isci.2020.100972>
- Varasteh, T., Moreira, A. P. B., Silva Lima, A. W., Leomil, L., Otsuki, K., Tschoeke, D., ... Thompson, F. (2020). Genomic repertoire of *Mameliella alba* Ep20 associated with *Symbiodinium* from the endemic coral *Mussismilia braziliensis*. *Symbiosis*, 80(1), 53–60. <https://doi.org/10.1007/s13199-019-00655-x>
- Vega Thurber, R., Payet, J. P., Thurber, A. R., & S Correa, A. M. (2017). Virus–host interactions and their roles in coral reef health and disease. *Nature Publishing Group*. <https://doi.org/10.1038/nrmicro.2016.176>
- Vigil-Stenman, T., Ininbergs, K., Bergman, B., & Ekman, M. (2017). High abundance and expression of transposases in bacteria from the Baltic Sea. *ISME Journal*, 11(11), 2611–2623. <https://doi.org/10.1038/ismej.2017.114>
- Vikas, M., & Dwarakish, G. S. (2015). Coastal Pollution: A Review. *Aquatic Procedia*, 4(Icwrcoe), 381–388. <https://doi.org/10.1016/j.aqpro.2015.02.051>
- Visram, S., Yang, M. C., Pillay, R. M., Said, S., Henriksson, O., Grahn, M., & Chen, C. A. (2010). Genetic connectivity and historical demography of the blue barred parrotfish (*Scarus ghobban*) in the western Indian Ocean. *Marine Biology*, 157(7), 1475–1487. <https://doi.org/10.1007/s00227-010-1422-8>
- Wairimu, E., & Khainga, D. (2017). Kenya’s Agenda in Developing the Blue Economy. Retrieved March 24, 2018, from <http://kippra.or.ke/kenyas-agenda-in-developing-the->

- Wakida, F. T., & Lerner, D. N. (2005). Non-agricultural sources of groundwater nitrate: A review and case study. *Water Research*, 39(1), 3–16. <https://doi.org/10.1016/j.watres.2004.07.026>
- Walsh, K., Haggerty, J. M., Doane, M. P., Hansen, J. J., Morris, M. M., Moreira, A. P. B., ... Dinsdale, E. A. (2017). Aura-biomes are present in the water layer above coral reef benthic macro-organisms. *PeerJ*, 5, e3666. <https://doi.org/10.7717/peerj.3666>
- Wardell, L. J. (1995). Potential for bioremediation of uranium-contaminated aquifers.pdf. *Journal of Soil Contamination*, 4(2), 111–121.
- WATERBURY, J. B., WATSON, S. W., GUILLARD, R. R. L., & BRAND, L. E. (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature*, 277(5694), 293–294. <https://doi.org/10.1038/277293a0>
- Weber, L., Gonzalez-Díaz, P., Armenteros, M., & Apprill, A. (2019). The coral ecosphere: A unique coral reef habitat that fosters coral–microbial interactions. *Limnology and Oceanography*, 64(6), 2373–2388. <https://doi.org/10.1002/lno.11190>
- Weber, L., González-Díaz, P., Armenteros, M., Ferrer, V. M., Bretos, F., Bartels, E., ... Apprill, A. (2020). Microbial signatures of protected and impacted Northern Caribbean reefs: changes from Cuba to the Florida Keys. *Environmental Microbiology*, 22(1), 499–519. <https://doi.org/10.1111/1462-2920.14870>
- Wegley, L., Edwards, R., Rodriguez-Brito, B., Liu, H., & Rohwer, F. (2007). Metagenomic analysis of the microbial community associated with the coral *Porites astreoides*. *Environmental Microbiology*, 9(11), 2707–2719. <https://doi.org/10.1111/j.1462-2920.2007.01383.x>
- Wemheuer, F., Von Hoyningen-Huene, A. J. E., Pohlner, M., Degenhardt, J., Engelen, B., Daniel, R., & Wemheuer, B. (2019). Primary production in the water column as major structuring element of the biogeographical distribution and function of archaea in deep-sea sediments of the central Pacific Ocean. *Archaea*, 2019. <https://doi.org/10.1155/2019/3717239>
- Wickham, H. (2016). *ggplot2-Elegant Graphics for Data Analysis* (2nd ed.). New York: Springer-Verlag. <https://doi.org/10.18637/jss.v077.b02>
- Wilkins, S. (2015). Structure and mechanism of ABC transporters. *F1000Prime Reports*, 7. <https://doi.org/10.12703/P7-14>
- Williamson, S. J., Allen, L. Z., Lorenzi, H. A., Fadrosch, D. W., Bami, D., Thiagarajan, M., ... Venter, J. C. (2012). Metagenomic Exploration of Viruses throughout the Indian Ocean.

- PLoS ONE*, 7(10). <https://doi.org/10.1371/journal.pone.0042047>
- Won, N. Il, Kim, K. H., Kang, J. H., Park, S. R., & Lee, H. J. (2017). Exploring the impacts of anthropogenic disturbance on seawater and sediment microbial communities in Korean coastal waters using metagenomics analysis. *International Journal of Environmental Research and Public Health*, 14(2). <https://doi.org/10.3390/ijerph14020130>
- Wuertz, S., Wang, D., Reischer, G. H., & Farnleitner, A. H. (2011). *Microbial Source Tracking: Methods, Applications, and Case Studies*. <https://doi.org/10.1007/978-1-4419-9386-1>
- Wynberg, R. (2016). *Marine Genetic Resources in the Western Indian Ocean and Bioprospecting in the Western Indian Ocean. Regional State Coast Report*.
- Zhang, W., Cao, S., Ding, W., Wang, M., Fan, S., Yang, B., ... Zhang, Y. Z. (2020). Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics. *Microbiome*, 8(1), 47. <https://doi.org/10.1186/s40168-020-00826-9>
- Zhao, X., Schwartz, C. L., Pierson, J., Giovannoni, S. J., McIntosh, J. R., & Nicastro, D. (2017). Three-Dimensional Structure of the Ultraoligotrophic Marine Bacterium “Candidatus Pelagibacter ubique.” *Applied and Environmental Microbiology*, 83(3). <https://doi.org/10.1128/AEM.02807-16>
- Zinger, L., Amaral-Zettler, L. A., Fuhrman, J. A., Horner-Devine, M. C., Huse, S. M., Welch, D. B. M., ... Ramette, A. (2011). Global patterns of bacterial beta-diversity in seafloor and seawater ecosystems. *PLoS ONE*, 6(9), 1–11. <https://doi.org/10.1371/journal.pone.0024570>
- Zobell, C. E., & Eltham, C. B. F. (1938). Bacteria as food for certain marine invertebrates. *Journal of Marine Research*, 1, 312–327.

APPENDIX A: DOES PROTECTION OF MARINE AREAS SAFEGUARD CORAL REEFS FROM HUMAN-SOURCE POLLUTION?

Published in *Frontiers in Environmental Science*, 7, 89

<https://doi.org/10.3389/fenvs.2019.00089>

Abstract

Marine biodiversity is under increasing threat as the area covered by corals diminishes under pressure from climate change and human activities, most of which lead to marine pollution. In Kenya, marine protected areas (MPAs) are the key strategy used to protect coral reefs and biodiversity. However, MPAs' effectiveness in preventing pollution of the reefs has not been specifically assessed.

We determined if the levels of surrogates of human-source pollution, *i.e.* *E. coli* and nutrient concentrations on Kenyan coral reefs, varied with increasing levels of marine protection at the Kilifi creek (least protection), Malindi Reserve (moderate protection) and Kuruwitu Conservancy (strictest protection). The most probable number (MPN) of *E. coli* was estimated by serial dilution while nitrate and orthophosphate concentrations were determined spectrophotometrically.

As protection increased from “least”, to “moderate” and “strictest”, *E. coli* concentrations (MPN/100 mL) decreased from 29, to 16 and undetectable, while mean orthophosphate concentrations increased from 0.326, to 0.422, and 0.524 mg/L, respectively. Mean nitrate concentrations, on the other hand, showed no trend with protection.

These results suggest the potential of marine protection to mitigate coral reef pollution, especially from microbes. They also point to the possibility that multiple sources of pollution exist on which marine protection may have little or no effect. Significantly, this pilot study indicates the need for improved study design to definitively determine role MPAs may play in protecting against pollution.

Keywords: MPA (marine protected area), *E. coli*—*Escherichia coli*, nutrients (nitrogen and phosphorus), coral reefs, pollution

Introduction

The Western Indian Ocean (WIO) hosts the second largest coral reef biodiversity globally (Obura, 2012) and holds potential wealth for the region due to the significant economic role that oceans play in jobs, tourism and fisheries. Surprisingly, the blue economy of the WIO region remains underdeveloped.

