*Original Research Article*

Assessing Microbial Diversity at Watering Points Along the Isiukhu River, Kenya Using 16S rRNA Sequencing

.

ABSTRACT

|  |
| --- |
| **Aims:** 1) assess microbial diversity in water and sediment samples from watering points and springs along the upper reaches of river isiukhu, kakamega county, kenya, 2) utilize advanced molecular techniques, specifically 16s rRNA sequencing, to understand microbial communities, including both cultivable and non-cultivable microbes and 3) evaluate the influence of anthropogenic and environmental factors on microbial assemblages in freshwater ecosystems.  **Study design:** The research employs a cross-sectional study design with random sampling of microbial communities from multiple locations along the river.  **Place and duration of study:** the study was conducted in kakamega county, kenya, along the upper reaches of river isiukhu, between GPS coordinates 34.8372 and 0.2970 in the forest, and 34.7448 and 0.2557 downstream.  **Methodology:** Nine sampling points were established for collecting water and sediment samples, resulting in 54 water samples and 18 sediment samples. Genomic DNA was extracted directly from 1 liter of water and sediment samples. 16s rRNA sequencing was performed on the extracted DNA, with sequence reads generated from an illumina miseq run. Taxonomy assignment was conducted using qiime pipeline scripts, matching against the ribosomal database project (RDP) and green genes. Data visualization was performed using the emperor online tool.  **Results:** The analysis identified 1,549 operational taxonomic units (OTUS), comprised of 99% bacteria and 1% archaea. Forty phyla were detected, with predominant groups being proteobacteria (34%), bacteroidetes (17%), firmicutes (6%), and verrucomicrobia (5%). Sites ranked by faith’s phylogenetic diversity index included spring 1 at lianila (243) as the highest and the confluence site (43) as the lowest.  **Conclusion:** The study reveals significant microbial diversity along river isiukhu, highlighting the impact of anthropogenic and environmental factors on microbial assemblages, and contributing to water conservation and management efforts. Anthropogenic factors could help explain the spatial variability in the composition of microbes seen here. This study begins to shed more light into the richness in freshwater microbiomes thus availing important data for monitoring and conservation of environment. |

*Keywords: Microbial Diversity, 16S rRNA Sequencing, Freshwater Ecosystems, River Isiukhu*

1. INTRODUCTION

The aquatic environment contains a variety of microorganisms, such as viruses, bacteria, cyanobacteria, algae, fungi, and protozoa. Therefore, it is necessary to monitor the microbiological quality of water. The problem mainly concerns the sources that are intended to supply us with drinking water, but also rainwater, which is increasingly being considered to be an alternative water source. Accurate and fast detection of microbial cells is a constant challenge across a broad spectrum of research and application fields. This challenge covers issues as diverse as obtaining quantitative information on specific microbial populations in natural surface water (Zamorska et al.,. 2023; Kalsoom et al., 2020). Microbes composed of archaea, algae, bacteria, protozoa, fungi and small metazoa representing the prevalent and diverse assemblage across ecosystems have crucial roles in the functioning of the aquatic ecosystem (Head & Gray, 2016; Kumar et al., 2021). As such, microbial existence in drinking water is a key factor fueling the prevalence of water-borne diseases (Paruch et al., 2019). Knowledge of the spatial patterns of microbial abundance and diversity in oceans and temperate lakes exists (Liu et al., 2013; Wee et al., 2021). However, a few studies have focused on patterns of microbial community composition in rivers. Microbial communities in the river ecosystems are driven by various interacting factors and processes, with environmental factors playing an imperative role in shaping the microbial communities’ composition (Douglas et al., 2021; Sánchez et al., 2021). However, spatial variation in the dispersal and abundance of microbes is an inherent property of ecological systems thus, insufficient knowledge of microbial spatial variation can hinder the effective assessment of the relative importance of environmental factors in driving microbial community succession not only in structure but also in function. The concept that microorganisms are ubiquitous in distribution and can proliferate in any habitat that supports their growth has been a long-standing notion in microbial ecology. In general, patterns of microbial diversity are correlated with habitat conditions, owing to varying degrees of habitat preference, and adaptability of different groups of microorganisms. In particular, salinity is shown to be a major factor relating microbial communities (Dalmacio et al., 2019).

Human activities particularly urbanization, waste disposal and agricultural practices have greatly increased inputs of microbial and other pollutants into terrestrial and aquatic habitats (Huang et al., 2020; Valverde et al., 2021). Previous studies investigating environmental colonization by microbes have mainly employed culture methods. However, this method limits the fundamental appreciation of the diversity of interactions within these systems. Nonetheless, with the development of technology, studies exploring environmental bacteria in a culture-independent manner by isolating DNA from environmental samples and transforming it into large insert clones do exist (Kim et al., 2010). In addition, bacterial identification using DNA-based tools, such as polymerase chain reaction (PCR), is becoming increasingly popular due to their specificity and short turn-around time, compared to the traditional culture-based methods (Parks et al., 2020). These molecular methods also allow the detection and identification of "viable but unculturable" bacteria which though are metabolically-active they happen to be non-dividing (Herrera & Cockell, 2007). With the introduction of molecular methods based on sequencing of the ribosomal RNA (rRNA) genes, it became possible to make complete analyses of microbial diversity in an environmental sample (Rajendhran & Gunasekaran, 2011).

River Isiukhu is one of the largest rivers in Kakamega County, Kenya and it’s an essential source of drinking water not just for humans but a watering point for animals, agriculture, source of industrial water, and laundry point as well as sewage disposal (Oremo et al., 2020). The upper stream of this river percolates through one of the most agricultural-rich regions comprising both large and small-scale farming systems within the county. The excess input of nitrogen and phosphorus from intensive agricultural activities has significantly impacted the water quality in the upper sections of this river in recent years. Even though microbial composition in this river is not well documented, existing studies have suggested that numerous anthropogenic activities along the river may influence microbial diversity (Onyango & Angienda, 2010).

This study used targeted metagenomics; small-scale metagenome of the 16S rRNA-based methods to explore the spatial and temporal diversity of planktonic bacterial and microbial eukaryotic communities at the water source(s) at different watering points along the upper reaches (Watershed) of River Isiukhu. The current study set

2. material and methods

**2.1 Study site**

The study was conducted along River Isiukhu, whose source is Nandi hills and flows through Kakamega tropical rain forest in Kakamega County-Kenya (Latitude: 0° 16' 60.00" N Longitude: 34° 45' 0.00" E). It is located in the tropical zone, which makes the river subject to seasonal changes in hydrology and aquatic environmental conditions. The average annual rainfall of Kakamega County is 1800mm per annum, and is bimodally distributed with peaks in April-June and August-October. The driest months are from December-February. Temperatures range from a minimum of 10.3°C to a maximum of 30.8°C with an average of 20.5°C. The area is covered by phonolites and also tertiary volcanic rocks-olivine basalts and nepheline which is fertile that allow growth of planktons. Kakamega County is densely populated with 1.66 million people and an area of 3,033.8km² (KPHC, 2009). The river drains land cover of forest, mixed agriculture and periurban from upstream to downstream before joining River Nzoia. Anthropogenic activities conducted along the river include small scale agriculture, livestock watering, laundry, bathing and sewage disposal.

