COMPARATIVE METAGENOMIC PROFILING OF MICROBIAL COMMUNITIES IN HUMAN-IMPACTED AND PRISTINE STREAM WATER IN OLUGBADE VILLAGE, ISEYIN, OYO STATE

Olowe Busayo Mutiat* and Raheem Olaniyi

Address(es):
*Department of Biological Sciences, Microbiology Unit, School of Pure and Applied Sciences, College of Science, Bamidele Olumilua University of Education, Science and Technology, P. M. B. 250, Ikere, Ekiti State, Nigeria

ABSTRACT

Stream water, a major water source in rural communities, being impacted by human activities results to disturbance of the microbial ecosystem balance, pollution and impacts community health depending on such water. The study aimed to compare the metagenomic profiling of microbial communities in human-impacted and pristine stream water in Olugbade village, Oyo State, Nigeria. Human-impacted (sample A) and pristine (sample B) areas of the Stream were sampled following standard collection method. DNA extraction from the samples was performed using CTAB method and estimated through the Qubit ds DNA high sensitivity kit. Samples were sequenced on the Sequel system by PacBio (www.pacb.com). Raw subreads were processed through the SMRTLink (v9.0) Circular Consensus Sequences (CCS) algorithm to produce highly accurate reads (>QV40). These highly accurate reads were then processed through vsearch (https://github.com/torognes/vsearch) and taxonomic information was determined based on Qualitative Insights Into Microbial Ecology (QIMME2) version 2018.6.0. Results showed that organisms identified were 100% bacteria where sample A had 8,198 reads and B had 9,827 reads. The taxonomic distribution revealed 23 and 43 phyla, 52 and 109 classes, 91 and 170 orders, 108 and 212 families, 211 and 336 genus and 277 and 455 species for sample A and B respectively. The study revealed that there is a statistical difference between human-impacted and pristine samples (p-value < 0.05). Sample B had more bacterial community than sample A thereby, showing that many of the bacterial naturally occurring in sample A are extinct or displaced due to different anthropogenic activities occurring there.

Keywords: Pristine Stream water, Human-impacted Stream water, DNA, 16S Metagenomics, Taxonomic analysis

INTRODUCTION

Stream water source has ever been an indispensable natural resource on which rural communities depends. It is worthy of note that this source of water is prone to contamination or pollution. Pollution or contamination of surface water bodies by human activities constitute a great health concerns and threat to the ecosystem balance (Anh et al., 2021; Rosca et al., 2020; Neamtu et al., 2021; Zhang et al., 2021). These pollutants or contaminants released into water bodies are responsible for various health challenges most especially water borne infections which is one of the leading cause of deaths in every communities especially in sub-urban and rural environments (Afroz et al., 2014; Tranakakis et al., 2020). Another challenge of the impact of human activities on streams is the displacement of beneficial microbes, present in non-disturbed environment, which can help in degradation of pollutants in the environment; contributing to the economic development of a nation and maintaining balance in the ecosystem. These call for a great concern.