The oceans' economic value depends on healthy coral reefs (Obura et al., 2017), which cover only 0.1% of the ocean floor but host at least 25% of all marine biodiversity (Hoegh-Guldberg, et al, 2017), provide goods and services such as seafood, recreational possibilities, coastal protection and aesthetic and cultural benefits (Moberg & Folke, 1999). However, coral reef ecosystems are increasingly under threat mainly from climatic changes as evident from mass coral bleaching and death following episodes of elevated sea surface temperatures (Hughes et al., 2017). The situation is aggravated by overfishing and pollution (Hughes et al., 2003) from human activities on the coast zone leading to decrease in abundance of coral reef species. While corals may recover from bleaching episodes (Hughes et al., 2003; West & Salm, 2003), nutrient influx, especially of nitrate and phosphate, is thought to increase susceptibility to bleaching (Wiedenmann et al., 2013), and compromise this resilience (Hall et al., 2018). The scale of human impacts is, therefore, bound to intensify over time as temperatures are expected to keep rising, while the human population in coastal zones increases, attracted by the goods and services of the marine ecosystems (Neumann et al, 2015).

Over the past 50 years, Kenya has mitigated anthropogenic impacts on its coastal and marine resources, through MPAs as the primary marine ecosystems management strategy (Samoilys & Obura, 2011). Two classes of MPAs, differing in the level of protection, were established and enforced by the Kenya Wildlife Services (KWS): marine parks, “no-take” areas protected from all forms of consumptive utilization, and marine reserves which are contiguous to parks acting as buffers. Here, artisanal, but not commercial fishing, is allowed (Tuda & Omar, 2012). Owing to the success of MPAs, local communities are increasingly establishing and enforcing conservation areas termed as “locally managed marine areas” (LMMAs) which adopt the “no-take” protection management plan, like marine parks. Most LMMAs are administered with financial and technical support from government of non-governmental organizations (NGOs).

The primary goal for implementation of MPAs and LMMAs is biodiversity conservation and the regulation of fisheries. These efforts have generally been successful, especially in increased

and improved fish populations (Ransom & Mangi, 2010; Samoilys & Obura, 2011). While these strategies seem effective in curbing the direct impacts of human-related pressures such as overfishing and fisheries-related damages, it is necessary to interrogate their effectiveness on subtler effects that may result from spatially removed urbanization, development, agriculture and industrialization. Globally, these activities are major drivers of marine pollution (Islam & Tanaka, 2004) posing health risks to fishery products, consumers and recreational users on the coasts (Kilinc & Besler, 2014), as well as leading to the death of corals and other reef organisms (Hipsey et al, 2008).

We conducted a pilot study to assess whether area-based marine protection may have potential to curb human-source pollution. In coral reefs under protection of differing levels, we measured concentrations of surrogates for human-source pollution including, *E. coli*, a conventional indicator for faecal pollution and the subsequent public health risk from pathogenic microorganisms (Fremaux et al., 2009). We also measured nitrate and phosphate as they are known to be increased in marine environments by human activities (Jickells, 1998; Sayadi et al., 2016; Wiedenmann et al., 2013). Furthermore, nitrate and phosphate are important for the optimum physiological functioning of coral symbionts and in determining coral reef resilience (Rosset et al., 2017) whereby their enrichment is thought to increase susceptibility to bleaching (Wiedenmann et al., 2013), compromise resilience, promote algal dominance over corals (Hall et al., 2018) and suppress calcification (Kinsey & Davies, 1979).

Materials and Methods

Sampling was done between March and April 2018 during the long rains season at three sites on the northern coast of Kenya: 1) Kuruwitu Conservancy, an LMMA administered as a “no-take” area since 2006, had the “strictest” protection safeguarded from all forms of consumptive utilization and human activities, save for research and tourism; 2) Malindi Marine Reserve, established in 1968 (Lambo & Ormond, 2006), was considered “moderate” protection. The reserve is contiguous to Malindi Marine Park where it acts as a buffer to the “no-take” area. Here, artisanal fishing as well as research and tourism is allowed, but commercial forms of utilization are prohibited (Tuda & Omar, 2012); and 3) Kilifi creek is not protected hence was designated the “least protected”. It is open to fishing, recreation and experiences effluent input from surrounding hotels and residential houses.

Duplicate samples of 1L seawater each were collected in the morning between 08:00 -12:00 at 1 - 2 meters depth at low tide within 15-20 cm of the dominant species of coral (*Acropora* spp.) 500 - 2000 m from the shore. The samples were then transported on ice to the laboratory for processing within 3 hours post sampling.

Nitrate concentration was determined with ultraviolet spectrophotometry at 275 nm. Phosphate concentration was derived from orthophosphate colorimetric measurement at 690 nm using the Stannous Chloride method (APHA et al., 1999). Concentrations of viable *E. coli* cells were estimated by a three-tube most probable number (MPN) method as described by Vincy et al (Vincy et al., 2015).

Results

E. coli concentrations decreased with increased levels of protection. At Kuruwitu, the site with the strictest protection, *E. coli* was not detected (<3 MPN/100 mL), whilst Malindi (moderate protection) recorded 16 MPN/100 mL and Kilifi (no protection) 29 MPN/100 mL. Between-sites differences of *E. coli* density were not statistically significant (Fig 3.3.1.).

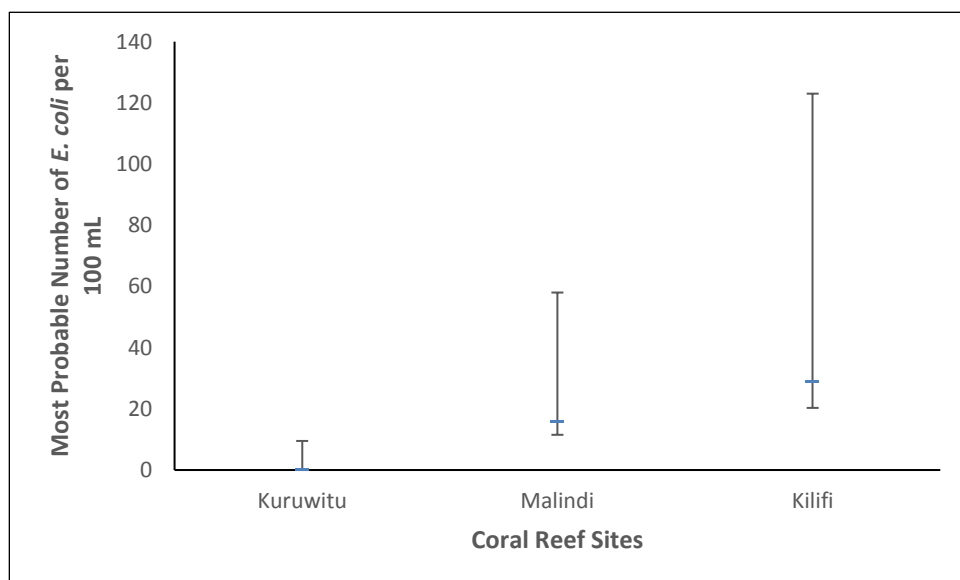


Figure 1: *E. coli* concentrations (MPN/100 mL) in water sampled at three different coral reef sites. Error bars indicate 95% confidence intervals.

Orthophosphate mean concentrations differed significantly among sites, increasing with the degree of marine protection (Table 3.3.1.): the highest concentration, 0.524 mg/L, was observed at Kuruwitu followed by Malindi (0.422 mg/L) and Kilifi (0.326 mg/L) ($P < 0.01$). Nitrate mean concentrations also differed significantly among sites but did not correlate with the level of protection. Kuruwitu registered 0.566 mg/L, Kilifi 0.188, and Malindi, 1.402 ($P < 0.01$).

Table 1: Nutrient mean concentrations for in water samples collected at from the coral reef sites. Between-sites comparisons tested by ANOVA.

Nutrient	Kuruwitu	Malindi	Kilifi	<i>P</i>
ORTHOPHOSPHATE				
(mg/L \pm SD)	0.523 \pm 0.018	0.421 \pm 0.001	0.325 \pm 0.004	<0.01
NITRATE				
(mg/L \pm SD)	0.565 \pm 0.007	0.187 \pm 0.001	1.401 \pm 0.004	<0.01

Discussion

The coral reef sites studied differed in both surrogates of human-source pollution investigated. As expected, *E. coli* concentrations decreased as the level of protection increased. Conversely, mean orthophosphate concentrations increased with increasing protection while mean nutrient concentration did not correlate with protection. These observations can be explained by the wide and differing range of human activities taking place around each study site, the level of marine protection and the proximity of each sampling site to a town.

In urban settings, wastewater and solid waste disposal is the principal cause of microbial pollution and nutrient loading (particularly nitrate) in the ocean (Okuku et al., 2011; Wakida & Lerner, 2005). The Kilifi site, which registered the highest *E. coli* and nitrate concentrations, is a creek dividing two business centers, Mnarani and Kilifi towns, which together constitute the capital of Kilifi County. The creek is surrounded by resorts and residential houses leaking and/or disposing wastewater into the ocean, and the beach is frequented by foreign tourists and residents, and holiday makers for recreation. In contrast, Malindi Marine Park and Kuruwitu Conservancy have limited, if any, exposure to urban pollution due to their distance from towns and levels of protection; the Malindi site (moderate protection) is about 10 km from the town, and Kuruwitu (highest protection) about 23 km from Kilifi town.