**2.2 Research design**

The across sectional study design was confined to the upper reaches of River Isiukhu between GPS locations 34.8372 and 0.2970 in the forest and 34.7448 and 0.2557 downstream.

**2.3 Sampling technique**

A completely randomized sampling design was used, where a total of 72 samples (54 water and 18 sediments) were collected in sterile containers in duplicate localities, directly from nine randomly selected points (i.e. springs, river watering points and wastewater 17 entry points). Sampling lasted for three months between August and October, and sampling was done in six sessions. During the fifth and sixth sessions, samples were included for 16S rDNA analysis. A hand-held GPS was used to collect GIS data to locate sampling points on the topographic map of the study area (Figure 1).



Figure 1: GIS map of the study area showing the various sampling sites (s1-s5) and (sp1-sp4) in the upper reaches of river Isiukhu. This map was drawn to scale using ARC GIS software.

**2.4 Characteristic of water and sediment sampling sites**

The inclusion of sampling site was such that it was accessed watering point with greater than 10 households along River Isiukhu. The sampling sites comprised the forest area the confluence of Lianila stream and Isiukhu River (assessed bacterial load of Lianila stream that drains the treated sewage effluent and farm lands, S2), Savona resort area (quantified the bacterial contribution from the recreation area, S3), Amalemba (determined the bacterial contribution of the small-scale mixed agriculture and peri-urban area, S4) and wastewater entry point (established the bacterial load from treated sewer waters in the area, S5). The springs sampled included four sites namely: Confluence-Lianila (SP1), Resort (SP2), Amalemba (SP3) and Sewage (SP4) (Table 1). All springs were located along the River Isiukhu, to establish their bacterial contamination. In each sampling site, a total of six water samples were collected. In addition, a sediment sample was collected from each site during the fifth and sixth sessions. Samples collected during fifth and sixth sessions included both coliform enumeration and 16S rRNA sequencing and were in addition labelled as either Early (E) or Late (L) denoting early October 2015 or late October 2015. These are referred to in sections with sequencing results.

Table 1: Sampling Points and associated labels

|  |  |  |  |
| --- | --- | --- | --- |
| Sampling site | Water Sample labeling | Sediment Sample labeling | Coordinates Y/X |
| Forest (Site 1) | S1w | S1s | 32842/704289 |
| Confluence of Lianila stream and Isiukhu River (Site 2) | S2w | S2s | 30860/697837 |
| Savona resort area (site 3) | S3w | S3s | 30520/697671 |
| Amalemba area (Site 4) | S4w | S4s | 28124/694710 |
| Waste water entry point (Site 5) | S5w | S5s | 28232/694051 |
| Spring 1 (site 1 area) | SP1w | SP1s | 30806/697772 |
| Spring 2 (site 2 area) | SP2w | SP2s | 30696/697601 |
| Spring 3 (site 3 area) | SP3w | SP3s | 27988/694427 |
| Spring 4 (site 4 area) | SP4w | SP4s | 28377/694324 |

\*Sampling Points

**2.5 Sample preparation and DNA extraction**

The samples collected from the river were pooled for each sampling point and diluted at 1:100, while samples collected from the springs were not diluted. The biomass from each of the water samples one liter was filtered via nitrocellulose filters 47 mm diameter with a pore size of 0.45 μm by vacuum filtration. Total DNA was extracted from both the filtrates and sediments (0.25g) using Powerlyzer Power Soil DNA Isolation Kit (MOBIO Laboratories Inc. California - USA) as per the manufacturer’s instructions. The extracted DNA was quantified using the NanoDropTM-2000 spectrophotometer at an absorbance of 260/280 and 260/230. The system works based on the principle of Ultraviolet light absorbance whereby nucleic acids absorb UV light at 230 nm. DNA samples were considered pure if the 260/280 ratio was ~1.8 and the 260/230 ratio ranged 1.8-2.2. The possible reason could be that the presence of contaminants such as protein and phenol lower the 260/280 ratio, while contamination with chemotropic salts lowers the 260/230 absorbance. The extracted DNA samples were stored at -80˚C awaiting Polymerase Chain Reaction (PCR) amplification and sequencing procedures for the 16S rDNA, which were done at the Helmholtz Centre for Infection Research Group Microbial Communication, in Braunschweig, Germany.

**2.6 PCR amplification of 16S rRNA**

The 2-step PCR amplification in the V5-V6 region of 16S rRNA gene was done using optimized protocol designed for high-throughput by next-generation sequencing (NGS). The forward PCR primer used contained an Illumina adapter, a linker, and a gene primer, while the reverse PCR primer contained an Illumina adapter, a barcode, a linker, and a gene primer. All the samples received the same forward primer, but each sample received a different barcoded reverse primer. A “mapping file” was used to track the amplified samples with the respective barcoded primer.

The extracted DNA templates were amplified in a 20 μl mixture containing 4 μl of 5x Primer Star Buffer (TAKARA BIO INC. Code No. R010A), 1.6 μl dNTP Mixture, 1 μl DMSO 5%, 0.2 μl Prime Star HS DNA Polymerase, 0.4 μl of Illumina 807 Fw (5’ – GGATTAGATACCCBRGTAGTC – 3’) primer and 0.4 μl of Illumina 1050 Rev (5’–AGYTGDCGACRRCCRTGCA– 3’) primer, 11.4 μl of ultrapure sterile water (MilliQ water) and 1 μl of DNA (5 ng DNA). The amplification of the DNA templates was carried out in a Real-Time PCR System (Applied Biosystems, UK). PCR cycling conditions were: initial denaturation at 98 °C for 3 min, followed by 35 cycles of denaturation at 95°C for 10 s, annealing at 51 °C for 10 s, extension at 72°C for 45 seconds and a final extension step at 72°C for 2 min, with samples held at 4°C (Bohorquez et al., 2011).

