



Cross-Sectional Variations in Structure and Function of Coral Reef Microbiome With Local Anthropogenic Impacts on the Kenyan Coast of the Indian Ocean

Sammy Wambua^{1,2*}, Hadrien Gourel³, Etienne P. de Villiers^{4,5}, Oskar Karlsson-Lindsjö³, Nina Wambiji⁶, Angus Macdonald⁷, Erik Bongcam-Rudloff³ and Santie de Villiers^{1,8}

OPEN ACCESS

Edited by:

Rodrigo Gouvea Taketani,
Rothamsted Research,
United Kingdom

Reviewed by:

Cynthia B. Silveira,
University of Miami, United States
Juline Walter,
Federal University of Rio de Janeiro,
Brazil

*Correspondence:

Sammy Wambua
sammywambua@gmail.com

Specialty section:

This article was submitted to
Aquatic Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 26 February 2021

Accepted: 31 May 2021

Published: 23 June 2021

Citation:

Wambua S, Gourel H, de Villiers EP, Karlsson-Lindsjö O, Wambiji N, Macdonald A, Bongcam-Rudloff E and de Villiers S (2021) Cross-Sectional Variations in Structure and Function of Coral Reef Microbiome With Local Anthropogenic Impacts on the Kenyan Coast of the Indian Ocean. *Front. Microbiol.* 12:673128. doi: 10.3389/fmicb.2021.673128

¹ Pwani University Bioscience Research Centre (PUBReC), Pwani University, Kilifi, Kenya, ² Department of Biological Sciences, Pwani University, Kilifi, Kenya, ³ Department of Animal Breeding and Genetics, Swedish University of Agricultural Sciences, Uppsala, Sweden, ⁴ KEMRI-Wellcome Trust Research Programme, Kilifi, Kenya, ⁵ Nuffield Department of Medicine, University of Oxford, Oxford, United Kingdom, ⁶ Kenya Marine and Fisheries Research Institute, Mombasa, Kenya, ⁷ School of Life Sciences, University of KwaZulu-Natal, Durban, South Africa, ⁸ Department of Biochemistry and Biotechnology, Pwani University, Kilifi, Kenya

Coral reefs face an increased number of environmental threats from anthropomorphic climate change and pollution from agriculture, industries and sewage. Because environmental changes lead to their compositional and functional shifts, coral reef microbial communities can serve as indicators of ecosystem impacts through development of rapid and inexpensive molecular monitoring tools. Little is known about coral reef microbial communities of the Western Indian Ocean (WIO). We compared taxonomic and functional diversity of microbial communities inhabiting near-coral seawater and sediments from Kenyan reefs exposed to varying impacts of human activities. Over 19,000 species (bacterial, viral and archaeal combined) and 4,500 clusters of orthologous groups of proteins (COGs) were annotated. The coral reefs showed variations in the relative abundances of ecologically significant taxa, especially copiotrophic bacteria and coliphages, corresponding to the magnitude of the neighboring human impacts in the respective sites. Furthermore, the near-coral seawater and sediment metagenomes had an overrepresentation of COGs for functions related to adaptation to diverse environments. Malindi and Mombasa marine parks, the coral reef sites closest to densely populated settlements were significantly enriched with genes for functions suggestive of mitigation of environment perturbations including the capacity to reduce intracellular levels of environmental contaminants and repair of DNA damage. Our study is the first metagenomic assessment of WIO coral reef microbial diversity which provides a much-needed baseline for the region, and points to a potential area for future research toward establishing indicators of environmental perturbations.

Keywords: coral reef microbiota, microbiome, COGs, human impacts, environmental stressors

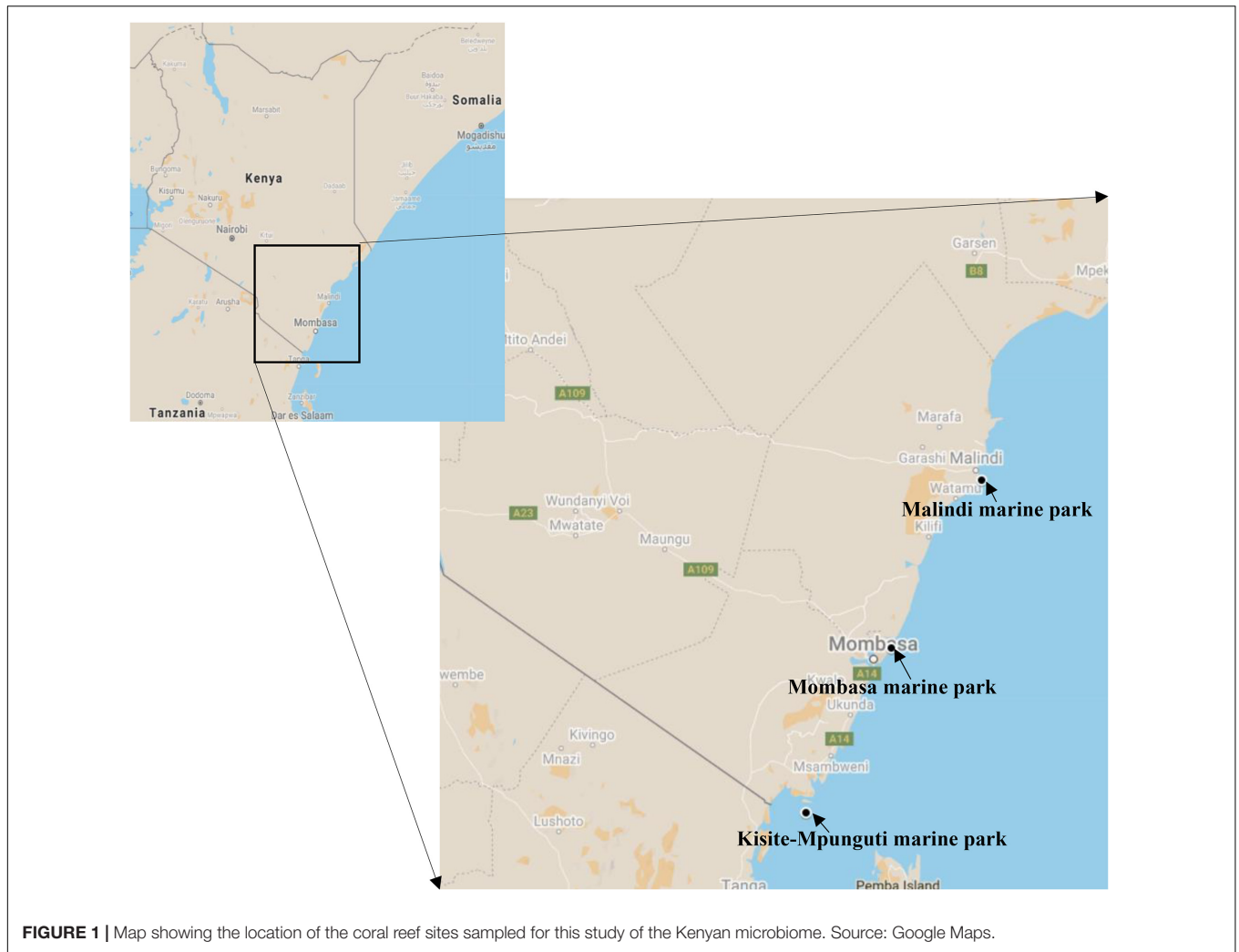
INTRODUCTION

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.

The Western Indian Ocean (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 with respect to the associated microbial ecology (Díez et al., 2016). Indeed, the *Web of Science* records only four papers, published in the last decade (2010 – 2020) on genomic assessment of coral reef microbes in the WIO, compared to 17 in the Caribbean Sea and 114 in the Great Barrier Reef (GBR). Because of the broad variety of biogenic habitats and oceanographic conditions, the region's waters encompass 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 (Van Der Elst et al., 2005; Obura, 2017), the WIO and its coastal ecosystems are essential to the region's economy. It is feared that with the rapidly increasing development and utilization of coastal zones (Neumann et al., 2015), the WIO's resources are overwhelmed, overexploited and poorly managed, putting its wealth at risk. Indications of distress on the region's ecosystems are noted with increasing frequency characterized by reduced fish catches, diminishing mangrove coverage, and declining coral reef cover (Van Der Elst et al., 2005; Obura, 2017) consequently threatening the livelihoods of dependent coastal communities. Protected areas are the principal tool of marine management in the region – the WIO countries have protected about 2.4 per cent of their marine area (Obura, 2017). Other strategies extensively applied by government institutions, Non-Governmental Organizations (NGO) and communities throughout the WIO to conservation coral reefs are reviewed by Hattam et al. (2020) and include the introduction of alternative livelihoods, coral gardening and payment for ecosystem services (PES) schemes. Monitoring of coral reef status in the WIO is undertaken under the Global Coral Reef Monitoring Network (GCRMN) umbrella, and has mainly involved visual estimations of hard coral cover, fish abundance and coral bleaching (Obura et al., 2019). Technological advancements including autonomous vehicles facilitating underwater, surface, and aerial surveys, and satellites, as well as structure from motion image processing, and acoustic techniques, are set to greatly improve monitoring of coral reef status. Advancements and declining costs of high-throughput sequencing technologies aided by autonomous reef monitoring systems has also improved collection of diversity data on cryptic marine communities for coral reef monitoring (Ransome et al., 2017). Marine microbes are thought to have significant potential as a cryptic community due to their rapid response to environmental change (Glasl et al., 2017).

Shifts in the composition and diversity of microbial communities may be a good indicator for a marine ecosystem's health and for predicting stress responses (Won et al., 2017). Current efforts exploring the potential of microbes as indicators for coral reef ecosystem stressors often correlate taxonomic profiles with biogeographic variation (Glasl et al., 2017, 2018, 2019; Roitman et al., 2018). However, functional profiling may also be informative of environmental disturbance 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 and Pomati, 2020). Metagenomics provides the opportunity to access microbial community taxonomic and genomic content, as well as their functional potential, which can serve as indicators of the health in an ecosystem (Bellwood et al., 2019). For instance, comparative metagenomics studies have demonstrated that microbial diversity is influenced by the local environment (Behzad et al., 2016) leading to the hypothesis that unique environments harbor unique microbial diversity as well as unique metabolic pathways. Furthermore, Kelly et al. (2014) has shown that, at the ecosystem level, benthos determine core microbial taxa which then adapt to local oceanographic conditions by selecting for advantageous metabolic genes.