Therefore, sources of water must be monitored regularly to determine whether they are in sound health or not. To monitor or investigate the microbial quality of drinking water, a wide range of methods are used, which include culture-dependent and culture-independent methods (He et al., 2022). Culture-based methods are not only time consuming and laborious, but also, they are proved to have limitation as they under-represent the microbial community of any sample. With this procedure only a few bacterial isolates can be studied at a time, whereas, culture-independent methods can be present in such water bodies and especially the effluent-receiving waterbodies (Kori et al., 2019). Another problem, associated with culture-based systems, is the culturing time, which may take from 1 to 2 days for fast-growing bacteria to several weeks for slow-growing species (Chakraborty et al., 2022). Whereas, culture-independent methods, which mainly came into practice after the advent of next-generation sequencing (NGS) and metagenomics techniques have the ability to deeply sight microbial communities present in any environmental sample (Dubey et al., 2022). The outbreak of next-generation sequencing circumvents these limitations because it is a culture-/amplification independent technique. Therefore, these technologies allow for a deeper insight into the genomic information of most bacteria, leading to the detection of unculturable microorganisms (Nogueira & Botelho, 2021). This metagenomic approach also provides information regarding the presence, absence, or modification of the genes responsible for different functions in the ecosystem. Metagenomics has emerged in the last decade as a promising centrepiece that bookends attempts to analyse the multiple genomes contained within a microbial niche or biome (Dubey et al., 2022). Thus, instead of collecting live microorganisms from a microbial community to be cultured or observed in the laboratory, the isolation of DNA directly from a sample can provide information related to the diversity of the microorganisms thriving in certain areas and can inclusively reveal information related to their functions and biological roles. Until recently, the classic bacteriological culture was the standard procedure to study bacterial communities, but with the advent of genomic approaches, namely metagenomics, a new window to study the world of microorganisms has opened. Several metagenomic studies have been reported on microbial communities of rivers, streams, sediments and water treatment plants in many countries (Ibeke et al., 2016; Abia et al., 2018; Reddy & Dubey, 2019; Parida et al., 2022; Obieze et al., 2022). But there is little or no information of comparative assessment of microbial communities of pristine and human impacted water bodies in Nigeria. And since water quality is a function of any community health, the study was predicated on concerns for the quality of stream water available to Olugbade villagers at Iseyin, Oyo state, Nigeria. Studying this stream will provide understanding of the microbial ecology of the stream available for consumption through metagenomic analysis which is important in characterizing microbial pathogens and remediation in case of contamination/infection. Also, metagenomics profiling of the stream will reveal many unculturable bacteria which are potential causative agents of many waterborne infection. Moreover, the organisms that will be detected, which is a function of source, will help in formulating policies that will help in protecting the quality of the stream. The stream water available in Olugbade village is the major source of water available for drinking, domestic and farming purposes. The community preserved the upper part of the stream exclusively for drinking while the downstream for other activities/purposes. These two different points of stream are far apart. The extent of human activities on the microbial ecology of the downstream remained unknown. And if findings on this is not carried out, the community may not be able to safeguard the stream healthiness and also the community. This is because the health of the community is a function of the stream healthiness. Therefore, this study aimed to study and compare the microbial communities of both the pristine stream and human-impacted stream in order to determine the extent of human activities on Olugbade Stream.
**MATERIAL AND METHODS**

**Sample Area**

The study area considered for this study was Olugbade Stream consumed and used by residents of Olugbade village, Iseyin, Oyo State. The stream is located on the coordinate of 7°57′35″N and 3°34′18″E approximately 100 km North of Ibadan, Oyo State, Nigeria. The residents of the community preserved the upper part of the stream (7°57′30″N and 3°44′23″E) for consumption and as such referred to it as a pristine stream water. In this area, the villagers do not allow any activities such as swimming, washing, farming and all sorts of domestic or farm activities to take place. Downstream the pristine, is the human impacted area (7°57′35″N and 3°34′18″E) where all sort of activities occurs. This stream has the capacity to supply and meet the daily needs of the residents.

**Sample Collection**

Stream water samples were collected at two points of Olugbade village streams; the pristine (Sample B) and the human-impacted points (Sample A). A 5000 ml of replicate samples were collected from each point into a sterile container and transported on ice to Inqaba Biotec laboratory, Ibadan for analyses.

**DNA Extraction and Estimation**

The extraction of total DNA was carried out following the CTAB DNA extraction technique employed by Kori et al. (2019). A 500 ml of the stream water samples was filtered through sterile 0.45μm cellulose acetate+ CN gridded filter papers. Filters were transferred to the sterile 50 ml falcon tubes and were stored at −4°C until further process. DNA extraction was performed using CTAB method (Nishiguchi et al., 2002) and the DNA was estimated through the Qubit ds DNA high sensitivity kit (Invitrogen Inc., USA).

**Amplipac Library Preparation and Illumina Sequencing**

For each sample, an Illumina library was prepared from total DNA using the TruSeq kit v2 (Illumina, Inc., San Diego, CA, USA) following the manufacturer's specifications with an average fragment size of 500 bp. The sequencing was performed on the NextSeq500 (Illumina, Inc., San Diego, CA, USA) platform with a 150-cycle configuration, generating paired-end reads with a length of 75 bp.

**Bioinformatics and Statistical Analyses**

Samples were sequenced on the Sequel system by PacBio (www.pacb.com). Raw subreads were processed through the SMARTTrinQ (v9.0) Circular Consensus Sequence (CCS) algorithm to produce highly accurate reads (>QV40). These highly accurate reads were then processed through versace (https://github.com/toroges/vsearch) and taxonomic information was determined based on Qualitative Insights Into Microbial Ecology (QIMME2) version 2018.6.0 (https://qiime2.org/). Report generation command used was: $create_vsearch_single_sample_pdf_report_pacbio.py create_vsearch_single_sample_pdf_results_pacbio.py A-M13_bc1002_F--M13_bc1049_R withheld_sreads_otu_table.tsv. The hypothesis formulated to compare the difference between the human-impacted and pristine water sample was tested on the Phylum classification of Sample A and B using regression analysis since other downstream taxonomic classifications emerged from the phylum.