**2.7 Illumina sequencing and sequence library processing**

Illumina sequencing works on the principle that a DNA polymerase catalyzes the incorporation of fluorescent labeled deoxyribonucleotide triphosphates (dNTPs) into a DNA template strand during consecutive cycles of DNA synthesis. The nucleotides are identified by fluorophore excitation at the point of incorporation in each cycle. This process occurs across millions of fragments in a massively parallel fashion whilst delivering a high yield of error-free reads. The 16S rDNA sequencing was performed on an Illumina Miseq machine. The reads generated were paired-end 250bp (Chaitankar et al., 2016).

Quality control of the Miseq Illumina sequencing reads was done using next-generation sequencing (NGS) Tool kit v2.3. Filter parameters were set at a cut-off quality (Phred) score of 20 with a cut-off read length for high quality set at 97% to exclude flawed sequences due to sequencing error. Primers were not trimmed off during quality control of the reads. High-quality reads were written into a separate file. The high-quality forward and reverse reads were merged using pandaseq and default settings were used. The merged fasta reads were quality filtered and reads assigned to their samples source and metadata mapping file using Qiime scriptsplit\_libraries\_fastq.py. Reads that were too short after quality truncation were discarded. The sequences were clustered to operational taxonomic units (OTUs) using an open-reference OTU picking protocol in qiime pipeline (Caporaso et al., 2010, Rideout et al., 2014) and against the Green Genes, Ribosomal Database Project (RDP) and NCBI 16S rRNA microbial databases. Representative sequences from each OTUs were picked to generate the OTU table and checked for chimera with QIIME via Chimera Slayer (Qu et al., 2020).

**2.8 Sequence data analyses**

The Quantitative Insights into Microbial Ecology (QIIME) pipeline was used for the sequence data analysis (available at: http://qiime.sourceforge.net/). QIIME was preferred because it supports a wide range of microbial community analyses and visualizations, including histograms of within and between sample diversity (Kuczynski et al., 2011).

**2.8.1 Trimming barcodes**

The QIIME script “split\_libraries.py” was used to examine the sequences for barcodes and primer; with subsequent trimming of the barcodes and the primers from the sequences. The “Fasta file” containing the raw sequencing data and the “mapping file” linking the barcodes and the primers to the samples were used to add a header indicating where each sequence came from (Mäki et al., 2016).

**2.8.2 Picking operational taxonomic units (OTUs)**

Operational Taxonomic Unit picking was done, so as to analyze similar sequences (≥97% identical) as one hence minimizing the number of sequences analyzed. Obtained sequences were clustered into Operational Taxonomic Units (OTUs) using the QIIME script “pick\_otus.py” based on sequence similarity thus discriminating between specific microorganisms namely bacteria, archaea and microbial eukarya (Lim et al., 2021).

Taxonomic assignment was performed by mapping OTU representative tags to the Ribosomal Database Project (RDP), Green Genes and NCBI 16S rRNA microbial databases using the script “assign\_taxonomy.py” (Kuczynski et al., 2011).

Multiple sequence alignments were done to build phylogenetic trees and test for evolutionary distances. The script “align\_seqs.py” was used for multiple sequence alignment. Since the aligned sequences contained numerous gaps due to hyper variable regions in the 16S rRNA gene sequences, “Lane mask” was used for filtering using the script “filter\_alignment.py” to build accurate trees. The script “make\_phylogeny.py” was used to build trees using the “maximum likelihood” algorithm which is suitable for 16S rRNA sequences (Chowdhury & Garai, 2017).

**2.8.3 Alpha Diversity**

To systematically look at alpha diversity, multiple rarefactions were performed at three different depths: 350 sequences, 3500 sequences and 20000 sequences using the QIIME script “multiple\_rarefactions.py”. This was done to overcome the challenge of deeper sequencing that yields more species, giving different yields for each sample. Alpha diversity was calculated using the QIIME script “alpha\_diversity.py”. The Chao1 metric was preferred in this case to capture information (richness and evenness) on an array of species that are less detectable such as singletons (species sampled once) and doubletons (species sampled twice) (Walters & Martiny, 2020).

**2.8.4 Beta Diversity**

The workflow script “jacknifed\_beta\_diversity.py” in QIIME was used to perform beta diversity analysis (Kuczynski et al., 2011). In this case, the beta diversity metric was used to compare and calculate the distance between the samples. Generally, a beta diversity distance matrix is computed for the full dataset followed by multiple rarefactions at a single sequence depth of 350, since most samples had a little fewer than 400 sequences per sample total, with 377 being the minimum number of sequences in a sample. In addition, using 350 as the sequencing depth created a large collection of distance matrices; hence, the need for a depth that is significantly smaller than the number of sequences in the smallest sample (i.e. samples having fewer sequences than the rarefaction depth will be left out). Higher rarefaction depth would make every rarefied OTU table similar with no difference in the distance matrices.

The distance matrices were computed using the unweighted unifrac for all the rarefied OTU tables. The unweighted unifrac is a qualitative distance metric hence it is most informative since the aim is to assess differences in microbial communities about physical parameters in the environment. Therefore, we do not consider abundance information (as in weighted unifrac) which will greatly obscure important variation patterns in the present taxa (Lozupone et al., 2007). The Principal Coordinate Analysis (PCoA) plot, the cladogram and biplot were created and were visualized using an online tool, Emperor (http://biocore.github.io/)

**2.9 Ethical Consideration.**

Apparently the microbial diversity study itself did not inherently require ethical approval from Masinde Muliro University of Science and Technology institutional review ethical committee (MMUST IREC), since no aspects of the research engaged specific environmental regulations.

3. results and discussion

**3.1 16S rRNA Amplification and Sequencing**

PCR amplification was successful and low-quality sequences and dubious amplicons were excluded from the analysis. The gel photos are as represented in the plates in (figure 2) and Sequences of 16S rRNA in 36 Samples in (table 2).

