Metagenomics analysis ACFC Wastewater

Floris Menninga

2025-03-15

Table of Contents

library(tidyverse)

15-03-2025:

## **Introduction**

In the wastewater of a Kenyan water treatment plant, there can be many organisms whose presence influences the water quality by their production of toxins like those produced by algal blooms from Cyanobacteria. **1** These algal blooms are a global threat to freshwater systems like the Nyando River. Using more traditional chemical analyses, an abundance of nitrogen, phosphorus and potassium was determined.

It is imperative to measure those toxins or in this case genetic material from the micro-organisms that produce them. This wastewater is directly exhausted into the Nyando River, a basin covering about 3,590 square kilometers. The population density at the basin is higher than the national average, this indicates it’s importance. The life expectancy is very low with an average of 37.7 years for males and 42.9 years for females. Enhancing water quality could contribute to a higher lifespan for the people and other organisms that make use of these wetlands.

This water provides food, stores energy and is crucial for biodiversity. **2** These resources are threatened by wastewater from factories and treatment plants.

In the wastewater treatment facility is among other process steps a digester, with a lagoon where the wastewater is expelled into. The influence of this digester on the microbial diversity and number will be determined.

There are a total of three samples, the first one was taken before the water enters the digester, the second inside of the digester and the last one in the lagoon before entering the Nyando River. We will make a comparison between these samples and we will determine if the amount of bacteria or other micro-organisms exceeds the limits for reclaimed water before and after entering the digester. Toxin producing algae can be present in the sample that exited the digester only if this was also the case before entering it. The scope of this article is limited to determining the influence of the digester in the wastewater treatment facility. So even if there already are contaminants, as long as their numbers are not increasing, the treatment process is not at fault.

Metagenomics is DNA based and can provide information about what organisms are present in the sample, this can be taxonomic and phylogenetic information. **4** Metagenomics differs from whole genome sequencing which refers to the sequencing of a single genomes.

To do this, a bio-informatics pipeline was constructed to compare these samples in which a taxonomic classification will be applied on the samples using the Kraken2 tool to check the presence of known algal bloom causing Cyanobacteria. **1** Given the results from incubating the samples on agar plates, we hypothesize that there are toxin producing Cyanobacteria in the samples but that the digester doesn’t make a difference for the diversity and number of those organisms.

In addition to the phylogenetic classification, there will also be an anti-biotic resistance test to determine to what antibiotic resistance genes the micro-organisms have. This will provide an overview of means to combat the bacteria found in the sample more effectively. **7**

To first assemble the microbial genomes from the metagenomic dataset, several tools can be used like Metabat, Concoct and Maxbin4. Identifying potentially up to 2000 micro-organisms in the sample can be challenging given that their average genome size would be 4 Mpb 8 Gpb of reads would have to be obtained to get an average coverage of 1x **5**. The diversity of the microbiome will be quantified using the alpha diversity measure. This is a formula that can be applied to calculate this measure for diversity. **6**.

Biosynthetic gene clusters (BGCs) will also be detected by mapping the reads from the Oxford Nanopore sequencer to the MiBIG database. **3** The results from that analysis will show the biosynthetic potential of the organisms to produce toxins.

# **Pipeline**

The different steps of the pipeline are explained below.

### **1. Quality control**

To check the quality of the raw sequence data, Fastplong was used. This tool uses

**1.1 Results**

### **2. Taxonomic classification**

## **Citations**

**To-do: use .bib file for this…**

1. <https://www.zotero.org/groups/5870913/acfc_afvalwater/items/WLVUDDXJ/attachment/RHY85NTI/reader>
2. <https://www.zotero.org/groups/5870913/acfc_afvalwater/items/DW2JTKTX/reader>
3. biosynthetic gene clusters. Nucleic Acids Res 51:D603–D610. <https://> doi.org/10.1093/nar/gkac1049
4. <https://www.zotero.org/groups/5870913/acfc_afvalwater/items/F3R5L7PV/attachment/6YZYPRXC/reader>
5. is the treated wastewater safe to reuse for agricultural irrigation? Water Res 73:277–290. <https://doi> .org/10.1016/
6. Best practices for analysing microbiomes. Nat Rev Microbiol 16:410 – 422. <https://doi.org/10.1038/s41579-018-0029-9>.
7. Platforms for elucidating antibiotic resistance in single genomes and complex metagenomes. Environ Int 138:105667. <https://doi.org/10.1016/j.envint.2020.105667>.

Fastplong Shifu Chen. 2023. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta 2: e107. <https://doi.org/10.1002/imt2.107>

Taxonomic\_classifier\_tool\_comparison Ye, S. H., Siddle, K. J., Park, D. J., & Sabeti, P. C. (2019). Benchmarking Metagenomics Tools for Taxonomic Classification. Cell, 178(4), 779–794. <https://doi.org/10.1016/j.cell.2019.07.010>