The Kuruwitu site shares a beach with several affluent residential homes with swimming pools, lush lawns utilizing fertilizers, as well as commercial farms. The conservancy administration reported seeking the intervention of National Environmental Management Agency (NEMA) to stop residents from discharging swimming pools wastewaters directly into the ocean during the initiation phase of the LMMA. This suggests that the level of marine protection may also explain the variations in *E. coli* concentrations observed.

Agriculture-related activities are the main contributors of both orthophosphates and nitrate in rural settings (Coulter, Kolka, & Thompson, 2004). We expected Malindi (where the Sabaki river meets the ocean) to contain higher mean nutrient concentrations than Kuruwitu since the Sabaki river runs through a catchment area with farming activities. However, the opposite was observed. Run-off from Kuruwitu lawns and commercial farms may explain the elevated orthophosphate and nitrate concentrations compared to Malindi where little or no farming occurs in the immediate vicinity of the sampling site. Furthermore, the Sabaki river has been shown to have the lowest discharge in the ocean during the north-eastern to south-eastern Monsoon transition period (March-April), which was when this study sampled. Most river sediment was deposited along the beach and intertidal area, and not in the neritic zone where this study sampled (Munyao, Tole, & Jungerius, 2003). These observations support our results, assuming that nutrients are deposited in a similar manner as sediments.

There are no universal *E. coli* and nutrients references for tropical coral reefs but in comparison to local standards elsewhere, the levels observed in this study are of public and environmental health concern. Although the *E. coli* concentration was lower than the threshold allowable for bathing beaches in parts of Europe (Baudart, et al., 2009), it may have been much higher at the shore close to the point where the pollutants entered the sea as bacterial numbers would decrease as they moved towards the open ocean due to tidal dilution and the bactericidal activity of seawater (Vaccaro et al., 1948). The nutrient concentrations assessed were also high, consistent with eutrophication (Okuku et al., 2011) or may have been the result of resuspended sediments (which accumulate nutrients at higher concentrations than seawater) following disturbances caused by rains and sampling activities. This finding is alarming and needs to be confirmed by long-term follow-up study with supplementary data on possible impacts of the nutrient elevations on reef biodiversity.

This pilot study indicated that human activities around the Kenyan Coastal zone likely increase pollution that may negatively affect coral reef health. An exact cause for the observed variations may not be conclusively specified yet, but the likelihood of marine protection in mitigating against microbial pollution is suggested. This study, therefore, confirms the need to research the topic, and informs on critical areas of consideration in designing an effective study to definitively determine the role, if any, of MPAs in curbing marine pollution.

References

- APHA, AWWA, & AEF. (1999). Standard Methods for the Examination of Water and Wastewater (20th Edn). Washington DC: *American Public Health Association*.
- Baudart, J., Servais, P., Paoli, H. De, Henry, A., & Lebaron, P. (2009). Rapid enumeration of *Escherichia coli* in marine bathing waters : potential interference of nontarget bacteria. *Journal of Applied Microbiology*, 107, 2054–2062. <https://doi.org/10.1111/j.1365-2672.2009.04392.x>
- Coulter, C. B., Kolka, R. K., & Thompson, J. a. (2004). Water quality in agricultural, urban, and mixed land use watersheds. *Journal of the American Water Resources Association*, 40(6), 1593–1601. <https://doi.org/10.1111/j.1752-1688.2004.tb01608.x>
- Fremaux, B., Gritzfeld, J., Boa, T., & Yost, C. K. (2009). Evaluation of host-specific *Bacteroidales* 16S rRNA gene markers as a complementary tool for detecting fecal pollution in a prairie watershed. *Water Research*, 43(19), 4838–4849. <https://doi.org/10.1016/j.watres.2009.06.045>
- Hall, E. R., Muller, E. M., Goulet, T., Bellworthy, J., Ritchie, K. B., & Fine, M. (2018). Eutrophication may compromise the resilience of the Red Sea coral *Stylophora pistillata* to global change. *Marine Pollution Bulletin*, 131, 701–711. <https://doi.org/10.1016/j.marpolbul.2018.04.067>
- Hipsey, M. R., Antenucci, J. P., & Brookes, J. D. (2008). A generic, process-based model of microbial pollution in aquatic systems. *Water Resources Research*, 44(7). <https://doi.org/10.1029/2007WR006395>
- Hoegh-guldberg, O., Poloczanska, E. S., Skirving, W., & Dove, S. (2017). Coral Reef Ecosystems under Climate Change and Ocean Acidification. *Frontiers in Marine Science*, 4(May). <https://doi.org/10.3389/fmars.2017.00158>
- Hughes, T. P., Kerry, J. T., Álvarez-Noriega, M., Álvarez-Romero, J. G., Anderson, K. D., Baird, A. H., ... Wilson, S. K. (2017). Global warming and recurrent mass bleaching of corals. *Nature*, 543(7645), 373–377. <https://doi.org/10.1038/nature21707>

- Hughes T.P., Baird A.H., Bellwood D.R., Card M., Connolly S.R., Folke C., ... Roughgarden J. (2003). Climate change human impacts and the resilience of coral reefs. *Science*, 301(August), 929–934.
- Islam, M. S., & Tanaka, M. (2004). Impacts of pollution on coastal and marine ecosystems including coastal and marine fisheries and approach for management: A review and synthesis. *Marine Pollution Bulletin*, 48(7–8), 624–649. <https://doi.org/10.1016/j.marpolbul.2003.12.004>
- Jickells, T. D. (1998). Nutrient Biogeochemistry of the Coastal Zone. *Science*, 281(July), 217–221.
- Kilinc, B., & Besler, A. (2014). The occurrence of enteric bacteria in marine environment and pollution ., 3, 39–43.
- Kinsey, D. W., & Davies, P. J. (1979). Effects of elevated nitrogen and phosphorus on coral reef growth. *Limnology and Oceanography*, 24(5), 935–940.
- Lambo, A. L., & Ormond, R. F. G. (2006). Continued post-bleaching decline and changed benthic community of a Kenyan coral reef. *Marine Pollution Bulletin*, 52(12), 1617–1624. <https://doi.org/10.1016/j.marpolbul.2006.05.028>
- Moberg, F., & Folke, C. (1999). Ecological goods and services of coral reef ecosystems. *Ecological Economics*, 29(2), 215–233. [https://doi.org/10.1016/S0921-8009\(99\)00009-9](https://doi.org/10.1016/S0921-8009(99)00009-9)
- Munyao, T., Tole, M. P., & Jungerius, P. D. (2003). Sabaki River transport and deposition in the Indian Ocean. *Recent Advances in Coastal Ecology, Studies from Kenya*, 119–132.
- Neumann, B., Vafeidis, A. T., Zimmermann, J., & Nicholls, R. J. (2015). Future coastal population growth and exposure to sea-level rise and coastal flooding - A global assessment. *PLoS ONE*, 10(3). <https://doi.org/10.1371/journal.pone.0118571>
- Obura, D. (2012). The Diversity and Biogeography of Western Indian Ocean Reef-Building Corals. *PLoS ONE*, 7(9). <https://doi.org/10.1371/journal.pone.0045013>
- Obura, D., Gudka, M., Bijoux, J., Freed, S., Gian, S. B., Maharavo, J., ... Ahamada, S. (2017). Coral reef status report for the Western Indian Ocean (2017) *Global Coral Reef Monitoring Network*.
- Okuku, E. O., Ohowa, B., Mwangi, S. N., Munga, D., Kiteresi, L. I., Wanjeri, V. O., ... Kilonzo, J. (2011). Sewage pollution in the coastal waters of Mombasa City, Kenya: A norm rather than an exception. *International Journal of Environmental Research*, 5(4), 865–874.
- Ransom, K. P., & Mangi, S. C. (2010). Valuing recreational benefits of coral reefs: The case of

- Mombasa Marine National Park and Reserve, Kenya. *Environmental Management*, 45(1), 145–154. <https://doi.org/10.1007/s00267-009-9402-9>
- Rosset, S., Wiedenmann, J., Reed, A. J., & D'Angelo, C. (2017). Phosphate deficiency promotes coral bleaching and is reflected by the ultrastructure of symbiotic dinoflagellates. *Marine Pollution Bulletin*, 118(1–2), 180–187. <https://doi.org/10.1016/j.marpolbul.2017.02.044>
- Samoilys, M. A., & Obura, D. O. (2011). Marine conservation successes in Eastern Africa. *Cordio East Africa*, (December), 1–28.
- Sayadi, M. H., Ahmadpour, N., Fallahi Capoorchali, M., & Rezaei, M. R. (2016). Removal of nitrate and phosphate from aqueous solutions by microalgae: An experimental study. *Global Journal of Environmental Science Management*, 2(4), 357–364. <https://doi.org/10.22034/gjesm.2016.02.04.005>
- Tuda, A., & Omar, M. (2012). Protection of Marine Areas in Kenya. *The George Wright Forum*, 29(1), 43–50. Retrieved from <http://www.georgewright.org/291tuda.pdf>
- Vaccaro, R. F., Briggs, M. P., Carey, L., & Ketchum, B. H. (1950). Viability of *Escherichia coli* in Sea Water. *American Journal of Public Health*, 40(10), 1257–1266.
- Vincy, M. V., Brilliant, R., & Pradeepkumar, A. P. (2017). Prevalence of indicator and pathogenic bacteria in a tropical river of Western Ghats, India. *Applied Water Science*, 7(2), 833–844. <https://doi.org/10.1007/s13201-015-0296-9>
- Wakida, F. T., & Lerner, D. N. (2005). Non-agricultural sources of groundwater nitrate: A review and case study. *Water Research*, 39(1), 3–16. <https://doi.org/10.1016/j.watres.2004.07.026>
- West, J. M., & Salm, R. V. (2003). Resistance and Resilience to Coral Bleaching: Implications for Coral Reef Conservation and Management. *Conservation Biology*, 17(4), 956–967. <https://doi.org/10.1046/j.1523-1739.2003.02055.x>
- Wiedenmann, J., D'Angelo, C., Smith, E. G., Hunt, A. N., Legiret, F. E., Postle, A. D., & Achterberg, E. P. (2013). Nutrient enrichment can increase the susceptibility of reef corals to bleaching. *Nature Climate Change*, 3(2), 160–164. <https://doi.org/10.1038/nclimate1661>