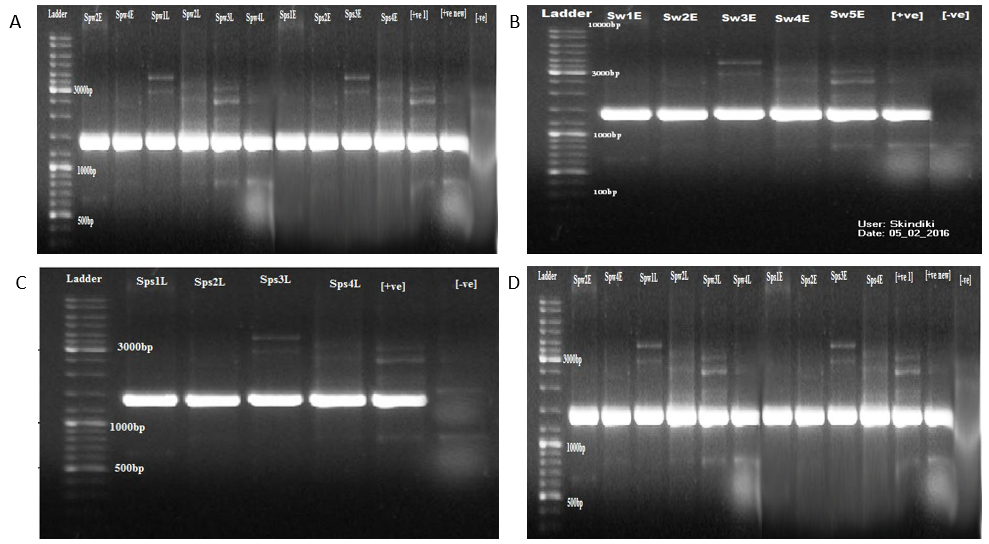


Figure 2: Gel Photos Showing the 1500 Base-Pair Bands of the 16S rRNA Coding Region . A-D shows the 1500 base-pair bands of the 16S rRNA coding region, the ladder, the controls and the samples as mounted on the gel. Out of 36 water and sediment samples, only 34 were sequenced as two samples; Spring 1 water early (SP1w\_E) and Spring 3 water early (SP3w\_E) were negative for PCR products. Two other samples namely, S2s\_L (n=377) and S3s\_L (n=549) had too low reads to be analyzed in a meaningful way (i.e. less than 10% compared to the others) (See Table 2). The 34 samples that were studied yielded a total number of 640,775 16S rRNA sequence counts.

Table 2: Sequences of 16S rRNA in 36 Samples

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Time point** | **Site** | **River**  **Water (Sw)** | **River**  **Sediment (Ss)** | **Site** | **Spring**  **Water**  **(SPw)** | **Spring**  **Sediment (SPs)** |
| **Early**  **October 2015** | Site 1 | 20558 | 23998 |  | - | - |
| Site 2 | 18426 | 25525 | Spring 1 | No Amplicon | 21376 |
| Site 3 | 21661 | 9594 | Spring 2 | 14380 | 23372 |
| Site 4 | 21521 | 22752 | Spring 3 | No Amplicon | 21277 |
| Site 5 | 21600 | 23871 | Spring 4 | 14845 | 23469 |
| **Late**  **October 2015** | Site 1 | 22991 | 30673 |  | - | - |
| Site 2 | 23393 | 377 | Spring 1 | 13449 | 20390 |
| Site 3 | 21324 | 549 | Spring 2 | 22039 | 16572 |
| Site 4 | 22744 | 14841 | Spring 3 | 10244 | 19493 |
| Site 5 | 21046 | 20472 | Spring 4 | 11929 | 20024 |

\*The number of 16s rRNA sequences obtained from each sample (either water or sediment) collected from different sampling sites. The table gives details of the 36 samples at two time points denoted as early October 2015 and late October 2015. The minimum number of sequences recorded in a site was 377 (sampling point 2 sediment late (S2s\_L), while the maximum number of sequences recorded in a site were 30,673 (sampling point 1 sediment late (S1s\_L), with 18,846 being the mean sequences per sample. The raw sequence chromatograms, sequence contigs and aligned sequences are shown in appendices VI, VII and VIII respectively. The 640,775 amplified sequences were clustered into 1,549 operational taxonomic units (OTUs).

**3.2 Microbial Assemblages**

The Quantitative Insights into Microbial Ecology (QIIME) pipeline was used for the sequence data analysis as presented in the subsequent sections. Two domains namely, Bacteria (99%, n=1531) and Archaea (1%, n=18), which are prokaryotes were identified. The bacterial domain had 35 representative phyla while, the archaea domain had four representative phyla, namely Euryarchaeota, Thaumarchaeota, Crenarchaeota and Aenmarchaeota as represented in (Figure 2). There were 12 major phyla accounting for more than 1% of the total phyla observed. These major phyla include: Proteobacteria (34%), Bacterioidetes (18%), Firmicutes (6%), Verrucomicrobia (5%), Acidobacteria (5%), Planctomycetes (4%), Chloroflexi (4%), Pacubacteria (4%), Spirochaetes (2%), Ignavibacteriae (2%) and Candidate division WPS-1 (1%) A large number of the phyla (65%) were rare species representing less than 1% of all the sequences (Figure 3).

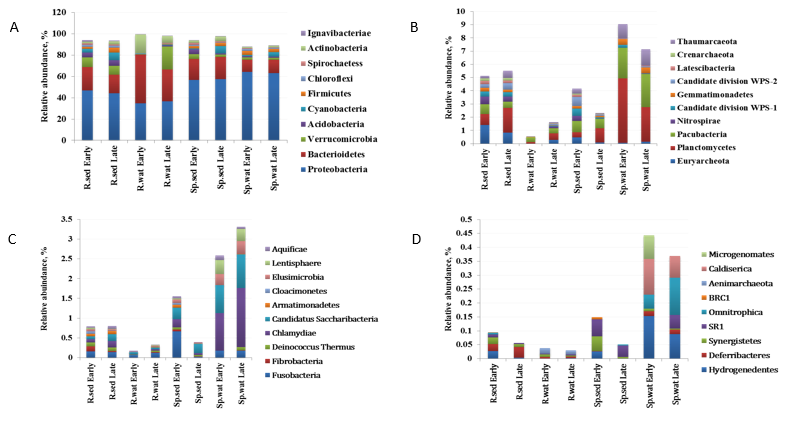


Figure 3: Relative Abundance of Bacterial Phyla on Pooled Samples Early (early October 2015) and Late (late October 2015). Different colours represent types of taxa and their relative abundance.

**3.3 The Major Classes of the Six Phyla**

The classes of the six major phyla are summarized in (Figure 4). In the phylum Proteobacteria, Betaproteobacteria (39%) was the most abundant followed by Deltaproteobacteria (27%), Gammaproteobacteria (17%), Alphaproteobacteria (15%) and Epsilonproteobacteria (2%). The second most abundant phyla was Bacteriodetes with the classes Sphingobacteria (38%), Flavobacteria (21%), Cytophaga (17%), Bacterioidetes (17%), Incertiaesedis (15%) and Bacterioide (9%). Amongst the Firmicutes, Clostridia (73%) was the most abundant class, followed by Baccili (13%), Negativicutes (13%) and Erysipelotrichia (2%). Verrucomicrobia was the fourth most abundant phyla with the following classes: Sub-division 3 (37%), Verrucomicrobiae (30%), Opitutiae (18%), Spartobacterium (13%) and Sub-dividion 5 (1%). Acidobacteria had two classes; Acidobacteriae (80%) and Holophaga (20%). Likewise, the phylum Actinobacteria had two representative classes with the most abundant class being Actinobacteria (96%) and Thermophila (4%).

