Bo Maxwell Stevens

12/12/2023

# Sequencing report

## Sample summary

245 samples were extracted for DNA from the 1, 5, 6 Nitrogen treatments and Water 1 and 2 treatments (Table 1). Overall, 242 samples were included in the final mapping\_file\_taxonomy.csv. One sample was dropped from the mapping file because there was no corresponding qPCR data. Two samples were dropped because there was no corresponding sequencing file.

Table 1. Summary of samples sequenced for each date and treatment. 245 samples were extracted for DNA from the 1, 5, 6 Nitrogen treatments and Water 1 and 2 treatments. One sample was dropped from the mapping file because there was no corresponding qPCR data. Two samples were dropped because there was no corresponding sequencing file.

|  |  |  |  |
| --- | --- | --- | --- |
| Sample date | Nitrogen | Water | Count |
| 6/15/2021 | 1 | 1 | 6 |
| 6/15/2021 | 1 | 2 | 5 |
| 6/15/2021 | 5 | 1 | 6 |
| 6/15/2021 | 5 | 2 | 6 |
| 6/15/2021 | 6 | 1 | 6 |
| 6/15/2021 | 6 | 2 | 6 |
| 6/30/2021 | 1 | 1 | 5 |
| 6/30/2021 | 1 | 2 | 6 |
| 6/30/2021 | 5 | 1 | 4 |
| 6/30/2021 | 5 | 2 | 6 |
| 6/30/2021 | 6 | 1 | 6 |
| 6/30/2021 | 6 | 2 | 6 |
| 7/27/2021 | 1 | 1 | 6 |
| 7/27/2021 | 1 | 2 | 6 |
| 7/27/2021 | 5 | 1 | 6 |
| 7/27/2021 | 5 | 2 | 6 |
| 7/27/2021 | 6 | 1 | 6 |
| 7/27/2021 | 6 | 2 | 5 |
| 8/10/2021 | 1 | 1 | 6 |
| 8/10/2021 | 1 | 2 | 5 |
| 8/10/2021 | 5 | 1 | 6 |
| 8/10/2021 | 5 | 2 | 5 |
| 8/10/2021 | 6 | 1 | 6 |
| 8/10/2021 | 6 | 2 | 6 |
| 9/23/2021 | 1 | 1 | 3 |
| 9/23/2021 | 1 | 2 | 5 |
| 9/23/2021 | 5 | 1 | 6 |
| 9/23/2021 | 5 | 2 | 6 |
| 9/23/2021 | 6 | 1 | 6 |
| 9/23/2021 | 6 | 2 | 5 |
| 5/26/2022 | 1 | 1 | 3 |
| 5/26/2022 | 1 | 2 | 3 |
| 5/26/2022 | 5 | 1 | 3 |
| 5/26/2022 | 5 | 2 | 3 |
| 5/26/2022 | 6 | 1 | 3 |
| 5/26/2022 | 6 | 2 | 3 |
| 6/22/2022 | 1 | 1 | 3 |
| 6/22/2022 | 1 | 2 | 3 |
| 6/22/2022 | 5 | 1 | 3 |
| 6/22/2022 | 5 | 2 | 2 |
| 6/22/2022 | 6 | 1 | 3 |
| 6/22/2022 | 6 | 2 | 3 |
| 7/20/2022 | 1 | 1 | 3 |
| 7/20/2022 | 1 | 2 | 2 |
| 7/20/2022 | 5 | 1 | 3 |
| 7/20/2022 | 5 | 2 | 2 |
| 7/20/2022 | 6 | 1 | 3 |
| 7/20/2022 | 6 | 2 | 3 |
| 9/1/2022 | 1 | 1 | 3 |
| 9/1/2022 | 1 | 2 | 3 |
| 9/1/2022 | 5 | 1 | 3 |
| 9/1/2022 | 5 | 2 | 3 |
| 9/1/2022 | 6 | 1 | 3 |
| 9/1/2022 | 6 | 2 | 2 |
| 9/27/2022 | 1 | 2 | 3 |
| 9/27/2022 | 5 | 2 | 3 |
| 9/27/2022 | 6 | 2 | 3 |

## Methods

DNA was extracted from 0.25 g soil samples from each plot using the Qiagen DNeasy Powersoil Pro Kit (Qiagen, Germantown, MD). The extraction process was carried out using a fully automated Qiagen QIAcube robot with a 10-min vortex lysis step. DNA was quantified fluorometrically with the Invitrogen dsDNA HS Assay Kit on a Qubit Flex (Life Technologies, Carlsbad, CA). PCR amplifications were performed on each DNA sample using 27F/1492R (Lane, 1991; Muyzer et al., 1995), which targets the full-length 16S rRNA gene.