Coral reef microbial communities inhabit sediments, overlying water column, and benthic invertebrates such as corals and sponges (Bourne and 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 and Webster, 2013; Vanwonderghem and Webster, 2020). These interactions within the coral ecosphere are known to influence the composition of coral-associated microorganisms (Weber et al., 2019) through the benthos releasing host-specific microbes into the surrounding water, or producing dissolved organic matter which stimulates the activity of specific microbes within the surrounding water layer (Barott and Rohwer, 2012; Silveira et al., 2017; Walsh et al., 2017). 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 (Dinsdale et al., 2008; Bruce et al., 2012). 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 ecosphere (McDole et al., 2012; Bourne and Webster, 2013; Kelly et al., 2014; Tout et al., 2014; Walsh et al., 2017; Glasl et al., 2019). To approximate the anthropogenic impact on composition and function of coral reef microbiota, we compared cross-sectional metagenomes of near-coral seawater and sediment samples from three reefs subject to varied suites of environmental perturbations on the Kenyan coast of the WIO. This is the first shotgun metagenomic assessment of coral ecosphere on the WIO – previous studies explored the open water microbiome (Sunagawa et al., 2015) and virome (Williamson et al., 2012).



MATERIALS AND METHODS

Study Sites and Human Activities

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 of cosmopolitan urban and rural settlements along the coast of Kenya. These sites are spread across three of the five counties on the Kenyan coast.

We investigated coral ecosystem within three of the five Kenya marine national parks i.e., Malindi ($3^{\circ}15'35.1''S$ $40^{\circ}08'40.0''E$), Mombasa ($3^{\circ}59'45.7''S$ $39^{\circ}44'50.1''E$) and Kisite-Mpunguti ($4^{\circ}42'54.0''S$ $39^{\circ}22'23.8''E$) (Figure 1). 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 snorkeling among other recreational touristic activities. It also experiences significant year-round discharge of freshwater and sediments with different types of nutrients from the Sabaki River which runs through a catchment area dominated by agricultural settlements (Katwijk et al., 1993;

Munyao et al., 2003; Kitheka, 2019; Wanjeri et al., 2021). 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 four meters. Mombasa Marine National Park was established in 1986 (Ngugi, 2002) with restrictions of protection being enforced commencing in 1991 (Tuda and Omar, 2012) and is the most visited of Kenya’s marine parks by both local and international tourists (Owens, 1978). It was considered the “polluted” site; owing to its proximity to the urbanized touristic city, the park experiences pollution from hotels, hospitals, domestic and industrial waste disposal (Mwangi et al., 2001; Okuku et al., 2011). 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 and Tessema, 2001). There is no river runoff affecting the site (Okuku et al., 2019) and the most common activity here is snorkeling.

Sampling

Cross-sectional sampling and *in situ* measurements were done in November and December of 2016 in all coral reef sites, 200–500 m from the shore 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 (McClanahan et al., 2007). A total of six samples were collected, three near-coral seawater, and three sediments. From each site, approximately 5L of seawater, and 10 cm column of sediments collected at the base of the same coral colony with a 10 mL syringe barrel. To assess inter-site variations at the time of sampling physicochemical parameters (water temperature, pH, salinity, and dissolved oxygen) were determined *in situ*, using portable multiprobe water quality meters as per the manufacturer's instructions (YSI Inc., Yellow Springs, OH, United States). Finally, seawater samples were collected in triplicate 50-mL, for nutrient analysis. Samples were transported on ice to the laboratory for processing, typically within 3 h of sampling.

DNA Isolation

From each site, 4 L of seawater was vacuum-filtered (VWR, West Chester, PA, United States) through a 0.2- μ m pore size membrane (Pall Corporation, Port Washington, WI, United States) 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, United States). 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).

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 h instead of the maximum recommended 16 h. Sequencing libraries were prepared from 1 μ g 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 350 bp 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 Fragment Analyzer 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 (Bio-Rad) before cluster generation and sequencing. A 200 pM solution of the individual sequencing libraries was subjected to cluster generation and paired-end sequencing with 150 bp 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, Sweden¹.

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). 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. Trimmed sequences were deposited to the European Nucleotide Archive under the project accession # PRJEB30838.

Taxonomic profiling was performed by applying the Lowest Common Ancestor (LCA) pipeline based on the Kraken tool, against the NCBI non-redundant ("nr") database, enhanced by DIAMOND. Functional annotation was done using the clusters of orthologous groups (COG) of proteins database (Tatusov et al., 2000, 2001) with a BLASTX search of reads vs. the COG database applying MGX pipeline defaults (*E*-value cutoff $1e^{-5}$).

Statistical Analyses

Output files from the MGX metagenomics analysis platform were exported to the R statistical environment (R Core Team, 2020) for analysis and visualization. Taxa and metadata files were merged using *phyloseq* version 1.28.0 (McMurdie and 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 package *vegan* 2.5–6 (Oksanen et al., 2019).

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 the Shannon index based on COG richness and evenness (Johnson and Pomati, 2020). Bray-Curtis distance based on the relative abundances was calculated to represent functional composition variation among the samples (Bray and Curtis, 1957), and 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 Kruskal–Wallis test for

¹www.sequencing.se

TABLE 1 | Measurements of physicochemical characteristics of the three coral reefs within the marine national parks at the time of sampling for metagenomic sequencing.

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

TABLE 2 | Nutrient levels measurements of the three coral reefs within the marine national parks at the time of sampling for microbiome sequencing.

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

Between-sites comparisons tested by ANOVA.

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 method (Benjamini and Hochberg, 1995).

Comparison of beta diversity between groups (both taxonomic and functional) was assessed by permutational multivariate analysis of variance (PERMANOVA) (Anderson, 2001) using the *adonis* test based on Bray-Curtis distances with 999 permutations in the package *vegan*.

RESULTS

Inter-Site Variations in Coral Reefs

Physicochemical parameters characteristics are presented in **Table 1**. Measurements were comparable across the sites except for temperature and salinity recorded in Malindi that seemed to vary 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.

The nutrients assessed, except for nitrites, were highest at Malindi, the “exploited” site, and lowest at Kisite-Mpunguti, the “baseline” site (**Table 2**). Although observed, differences in nutrient concentrations between the study sites were not statistically significant (ANOVA).

Microbial Diversity

A total of 211,966,893 reads from 6 samples were processed, 43% of which were classified taxonomically and functionally. Taxonomic assignment showed that four superkingdoms were

represented, the majority (95.8% of the assigned reads) being Bacteria, followed by Eukaryota (1.7%), Archaea (1.3%), and Viruses (1.2%). 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 2A**) for most of the sequences from both datasets began to level off suggesting reasonable coverage of the microbial communities characterized with the processed reads. In all, 19,630 species (18,271 bacteria, 801 viruses and 558 archaea) belonging to 74 phyla were identified. Overall, Proteobacteria was the most dominant phylum with over 50% overall relative abundance (**Figure 2B**) followed by Bacteroides (20%) and Cyanobacteria (11%). Except for Kisite-Mpunguti metagenomes, Cyanobacteria was the most dominant phylum in the sediment sequences.

Bacteria

A total of 18,271 bacterial species assigned to 53 phyla were classified from near-coral seawater and sediment samples. Classes belonging to Proteobacteria and Bacteroidetes phyla had the highest relative abundances. Among the Proteobacteria classes, Gammaproteobacteria, Betaproteobacteria and Deltaproteobacteria were the most dominant. Other classes that had a relative abundance higher than 1% belonged to the phyla Actinobacteria (Actinobacteria) and Firmicutes (Bacilli, Clostridia and Planctomycetes). Further inspection was made of the phyla Proteobacteria, Bacteroidetes and Actinobacteria, which are considered copiotrophic i.e., inhabiting nutrient-rich environments. In all, 13,903 copiotrophic species were assigned, the distribution of which did not cluster by site (**Figure 3A**) (PERMANOVA, *p* = 0.798). The “baseline” site had the highest number of unique species and the “polluted” site the lowest (**Figure 3B**).

The first ten most abundant orders of the copiotrophic phyla assigned belonged to Proteobacteria and Bacteroidetes. Sequences assigned to Rhodobacterales (30.5%), Pelagibacterales (22.7%) orders of Proteobacteria, and the order Flavobacteriales (17.2%) of Bacteroidetes, were the most relatively abundant. By comparison, the relative abundance of each copiotrophic order was highest in either the “exploited” or “polluted” site except the orders Pelagibacterales, Cellvibrionales and Xanthomonadales whose relative abundances were highest in the “baseline” site (**Figure 4A**).

The relative abundances for sequences assigned to the common fecal indicator bacteria (FIB) genera (*Escherichia*, *Salmonella*, and *Campylobacter*) as well the alternative ones employed for microbial source tracking (i.e., *Bacteroides* and *Prevotella*) were less than 1% each (**Figure 4B**). 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.