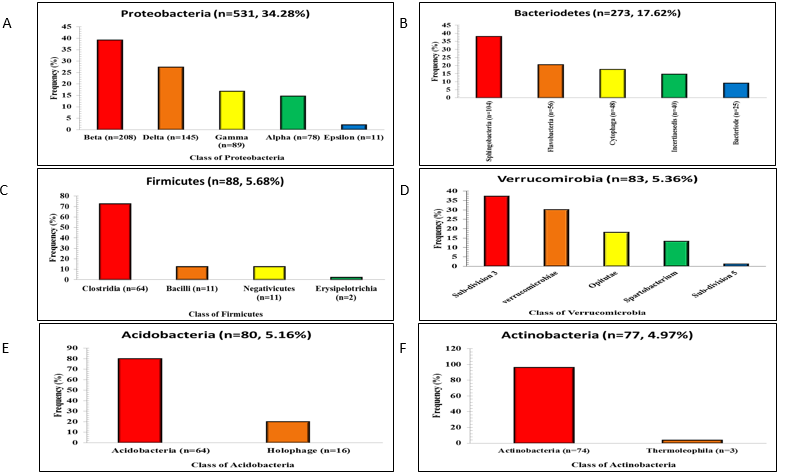
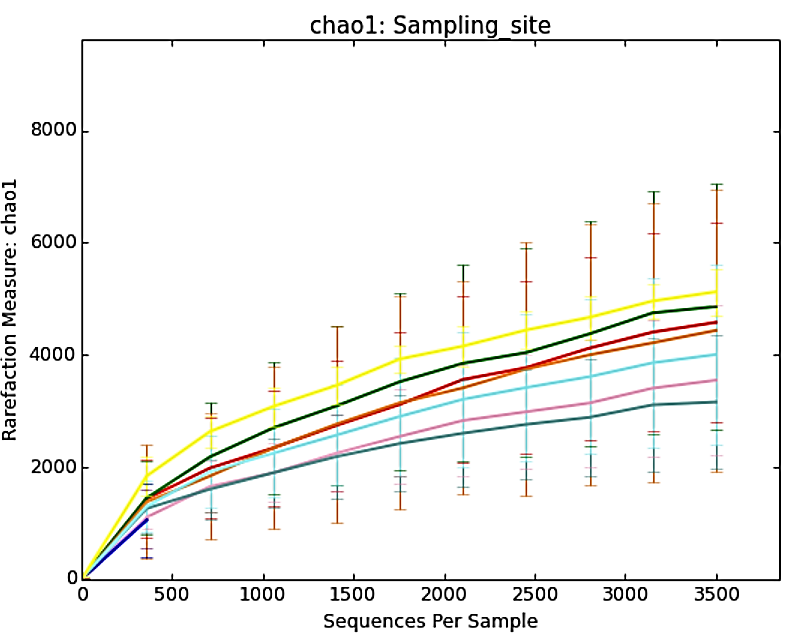


Figure 4: Major Phyla with Representative Classes.

**3.4 Alpha Diversity**

The Chao1 metric was used in determining alpha diversity (local species pool) in terms of species richness and species evenness. Richness in this case denote the number of species within a sample, while evenness is the relative abundance of each of the different species in a sampling area.

From the rarefaction curves above, the Chao1 metric measures diversity as a function of both species’ richness and evenness. The spring at Lianila (SP1w & SP1s) was the richest and most even in species abundance (Figure 5). This was followed by the river sampling site at the Kakamega sewage entry (S5w & S5s); the river sampling sites at Amalemba (S4w & S4s); the river samples at the forest (S1w & S1s); the spring at the resort (SP2w & SP2s); the spring water at Amalemba (SP3w & SP3s) and the spring water at the sewage (SP4w & SP4s). Overall, river water samples at the wastewater entry point and Amalemba point were richer in microbial communities compared to spring samples, except for the spring samples at confluence of Lianila stream and River Isiukhu which had the most abundant OTUs. The river sections at the forest were used as the reference site due to the minimal human activities.



KEY

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Forest |  | Confluence Lianila-Isiukhu |  | Resort |  | Amalemba |  | Kakamega Sewage Entry |
|  |  |  | Spring 1 Lianila |  | Spring 2 Resort |  | Spring 3 Amalemba |  | Spring 4 Sewage |

Figure 5: Rarefaction Curve of Chao1 against Sequences per Sample

**3.5 Beta Diversity**

Using the Beta diversity metric, the distances/ dissimilarity between pairs of samples from different sampling sites were compared. Areas having similar phylogenetic diversity were clustered together as shown in the PCoA plot (Figure 6). The same has also been presented in the Cladogram shown in figure 6 for easier visualization due to the many sampling sites. The colours in both figures represent the clusters of samples with similar phylogenetic diversity of microbial species. The taxa that would be more prevalent in the different areas of the Principal Coordinate Analysis (PCoA) plot, thus influencing the clustering was also investigated. The Bray Curtis index used to create a biplot shown in Figure 6:C. From the Figure 6: C, the four major taxa influencing the variation are the Betaproteobacteria, Cytophagia, Actinobacteria and Deltaproteobacteria.

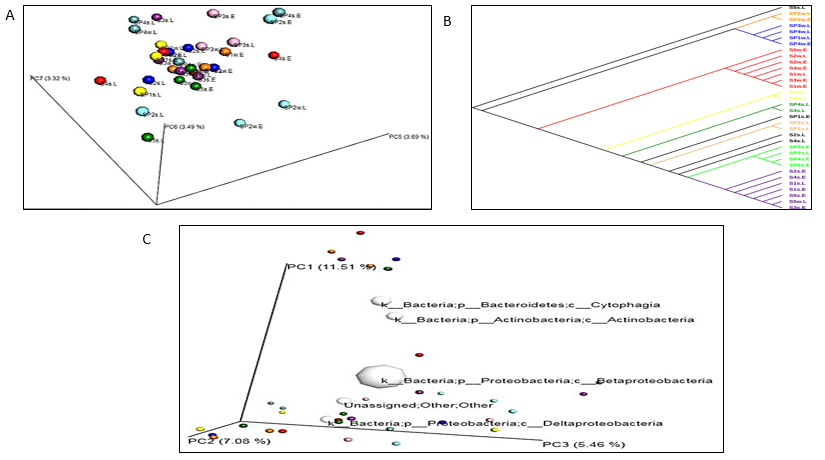


Figure 6: A ,B, shows principal Coordinates and Cladogram Showing analysis Plot and Clustering Based on Similarity and C; Biplot Showing the Major O TUs Influencing the Clustering of the Samples.

