

Samuel L. Peoples
BBIO 393: Computational Biology
Project 1 Proposal

1. Overview

This project will calculate the beta diversity between samples from [Jeff Gordon's A Core Gut Microbiome in Obese and Lean Twins](#). Given an input [.BIOM file](#), each sample will be compared with all others, and a beta diversity distance matrix will be generated. Because the mapping file will have names associated with the samples from this output, the result will facilitate further analysis in Project 2.

2. Inputs

The script will call a single [.BIOM file](#), which will compare the beta diversity between 281 samples. If used, the [mapping file](#) is composed of 39 features, with 281 samples.

3. Approach

I will use a virtual machine running an Ubuntu Qiime distribution. This will have the necessary packages to access the OTU tables. I will then implement the UniFrac metric as described in the beta-diversity resources. The output table will be saved to an appropriately formatted file for analysis in the second project. The script will be saved to a *.py* file, as well as annotated in a Jupyter Notebook containing properly cited (APA) resources.

4. References

Qiime: <http://qiime.org/>
Qiita: <https://qiita.ucsd.edu/>
Gut Microbiome Dataset: <https://qiita.ucsd.edu/study/description/77>
Biom-Format: http://biom-format.org/documentation/biom_format.html
Lozupone/ Knight's UniFrac: <http://aem.asm.org/content/71/12/8228.full>
Qualitative/ Quantitative: <https://www.ncbi.nlm.nih.gov/pubmed/17220268>
Weighted UniFrac: <https://liorpachter.wordpress.com/2013/09/18/unifrac-revealed/>