**RESULTS**

From the two samples, a total of 100% kingdom bacteria was detected and analysed with sample A giving a total of good quality reads of 8,198 and sample B, reads of 9,827 which collectively gives reads of 18,025 for the two samples. It is observed that sample B had a higher read than sample A (Table 1).

<table>
<thead>
<tr>
<th>Kingdom</th>
<th>Read count</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample A</td>
<td>Bacteria</td>
<td>8198.0</td>
</tr>
<tr>
<td>Sample B</td>
<td>Bacteria</td>
<td>9827.0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>18,025.0</td>
</tr>
</tbody>
</table>

The bacterial taxonomic classification of the bacterial communities of the samples from human impacted (sample A) and pristine (sample B) is depicted in Figure 1.

![Figure 1](https://example.com/figure1.png)  
**Figure 1** Taxonomic distribution of bacterial community of human streams of Olugbade village impacted and pristine

The figure reveals that sample B comprised more bacterial communities at all level of taxonomic classifications than sample A. (Figure 1). The similarity search of 18,025 reads resulted in a similarity with 23 bacterial phyla. Proteobacteria in the two samples was found to be most abundant with 69.58 % in sample A and 55.64% in sample B. From Figure 2, it is observed that Sample B had 21 more phyla which are not present in sample A. Also, the figure depicts several phyla with varying read counts. Furthermore, about 0.30% of the bacteria were unknown in sample A and 1.27% in Sample B. Figure 3 and 4 show the percentage of occurrence of bacterial phylum in sample A and B. In Figure 3, the phyla Proteobacteria, Bacteroidetes, Verrucomicrobia were the phyla groups with >5% occurring in sample A, while Proteobacteria, Actinobacteria, Bacteroidetes and Planctomycetes were phyla with >5% in sample B (Figure 4).

At the class level, the 18,025 reads produced varied bacterial classes with different percentage composition (Figure 5). From among the classes identified, 50 classes were common to both sample A and sample B. The abundant classes of bacterial in the two samples included Betaproteobacteria, Alphaaluproteobacteria, Gammaproteobacteria, Verrucomicrobia, Actinobacteria, Bacilli and cytophagia with >2.00%. The percentage distributions of the classes in sample A and B are shown in Figure 6 and 7.

In the two samples, different numerous types of orders were detected, however, due to the low abundance of some taxonomic order groups, they were grouped into “others” group. An attention was focused mainly on the abundant bacterial order in the samples. The top order classification identified in sample A included Burkholderiales (42.78%), Sphingomonadales (5.33%), Pseudomonadales (4.77%), Verrucomicrobiiales (4.14%), Actinomycetales (3.92%), Cytophagales (3.12%) and others (32.57%) (Figure 8). While in sample B, the “others” group (46.10%) represented the largest percentage (Figure 9). Some order classification in sample B included Burkholderiales (23.21%), Actinomycetales (3.75%), Rhizobiales (3.53%), Myxococcales (3.02%), Sphingomonadales (2.62%) (Figure 9).

The family classification showed 108 and 212 diverse families in sample A and B respectively. However, due to the low percentages of several families present in the samples, attention was focused on the families with higher percentages. These included the Comamonadaceae (38.62%), unknown (10.39%), Sphingomonadaceae (4.48%), Verrucomicrobiaceae (4.14%), Moraxellaceae (3.92%), Streptococccaceae (2.68%), Gemmatimonadaceae (2.5%) and others (33.28%) in Sample A (Figure 10). In sample B, the families on which attention was focused included Comamonadaceae (21.79%), Sphingomonadaceae (2.32%), Rhodocyclaceae (1.98%), Cytophagaceae (1.93%), Moraxellaceae (1.68%), Chitinophagaceae (1.51%), Unknown (34.63%) and others (34.16%) (Figure 11). Comparing Figures 10 and 11, it was observed that Comamonadaceae, Sphingomonadaceae and Moraxellaceae are more in percentage in sample A than Sample B while the Unknown genus in B are more than the Unknown in Sample A.