**4.0 Discussion**

Findings from this study reveal significantly abundant and diverse microbial assemblages at watering points and springs in the upper reaches of River Isiukhu in Kakamega County. The estimated richness of species (mean 18,846) per sample supersedes that which can be obtained using conventional culture methods. The detection of 1,549 OTUs belonging to thirty-nine microbial phyla facilitated an all-inclusive investigation of microbial communities in the different niches along the watering points and springs. Consistent with existing studies, this approach has revealed that sequencing of the V5-V6 of the 16S rRNA genes provides detailed information on the composition and abundance of bacterial communities (Azaroual et al., 2022). The finding showed that bacteria in River Isiukhu are dominated by common freshwater groups: Proteobacteria, Bacteroidetes, Firmicutes, Verrucomicrobia, Acidobacteria, Actinobacteria (Figure 2: A). This finding agrees with other studies that of most dominant heterotrophic bacterial groups in freshwater ecosystems belong to Bacteroidetes and Proteobacteria (Osterholz et al., 2018). However, it is important to note that comparable studies using high-throughput techniques to investigate the diversity of freshwater bacteria are scarce. Although the few that exist like that of Jordaan and Bezuidenhout carried out a similar study in the River Vaal, South Africa and also found that Proteobacteria (Beta, Alpha and gamma), Bacteroidetes and Actinobacteria were the dominant bacterial groups found in river water(Jordaan & Bezuidenhout, 2016). Also, in the River Mississippi, found that approximately 93% of (Illumina) reads belonged to Proteobacteria, Actinobacteria, Bacteroidetes, Cyanobacteria and Verrucomicrobia (Staley et al., 2016; Tang et al., 2020).

Amongst the four major phyla, Proteobacteria was the most abundant phyla in the springs and sampling water points. Faecal Coliforms are classified in phylum Proteobacteria and have been associated with contamination and a myriad of human diseases (Rizzatti et al., 2017). Five divisions of the Proteobacteria were identified with Betaproteobacteria, Deltaproteobacteria, Gammaproteobacteria and Alphaproteobacteria in abundance. This finding is consistent with that which was conducted in Lake Magadi in Kenya (Rizzatti et al., 2017).

The second most abundant phyla were Bacterioidetes with the major representative divisions being Sphingobacteria, Flavobacterium, Cytophaga, Incertiaesedis and Bacterioide. Both Proteobacteria and Bacterioidetes comprise a myriad of gram-negative bacteria that colonize the guts of warm-blooded animals including humans. Livestock were also observed to drink directly at the same water points, not provided with water troughs hence a potential source of coliform contamination to the waters.

The third most abundant phylum was the Firmicutes, which is usually most abundant in swine faecal matter as they are a major constitute of the gut microbiome in pigs. The major classes of Firmicutes observed during the study include the Clostridia, Bacilli, Negativicutes and Erysipelotricia. One of the possible sources of contamination of the river water with the phlyla Firmicutes could be attributable to the peri-urban activities such as disposal of household and municipal wastes consisting of solid and semi-solid wastes from markets and slaughter houses(Mwaikono et al., 2016).

**4.1 Microbial richness and diversity indices**

Rarefaction measures at a single sequence depth were conducted to compare microbial communities using alpha and beta diversity indices. River samples at the wastewater entry points and Amalemba exhibited a higher abundance of microbial communities compared to the forest area which serves as the control point. This observation is highly likely due to the influence of human activities in these areas. Amalemba area is highly populated with a myriad of activities including small-scale mixed agriculture, livestock rearing, laundering and even bathing in the river. The extensive use of manure, fertilizers and chemicals to manage agricultural and livestock pests including the poor drainage system in Amalemba facilitate the flourishing of diverse microbial communities. The wastewater entry points comprise an assortment of water from the municipal sewage treatment plant and water draining from the slaughterhouses in the area. Treated sewer water is safe for the environment as a majority of the contaminants are removed by biological, chemical and physical processes. However, treated sewer water is rich in organic matter and essential macronutrients (Wilkinson et al., 2022). Consistent with this study, treated sewer water had the highest amounts of total dissolved solids (TDS). Effluents from slaughterhouses were observed to mix with the treated water from the sewage treatment plant at the sewage entry point. Wastewater from slaughterhouses in Kakamega County and Kenya is rarely treated prior to disposal. Considering that this water is used to wash off animal carcasses and internal organs (Stets et al., 2014) . Wastewater from abattoirs are likely to be rich in microbiota that previously inhabited these animals (Jabari et al., 2016). Currently, no study has been published in Kenya that assessed the microbial assemblages in wastewaters from slaughter-houses. The mixing of nutrient rich, treated sewer water likely provides the necessary optimum nutritional and growth requirements for the microorganisms in the microbe-rich wastewater from the abattoirs. Hence, the mixing of waters from these two sources can be attributed to the extravagant microbial communities in the sewage entry point compared to the forest area which is devoid of heavy human encroachment (Douglas et al., 2021; Stets et al., 2014; Valverde et al., 2021).

All spring sampling points downstream were less diverse and had fewer species compared to the forest area and the river water except for the spring (SP1) at the confluence of River Isiukhu and Lianila stream which was upstream. There are two possible explanations for this observed phenomenon. First, at the confluence of river Isiukhu and Lianila stream, there is cross-contamination of the river and stream water thus explaining why the microbial load at this area was highest. Second, as the stream water continues to flow downstream, it undergoes through a series of natural filtration systems including underground rocks, gravel and sand, thus becoming pure in the process. Moreover, most sections of the streams were protected from humans.

5. Conclusion

Through the utilisation of 16s rRNA sequencing, this study revealed a wide range of archaeal and bacterial populations thus highlighting the limitations traditional culture-based methods have on this identification. The abundance and diversity of microbial communities identified in this study could be shaped by human activities and other environmental conditions. The use of 16s rRNA sequencing ensured that there is a broader identification of a broader spectrum of organisms in comparison with the conventional culture methods. Here, proteobacteria, Bacteroides and firmicutes were the dominant phyla. Anthropogenic factors could help explain the spatial variability in the composition of microbes seen here. This study begins to shed more light on the richness in freshwater microbiomes thus availing important data for monitoring and conservation of the environment.