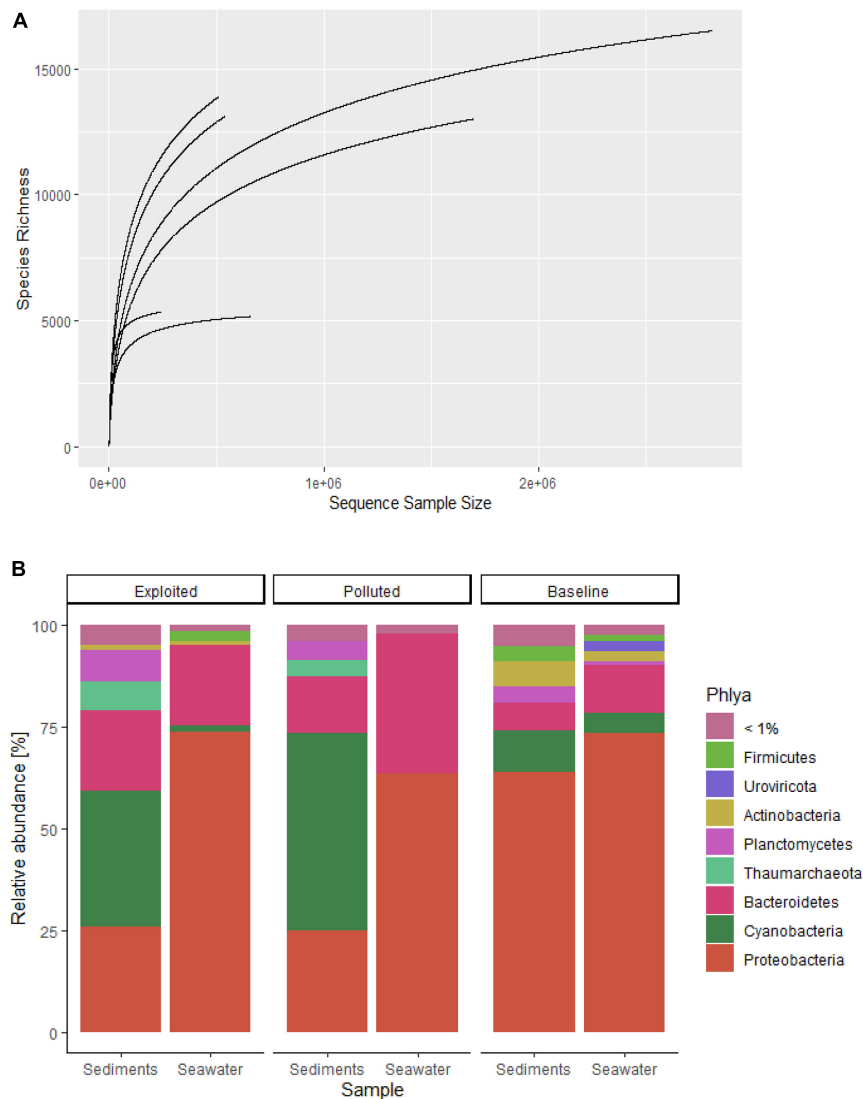


FIGURE 2 | (A) Rarefaction curves for community species richness, and **(B)** stacked charts of the distribution of community phyla with relative abundance > 1% sequenced from near-coral seawater and sediment samples from Kenya reefs at Malindi (Exploited), Mombasa (Polluted) and Kisite-Mpunguti (Baseline) Marine National Parks.

Archaea

For the Archaea superkingdom, 558 species were assigned spread in eight phyla. Thaumarchaeota (88.5%) and Euryarchaeota (10.9%) were the most dominant phyla with the other six phyla accounting for only 0.7% of the sequences assigned. The distribution of archaeal orders with relative abundances above 1% is presented in **Figure 5A**. The order Nitrosopumilales recorded the highest relative abundance (>79%) followed by Methanosarcinales (5.1%). Sequences from the “baseline” site showed higher relative abundance for all orders except Nitrosopumilales (31%) for which the “exploited” and “polluted” site sequences were three-fold higher. Non-metric multidimensional scaling did not show any site-specific structuring of the Archaea superkingdom (PERMANOVA, $p = 0.733$).

Viruses

A 801 viral species, harbored in 11 phyla, were assigned to sequences from the near-coral seawater and sediment samples. Uroviricota (88.8%) dominated the sequences assigned. Only three other phyla i.e., Nucleocytoviricota (6.5%), Cressdnaviricota (2.5%), and Phixviricota (1.2%) had relative abundances of >1%.

Coliphages (viruses that infect coliform bacteria) constituted 85% of the sequences assigned at the family level. Myoviridae was the most dominant family (79.8%) (**Figure 5B**) and had higher relative abundance in the “baseline” (82.2%) than in the “exploited” (50.3%) and the “polluted” (69.7%) site metagenomes. The other coliphage families, Podoviridae, Siphoviridae and Microviridae were generally higher in the “polluted” (6.2, 5.9, and 5.5%) and “exploited” (2.4, 2.7, and 16.3%), than in the

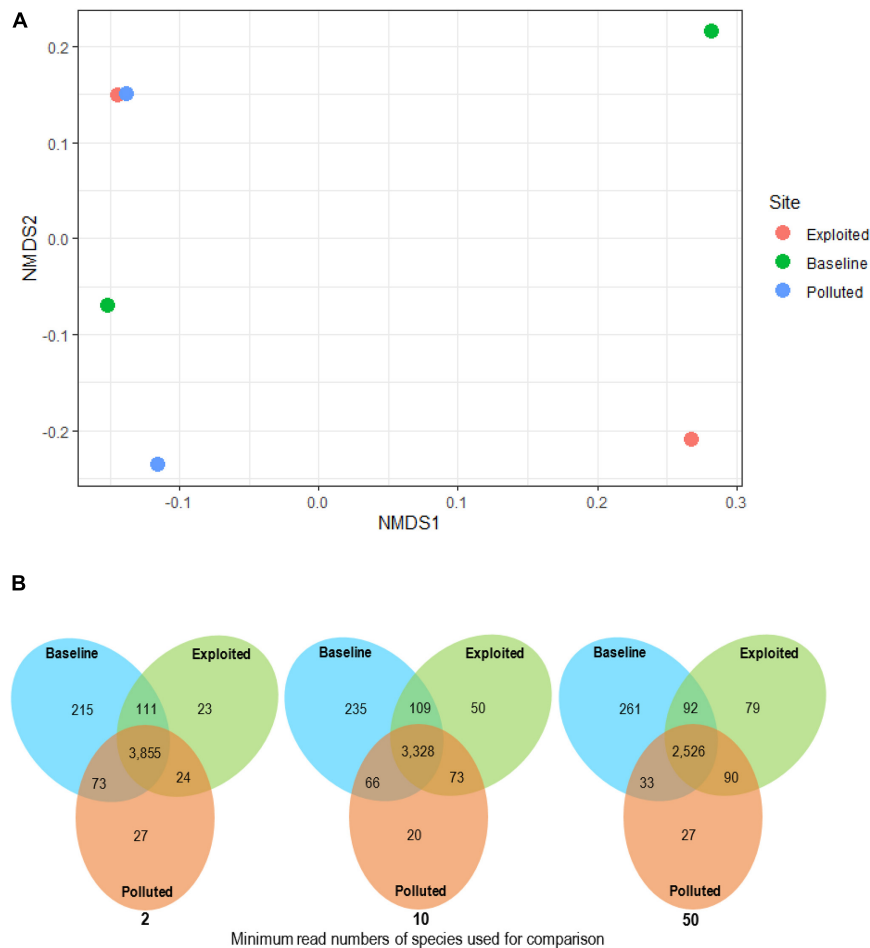


FIGURE 3 | (A) Bacterial community composition analyzed with non-metric multidimensional scaling (nMDS) plots using Bray-Curtis dissimilarity for near-coral seawater and sediment samples from Malindi (peach circle; exploited), Mombasa (aqua circles; polluted) and Kisite-Mpunguti (green circle; baseline) national marine parks. For this nMDS plot, stress = 0.05, $r^2 = 0.998$, and **(B)** venn diagrams showing the distribution of microbial species across study site metagenomes based on species with more than 2, 10, 50 reads.

“baseline” (2.0, 1.6, and < 0.1%) site sequences. 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 five-fold higher) in the “baseline” site sequences than in either the “exploited” or “polluted” site sequences.

Microbiome Function

A total of 4,584 clusters of orthologous groups of proteins (COGs) belonging to 24 COG categories were classified. Overall, the categories of amino acid metabolism and transport (E), general functional prediction only (R), energy production and conversion (C), and translation (J) had the most relative abundance together constituting 41% of all sequences. In contrast, categories of nuclear structure (Y, <0.01%), RNA processing and modification (A, 0.02%), cytoskeleton (Z, 0.05%), and chromatin structure and dynamics (B, 0.08%) had the least relative abundances. Between sequences from the reef sites (**Figure 6**), no statistically significant differences in relative

abundances were observed (Kruskal-Wallis, $p = 0.997$), except for subtle variations in a number of COG categories including amino acid transport and metabolism (E), energy production and conversion (C), translation (J), carbohydrate transport and metabolism (G), and signal transduction (T) where the “baseline” site sequences seemed to have higher relative abundance compared to the “exploited” and “polluted” site sequences. In contrast, the relative abundances for the categories of general functional prediction only (R), replication and repair (L), cell wall membrane biogenesis (M), lipid metabolism (I), coenzyme metabolism (H), and secondary structure (Q) where higher in either “exploited” or “polluted” site sequences than in the “baseline” site sequences.

The rarefaction curves (**Figure 7A**) shows the number of prevalent COGs identified with increasing sample number. The extent of sequencing showed that most COGs were recovered. The average per sample COG richness was estimated to be $4,111 + 205$. The most prevalent COGs in all metagenomes corresponded to signal transduction histidine kinase (COG0642,