APPENDIX B: MOLECULAR ECOLOGY OF CORAL REEF MICROORGANISMS IN THE WESTERN INDIAN OCEAN COAST OF KENYA

Submitted to *Ecology and Evolution*
<https://doi.org/10.1101/2020.03.04.976431>

Abstract

Coral reefs face increased environmental threats from anthropomorphic climate change and pollution, from agriculture, industries and tourism. They are economically vital for many people worldwide, and harbour a fantastically diverse ecosystem, being the home for many species of fish and algae. Surprisingly little is known about the microbial communities living in and in the surrounding of coral reefs. Here we employ high throughput sequencing for investigating the bacteria living in the water column and upper sediment layer in close proximity to coral reefs on the Kenyan coast of the West Indian Ocean. We show that while the read-level taxonomic distribution of bacteria is similar with ones obtained from 16S metabarcoding, whole metagenome sequencing provides valuable functional insights not available with 16S metabarcoding. We find evidence of pollution, marked by the presence of *Vibrio* and more importantly the presence of antibiotic resistance notably to vancomycin, that we attribute to the use of avoparcin in agriculture. Additionally, 175 bacterial genomes not previously sequenced were discovered.

Our study is the first whole-metagenome study from the West Indian Ocean, provides a much-needed baseline to study microbes surrounding coral reefs under different conditions as well as the microbiome of coral reefs.

Keywords: *Coral reefs, Sequencing, Metagenomics, West Indian Ocean*

Introduction

Coral reefs are one of the most biodiverse ecosystems in the world, thereby providing vast ecological and socio-economic resources. However, coral reefs and consequently their invaluable ecosystem services are increasingly under threat globally due to climate change and a range of other human-related pressures, such as pollution from agriculture, industries and tourism-related activities. These stressors lead to coral degradation the onset of which is most often marked by bleaching following the expulsion of the symbiotic algae.

Addressing the challenges leading to coral death requires a comprehensive understanding of corals and their interactions with other members of the reef ecosystems. Appreciation of this fact has led to increased interest in marine microbiology research as microorganisms are thought to be critical for reef ecosystem processes including coral homeostasis, nutrition and protection against disease (Godoy-Vitorino et al., 2017). Microbial communities are also known to respond and adapt quickly to disturbance (Ainsworth et al., 2010). Therefore, studying coral reef-associated microorganisms, as well as microorganisms living in close proximity of coral reefs, holds the potential to help in improving the capacity to predict responses of coral reef ecosystems to changing environmental conditions.

The advancement of DNA sequencing and analysis technologies, allowing sequencing DNA directly without the need of cultivating organisms in laboratory settings, has augmented our understanding of the complexity and diversity of natural microbial populations (Biller et al., 2018). Most oceanic and costal surveys to date have used 16S metabarcoding, which provides precious insights about population complexity and diversity, but only a broad overview of the taxonomic distribution across samples with (i) limited resolution, especially for poorly characterised samples such as environmental ones and (ii), no or poor functional and metabolic insights of the sequenced communities (Poretsky et al., 2014; Rausch et al., 2019). Whole Metagenome sequencing (WMS) on the other hand, has the potential for better taxonomic resolution, theoretically at the species level, even though as for 16S taxonomic classification is highly tied to the database used and the type and provenance of samples. Additionally, WMS also allows for reconstructing draft genomes from metagenomes, providing exceptional phylogenomic insights as well as opening up potential for functional annotation. Combined with recent advances in protein assembly directly from metagenomic data, the method shows great promise to harness metabolic pathways and functional knowledge from microbial communities.

The Western Indian Ocean (WIO), however, remains the least studied (Díez et al., 2016) of the oceans, despite hosting the second largest hotspot for coral biodiversity globally, partly due to the region's lack of technological capacity. Marine metagenomes, like most environmental metagenomes, are diverse, making them especially difficult and expensive to analyse and interpret. Here we deeply sequence microbiomes from the water column and the upper sediment layer of three coastal reefs in Kenya, in an attempt to give an excellent taxonomic overview of WIO coastal communities and to unravel the functional characteristics of those rich environments.

Methods

Study Sites

The study was conducted at three marine protected areas (MPA) covering three of the five counties on the coast of Kenya Indian Ocean (Fig 1). Consisting of fringing reefs, each of the three sites was selected for its distinct human activities. Located about 118 km north of Mombasa, Malindi Marine National Park and Reserve is the oldest MPA in Kenya, having been gazetted in 1968 (McClanahan, Kaunda-Arara, & Omukoto, 2010). Sampling was done close to the reserve (3°15'35.1"S, 40°08'40.0"E) where artisanal fishing is allowed. The marine park is famous for glass-bottom boat tours and snorkelling among other recreational touristic activities. It also experiences significant year-round discharge of freshwater and sediments from the Sabaki River which runs through a catchment area dominated with agricultural settlements (Munyao et al., 2003; van Katwijk et al., 1993). Mombasa Marine National Park and Reserve (3°59'45.7"S, 39°44'50.1"E) was established in 1986 (Ngugi, 2001)– with restrictions of protection being enforced commencing 1991 (Tuda & Omar, 2012) – and is, arguably, the most visited of Kenya's marine parks by both local and international tourists (Owens, 1978). Due to its proximity to the urbanised touristic city, the park experiences pollution from hotels, hospitals, domestic and industrial waste disposals (Okuku et al., 2011; Mwangi et al., 2001). Moreover, because Mombasa is the primary port serving inland eastern and central African countries, the park is also impacted by marine traffic activities including oil pollution and dredge-spoil dumping (Mwangi et al., 2001). Lastly on the southern coast, 90 km from Mombasa, the Kisite-Mpunguti site (4°42'54.0"S, 39°22'23.8"E) is a protected area that comprises of Kisite Marine National Park and Mpunguti Reserve created in 1973 and gazetted in 1978. It is bordered by sparsely populated coral islands and experiences the least human activities because it is 11 km offshore (Emerton & Tessema, 2001).

Field Procedures

Sampling was done between 2016 and 2017 within the coral reefs, 200 - 500 m from the shore, at a depth of 1-2 m during low tides in the morning hours. At each site, seawater was collected in 5 L water bottles within 10 - 20 cm of a colony of *Acropora* spp. - the dominant coral species at the sites - for microbial DNA isolation. Additional triplicate 50 mL seawater samples were collected in disposable centrifuge tubes for nutrient analysis. A 10 mL syringe barrel was used to collect 2 cm column of sediment at the base of each sampled coral colony, 0.25 g of which was suspended in a bead tube containing inhibitor-dissolving and nucleic acid-preserving buffer. Samples were transported on ice to the laboratory for processing, typically within 3 hours of sampling.

Physicochemical parameters including water temperature, pH, salinity, and dissolved oxygen were determined *in situ*, at the time of sampling within the coral reefs, using portable multiprobe water quality meters, per manufacturer's instructions (YSI Inc., Yellow Spring, OH).

Laboratory Procedures

Nutrient testing

Spectrophotometry methods were used to determine seawater concentrations of nitrate (NO_2^- - N), nitrite (NO_3^- - N), ammonium (NH_4^+ - N) and phosphate (PO_4^{3-} - P) nutrients (Supplementary Table 1 and 2) as described in Ongore et al., 2013 (Ongore et al., 2013).

DNA isolation

For each site, 4 L of seawater was vacuum-filtered (VWR, West Chester, PA, USA) through a 0.2-um pore size membrane (Pall Corporation, Port Washington, NY, USA) to capture microbial cells which were then added to a bead tube with lysis buffer. PowerWater® DNA isolation kits were used to isolate microbial DNA from seawater samples while sediment samples were extracted with PowerSoil® DNA isolation kits according to the manufacturer's instructions (Mo Bio, Inc., Carlsbad, CA, USA). Quality and quantity of DNA were checked by Nanodrop and suitability for sequencing of DNA samples for metagenomics analysis was confirmed with 1% agarose gel (Rohwer, Seguritan, Azam, & Knowlton, 2002).