The genus classification revealed 211 and 336 several general in sample A and B respectively. The top genus classification is presented in Figure 12 and 13 for sample A and B respectively. The Unknown in Sample B (63.71%) are more than Unknown (34.63%) in Sample A. Another striking difference observed in both samples is that, some genera in top percentages in sample A are not represented in sample B. The specie classification results gave a total of 277 and 455 species in sample A and B respectively. The unknown species in sample B (Figure 15) were...
more than in sample A (Figure 1). It was also observed that some microorganisms identified in sample A indicated human impact on the water source.

**Figure 2** Relative abundances of identified bacterial phyla in samples A and B

**Figure 3** Percentage phyla distribution in sample A
Figure 4 Percentage phyla distribution in sample B

Figure 5 Relative abundances of identified bacterial class in samples A and B
**Figure 6** Top Percentage distribution of bacterial classes in Sample A

- Betaproteobacteria (47.50%)
- Alphaproteobacteria (10.75%)
- Gammaproteobacteria (9.89%)
- Verrucomicrobiae (4.14%)
- Actinobacteria (3.93%)
- Bacilli (3.31%)
- Cytophagia (3.12%)
- Other (17.37%)

**Figure 7** Top Percentage distribution of bacterial classes in Sample B

- Alphaproteobacteria (9.69%)
- Gammaproteobacteria (6.88%)
- Betaproteobacteria (31.83%)
- Deltaproteobacteria (6.21%)
- Planctomycetia (4.00%)
- Actinobacteria (3.83%)
- Other (31.58%)

**Figure 8** Top Percentage distribution of bacterial order in Sample A

- Burkholderiales (42.78%)
- Sphingomonadales (5.33%)
- Pseudomonadales (4.77%)
- Verrucomicrobiales (4.14%)
- Actinomycetales (3.92%)
- (3.38%)
- Cytophagales (3.12%)
- Other (32.57%)
**Figure 9** Top Percentage distribution of bacterial order in Sample B

- **Burkholderiales (23.21%)**
- **Actinomycetales (3.75%)**
- **Rhizobiales (3.53%)**
- **Myxococcales (3.02%)**
- **Sphingomonadales (2.62%)**
- **Unknown (7.38%)**
- **Other (46.10%)**

**Figure 10** Top Family classification of sample A

- **Comamonadaceae (38.62%)**
- **Sphingomonadaceae (4.48%)**
- **Verrucomicrobiaceae (4.14%)**
- **Moraxellaceae (3.92%)**
- **Streptococcaceae (2.68%)**
- **Gemmata monadaceae (2.50%)**
- **Other (33.28%)**

**Figure 11** Top Family classification of sample B

- **Unknown (34.63%)**
- **Sphingomonadaceae (2.32%)**
- **Rhodocyclaceae (1.98%)**
- **Cytophagaceae (1.93%)**
- **Moraxellaceae (1.68%)**
- **Chitinophagaceae (1.51%)**
- **Other (34.16%)**
**Figure 12** Top genus classification in sample A

**Figure 13** Top genus classification in sample B

**Figure 14** Top species classification in sample A
Table 2 Regression analysis showing a significance difference between the human-impacted and pristine water samples from Olugbade Village, Iseyin, Oyo State

<table>
<thead>
<tr>
<th>Regression Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.989103</td>
</tr>
<tr>
<td>R Square</td>
<td>0.978325</td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.977796</td>
</tr>
<tr>
<td>Standard Error</td>
<td>130.1787</td>
</tr>
<tr>
<td>Observations</td>
<td>43</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
<td>SS</td>
<td>MS</td>
<td>F</td>
<td>Significance F</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
<td>----------------</td>
<td>--------</td>
</tr>
<tr>
<td>Regression</td>
<td>1</td>
<td>31360713</td>
<td>31360713</td>
<td>1850.571</td>
<td>9.65E-36</td>
</tr>
<tr>
<td>Residual</td>
<td>41</td>
<td>694806.6</td>
<td>16946.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>32055520</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Lower 95.0%</th>
<th>Upper 95.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-44.1534</td>
<td>-2.14454</td>
<td>0.037966</td>
<td>-85.7333</td>
<td>-2.57359</td>
<td>-85.7333</td>
<td>-2.57359</td>
</tr>
<tr>
<td>X Variable 1</td>
<td>1.027434</td>
<td>43.01827</td>
<td>9.65E-36</td>
<td>0.9792</td>
<td>1.075668</td>
<td>0.9792</td>
<td>1.075668</td>
</tr>
</tbody>
</table>

