

TJHSST Computer Systems Lab Senior  
Research Project  
Statistical Analysis of Mouse Gut Microbiota  
Proposal  
2009-2010

Alex Tran

October 27, 2009

**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.

**Keywords:**Computational Genomics, Genomics, Gut Microbiota, Statistics, Microbiota

# **1 Introduction - Elaboration on the problem statement, purpose, and project scope**

## **1.1 Scope of Study**

To use genomics to observe and document the gut microbial community in the fecal matter of mice during all stages of analysis. From the sequencing of genetic material to the processing of the microbial community. To create an all encompassing suite to streamline the analysis of any microbial community.

## **1.2 Expected 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. Implemented Gui utilization that will display statistical information on each run, which can be used as visuals in a report, and to be able to do a full run from the ground up given 16sRNA sequencing data that can accurately and efficiently identify individual microbiota within the community.

## **1.3 Type of research**

This study will largely fall under pure applied research in that its primary purpose is to create an application, which streamlines many prediscovered methods for computational genomics into one efficiently designed, and user friendly package.

# **2 Background and review of current literature and research**

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 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.)

### **3 Procedures and Methodology**

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 is being 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 becomes a problem. It is primarily going to be programmed in Python. 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 is going to be focused on user friendliness in the graphical interfaces, which can be used as visuals in the final paper.

### **4 Expected 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. I plan to utilize a gui that will display statistical information on each run, which can be used as visuals in a report.

The results from this study would be of great use 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.

## **5 Bibliography**

Peter J. Turnbaugh and Jeffrey I. Gordon, "The core gut microbiome, energy balance, and obesity." Center for Genome Sciences, Washington University School of Medicine, St. Louis MO, 2009.

Peter J. Turnbaugh, Ruth E. Ley, Micah Hamady, Claire M. Fraser-Liggett, Rob Knight Jeffrey I. Gordon, "The Human Microbiome Project."