Library preparation and DNA sequencing

Sequencing libraries were prepared from 1µg of DNA according to the manufacturers' preparation guide # 15036187 using the TruSeq DNA PCR-free library preparation kit (20015962/3, Illumina Inc.).

Briefly, the DNA was fragmented using a Covaris E220 system, aiming at 350bp fragments. Resulting DNA fragments were end-repaired, and the 3' end adenylated to generate an overhang. Adapter sequences were ligated to the fragments via the A-overhang and the generated sequencing library was purified using AMPure XP beads (Beckman Coulter). The quality of the library was evaluated using the FragmentAnalyzer system and a DNF-910 kit. The adapter-ligated fragments were quantified by qPCR using the Library quantification kit for Illumina (KAPA Biosystems/Roche) on a CFX384Touch instrument (BioRad) before cluster generation and sequencing.

A 200 pM solution of the individual sequencing libraries was subjected to cluster generation and paired-end sequencing with 150bp read length using an S2 flowcell on the NovaSeq system (Illumina Inc.) using the v1 chemistry according to the manufacturer's protocols.

Base-calling was done on the instrument by RTA 3.3.3 and the resulting .bcl files were demultiplexed and converted to fastq format with tools provided by Illumina Inc., allowing for one mismatch in the index sequence.

Sequencing was performed by the NGI SNP&SEQ Technology Platform in Uppsala, Sweden www.sequencing.se.

Bioinformatics Analyses

The raw Illumina reads were trimmed at Q5 threshold (Macmanes, 2014), and the adapters were removed using fastp v0.19.5 (Chen, Zhou, Chen, & Gu, 2018). Trimmed sequences were deposited to the European Nucleotide Archive under the study accession PRJEB30838.

The trimmed reads were assigned a taxonomic classification using a combination of Kraken v2.0.8 and bracken v2.2 against the nt database using default parameters (Lu, Breitwieser, Thielen, & Salzberg, 2017; Wood, Lu, & Langmead, 2019). Rarefaction curves were computed using R and vegan v2.5 (Oksanen et al., 2019; R Core Team, 2018).

The samples were assembled using megahit v1.1.4 with the options --k-min 27 --k-max 147 --k-step 10 (Li, Liu, Luo, Sadakane, & Lam, 2015). The reads were then mapped to the assemblies with bowtie v2.2.9 (Langmead & Salzberg, 2012) using default parameters and binned into draft genomes with metabat v2.11.1 with option --minContig 1500 (Kang et al., 2019). The genomes bins were then quality checked and refined with checkm v1.0.7 and refinem v0.0.24 (Parks, Imelfort, Skennerton, Hugenholtz, & Tyson, 2015; Parks et al., 2017) (scripts and refining parameters are available at <https://osf.io/5fzqu/>), and the best genome bins were annotated using prokka v1.10 (Seemann, 2014) and eggno-mapper v1.0.3 (Huerta-Cepas et al., 2017), as well as placed phylogenetically using gtdbtk v0.3.2 (Chaumeil et al., 2019). The tree figure was generated using ggtree v1.14.6 (Yu et al., 2017).

The trimmed reads were also assembled directly at the protein level using plasm (commit 26b5d6625a2fbef4cfaab4bfaa99b1682d35921c) (Steinegger et al., 2018). The resulting assemblies were then clustered at 40, 50 and 90% identity using cd-hit v4.7 (Fu et al., 2012) and annotated using eggno-mapper v1.0.3.

Results

Taxonomic distribution of sediment and water associated organisms

Of the 4.2 billion reads obtained, 608 million were classified as bacteria, 16 million as archaea, 12 million as viruses, 537 million as eukaryotic sequences and 3 billion remained unclassified. The bacterial species richness was estimated at around 15000 for all samples (Supplementary Fig 1), which is high but within reasonable magnitudes according to previously published large metagenomes (Rodriguez-R & Konstantinidis, 2014). The vast majority of bacterial sequences were identified as proteobacteria (Figure 2). The second most abundant phyla were *Bacteroidetes* in the water samples and *Cyanobacteria* in the sediment samples.

At the genera level, *Pseudomonas* was ubiquitous in all samples (Figure 2), and dominant in the sediment, whilst the diversity of Proteobacteria was found to be much higher in the water samples, constituting the five most abundant genera. A non-negligible fraction of *Vibrio* was also found in all sampling sites, mainly in the water samples. The most abundant cyanobacteria in the Sediment samples were *Nostoc*, *Stanieria*, *Cyanothece*, *Calothrix* and *Synechococcus* (with *Synechococcus* being also found in abundance in the water samples). In water, Bacteroidetes were dominated by Flavobacteriaceae, more specifically the genera

Flavobacterium, *Zunongwangia*, *Chryseobacterium* and *Capnocytophaga*. Lastly, abundant traces of *Bacillus*, *Paenibacillus*, *Lactobacillus* and *Streptococcus* were found.

The classified archaeal reads were mostly divided into two phyla: Euryarchaeota and Thaumarchaeota, with the former dominating the water samples and the latter the sediment samples. The Thaumarchaeota fraction is explained by the presence of *Nitrosopumilus*, a common organism living in seawater. The Euryarchaeota fraction is a bit more diverse but is mainly comprised of methanogens (Supplementary Fig 2).

Assembly and Functional annotation of proteins

A total of 21 million proteins were assembled, 12 million of which were unique proteins (after clustering at 90% identity). A total of 424 distinct KEGG pathways were identified in the data, which supports high diversity in function, especially in the water samples. Of note is that, in the 20 most abundant pathways, we found a total of 952017 genes associated with antibiotic biosynthesis, making it the second most abundant KEGG category after Biosynthesis of secondary metabolites (Fig. 3).

Amongst less abundant but still expressed pathways, we also find carbon, methane and nitrogen metabolism, as well as photosynthesis (Supplementary Fig 3). Additionally, the most abundant pathways associated with antibiotics were related to monobactam, streptomycin and vancomycin. (Supplementary Fig 4.).

Metagenome-assembled genomes

A total of 782 genome bins were recovered. Of those, 193 presented more than 50% completeness and less than 15% contamination according to the checkm results. One hundred seventy-eight of those were bacterial genomes, while 15 were classified as archaea.

Twenty-eight of the 178 bacterial genomes had a >95% match to an already published genome. The remaining 150 are either new strains or new species. All genomes were classified to at least the Phylum level. More than half of the recovered Bacteria were Proteobacteria, Bacteroidetes and Cyanobacteria (Fig. 4).

The recovered archaeal genomes all came from water samples. Of the 15, 7 were classified as known archaea from the order Poseidoniales, and eight are newly discovered archaeal species, all putatively placed in the Poseidoniaceae family.

Discussion

Bacterial communities from Indian Ocean reefs – and coastal waters – are critically understudied. Here we present a catalogue of the microorganisms present in the upper sediment layer and the water near coral reefs, as well as a catalogue of putative proteins and functions of said microorganisms.

This study presents – to the extent of our knowledge – the first metagenomes taken from the coastline of the West Indian Ocean, and offers a baseline for much needed further work on conservation and monitoring of the West Indian ocean coasts. We present and publish a catalogue of 12 million putative proteins and 193 draft genomes, including 175 previously unpublished bacteria. While it had been valuable also to investigate the coral microbiome itself, our study presents a solid baseline for monitoring water quality which we hypothesise may be a good proxy for coral health. Indeed, while physicochemical properties of coastal waters have remained stable in the region regardless of pollution status, bacterial communities may not be.

The taxonomic distribution of bacterial species, as presented in this study, while diverse, is consistent with previously published coastal metabarcoding studies (Kelly et al., 2014). The presence of *Vibrio* in the water samples is worrying, especially given the presence of *Vibrio coralliilyticus* and *Vibrio owensii*. The former may be commensal at some water temperatures, but the pathogenicity of both species is well documented (Gibbin et al., 2019; Ushijima, Smith, Aeby, & Callahan, 2012).

Even though there were no obvious differences in the distribution of phyla between sites and sample types, variations were observed in the abundances of three ecologically important genera. *Candidatus* Pelagibacter, *Prochlorococcus* and *Synechococcus*, which are considered to be ubiquitous in marine environments (Biller, Berube, Lindell, & Chisholm, 2015; Morris et al., 2002; Ruffing, Jensen, & Strickland, 2016), had higher abundances in water at the Kisite-Mpunguti site. *Candidatus* Pelagibacter is believed to be the most successful clade of organisms on Earth and, although a heterotroph, it is known to thrive at the low nutrient concentrations typical of open ocean conditions (Zhao et al., 2017). Overall, members of the

Alphaproteobacteria class are known to have higher relative abundance in habitats with higher coral cover than in nutrient-rich algae-dominated habitats. Besides *Candidatus Pelagibacter*, no differences were found in the distribution of genera from the class *Alphaproteobacteria*, between sites or samples. On the other hand, members of *Synechococcus* and *Prochlorococcus* are cyanobacteria that are considered the most important primary producers in the tropical oceans, responsible for a large percentage of the photosynthetic production of oxygen (Biller et al., 2015; Kim et al., 2018; Waterbury, Watson, Guillard, & Brand, 1979). Being autotrophs, members of these genera are usually found in great abundance in ocean zones low in nutrients (Dinsdale et al., 2008; Kelly et al., 2014). Also, cyanobacteria have high adaptive capacities for nutrients and light harvesting, a competitive advantage over other marine microorganisms (Louati et al., 2015). As such, *Synechococcus* and *Prochlorococcus* may have outcompeted the other microbes at the Kuruwitu site where its distance from the coast, settlement and increased human activities, may have had limited nutrients essential for their growth.