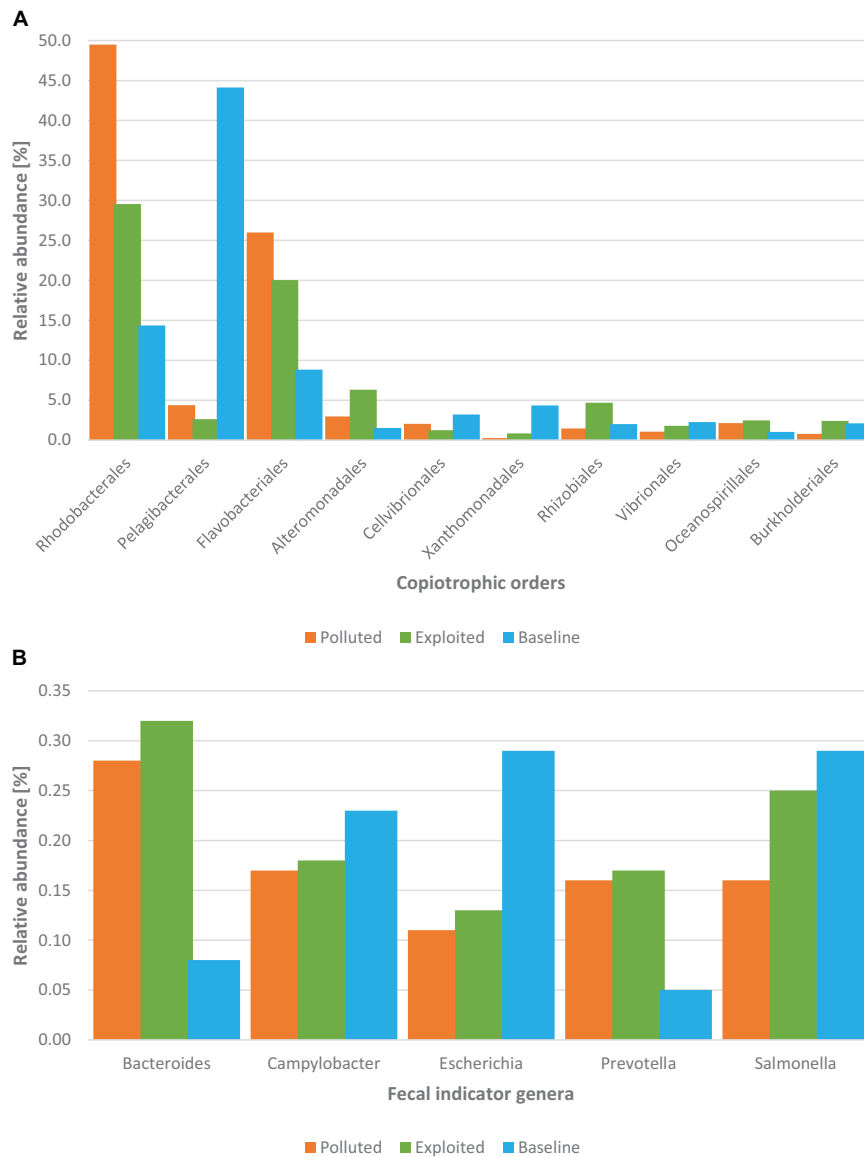


FIGURE 4 | Distribution of **(A)** the top ten most abundant orders of the copiotrophic phyla (Proteobacteria and Bacteroidetes), and **(B)** distribution of common FIB genera from metagenomes of near-coral seawater and sediment samples from Kenyan reefs at the national marine parks classified as “polluted” (Bronze; Mombasa), “exploited” (Olive; Malindi) and “baseline” (Sky; Kisite-Mpunguti).

0.80%), dehydrogenases with different specificities (COG1028, 0.75%), NAD-dependent aldehyde dehydrogenases (COG1012, 0.68%), and RTX toxins and related Ca²⁺ -binding proteins (COG2931, 0.61%). The “baseline” site sequences recorded 4,324 observed COGs compared to 4,000 (“exploited”) and 4,008 (“polluted”), with a Shannon index of 7.12, compared to 7.06 (“exploited”) and 7.09 (“polluted”) (**Figure 7B**).

The “baseline” sequences had 2,153 COGs that were significantly enriched, the most abundant of which were functions for energy production and conversion including, and transcription regulation 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 were dominated by predicted proteins related to defense 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 8**).

DISCUSSION

Microorganisms play a critical role to reef ecosystem health and resilience through mediating nutrient cycling, interactions

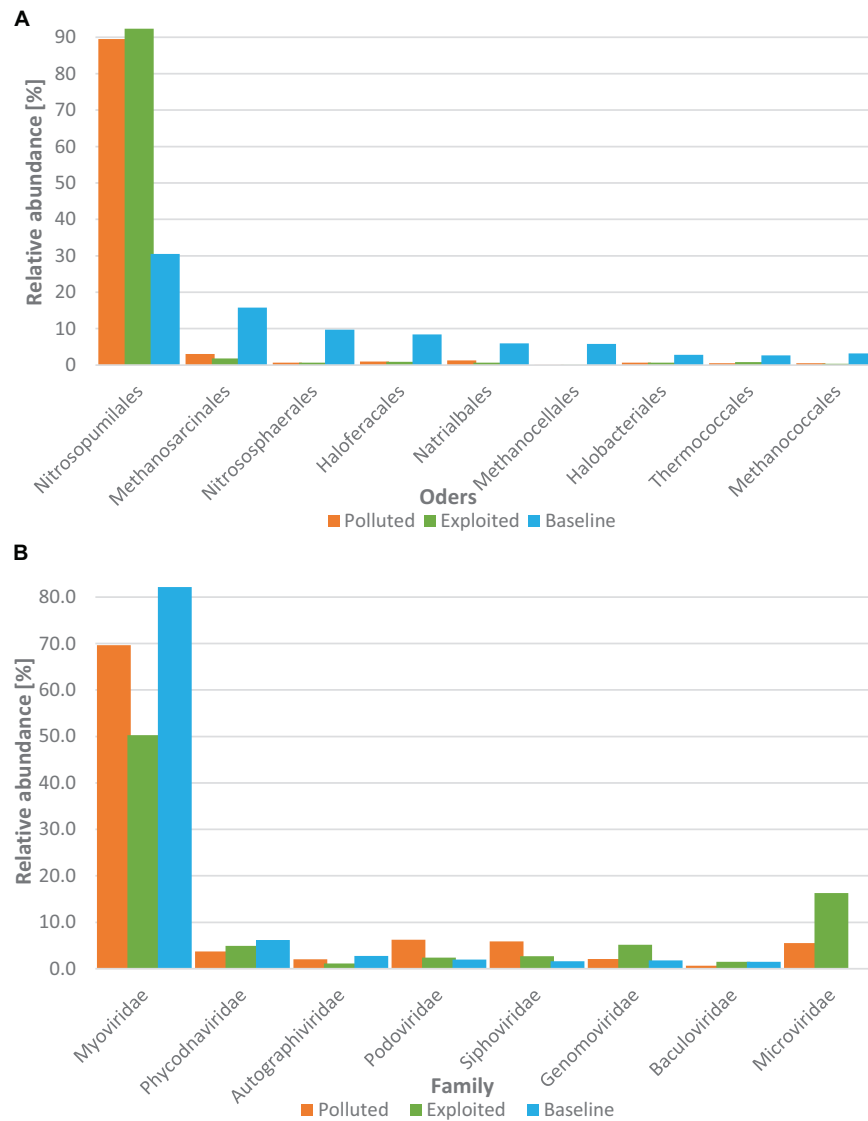
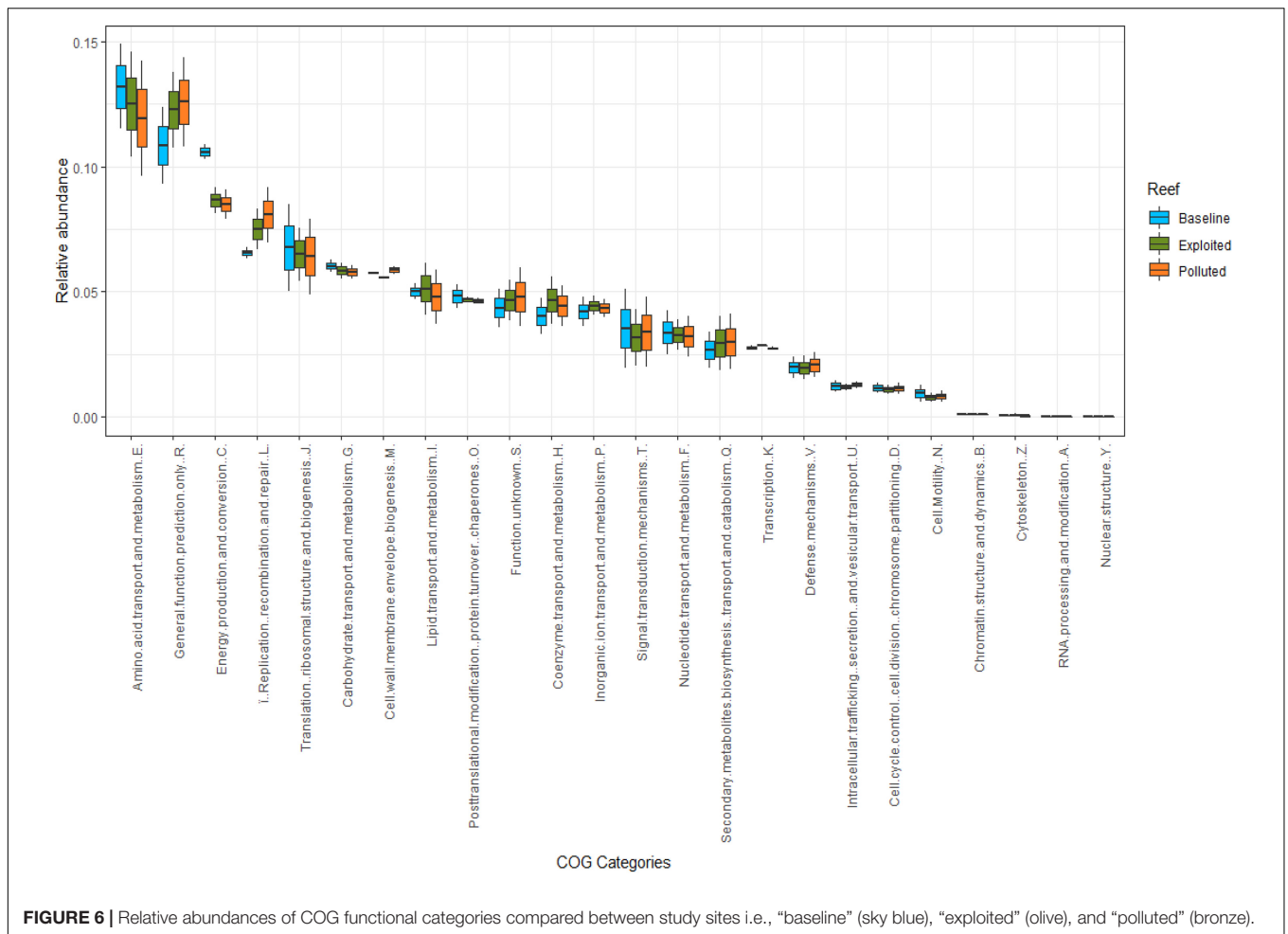


FIGURE 5 | Distribution of **(A)** archaeal orders and **(B)** viral families with overall relative abundance > 1% grouped by study site i.e., “polluted” (orange), “exploited” (olive) and “baseline” (sky).

with macro-organisms and provision of chemical cues influence the recruitment of diverse reef taxa. These processes may be impacted by environmental changes that cause microbial compositional and functional alterations which may, in turn, have consequences for the functioning of coral ecosystems (Glasl et al., 2018). Therefore, assessing changes in reef microbial communities and functional potential may provide early indicator of ecosystem impacts, and can help with development of diagnostic tools for monitoring shifts in coral reef health under different environmental states (Kisand et al., 2012; Bourlat et al., 2013; Glasl et al., 2017; Goodwin et al., 2017). We annotated near-coral seawater and sediment metagenomes from Kenyan coral reefs and compared microbial taxonomic and functional diversity across a gradient of the human activities the reefs are exposed to. Physicochemical

parameters were also estimated to assess inter-site variations at the time of sampling.

