

Statistical Analysis of Mouse Gut Microbiota

Alex Tran

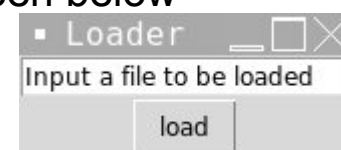
Computer Systems Lab 2009-2010

Abstract

The mouse gut microbial community is the population of bacteria, which inhabit the digestive track of a given population of mice. The ability to understand the microbial community has implications, which extend beyond mice and into factors of human obesity and the dietary needs of the human body. Computational genomics is an emerging field, which blends both biology and computer science, to analyze genes. This study seeks to utilize several methods emerging within computational genomics to analyze the gut microbiome of mice to observe the effects of varying diets on the gut microbial community, as well as create a universal and user friendly tool for researchers to utilize when studying any gut microbiome. The applications of this can extend past studying the gut microbiome, but can also be applied to any taxonomic group being counted for analysis, however several parts of the application are exclusively beneficial to the analysis of gut microbiota specifically those found in mice.

Development

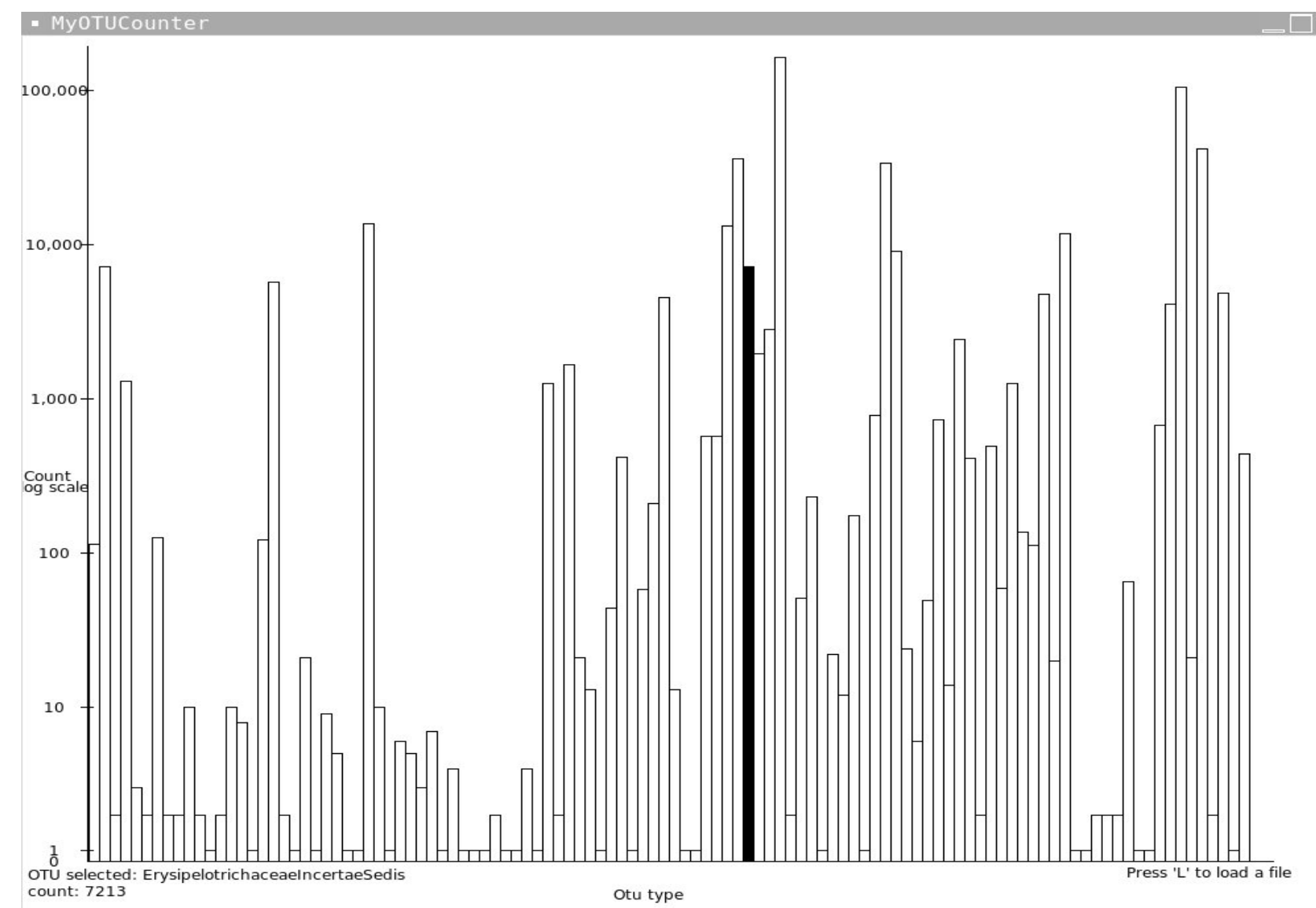
The program's end result will consist of several parts, which can be created individually in any order. Due to possible problems with time constraints it has been written in reverse depth from the most superficial part to the most abstract. This will allow for user friendly verification as well as a guaranteed presentable result if time had become a problem. It was primarily programmed in Python. Coding began by generating a simple taxonomic tree, and filling it with data points. It then moved onto generating a histogram of these data using Tkinter as a base. To test and verify results several premade computational genomics available tools that each perform a part of what the finalized project to be able to do as one universal tool will be used. Input data and 16sRNA sequences are to be received from the Gordon Lab at Washington University in St. Louis. A large part of the project was focused on user friendliness in the graphical interfaces. Utilizing several methods for user friendliness as seen below



Background

Other research has observed that general imbalances inside the gut microbiome can lead to differences in tendencies towards Obesity in mice (Turnbaugh, Gordon). It has also been observed that within various populations of mice the microbiome appears to be so diverse that there is little known correlation between an individual and the population of their microbiome, however it was also shown that though the population in the microbiome is diverse there are tendencies based on the diet of an individual mouse for an increased representation of certain strains of bacteria to be present within the microbial community (Turnbaugh, et. al.)

Though General imbalances have been observed in the gut microbiome of obese mice, there is still not enough information or analysis to say pinpoint what specific species of bacteria can lead to the tendency for mice to be obese. (Turnbaugh, Gordon) This lack of information is largely due to the inefficiency and inaccurate nature of current computational 16s sequencing techniques (Turnbaugh et. al.)



Results

To be able to present data from the microbial community in a way that is user friendly, efficient, and able to represent the data in an accurate way. To achieve this a gui was utilized that allows the display of statistical information on each run in a way that allows the user to more easily gather information from a sample than has ever before been possible. As a result of this analysis of data gathered from the gut microbiome can be done vastly faster than other archaic forms of analysis previously used. It cuts down on the amount of fact checking, and manual calculation that was previously necessary when doing analysis of this nature.

The results from this study are potentially very helpful to the field of genomics, because tools for analysis of gut microbial communities are a field that is in its early stages of development, and at this point most tools which are planned to be implemented into the end result application are only available separately, and very expensive.