The main strength of metagenomics over metabarcoding is the insight into function, provided by the protein assembly as well as the binning. We showed that the water samples the presence of many antibiotic-related pathways, for both biosynthesis and resistance. It is particularly striking that the majority of Metagenome-Assembled-Genomes (MAGs) from the *Bacteroidetes* phylum indicated the presence of the vancomycin resistance pathway. While surprising at first given vancomycin is not considered a first-line therapy antibiotic, it may be explained by the use of avoparcin as a food supplement in agriculture. Antibiotic resistant bacteria have been observed widely in aquatic environments following antibiotic contamination from wastewater treatment plants or agricultural runoffs (Schmieder & Edwards, 2012). Avoparcin is a glycopeptide antibiotic that is chemically very similar to vancomycin; there have been earlier concerns about the use of avoparcin in agriculture in various countries as well as Kenya (Bager, Madsen, Christensen, & Aarestrup, 1997; Nilsson, 2012; Raphael, Sam, Anne, Peter, & Samuel, 2017), and it would be reasonable to think it would be the cause of vancomycin resistance gene clusters in coastal waters.

Most assembled proteobacteria showed signs of being autotrophic, presenting carbon fixation and metabolism pathways. Some also seemed to be able to fix nitrogen. The cyanobacteria retrieved from the sediment layer exhibited photosynthetic pathways. These organisms may contribute a great deal of nutrient exchange in the whole ecosystem, even though they are not living in direct symbiosis with corals, their contribution cannot be ruled insignificant.

Lastly, quorum sensing was found to be one of the most abundant pathways. While relatively little is known about it, bacteria use quorum sensing as a way to chemically communicate with each other, which has its importance in nutrient cycling across and within microbial communities (DeAngelis, Lindow, & Firestone, 2008).

The difference in taxonomic distribution between the read-level classification and the genome binning is quite striking in a few aspects. Almost no draft genomes from the sediment samples passed the threshold for draft genomes of acceptable quality, and this could be explained by a combination of factor and biases at different levels of the experiment, leading to assemblies of poorer quality. All sediment samples resulted in more prominent and more fragmented assemblies (Supplementary table 3) than the water samples. Metagenomes are notoriously tricky to assemble, due to their variable genome coverage and non-clonal nature (Breitwieser, Lu, & Salzberg, 2019), and additional factors such as unusual GC content or repeat-rich genomes may have played a role. The total size of the assemblies may also indicate that the sediment samples are more diverse and complex than their water counterpart, but that diversity and complexity may be poorly represented in our results, due to the incomplete nature of our biological databases and the imperfectability of our algorithms.

In conclusion, the pathways analysis, as well as the annotation of the draft genomes, provide insights into the putative nutrients exchange and other interactions between bacteria and the environment, including corals. Through nitrogen fixation, photosynthesis and carbon metabolism, water and sediment bacteria may prove valuable to nutrient cycling in a healthy reef. Additionally, we hypothesise that coastal bacterial communities are a potential health indicator for reefs, especially regarding antibiotic resistance, potentially from agricultural runoff, as well for opportunistic *Vibrio* pathogens. While metagenomics is expensive and inconvenient to use in a monitoring setting, our dataset may prove valuable in designing more targeted primer-based approaches to detect pollution in coastal communities. It is also possible that – with the advance of long-read portable sequencers such as the oxford nanopore MinION, the cost barrier for field sequencing for monitoring purposes drops dramatically, making metagenomics a viable approach for monitoring coastal communities. Long-read sequencing could also be used for obtaining full-length 16s sequences, which could potentially provide resolution up to the species level.

Acknowledgements

This work was carried out under research permit from the National Commission for Science, Technology and Innovation (NACOSTI). Kenya Wildlife Service (KWS) provided logistical support and granted permission to work in the marine parks and the National Environmental Management Authority (NEMA) provided access permits.

The authors would like to acknowledge support from Science for Life Laboratory, the National Genomics Infrastructure, NGI, and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure.

This work was supported by the Swedish Research Council, grant number 2015-03443_VR.

Data accessibility

Trimmed sequences were deposited to the European Nucleotide Archive under the study accession PRJEB30838.

Scripts and Refining parameters for the genome binning are available at <https://osf.io/5fzqu/>

Author contributions

SW designed the study, conducted field sampling and laboratory processing, drafted and revised the manuscript. HG designed the study, performed the bioinformatics analyses and drafted the manuscript. EV, OKL, NW, EBR, AMD, SV helped design the study and edited the manuscript.

References

- Ainsworth, T. D., Thurber, R. V., & Gates, R. D. (2010). The future of coral reefs: a microbial perspective. *Trends in Ecology & Evolution*, 25(4), 233–240. doi: 10.1016/j.tree.2009.11.001
- Bager, F., Madsen, M., Christensen, J., & Aarestrup, F. M. (1997). Avoparcin used as a growth promoter is associated with the occurrence of vancomycin-resistant *Enterococcus faecium* on Danish poultry and pig farms. *Preventive Veterinary Medicine*, 31(1–2), 95–112. doi: 10.1016/s0167-5877(96)01119-1
- Biller, S. J., Berube, P. M., Dooley, K., Williams, M., Satinsky, B. M., Hackl, T., ... Chisholm, S. W. (2018). Marine microbial metagenomes sampled across space and time. *Scientific Data*, 5, 180176. doi: 10.1038/sdata.2018.176
- Breitwieser, F. P., Lu, J., & Salzberg, S. L. (2019). A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics*, 20(4), 1125–1136. doi: 10.1093/bib/bbx120

- Chaumeil, P.-A., Mussig, A. J., Hugenholtz, P., & Parks, D. H. (2019). GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics*. doi: 10.1093/bioinformatics/btz848
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. doi: 10.1093/bioinformatics/bty560
- DeAngelis, K. M., Lindow, S. E., & Firestone, M. K. (2008). Bacterial quorum sensing and nitrogen cycling in rhizosphere soil. *FEMS Microbiology Ecology*, 66(2), 197–207. doi: 10.1111/j.1574-6941.2008.00550.x
- Díez, B., Nylander, J. A. A., Ininbergs, K., Dupont, C. L., Allen, A. E., Yooseph, S., ... Bergman, B. (2016). Metagenomic Analysis of the Indian Ocean Picocyanobacterial Community: Structure, Potential Function and Evolution. *PloS One*, 11(5), e0155757. doi: 10.1371/journal.pone.0155757
- Emerton, L., & Tessema, Y. (2001). Economic constraints to the management of marine protected areas: the case of Kisite Marine National Park and Mpunguti Marine National Reserve, Kenya. Retrieved from <https://www.cbd.int/financial/values/kenya-ecoparks.pdf>
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23), 3150–3152. doi: 10.1093/bioinformatics/bts565
- Gibbin, E., Gavish, A., Krueger, T., Kramarsky-Winter, E., Shapiro, O., Guiet, R., ... Meibom, A. (2019). *Vibrio coralliilyticus* infection triggers a behavioural response and perturbs nutritional exchange and tissue integrity in a symbiotic coral. *The ISME Journal*, 13(4), 989–1003. doi: 10.1038/s41396-018-0327-2
- Godoy-Vitorino, F., Ruiz-Diaz, C. P., Rivera-Seda, A., Ramírez-Lugo, J. S., & Toledo-Hernández, C. (2017). The microbial biosphere of the coral *Acropora cervicornis* in Northeastern Puerto Rico. *PeerJ*, 5, e3717. doi: 10.7717/peerj.3717
- Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., & Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, 34(8), 2115–2122. doi: 10.1093/molbev/msx148
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., & Wang, Z. (2019). *MetaBAT 2*: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 7, e7359. doi: 10.7717/peerj.7359
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*