Environmental variables at the coral reefs sampled fell mostly within the ranges considered suitable for coral growth in tropical oceans (Kannapiran et al., 2008; Henkel, 2010; Guan et al., 2015; 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 occasioned by River Sabaki’s discharge into the ocean (Kitheka, 2019). 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, however, not statistically significant. 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 a cross-sectional survey, the reported measurements represented the quality of the coral reef water at the time of sampling, and not necessarily the stable state. A recent study modeling 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.

Taxonomic composition, especially of the most abundant taxa, was largely consistent with other marine microbiota (Williamson et al., 2012; Sunagawa et al., 2015; Godoy-Vitorino et al., 2017; Santoro et al., 2019). The estimate for species richness was within the same order of magnitude as the previous metagenomic estimate for global oceans (Sunagawa et al., 2015). Although the majority of assignments were shared, there were sequences unique to each site whereby Kisite-Mpunguti (“baseline”) had

the highest number, reinforcing the hypothesis that rather than every microbe being ubiquitous, taxa are selected or structured by the environment (Flaviani et al., 2018). The seawater and sediment samples collected within coral reefs generally harbored microbial composition consistent with the typical taxa enriched by metabolic products of corals i.e., copiotrophic bacterial groups (including Gammaproteobacteria, Betaproteobacteria, and Bacteroidetes), ammonia oxidizing archaea (AOA) of the order Nitrosopumilales, and phages for the copiotrophic bacteria (coliphages) i.e., Podoviridae, Siphoviridae, and Microviridae (Sunagawa et al., 2010; Kelly et al., 2014; Godoy-Vitorino et al., 2017; Thurber et al., 2017; Weber et al., 2019). 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 and 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. 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

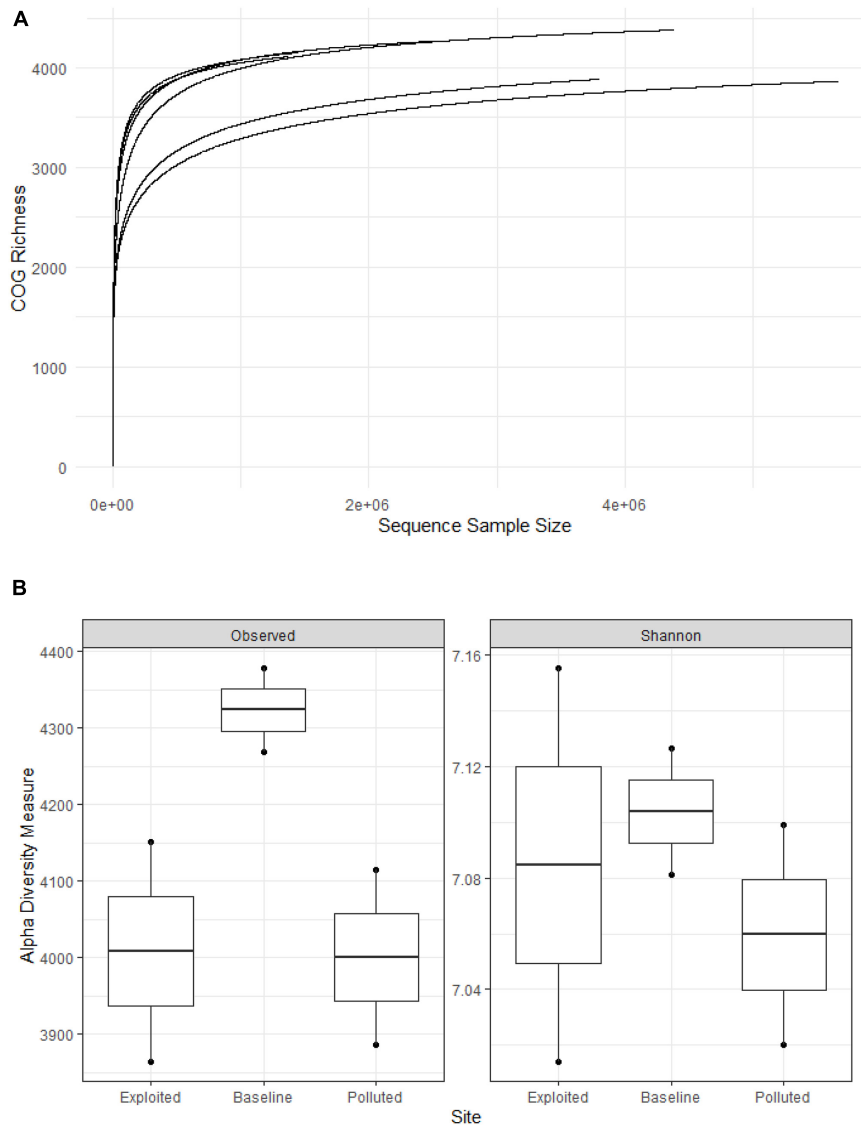
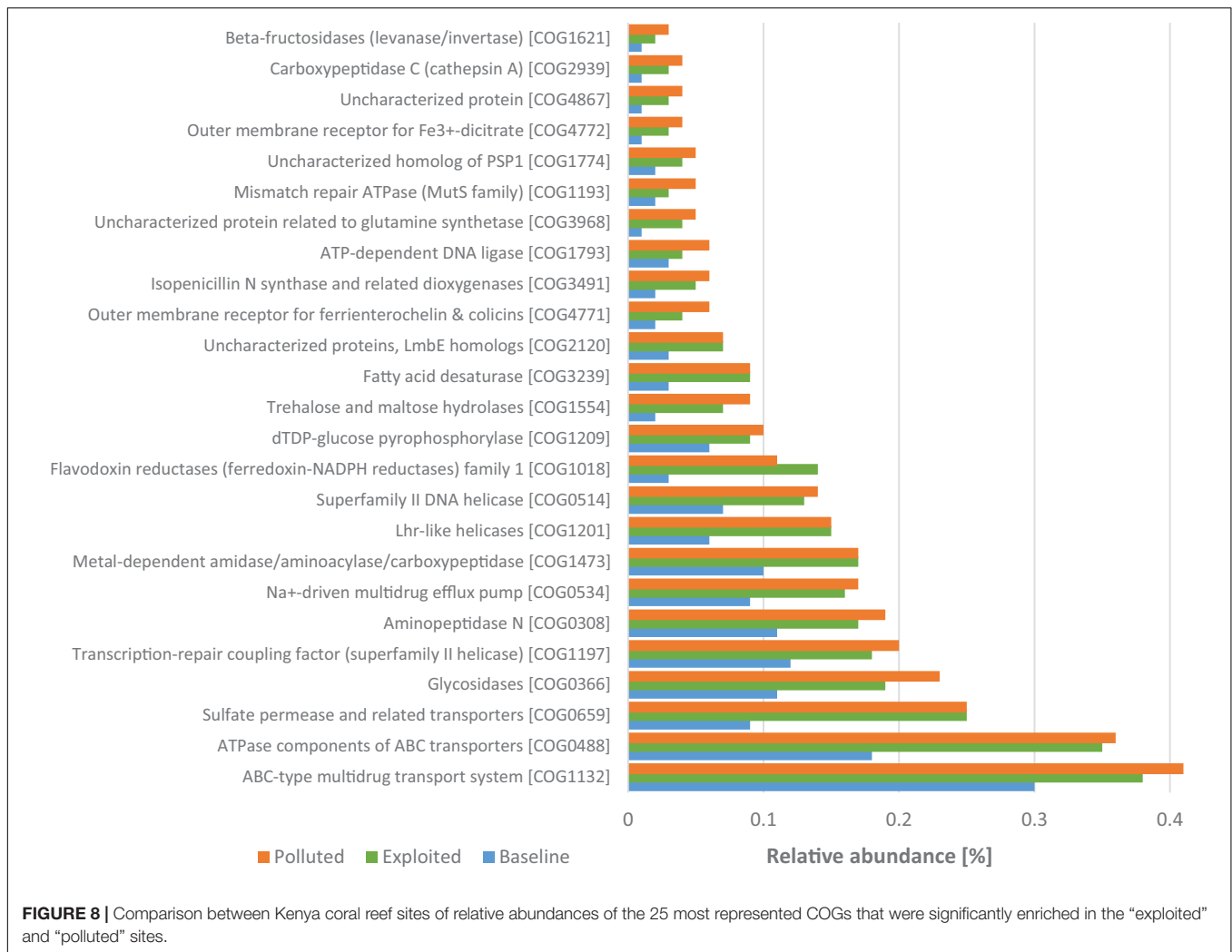


FIGURE 7 | (A) Rarefaction curves and **(B)** plots of alpha diversity metrics for clusters of orthologous groups of proteins (COG) richness on the near-coral seawater and sediment metagenomes from Kenya reefs on the WIO.

as fecal indicators (Buccheri et al., 2019). Also, the genera *Bacteroides* and *Prevotella* exhibited higher relative abundance in both the “exploited” and “polluted” than in the “baseline” site sequences – these genera are increasingly exploited as alternative groups to replace the less reliable traditional FIB. Contrary to expectations, the relative abundance for members of the order Pelagibacterales and the traditional FIB genera *Escherichia*, *Salmonella*, and *Campylobacter* were higher in the “baseline” than in the “exploited” and “polluted” site sequences (**Figure 4**). 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 this group in the “baseline” than in the “exploited” and “polluted” site sequences may have been, therefore, indicative

of differences in nutrient levels at the sites perhaps due to the surrounding human activities. As for the FIB genera, it is not clear why relative abundances did not correlate with human impact gradients at the coral reef sites. One possible explanation is what has been termed as “connectivity footprint” – the phenomenon that even MPAs thousands of kilometers away can be influenced by human activities due to dispersal by ocean currents (Robinson et al., 2017). The “baseline” site, which recorded highest relative abundances for the traditional FIB genera was sampled in December during the north-east monsoon (NEM) which is characterized by the convergence of the southward flowing Somali current (SC) with the northward flowing East African coastal current (EACC) and upwelling along the Kenyan coast (Heip et al., 1995; Mayorga-Adame et al., 2016). It is conceivable that the fecal bacteria originated elsewhere and



transported to the “baseline” site by the SC as the site is south to the “exploited” and “polluted” sites. However, if this was the case, similar relative abundances would have been observed of the other taxa, for instance copiotrophic bacteria and coliphages. An alternative explanation is that FIB detected at the “baseline” site were of sources not related to human fecal contamination. FIBs are shed in the feces of many different animals (Harwood et al., 2014), and naturalized or environmentally adapted strains of FIB can persist and multiply in a broad range of habitats far removed from their natural reservoir of the animal gut (Devane et al., 2020). It is for these reasons that traditional FIB genera are generally considered unreliable as fecal indicators.