Ethical approval

Apparently the microbial diversity study itself did not inherently require ethical approval from Masinde Muliro University of Science and Technology institutional review ethical committee (MMUST IREC), since no aspects of the research engaged specific environmental regulations.

Disclaimer (Artificial intelligence)

Option 1:

Author(s) hereby declare that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

Option 2:

Author(s) hereby declare that generative AI technologies such as Large Language Models, etc. have been used during the writing or editing of manuscripts. This explanation will include the name, version, model, and source of the generative AI technology and as well as all input prompts provided to the generative AI technology

Details of the AI usage are given below:

1.

2.

3.

References

Azaroual, s. E., kasmi, y., aasfar, a., el arroussi, h., zeroual, y., el kadiri, y., zrhidri, a., elfahime, e., sefiani, a., & meftah kadmiri, i. (2022). Investigation of bacterial diversity using 16s rrna sequencing and prediction of its functionalities in moroccan phosphate mine ecosystem. Scientific reports, 12(1), 3741. Https://doi.org/10.1038/s41598-022-07765-5

Bohorquez, l., delgado-serrano, l., lópez, g., osorio-forero, c., klepac-ceraj, v., kolter, r., junca, h., baena, s., & zambrano, m. (2011). In-depth characterization via complementing culture-independent approaches of the microbial community in an acidic hot spring of the colombian andes. Microbial ecology, 63, 103–115. Https://doi.org/10.1007/s00248-011-9943-3

Caporaso, j. G., kuczynski, j., stombaugh, j., bittinger, k., bushman, f. D., costello, e. K., fierer, n., peña, a. G., goodrich, j. K., gordon, j. I., huttley, g. A., kelley, s. T., knights, d., koenig, j. E., ley, r. E., lozupone, c. A., mcdonald, d., muegge, b. D., pirrung, m., reeder, j., sevinsky, j. R., turnbaugh, p. J., walters, w. A., widmann, j., yatsunenko, t., zaneveld, j., knight, r. Qiime allows analysis of high-throughput community sequencing data. Nat methods. 2010 may; 7(5):335-6. Doi: 10.1038/nmeth.f.303. Epub 2010 apr 11. Pubmed pmid: 20383131; pubmed central pmcid: pmc3156573.

Chaitankar, v., karakülah, g., ratnapriya, r., giuste, f., brooks, m., & swaroop, a. (2016). Next generation sequencing technology and genomewide data analysis: perspectives for retinal research. Progress in retinal and eye research, 55. Https://doi.org/10.1016/j.preteyeres.2016.06.001

Chowdhury, b., & garai, g. (2017). A review on multiple sequence alignment from the perspective of genetic algorithm. Genomics, 109(5–6), 419–431. Https://doi.org/10.1016/j.ygeno.2017.06.007

Douglas, m. M., lingappa, u. F., lamb, m. P., rowland, j. C., west, a. J., li, g., kemeny, p. C., chadwick, a. J., piliouras, a., schwenk, j., & fischer, w. W. (2021). Impact of river channel lateral migration on microbial communities across a discontinuous permafrost floodplain. Applied and environmental microbiology, 87(20), e0133921. Https://doi.org/10.1128/aem.01339-21

Head, i. M., & gray, n. D. (2016). Microbial biotechnology 2020; microbiology of fossil fuel resources. Microbial biotechnology, 9(5), 626–634. Https://doi.org/10.1111/1751-7915.12396

Herrera, a., & cockell, c. S. (2007). Exploring microbial diversity in volcanic environments: a review of methods in dna extraction. Journal of microbiological methods, 70(1), 1–12. Https://doi.org/10.1016/j.mimet.2007.04.005

Huang, l., bai, j., wen, x., zhang, g., zhang, c., cui, b., & liu, x. (2020). Microbial resistance and resilience in response to environmental changes under the higher intensity of human activities than global average level. Global change biology. Https://doi.org/10.1111/gcb.14995

Jabari, l., gannoun, h., khelifi, e., cayol, j.-l., godon, j.-j., hamdi, m., & fardeau, m.-l. (2016). Bacterial ecology of abattoir wastewater treated by an anaerobic digestor. Brazilian journal of microbiology, 47(1), 73–84. Https://doi.org/10.1016/j.bjm.2015.11.029

Jordaan, k., & bezuidenhout, c. (2016). Bacterial community composition of an urban river in the north west province, south africa, in relation to physico-chemical water quality. Environmental science and pollution research, 23. Https://doi.org/10.1007/s11356-015-5786-7

Kim, j. H., feng, z., bauer, j. D., kallifidas, d., calle, p. Y., & brady, s. F. (2010). Cloning large natural product gene clusters from the environment: piecing environmental dna gene clusters back together with tar. Biopolymers, 93(9), 833–844. Https://doi.org/10.1002/bip.21450

Kphc. (2009). Kenya population and housing census volume i: population by county and sub-county. Kenya national bureau of statistics. Https://www.knbs.or.ke/download/2019-kenya-population-and-housing-census-volume-i-population-by-county-and-sub-county/

Kuczynski, j., stombaugh, j., walters, w. A., gonzález, a., caporaso, j. G., & knight, r. (2011). Using qiime to analyze 16s rrna gene sequences from microbial communities. Current protocols in bioinformatics, chapter 10, unit 10.7. Https://doi.org/10.1002/0471250953.bi1007s36

Kumar, n., gupta, a. K., sudan, s. K., pal, d., randhawa, v., sahni, g., mayilraj, s., & kumar, m. (2021). Abundance and diversity of phages, microbial taxa, and antibiotic resistance genes in the sediments of the river ganges through metagenomic approach. Microbial drug resistance (larchmont, n.y.), 27(10), 1336–1354. Https://doi.org/10.1089/mdr.2020.0431

Lim, e. S., kim, j. J., sul, w. J., kim, j. S., kim, b., kim, h., & koo, o. K. (2021). Metagenomic analysis of microbial composition revealed cross-contamination pathway of bacteria at a foodservice facility. Frontiers in microbiology, 12, 636329

Liu, l., yang, j., yu, x., chen, g., & yu, z. (2013). Patterns in the composition of microbial communities from a subtropical river: effects of environmental, spatial and temporal factors. Plos one, 8(11), e81232. Https://doi.org/10.1371/journal.pone.0081232

Lozupone, c. A., hamady, m., kelley, s. T., and knight, r. (2007). Quantitative and qualitative diversity measures lead to different insights into factors that structure microbial communities. American society for microbiology, 73(5), 1576–1585. Https://doi.org/10.1128/aem.01996-06

Mäki, a., rissanen, a. J., & tiirola, m. (2016). A practical method for barcoding and size-trimming pcr templates for amplicon sequencing. Biotechniques, 60(2), 88–90. Https://doi.org/10.2144/000114380

Mwaikono, k. S., maina, s., sebastian, a., schilling, m., kapur, v., & gwakisa, p. (2016). High-throughput sequencing of 16s rrna gene reveals substantial bacterial diversity on the municipal dumpsite. Bmc microbiology, 16(1), 145. Https://doi.org/10.1186/s12866-016-0758-8

Onyango, d., & angienda, p. O. (2010). Epidemiology of waterborne diarrhoeal diseases among children aged 6-36 months old in busia—western kenya. World academy of science, engineering and technology, 37, 1158–1165.