- Methods*, 9(4), 357–359. doi: 10.1038/nmeth.1923
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., & Lam, T.-W. (2015). MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10), 1674–1676. doi: 10.1093/bioinformatics/btv033
- Lu, J., Breitwieser, F. P., Thielen, P., & Salzberg, S. L. (2017). Bracken: estimating species abundance in metagenomics data. *PeerJ Computer Science*, 3, e104. doi: 10.7717/peerj-cs.104
- Macmanes, M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Frontiers in Genetics*, 5, 13. doi: 10.3389/fgene.2014.00013
- McClanahan, T. R., Kaunda-Arara, B., & Omukoto, J. O. (2010). Composition and diversity of fish and fish catches in closures and open-access fisheries of Kenya. *Fisheries Management and Ecology*, 17(1), 63–76. doi: 10.1111/j.1365-2400.2009.00717.x
- Munyao, T. M., Tole, M. P., & Jungerius, P. D. (2003). Sabaki River sediment transport and deposition in the Indian Ocean. *Research Reports-African Studies Centre Leiden*, 119–132.
- Ngugi, I. (2001). Economic impacts of marine protected areas: A case study of the Mombasa Marine Park in Kenya. *WIOMSA Book Series*, (1), 507–516.
- Nilsson, O. (2012). Vancomycin resistant enterococci in farm animals - occurrence and importance. *Infection Ecology & Epidemiology*, 2. doi: 10.3402/iee.v2i0.16959
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., ... Wagner, H. (2019). vegan: Community Ecology Package. Retrieved from <https://CRAN.R-project.org/package=vegan>
- Okuku, E. O., Ohowa, B., Mwangi, S. N., Munga, D., Kiteresi, L. I., Wanjeri, V. O., ... Kilonzo, J. (2011). Sewage pollution in the Coastal waters of Mombasa City, Kenya: A norm Rather than an Exception. *International Journal of Environmental Research and Public Health*, 5(4), 865–874. doi: 10.22059/ijer.2011.444
- Ongore, C. O., Okuku, E. O., Mwangi, S. N., Kiteresi, L. I., Ohowa, B. O., Wanjeri, V. O., ... Kilonzi, J. (2013). Characterization of nutrients enrichment in the estuaries and related systems in Kenya coast. Retrieved from <http://www.cnstn.rnrt.tn/raf7009/wp-content/themes/Savarona/documents/raf700924032014/Ongore%20et%20al%202013.pdf>
- Owens, J. D. (1978). Coliform and escherichia coli bacteria in seawater around penang island, Malaysia. *Water Research*, 12(6), 365–370. doi: 10.1016/0043-1354(78)90101-X
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM:

- assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, 25(7), 1043–1055. doi: 10.1101/gr.186072.114
- Parks, D. H., Rinke, C., Chuvochina, M., Chaumeil, P.-A., Woodcroft, B. J., Evans, P. N., ... Tyson, G. W. (2017). Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, 2(11), 1533–1542. doi: 10.1038/s41564-017-0012-7
- Poretzky, R., Rodriguez-R, L. M., Luo, C., Tsementzi, D., & Konstantinidis, K. T. (2014). Strengths and limitations of 16S rRNA gene amplicon sequencing in revealing temporal microbial community dynamics. *PloS One*, 9(4), e93827. doi: 10.1371/journal.pone.0093827
- R Core Team. (2018). R: A Language and Environment for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Raphael, Z. S., Sam, L., Anne, W. M., Peter, G. W., & Samuel, K. (2017). Antimicrobial susceptibility patterns of *Staphylococcus aureus* and coagulase negative staphylococci isolated from humans in Nairobi, Kenya. *African Journal of Microbiology Research*, 11(18), 705–711. doi: 10.5897/AJMR2017.8481
- Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S., Dagan, T., Dierking, K., ... Baines, J. F. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome*, 7(1), 133. doi: 10.1186/s40168-019-0743-1
- Rodriguez-R, L. M., & Konstantinidis, K. T. (2014). Nonpareil: a redundancy-based approach to assess the level of coverage in metagenomic datasets. *Bioinformatics*, 30(5), 629–635. doi: 10.1093/bioinformatics/btt584
- Rohwer, F., Seguritan, V., Azam, F., & Knowlton, N. (2002). Diversity and distribution of coral-associated bacteria. *Marine Ecology Progress Series*, 243, 1–10. Retrieved from <https://www.int-res.com/abstracts/meps/v243/p1-10/>
- Seemann, T. (2014). *Prokka*: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068–2069. doi: 10.1093/bioinformatics/btu153
- Steinegger, M., Mirdita, M., & Söding, J. (2018). Protein-level assembly increases protein sequence recovery from metagenomic samples manifold (p. 386110). doi: 10.1101/386110
- Stephen Mwangi, David Kirugara, Melckzedek Osore, Joyce Njoya, Abdalla Yobe and Thomas Dzaha. (2001). State of Marine Pollution in Mombasa Marine National Park Marine Reserve and Mtwapa Creek.

- Tuda, A., & Omar, M. (2012). Protection of Marine Areas in Kenya. *The George Wright Forum*, 29(1), 43–50. Retrieved from <http://www.jstor.org/stable/43598974>
- Ushijima, B., Smith, A., Aeby, G. S., & Callahan, S. M. (2012). *Vibrio owensii* induces the tissue loss disease Montipora white syndrome in the Hawaiian reef coral *Montipora capitata*. *PloS One*, 7(10), e46717. doi: 10.1371/journal.pone.0046717
- van Katwijk, M. M., Meier, N. F., van Loon, R., van Hove, E. M., Giesen, W. B. J. T., van der Velde, G., & den Hartog, C. (1993). Sabaki River sediment load and coral stress: correlation between sediments and condition of the Malindi-Watamu reefs in Kenya (Indian Ocean). *Marine Biology*, 117(4), 675–683. doi: 10.1007/BF00349780
- Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with *Kraken 2*. *Genome Biology*, 20(1), 257. doi: 10.1186/s13059-019-1891-0
- Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree : an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution / British Ecological Society*, 8(1), 28–36. doi: 10.1111/2041-210X.12628

Tables and Figures

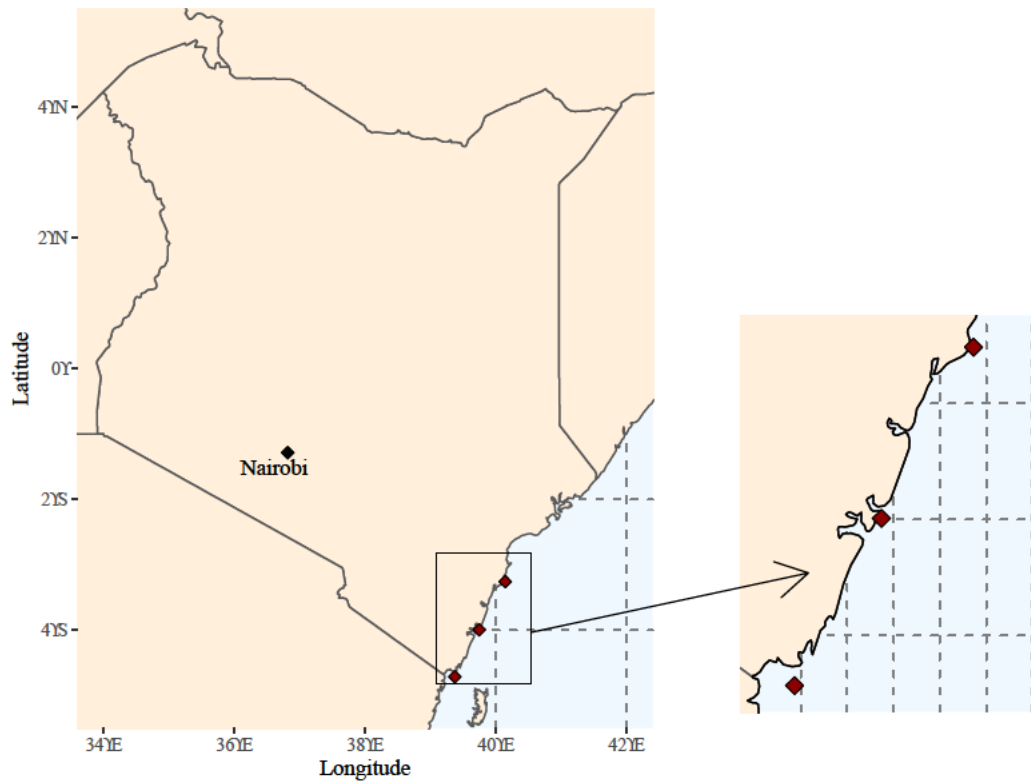


Figure1: Location of the sampling sites. The three sampling sites are indicated in red, respectively from top to bottom: Malindi, Mombasa and Kisite. Each of the three sites was selected for its human activities. Sampling was done in 2016 and 2017, 200 - 500 m from the shore, at a depth of 1-2 m during low tides in the morning hours.