Coliphages i.e., viruses that infect coliform bacteria, are considered better predictor of fecal pollution in seawater than FIBs (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 (Ebdon et al., 2012; Bari and Yeasmin, 2014). 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 and Rao, 2019). Cyanophages too recorded relative abundances commensurate to the bacteria they infect, cyanobacteria. Phages for *Prochlorococcus* and *Synechococcus* were up to five-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 (Waterbury et al., 1979; Biller et al., 2015; Kim et al., 2018) 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 correlation with human activity 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.

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). The predicted proteins for 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 the most abundant. Furthermore, the coral ecosystem sequences had overrepresentation of predicted proteins related to replication, recombination and repair especially transposases that play a role in adaptation to diverse environments (Vigil-Stenman et al., 2017), and predicted proteins for energy production and conversion which is an adaptation for high nutrient uptake and synthesis needed in low nutrient environments (Gudhka et al., 2015). These are likely strategies for optimizing rapid growth in the presence of labile nutrients – an adaptation of copiotrophs e.g., Gammaproteobacteria in environments of high levels of stress factors and nutrients (Kisand et al., 2012). These observations suggest that the three coral reefs had substantial levels of stressors and nutrients. Nonetheless, there were subtle differences across the human impacts’ gradient suggesting heterogeneity, possibly anthropogenically influenced.

Kisite-Mpunguti marine park (“baseline”) sequences were enriched, more than the “exploited” and “polluted” sequences, with protein groups essential for functions related to uptake and synthesis of nutrients and transcription – a common observation among oligotrophic organisms as an adaptation to low levels of nutrients (Lauro et al., 2009).

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 defense mechanisms such as transporters that prevent intracellular accumulation of toxic compounds (Wilkens, 2015), which serves as the major defense mechanism against antimicrobial compounds (Lubelski et al., 2007). For instance, a recent study assessing antibiotic resistance genes along a pollution gradient also found the highest abundance of transporters at 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). Together, these observations reflect hostile environmental conditions. 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 (Katwijk et al., 1993; Heip et al., 1995; Okuku et al., 2011; Ongore et al., 2013; Wanjeri et al., 2021), and not in Kisite-Mpunguti marine park which is situated farther from human settlements thereby potentially experiencing limited direct human impacts. Similar observations have been made, in a study comparing microbial functions of a highly polluted and protected marine environments (Kisand et al., 2012), and attributed to adaptations to an environment characterized by highly heterogeneous, variable and unfamiliar stimuli from anthropogenic source. 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.

CONCLUSION

Our results reveal subtle taxonomic and functional between-sites heterogeneity suggestive of human influence. The environmental significance of inter-site variations observed in this study needs to be confirmed through robust studies controlling for seasonal variations and spatial stability. Nonetheless our findings underscore the potential of microbial ecology in informing strategies of marine monitoring. We therefore recommend integration of microbial sampling in the management of national marine parks and the respective services they provide by both the national and county government agencies.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. Trimmed sequences are deposited to the European Nucleotide Archive under the study accession number PRJEB30838.

AUTHOR CONTRIBUTIONS

SW designed the study, conducted field sampling and laboratory processing, performed the bioinformatics analyses, and drafted and revised the manuscript. HG designed the study, coordinated sequencing, performed the bioinformatics analyses, and drafted the manuscript. EV, OK-L, NW, EB-R, AM, and SV helped design the study and edited the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was funded by the Swedish University for Agricultural Sciences and by the Swedish Research Council grant 2015-03443.

ACKNOWLEDGMENTS

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

parcs and the National Environment 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.

REFERENCES

- Alvarez-Yela, A. C., Mosquera-Rendón, J., Noreña-P, A., Cristancho, M., and López-Alvarez, D. (2019). Microbial diversity exploration of marine hosts at serrana bank, a Coral Atoll of the seaflower biosphere reserve. *Front. Mar. Sci.* 6:338. doi: 10.3389/fmars.2019.00338
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral. Ecol.* 26, 32–46. doi: 10.1111/j.1442-9993.2001.01070.pp.x
- Bari, M. L., and Yeasmin, S. (2014). *Water Quality Assessment: Modern Microbiological Techniques. Encyclopedia of Food Microbiology*, Second Edn. Amsterdam: Elsevier.
- Barott, K. L., and Rohwer, F. L. (2012). Unseen players shape benthic competition on coral reefs. *Trends Microbiol.* 20, 621–628. doi: 10.1016/j.tim.2012.08.004
- Bayer, B., Vojvoda, J., Offre, P., Alves, R. J. E., Elisabeth, N. H., Garcia, J. A. L., et al. (2016). Physiological and genomic characterization of two novel marine thaumarchaeal strains indicates niche differentiation. *ISME J.* 10, 1051–1063. doi: 10.1038/ismej.2015.200
- Behzad, H., Ibarra, M. A., Mineta, K., and Gojobori, T. (2016). Metagenomic studies of the Red Sea. *Gene* 576, 717–723. doi: 10.1016/j.gene.2015.10.034
- Bellwood, D. R., Streit, R. P., Brandl, S. J., and Tebbett, S. B. (2019). The meaning of the term ‘function’ in ecology: A coral reef perspective. *Funct. Ecol.* 33, 948–961. doi: 10.1111/1365-2435.13265
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Billler, S. J., Berube, P. M., Lindell, D., and Chisholm, S. W. (2015). Prochlorococcus: the structure and function of collective diversity. *Nat. Rev. Microbiol.* 13, 13–27. doi: 10.1038/nrmicro3378
- Bourlat, S. J., Borja, A., Gilbert, J., Taylor, M. I., Davies, N., Weisberg, S. B., et al. (2013). Genomics in marine monitoring: New opportunities for assessing marine health status. *Mar. Pollut. Bull.* 74, 19–31. doi: 10.1016/j.marpolbul.2013.05.042
- Bourne, D. G., and Webster, N. S. (2013). “Coral reef bacterial communities,” in *The Prokaryotes - Prokaryotic Communities and Ecophysiology*, eds E. Rosenberg, E. DeLong, S. Lory, E. Stackebrandt, and F. L. Thompson (Berlin: Springer), 163–187. doi: 10.1007/978-3-642-30123-0_48
- Bray, J. R., and Curtis, J. T. (1957). An ordination of the upland forest communities of southern wisconsin. *Ecol. Monogr.* 27, 325–349. doi: 10.2307/1942268
- Bruce, T., Meirelles, P. M., Garcia, G., Paranhos, R., Rezende, C. E., de Moura, R. L., et al. (2012). Abrolhos bank reef health evaluated by means of water quality, microbial diversity, benthic cover, and fish biomass data. *PLoS One* 7:36687. doi: 10.1371/journal.pone.0036687
- Buccheri, M. A., Salvo, E., Coci, M., Quero, G. M., Zoccarato, L., Privitera, V., et al. (2019). Investigating microbial indicators of anthropogenic marine pollution by 16S and 18S High-Throughput Sequencing (HTS) library analysis. *FEMS Microbiol. Lett.* 366:fnz179. doi: 10.1093/femsle/fnz179
- Buckley, R. J., Kramm, K., Cooper, C. D. O., Grohmann, D., and Bolt, E. L. (2020). Mechanistic insights into Lhr helicase function in DNA repair. *Biochem. J.* 477, 2935–2947. doi: 10.1042/BCJ20200379
- Burbano-Rosero, E. M., Ueda-Ito, M., Kisieliuss, J. J., Nagasse-Sugahara, T. K., Almeida, B. C., Souza, C. P., et al. (2011). Diversity of Somatic coliphages in coastal regions with different levels of anthropogenic activity in São Paulo state, Brazil. *Appl. Environ. Microbiol.* 77, 4208–4216. doi: 10.1128/AEM.02780-10
- Carlos, C., Pereira, L. B., and Ottoboni, L. M. M. (2016). Comparative genomics of *Paracoccus* sp. SM22M-07 isolated from coral mucus: insights into bacteria-host interactions. *Curr. Genet.* 63, 509–518. doi: 10.1007/s00294-016-0658-3
- Chen, J., McIlroy, S. E., Archana, A., Baker, D. M., and Panagiotou, G. (2019). A pollution gradient contributes to the taxonomic, functional, and resistome diversity of microbial communities in marine sediments. *Microbiome* 7:104. doi: 10.1186/s40168-019-0714-6
- Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890. doi: 10.1093/bioinformatics/bty560
- Devane, M. L., Moriarty, E., Weaver, L., Cookson, A., and Gilpin, B. (2020). Fecal indicator bacteria from environmental sources; strategies for identification to improve water quality monitoring. *Water Res.* 185:116204. doi: 10.1016/j.watres.2020.116204
- Diez, B., Nylander, J. A. A., Ininbergs, K., Dupont, C. L., Allen, A. E., Yooseph, S., et al. (2016). Metagenomic analysis of the indian ocean picocyanobacterial community: structure, potential function and evolution. *PLoS One* 11:e0155757. doi: 10.1371/journal.pone.0155757
- Dinsdale, E. A., Pantos, O., Smriga, S., Edwards, R. A., Angly, F., Wegley, L., et al. (2008). Microbial ecology of four coral atolls in the northern line islands. *PLoS One* 3:e1584. doi: 10.1371/journal.pone.0001584
- Ebdon, J. E., Sellwood, J., Shore, J., and Taylor, H. D. (2012). Phages of *bacteroides* (GB-124): A novel tool for viral waterborne disease control? *Environ. Sci. Technol.* 46, 1163–1169. doi: 10.1021/es202874p
- Emerton, L., and Tessema, Y. (2001). *Marine Protected Areas: the Case of Kisite Marine National Park and Mpunguti Marine National Reserve, Kenya*. Nairobi, Kenya: IUCN Eastern Africa Regional Office, 32.
- Flaviani, F., Schroeder, D. C., Lebrecht, K., Balestreri, C., Highfield, A. C., Schroeder, J. L., et al. (2018). Distinct oceanic microbiomes from viruses to protists located near the Antarctic Circumpolar current. *Front. Microbiol.* 9:1474. doi: 10.3389/fmicb.2018.01474
- Glasl, B., Bourne, D. G., Frade, P. R., Thomas, T., Schaffelke, B., and Webster, N. S. (2019). Microbial indicators of environmental perturbations in coral reef ecosystems. *Microbiome* 7:94. doi: 10.1186/s40168-019-0705-7
- Glasl, B., Bourne, D. G., Frade, P. R., and Webster, N. S. (2018). Establishing microbial baselines to identify indicators of coral reef health. *Microbiol. Australia* 39:42. doi: 10.1071/MA18011
- Glasl, B., Webster, N. S., and Bourne, D. G. (2017). Microbial indicators as a diagnostic tool for assessing water quality and climate stress in coral reef ecosystems. *Mar. Biol.* 164:91. doi: 10.1007/s00227-017-3097-x
- Godoy-Vitorino, F., Ruiz-Diaz, C. P., Rivera-Seda, A., Ramírez-Lugo, J. S., and Toledo-Hernández, C. (2017). The microbial biosphere of the coral acropora cervicornis in Northeastern Puerto Rico. *PeerJ.* 5:e3717. doi: 10.7717/peerj.3717
- Goodwin, K. D., Thompson, L. R., Duarte, B., Kahlke, T., Thompson, A. R., Marques, J. C., et al. (2017). DNA sequencing as a tool to monitor marine ecological status. *Front. Mar. Sci.* 4:e1002358–14. doi: 10.3389/fmars.2017.00107
- Guan, Y., Hohn, S., and Merico, A. (2015). Suitable environmental ranges for potential coral reef habitats in the tropical ocean. *PLoS One* 10:e0128831. doi: 10.1371/journal.pone.0128831
- Gudhka, R. K., Neilan, B. A., and Burns, B. P. (2015). Adaptation, ecology, and evolution of the halophilic stromatolite archaeon *Halococcus hamelinensis* inferred through genome analyses. *Archaea* 2015:241608. doi: 10.1155/2015/241608
- Harwood, V. J., Staley, C., Badgley, B. D., Borges, K., and Korajkic, A. (2014). Microbial source tracking markers for detection of fecal contamination in environmental waters: Relationships between pathogens and human health outcomes. *FEMS Microbiol. Rev.* 38, 1–40. doi: 10.1111/1574-6976.12031
- Hattam, C., Evans, L., Morrissey, K., Hooper, T., Young, K., Khalid, F., et al. (2020). Building resilience in practice to support coral communities in the Western Indian Ocean. *Environ. Sci. Policy* 106, 182–190. doi: 10.1016/j.envsci.2020.02.006