Extracted DNA samples were amplified in 60 µL PCR reactions containing 30 µL Phusion HSII (Thermo Scientific, Waltham, MA) master mix, 0.6 µL of each forward and reverse primer (10 µM concentration), 21.6 µL molecular grade H2O, and 6 µL soil DNA diluted 1:20 with nuclease-free water. Reactions were held at 98 °C for 30 s, with amplification proceeding for 25 cycles at 98 °C for 15 s, 50 °C for 15 s, and 72 °C for 60 s with a final extension at 72 °C for 5 min. The PCR products (PCR1) were purified using AMPure XP beads (Beckman Coulter, Indianapolis, IN).

Unique barcodes (EXP-PBC096, ONT, Oxford, UK) were added to both ends of the DNA fragments by PCR. These were 50 µL PCR reactions containing 25 µL Phusion HSII master mix, 19 µL H2O, 1 µL of forward/reverse barcodes, and 5 µL PCR1 product diluted 1:10 with nuclease-free water. Reactions were held at 98 °C for 30 s, with amplification proceeding for 15 cycles at 98 °C for 15 s, 62 °C for 15 s, and 72 °C for 60 s; a final extension at 72 °C for 5 min. The barcoded products of this PCR reaction were purified a second time using AMPure XP beads.

Barcoded amplicons from all samples were pooled and prepared for sequencing using the SQK-LSK109 Ligation Sequencing Kit (Oxford Nanopore Technologies, Oxford, UK). The library was loaded on a MinION flow cell FLO-MIN106D-R9 (Oxford Nanopore Technologies, Oxford, UK) per manufacturers’ protocol and sequencing was started with a runtime of 48 hours. All libraries included no template (H2O-only) negative controls and a mock community (ZymoBIOMICS Microbial Community DNA Standard D6305; Zymo Research, Irvine CA).

Sequences were base-called and demultiplexed using Guppy v6.0.6 (Oxford Nanopore Technologies, Oxford, UK). Except were otherwise noted, default parameters were used. Sequences were filtered based on length to between 1000 and 2000 bp and a minimum q-score of 70 using Filtlong v0.2.1 (Wick, 2017) and Cutadapt v3.2 (Martin, 2011). Chimeras were filtered using vsearch (Rognes et al., 2016), and taxonomy was assigned with minimap2 v2.22 (Li, 2018). Error-correcting was done with Emu v3.0.0 (Curry et al., 2022) which applies an expectation minimization algorithm to adjust taxonomic assignments using up to 50 sequence alignments per sequence read.

## Sequencing quality

Bacterial sequencing quality was good according to the expected mock community (Figure 1). Most negative control had little to no total reads (Table 2). However, total reads from sequencing was much higher in the third sequencing run (Figure 2).

Chart

Description automatically generated

Figure 1. Actual versus expected mock community results for bacterial sequencing.

Table 2. Reads in negative controls.

|  |  |
| --- | --- |
| Barcode | Total Reads |
| barcode22C | 0 |
| barcode30B | 3423 |
| barcode93A | 159 |
| barcode94A | 0 |
| barcode93B | 18575 |
| barcode94B | 0 |
| barcode93C | 11 |
| barcode94C | 0 |

Chart, bar chart

Description automatically generated

Figure 2. Total reads for each sequencing run.

## Reads removed by quality filtering

Of a total of 8,669,048 reads, 325,998 (3.7%) were removed during quality filtering and length filtering. A further 12,733 reads (0.2%) were removed with cutadapt. No further reads were removed by subsequent quality filtering. Including controls, the remain sample sizes were an average of 28,924.7 reads.

## Data availability

Code and all necessary files after Emu analysis are stored on GitHub (<https://github.com/bo-maxwell-stevens/nxwater_bacteria>). Sequences and bioinformatics code are stored on Juno in the akron project folder (/LTS/project/akron).