Table 2: Regression analysis showing a significance difference between the human-impacted and pristine water samples from Olugbade Village, Iseyin, Oyo State

Test of Research Hypothesis

Null Hypothesis (H₀): There is no significant difference between the human-impacted and pristine water sample
Alternative Hypothesis (H₁): There is a significant difference between the human-impacted and pristine water sample

Table 2 reveals that there is a significant difference between the human-impacted and pristine water sample (p<0.037). Therefore, the null hypothesis (H₀) is rejected while the alternate hypothesis (H₁) is accepted. This implies that there is a statistically significant effect of human activities on microbial community of the stream. However, a relationship, in terms of the types of microorganisms in the stream samples is high, positive and statistically significant at 0.05 level (r=0.989, p<0.037).

DISCUSSION

This study describes a metagenomic analysis of water samples collected from human impacted and pristine points of the stream consumed or used by Olugbade villagers in Iseyin, Oyo State. This study was informed as a result of the concern for water quality and safety in the village. Metagenomic analysis of total DNA recovered from the filtered stream water samples indicated different taxonomical distribution of typical freshwater microorganisms which include both culturable and non-culturable microorganisms. It is worthy of note that detection of both culturable and non-culturable microorganisms is essential to determine the quality and safety of any water body meant for consumption in any community. And in any water source, most bacterial populations in these environments are viable but not culturable (VBNC) using conventional standard bacteriological culture methods (Ding et al., 2017). Several naturally occurring microorganisms with beneficial potentials and potentially pathogenic bacterial species have been shown to enter the VBNC state in response to environmental stress which reduces their detection in the environment (Hamner et al., 2019). Finding from the study showed the detection and enumeration of different bacteria from both samples. Among which are bacteria which are of medical, environmental, economical and pharmaceutical importance. Comparing both samples, it was found that bacteria species detected and enumerated from sample A (Human-impacted) are also found in sample B (pristine stream) but in a reduced percentage composition. This might possibly be as a result of anthropogenic activities carried out in that water source. Also, many microorganisms implicated with human infection were detected in the water samples from the human-impacted stream (sample A). Examples of such included: Acidovorax sp responsible for Sepsis (Malkan, et al., 2009); catheter-associated bloodstream infection (Miller et al., 2010); Acinetobacter spp which has been implicated in

Figure 15: Top species classification in sample B
Cyanobacteria was another detected important phylum, though more abundant in sample B than in A. This group of bacteria is important as they produce toxins in aquatic environment (Valério et al., 2010). Some species of this phylum have been reported to produce microcystins, cytotoxins, and neurotoxins as secondary metabolites (Harada, 2004; Wiegand & Pflegmacher, 2005). These toxins are significant for economic losses. This study shows that the release of contaminants or pollutants, which is one of the main environmental consequences of anthropogenic activities, alters the biological characteristics of the stream environments, leading to modifications of the microbial community of the stream.

CONCLUSION

The study reflects the true nature of the microbial community of the stream at the moment of sampling which is a pointer to the safety of Olugbade stream. Comparatively, the study revealed a statistically significant difference between human impacted stream and pristine stream. This is substantiated with the findings that members of human and animal microbiota were overrepresented in the human-impacted streams. In addition, it was revealed that microorganisms of ecological, economical and medical importance were relatively low in percentage composition as compared with the pristine stream. All these findings are evident that are likely to affect the human activities on the streams. Thereby, threatening the microbial community of aquatic ecosystem which is extremely important for the maintenance and sustainability of this environment, as well as making the stream unfit for consumption. Hence, underscoring the need to adopt measures aimed at reducing impacts imposed by anthropogenic or human activities.

It is therefore recommended that frontier ethic, which is completely anthropocentric (human-centered) should be discarded because only the needs of humans are considered while using the streams. Rather a sustainable ethic, which is an environmental ethic by which people treat the earth as if its resources are limited, should be encouraged in the use of the stream. This ethic assures that the earth’s resources are not unlimited and that humans must use and conserve resources in a manner that allows their continued use in the future and that we suffer when the health of a natural ecosystem is impaired.

REFERENCES