- Heip, C., Hernminga, M., and de Bie, M. (1995). *Monsoons and Coastal Ecosystems in Kenya*, 5th Edn. Lieden: National Museum of Natural History.
- Henkel, T. P. (2010). *Coral Reefs*. Available Online at: <https://www.nature.com/scitable/knowledge/library/coral-reefs-15786954/> (accessed November 6, 2019).
- Houk, P., Comeros-Raynal, M., Sudek, M., Vaeoso, M., Mcguire, K., and Regis, J. (2020). Nutrient thresholds to protect water quality and coral reefs. *Mar. Pollut. Bull.* 159:111451. doi: 10.1016/j.marpolbul.2020.111451
- Jaenicke, S., Albaum, S. P., Blumenkamp, P., Linke, B., Stoye, J., and Goesmann, A. (2018). Flexible metagenome analysis using the MGX framework. *Microbiome* 6:76. doi: 10.1186/s40168-018-0460-1
- Johnson, D. R., and Pomati, F. (2020). A brief guide for the measurement and interpretation of microbial functional diversity. *Environ. Microbiol.* 22, 3039–3048. doi: 10.1111/1462-2920.15147
- Kandlikar, G. S., Gold, Z. J., Cowen, M. C., Meyer, R. S., Freise, A. C., Kraft, N. J. B., et al. (2018). Ranacapa: An R package and shiny web app to explore environmental DNA data with exploratory statistics and interactive visualizations. *F1000Research* 7:1734. doi: 10.12688/f1000research.16680.1
- Kannapiran, E., Kannan, L., Purushothaman, A., and Thangarajdou, T. (2008). Physico-chemical and microbial characteristics of the coral reef environment of the Gulf of Mannar marine biosphere reserve, India. *J. Environ. Biol.* 29, 215–222.
- Katwijk, M. M., Van Meier, N. E., Loon, R., Van Hove, E. M., Van Giesen, W. B. J. T., Velde, G., et al. (1993). Sabaki River sediment load and coral stress: correlation between sediments and condition of the Malindi-Watamu reefs in Kenya (Indian Ocean). *Mar. Biol.* 683, 675–676. doi: 10.1007/bf00349780
- Kelly, L. W., Williams, G. J., Barott, K. L., Carlson, C. A., Dinsdale, E. A., Edwards, R. A., et al. (2014). Local genomic adaptation of coral reef-associated microbiomes to gradients of natural variability and anthropogenic stressors. *Proc. Natl. Acad. Sci.* 111, 10227–10232. doi: 10.1073/pnas.1403319111
- Kim, Y., Jeon, J., Kwak, M. S., Kim, G. H., Koh, I., and Rho, M. (2018). Photosynthetic functions of *Synechococcus* in the ocean microbiomes of diverse salinity and seasons. *PLoS One* 13:e0190266. doi: 10.1371/journal.pone.0190266
- Kisand, V., Valente, A., Lahm, A., Tanet, G., and Lettieri, T. (2012). Phylogenetic and functional metagenomic profiling for assessing microbial biodiversity in environmental monitoring. *PLoS One* 7:43630. doi: 10.1371/journal.pone.0043630
- Kitheka, J. U. (2019). Salinity and salt fluxes in a polluted tropical river: The case study of the Athi river in Kenya. *J. Hydrol. Reg. Stud.* 24:100614. doi: 10.1016/j.ejrh.2019.100614
- Lauro, F. M., McDougald, D., Thomas, T., Williams, T. J., Egan, S., Rice, S., et al. (2009). The genomic basis of trophic strategy in marine bacteria. *Proc. Natl. Acad. Sci. U S A* 106, 15527–15533. doi: 10.1073/pnas.0903507106
- Lubelski, J., Konings, W. N., and Driessen, A. J. M. (2007). Distribution and physiology of ABC-type transporters contributing to multidrug resistance in bacteria. *Microbiol. Mol. Biol. Rev.* 71, 463–476. doi: 10.1128/mmb.00001-07
- MacManes, M. D. (2014). On the optimal trimming of high-throughput mRNA sequence data. *Front. Genet.* 5:13. doi: 10.3389/fgene.2014.00013
- Mayorga-Adame, G. C., Ted Strub, P., Batchelder, H. P., and Spitz, Y. H. (2016). Characterizing the circulation off the Kenyan-Tanzanian coast using an ocean model. *J. Geophys. Res. Oceans* 121, 1377–1399. doi: 10.1002/2015JC010860
- McClanahan, T. (1988). Seasonality in East Africa's coastal waters. *Mar. Ecol. Prog. Ser.* 44, 191–199. doi: 10.3354/meps044191
- McClanahan, T. R., Atewerberhan, M., Graham, N. A. J., Wilson, S. K., Ruiz Sebastián, C., Guillaume, M. M. M., et al. (2007). Western indian ocean coral communities: bleaching responses and susceptibility to extinction. *Mar. Ecol. Prog. Ser.* 337, 1–13. doi: 10.3354/meps337001
- McClanahan, T. R., Graham, N. A. J., Calnan, J. M., and Macneil, M. A. (2010). Toward pristine biomass: reef fish recovery in coral reef marine protected areas in Kenya. *Ecol. Appl.* 17, 1055–1067. doi: 10.1890/06-1450
- McDole, T., Nulton, J., Barott, K. L., Felts, B., Hand, C., Hatay, M., et al. (2012). Assessing coral reefs on a Pacific-wide scale using the microbialization score. *PLoS One* 7:e43233. doi: 10.1371/journal.pone.0043233
- McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One* 8:e61217. doi: 10.1371/journal.pone.0061217
- Munyao, T., Tole, M. P., and Jungerius, P. D. (2003). *Sabaki River transport and deposition in the Indian Ocean. Recent Advances in Coastal Ecology, Studies from Kenya*. Netherland: African Studies Centre, 119–132.
- Mwangi, S., Dzeha, T., and Kimathi, A. (2001). *State of Marine Pollution in Mombasa Marine ational Park Marine Reserve and Mtwapa Creek*.
- Neumann, B., Vafeidis, A. T., Zimmermann, J., and Nicholls, R. J. (2015). Future coastal population growth and exposure to sea-level rise and coastal flooding - A global assessment. *PLoS One* 10:e0118571. doi: 10.1371/journal.pone.0118571
- Ngugi, I. (2002). Economic impacts of marine protected areas?: a case study of the mombasa marine park (Kenya). *J. Soc. Sci. Grad. Stud. Assoc.* 1, 1–11.
- Obura, D. (2017). *Reviving the Western Indian Ocean Economy: Actions for a Sustainable Future*. Gland, Switzerland: WWF International.
- Obura, D. O. (2001). Kenya. *Mar. Pollut. Bull.* 42, 1264–1278. doi: 10.1016/S0025-326X(01)00241-7
- Obura, D. O., Aeby, G., Amornthammarong, N., Appeltans, W., Bax, N., Bishop, J., et al. (2019). Coral reef monitoring, reef assessment technologies, and ecosystem-based management. *Front. Mar. Sci.* 6:580. doi: 10.3389/fmars.2019.00580
- Okuku, O. E., Linet Imbayi, K., Gilbert Omondi, O., Veronica Ogolla Wayayi, W., Catherine Sezi, M., Mokeira Maureen, K., et al. (2019). “Decadal Pollution Assessment and Monitoring along the Kenya Coast” in *Monitoring of Marine Pollution*, ed. H. B. Fouzia (London: IntechOpen), 1–15.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., et al. (2019). *vegan: Community Ecology Package. R Package Version 2.5-6*.
- Okuku, E. O., Ohowa, B., Mwangi, S. N., Munga, D., Kiteresi, L. I., Wanjeri, V. O., et al. (2011). Sewage pollution in the coastal waters of Mombasa City, Kenya: A norm rather than an exception. *Int. J. Environ. Res.* 5, 865–874.
- Ongore, C. O., Okuku, E. O., Mwangi, S. N., Kiteresi, L. I., Ohowa, B. O., Wanjeri, V. O., et al. (2013). Characterization of nutrients enrichment in the estuaries and related systems in Kenya coast. *J. Environ. Sci. Water Res.* 2, 181–190.
- Owens, J. D. (1978). Coliform and *escherichia coli* bacteria in seawater around penang island, Malaysia. *Water Res.* 12, 365–370. doi: 10.1016/0043-1354(78)90101-X
- Ransome, E., Geller, J. B., Timmers, M., Leray, M., Mahardini, A., Sembiring, A., et al. (2017). The importance of standardization for biodiversity comparisons: A case study using autonomous reef monitoring structures (ARMS) and metabarcoding to measure cryptic diversity on Mo'orea coral reefs, French Polynesia. *PLoS One* 12:e0175066. doi: 10.1371/journal.pone.0175066
- Robinson, J., New, A. L., Popova, E. E., Srokosz, M. A., and Yool, A. (2017). Far-field connectivity of the UK's four largest marine protected areas: Four of a kind? *Earth's Future* 5, 475–494. doi: 10.1002/2016EF000516
- Rohwer, F., Seguritan, V., Azam, F., and Knowlton, N. (2002). Diversity and distribution of coral-associated bacteria. *Mar. Ecol. Prog. Ser.* 243, 1–10. doi: 10.3354/meps243001
- Roik, A., Röthig, T., Roder, C., Ziegler, M., Kremb, S. G., and 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:e0163939. doi: 10.1371/journal.pone.0163939
- Roitman, S., Joseph Pollock, F., and Medina, M. (2018). *Coral Microbiomes as Bioindicators of Reef Health*. Cham: Springer, 39–57.
- Santoro, A. E., Richter, R. A., and Dupont, C. L. (2019). Planktonic marine archaea. *Ann. Rev. Mar. Sci.* 11, 131–158. doi: 10.1146/annurev-marine-121916-063141
- Silveira, C. B., Cavalcanti, G. S., Walter, J. M., Silva-Lima, A. W., Dinsdale, E. A., Bourne, D. G., et al. (2017). Microbial processes driving coral reef organic carbon flow. *FEMS Microbiol. Rev.* 41, 575–595. doi: 10.1093/FEMSRE
- Skerker, J. M., Prasol, M. S., Perchuk, B. S., Biondi, E. G., and Laub, M. T. (2005). Two-component signal transduction pathways regulating growth and cell cycle progression in a bacterium: A system-level analysis. *PLoS Biol.* 3:e334. doi: 10.1371/journal.pbio.0030334
- Sullivan, M. B., Coleman, M. L., Weigle, P., Rohwer, F., and Chisholm, S. W. (2005). Three *Prochlorococcus* cyanophage genomes: signature features and ecological interpretations. *PLoS Biol.* 3:e144. doi: 10.1371/journal.pbio.0030144
- Sunagawa, S., Coelho, L. P., Chaffron, S., Kultima, J. R., Labadie, K., Salazar, G., et al. (2015). Structure and function of the global ocean microbiome. *Science* 348, 1261359–1261359. doi: 10.1126/science.1261359