Oremo, j., orata, f., owino, j., & shivoga, w. (2020). Assessment of available phosphates and nitrates levels in water and sediments of river isiukhu, kenya. Applied ecology and environmental sciences. Https://doi.org/10.12691/aees-8-3-7

Osterholz, h., kirchman, d. L., niggemann, j., & dittmar, t. (2018). Diversity of bacterial communities and dissolved organic matter in a temperate estuary. Fems microbiology ecology, 94(8). Https://doi.org/10.1093/femsec/fiy119

Parks, m., kedy, c., & skalla, c. (2020). Consistent patterns in 16s and 18s microbial diversity from the shells of the common and widespread red-eared slider turtle (trachemys scripta). Plos one, 15(12), e0244489. Https://doi.org/10.1371/journal.pone.0244489

Paruch, l., paruch, a. M., eiken, h. G., & sørheim, r. (2019). Faecal pollution affects abundance and diversity of aquatic microbial community in anthropo-zoogenically influenced lotic ecosystems. Scientific reports, 9(1), 19469. Https://doi.org/10.1038/s41598-019-56058-x

Qu, x., yeung, c., coleman, i., nelson, p. S., & fang, m. (2020). Comparison of four next generation sequencing platforms for fusion detection: oncomine by thermofisher, ampliseq by illumina, fusionplex by archerdx, and qiaseq by qiagen. Cancer genetics, 243, 11–18. Https://doi.org/10.1016/j.cancergen.2020.02.007

Rajendhran, j., & gunasekaran, p. (2011). Microbial phylogeny and diversity: small subunit ribosomal rna sequence analysis and beyond. Microbiological research, 166(2), 99–110. Https://doi.org/10.1016/j.micres.2010.02.003

Rideout, j. R., he, y., navas-molina, j. A., walters, w. A., ursell, l. K., gibbons, s. M., chase, j., mcdonald, d., gonzalez, a., robbins-pianka, a., clemente, j. C., gilbert, j. A., huse, s. M., zhou, h., knight, r., and caporaso, j. G. (2014) subsampled open-reference clustering creates consistent, comprehensive otu definitions and scales to billions of sequences. Peer journal, 2:e545 https://doi.org/10.7717/peerj.545

Rizzatti, g., lopetuso, l. R., gibiino, g., binda, c., & gasbarrini, a. (2017). Proteobacteria: a common factor in human diseases. Biomed research international, 2017, 9351507. Https://doi.org/10.1155/2017/9351507

Sánchez, á., vila, j. C. C., chang, c.-y., diaz-colunga, j., estrela, s., & rebolleda-gomez, m. (2021). Directed evolution of microbial communities. Annual review of biophysics, 50, 323–341. Https://doi.org/10.1146/annurev-biophys-101220-072829

Staley, d., negri, j., kean, j., laber, j., tillery, a., & youberg, a. (2016). Prediction of spatially explicit rainfall intensity-duration thresholds for post-fire debris-flow generation in the western united states. Geomorphology, 278. Https://doi.org/10.1016/j.geomorph.2016.10.019

Stets, m. I., etto, r., galvão, c., ayub, r., cruz, l., steffens, m., & barana, a. (2014). Microbial community and performance of slaughterhouse wastewater treatment filters. Genetics and molecular research : gmr, 13. Https://doi.org/10.4238/2014.june.16.3

Tang, h., mcguire, l. A., kean, j. W., & smith, j. B. (2020). The impact of sediment supply on the initiation and magnitude of runoff-generated debris flows. Geophysical research letters, 47(14), e2020gl087643. Https://doi.org/10.1029/2020gl087643

Valverde, a., cason, e. D., gómez-arias, a., bozkale, d., govender, d., riddell, e., & cowan, d. (2021). Pollution shapes the microbial communities in river water and sediments from the olifants river catchment, south africa. Archives of microbiology, 203(1), 295–303. Https://doi.org/10.1007/s00203-020-02035-2

Walters, k. E., & martiny, j. B. H. (2020). Alpha-, beta-, and gamma-diversity of bacteria varies across habitats. Plos one, 15(9), e0233872. Https://doi.org/10.1371/journal.pone.0233872

Wee, s. Y., edgcomb, v. P., beaudoin, d., yvon-lewis, s., & sylvan, j. B. (2021). Microbial abundance and diversity in subsurface lower oceanic crust at atlantis bank, southwest indian ridge. Applied and environmental microbiology, 87(22), e0151921. Https://doi.org/10.1128/aem.01519-21

Wilkinson, j. L., boxall, a. B. A., kolpin, d. W., leung, k. M. Y., lai, r. W. S., galbán-malagón, c., adell, a. D., mondon, j., metian, m., marchant, r. A., bouzas-monroy, a., cuni-sanchez, a., coors, a., carriquiriborde, p., rojo, m., gordon, c., cara, m., moermond, m., luarte, t., teta, c. (2022). Pharmaceutical pollution of the world’s rivers. Proceedings of the national academy of sciences of the united states of america, 119(8), e2113947119.

Zamorska, J., Karwowska, E., & Przystaś, W. (2023). Assessment of Microbiological Quality of Water Using Culture Methods, Flow Cytometry and Luminometry. Water, 15(23), 4077.

Kalsoom, M.; Rehman, F.; Shafique, T.; Junaid, S.; Khalid, N.; Adnan, M.; Zafar, I.; Tariq, M.A.; Raza, M.A.; Zahra, A.; et al. Biological importance of microbes in agriculture, food and pharmaceutical industry: A review. Innovare J. Life Sci. 2020, 8, 1–4.

M. M. Dalmacio, L., B. L. Ramirez, R. Estacio, I. Borlongan, J. M. Ramirez, K. V. Evangelista, E. Madlangbayan, F. Guillergan, and M. A. Kron. 2020. “Microbial Diversity Assessment in Milkfish Culture Ponds”. Asian Journal of Fisheries and Aquatic Research 5 (3):1-12. https://doi.org/10.9734/ajfar/2019/v5i330077.