
What we did:

1. Worked on Group Assignment 1

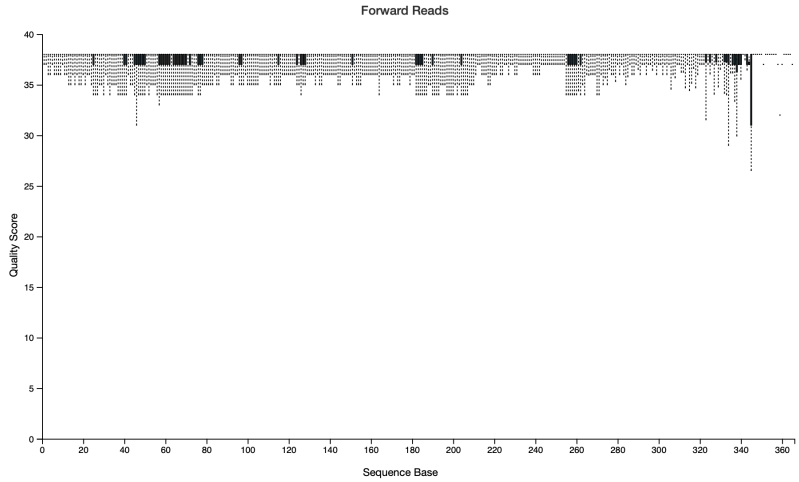
    a. Part 1 - Section A.
        i. Our research question investigates whether sex/gender modifies the relationship between histological stage of gastric disease progression and the composition of the gastric microbiome diversity using 16S rRNA sequencing data from patient samples. Microbiome composition will be evaluated using diversity metrics across various histological stages and compared between sex/gender groups.
    b. Part 1 - Section B
        i. Is this within expectations (example taken from assignment):

Term	Definition/Description
Site - Gastric Body	Categorical variable. Describes location where gastric biopsy was harvested from. Gastric body refers to the location of the stomach between the fundus and antrum.
Site - Gastric Antrum	Categorical variable. Describes location where gastric biopsy was harvested from. Gastric antrum refers to the location of the stomach after the body and right before the duodenum.

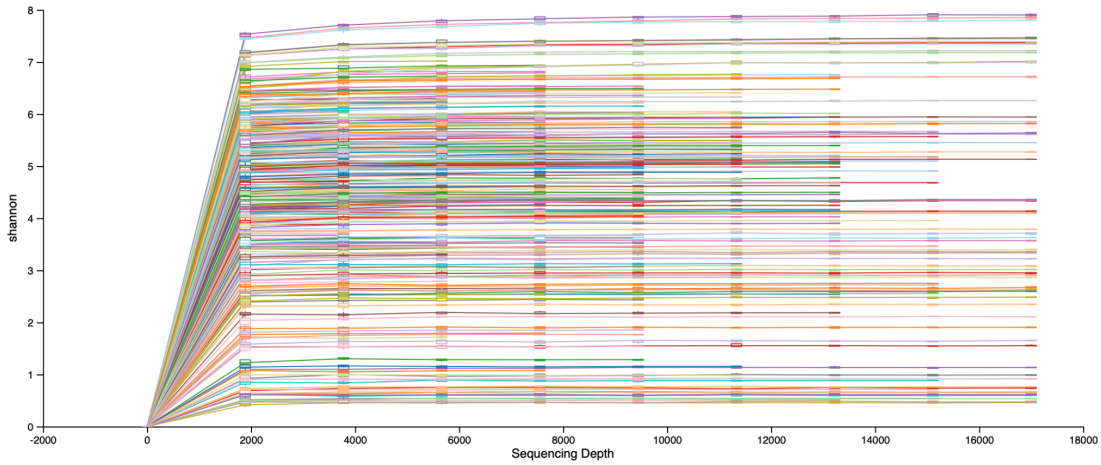
    c. Part 1 - Section C
        i. Will be completed before the due date (Friday).

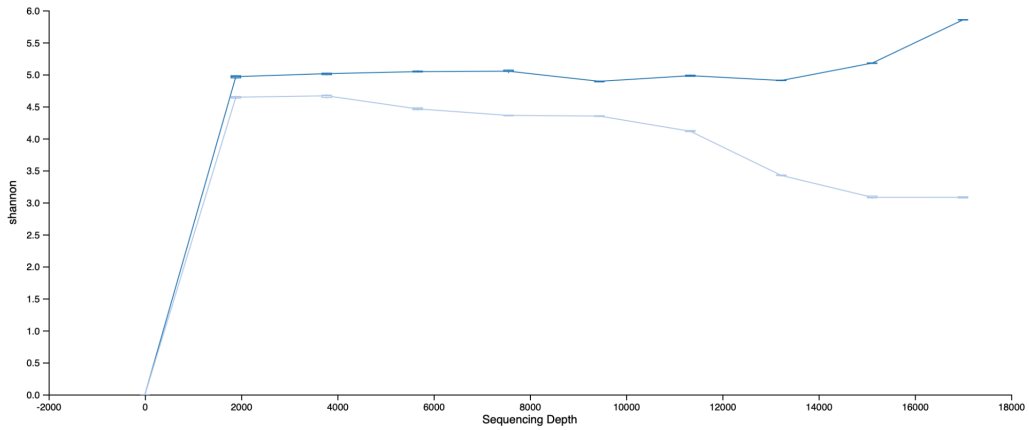
    d. Part 2 (Coding)
        - Chose to use DADA2 to denoise our dataset (denoising did not reduce sample size)

- **Read Quality Plot** to determine the truncation parameter:



  - Approximate distribution of samples based on the demultiplexed sequence length summary:
    - 2% 236 nts
    - 9% 252 nts
    - 25% 253 nts
    - 50% 253 nts
    - 75% 253 nts
    - 91% 254 nts
    - 98% 298 nts
- Truncating length was set to 0-253 (as most of our reads fall within this range)
- **Rarefaction Curve:**





- Rarefaction sampling depth chosen is 3145

What we want to discuss:

- Check in to see if research question (part A) is clear and within reasonable scope. In addition, if part B has been done correctly and if not how it could be improved.

- Make sure that the code was done properly and metrics were properly chosen (Section 2).
    - Specifically if the truncation length and distribution was chosen properly
- Check to see if we need to apply any additional filtering + what diversity metrics/files do we need to export from Qiime? Or are we mostly working in R for downstream analysis?