- Sunagawa, S., Woodley, C. M., and Medina, M. (2010). Threatened corals provide underexplored microbial habitats. *PLoS One* 5:e9554. doi: 10.1371/journal.pone.0009554
- Tao, P., and Rao, V. B. (2019). *Bacteriophage Vaccines. In Reference Module in Life Sciences*. Amsterdam: Elsevier.
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: A tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33
- Tatusov, R. L., Natale, D. A., Garkavtsev, I. V., Tatusova, T. A., Shankavaram, U. T., Rao, B. S., et al. (2001). The COG database: New developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res.* 29, 22–28. doi: 10.1093/nar/29.1.22
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. Available Online at: <https://www.r-project.org/> (accessed October 12, 2020).
- Thurber, R. V., Payet, J. P., Thurber, A. R., and Correa, A. M. (2017). Virus–host interactions and their roles in coral reef health and disease. *Nat. Rev. Microbiol.* 15, 205–216. doi: 10.1038/nrmicro.2016.176
- Tout, J., Jeffries, T. C., Webster, N. S., Stocker, R., Ralph, P. J., and Seymour, J. R. (2014). Variability in microbial community composition and function between different niches within a coral reef. *Microbiol. Ecol.* 67, 540–552. doi: 10.1007/s00248-013-0362-5
- Tuda, A., and Omar, M. (2012). Protection of marine areas in Kenya. *George Wright Forum* 29, 43–50.
- Van Der Elst, R., Everett, B., Jiddawi, N., Mwatha, G., Afonso, P. S., and Boule, D. (2005). Fish, fishers and fisheries of the Western Indian Ocean: Their diversity and status. A preliminary assessment. *Phil. Trans. R. Soc. A Math. Phys. Eng. Sci.* 363, 263–284. doi: 10.1098/rsta.2004.1492
- Vanwongterghem, I., and Webster, N. S. (2020). Coral Reef microorganisms in a changing climate. *IScience* 23:100972. doi: 10.1016/j.isci.2020.100972
- Varasteh, T., Moreira, A. P. B., Silva Lima, A. W., Leomil, L., Otsuki, K., Tschoeke, D., et al. (2020). Genomic repertoire of *Mameliella alba* Ep20 associated with Symbiodinium from the endemic coral *Mussismilia braziliensis*. *Symbiosis* 80, 53–60. doi: 10.1007/s13199-019-00655-x
- Vigil-Stenman, T., Ininbergs, K., Bergman, B., and Ekman, M. (2017). High abundance and expression of transposases in bacteria from the Baltic Sea. *ISME J.* 11, 2611–2623. doi: 10.1038/ismej.2017.114
- Walsh, K., Haggerty, J. M., Doane, M. P., Hansen, J. J., Morris, M. M., Moreira, A. P. B., et al. (2017). Aura-biomes are present in the water layer above coral reef benthic macro-organisms. *PeerJ*. 5:e3666. doi: 10.7717/peerj.3666
- Wanjari, V. W. O., Okuku, E. O., Barsanti, M., Schirone, A., Delbono, I., Owato, G., et al. (2021). Baseline radionuclide and heavy metal concentrations in sediments of Sabaki River estuary (Kenya, Indian Ocean). *Mar. Pollut. Bull.* 164:112033. doi: 10.1016/j.marpolbul.2021.112033
- Waterbury, J. B., Watson, S. W., Guillard, R. R. L., and Brand, L. E. (1979). Widespread occurrence of a unicellular, marine, planktonic, cyanobacterium. *Nature* 277, 293–294. doi: 10.1038/277293a0
- Weber, L., Gonzalez-Diaz, P., Armenteros, M., and Apprill, A. (2019). The coral ecosphere: A unique coral reef habitat that fosters coral–microbial interactions. *Limnol. Oceanogr.* 64, 2373–2388. doi: 10.1002/lno.11190
- Wickham, H. (2016). *ggplot2-Elegant Graphics for Data Analysis*, 2nd Edn. New York: Springer-Verlag.
- Wilkins, S. (2015). Structure and mechanism of ABC transporters. *F1000Prime Rep.* 7, 7–14. doi: 10.12703/P7-14
- Williamson, S. J., Allen, L. Z., Lorenzi, H. A., Fadrosch, D. W., Brami, D., Thiagarajan, M., et al. (2012). Metagenomic exploration of viruses throughout the Indian Ocean. *PLoS One* 7:e0042047. doi: 10.1371/journal.pone.0042047
- Won, N., Il, Kim, K. H., Kang, J. H., Park, S. R., and Lee, H. J. (2017). Exploring the impacts of anthropogenic disturbance on seawater and sediment microbial communities in Korean coastal waters using metagenomics analysis. *Int. J. Environ. Res. Public Health* 14:130. doi: 10.3390/ijerph14020130
- Zhang, W., Cao, S., Ding, W., Wang, M., Fan, S., Yang, B., et al. (2020). Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics. *Microbiome* 8:47. doi: 10.1186/s40168-020-00826-9
- Zhao, X., Schwartz, C. L., Pierson, J., Giovannoni, S. J., McIntosh, J. R., and Nicastro, D. (2017). Three-dimensional structure of the ultraoligotrophic marine bacterium “*Candidatus Pelagibacter ubique*”. *Appl. Environ. Microbiol.* 83:e2807–16. doi: 10.1128/AEM.02807-16

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Wambua, Gourelé, de Villiers, Karlsson-Lindsjö, Wambiji, Macdonald, Bongcam-Rudloff and de Villiers. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.