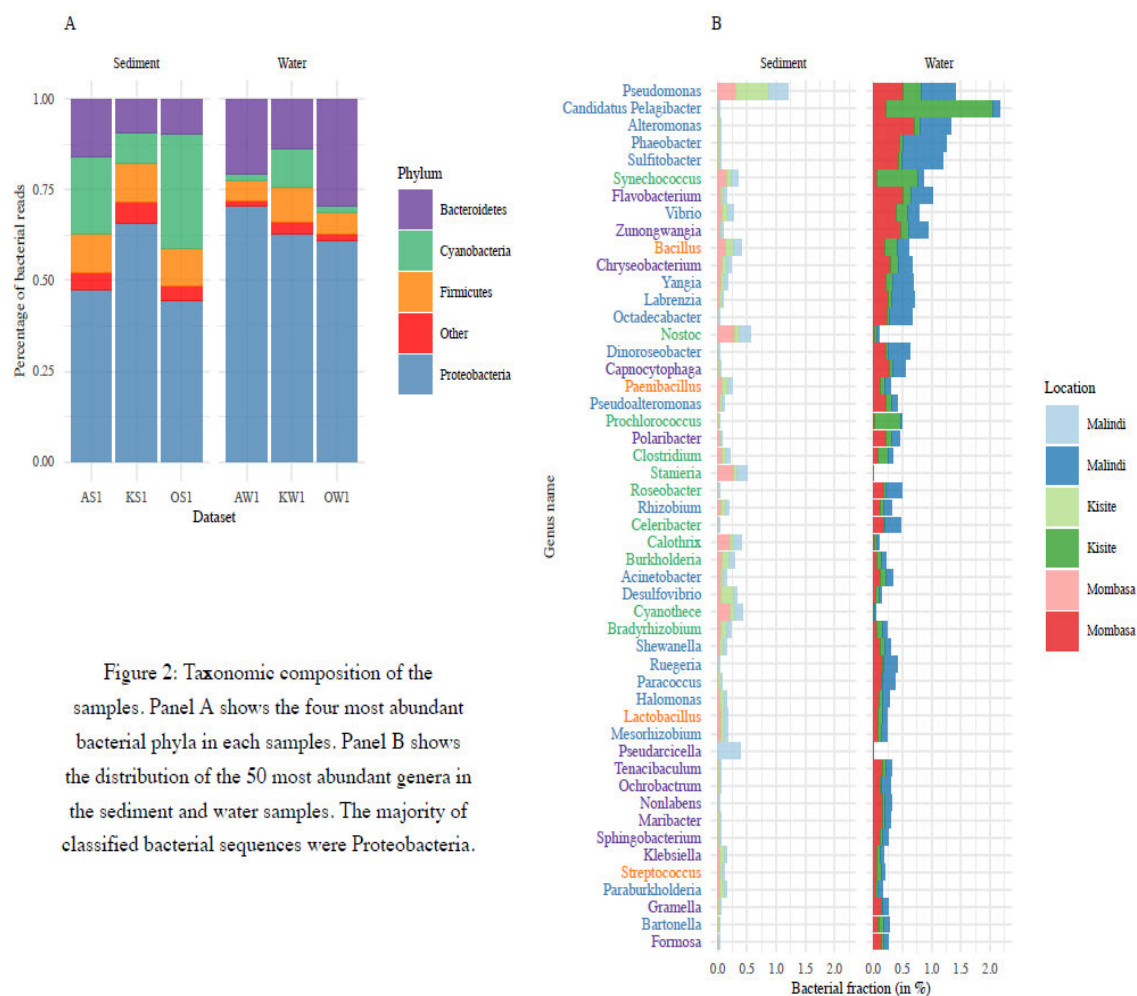


Figure 2: Taxonomic composition of the samples. Panel A shows the four most abundant bacterial phyla in each samples. Panel B shows the distribution of the 50 most abundant genera in the sediment and water samples. The majority of classified bacterial sequences were Proteobacteria.

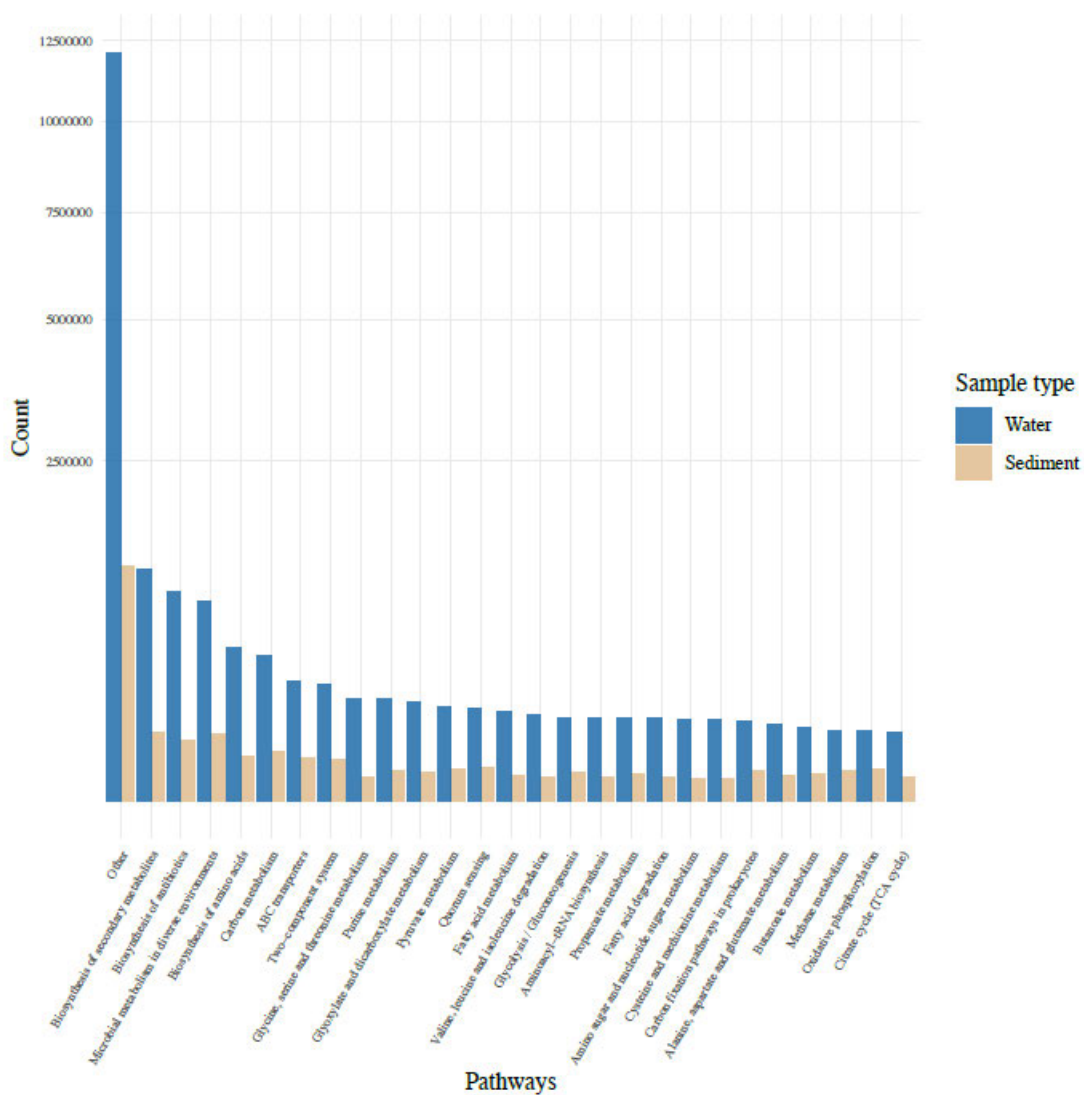
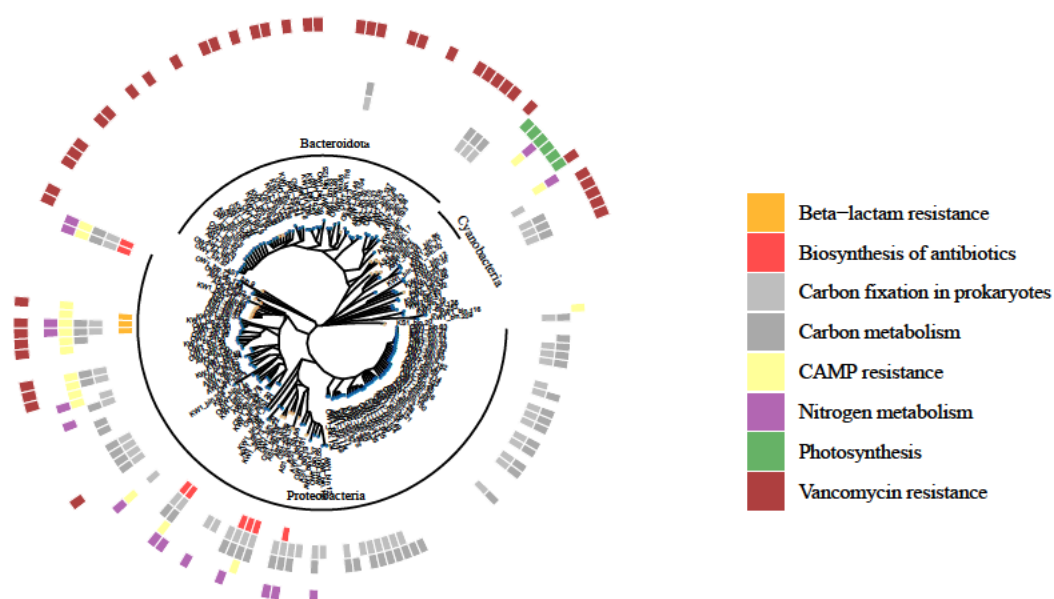


Figure 3: Distribution of proteins associated to KEGG pathways from the plasm protein assemblies. The 20 most abundant pathways for both sediment and water are displayed here.



ies.

The tree was generated with pplacer and plotted with ggtree. In blue are the bacteria recovered from the water samples and in brown from the sediment samples. The outer circles represent selected pathways that are present in the assemblies.