

# Report on analysis of time profile of microbial community by using one way ANOVA model

Abu Saleh Mosa Faisal

January 29, 2020

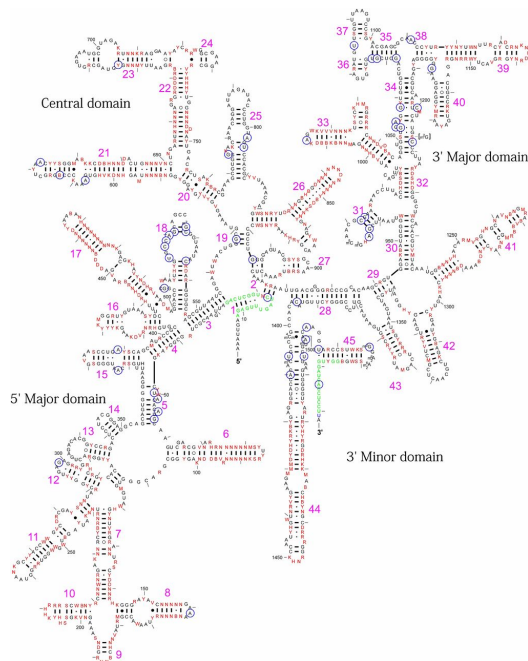
## 1 Introduction

We have a microbial community data set that identified by 16S rRNA gene. Here, we will detect the genera which are showing significant difference over time period. In our data set we have five time points(0 month, 5monmths, 23 months, 37 months, 68 months) where each of time points have eleven replicates except 68 months(five replicates).

### 1.1 16S rRNA

16S rRNA is a mitochondrial DNA specially used for species identification. 16S ribosomal RNA (or 16S rRNA) is the component of the 30S small subunit of a prokaryotic ribosome. 16S rRNA is also used for phylogenetic studies as it is highly conserved between different species. Moreover, 16S rRNA gene sequence analysis is standard method for taxonomy and identification.

Figure 1: Structure of 16S rRNA



## 2 Model

To find out the Genera which are changing significantly over time, we fit the following one way ANOVA model for each of the genus k:

$$W_{ij} = \mu + \alpha_i + \epsilon_{ij}$$

Here, c is the offset to help with 0 counts.(we took c=0.001)

i=1,2,3,4,5(1=0 month, 2=5 months, 3=23 months, 4=37 months, 5=68 months)  
j=1, ...,  $n_i$  (Here,  $n_i$  is the number of observation for the time point i.)

$W_{ijk}$  is number of observed normalized counts of  $j^{th}$  replicate of  $i^{th}$  time point for  $k^{th}$  genus.

After taking logs in each of the point, the result is normalized by dividing by the mean log reads at given time point. Similar to "RPM" normalization.

Here is the equation that used for normalization:

$$W_{ij} = \frac{\log(Y_{ijk}+c)}{\frac{\sum_{j=1}^{n_i} \sum_{k=1}^{129} (\log(Y_{ijk}+c))}{n_i k}}$$

$\alpha_i$ 's are fixed effects of factor groups at level i.

$\mu$  is combined population mean or overall effect for different genus abundances.

$\epsilon_{ij}$  is error term.

## 3 Hypothesis

Hypothesis that we are testing here is

Null Hypothesis  $H_0$ :  $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = \alpha_5$  (that is, the genus's average abundance is constant over time.)

Alternative Hypothesis  $H_1$ : not all  $\alpha_i$ 's are equal

## 4 Brief Description and Findings

Here, for time points 0, 5, 23, 37 months have eleven replicates for each genus. Only time point 68 months has five replicates for each genus.

We have 195 individual genus in our data set. Some of the genera contain zero in most of the time points. So, We have taken only those genus who have more than 50 total counts.

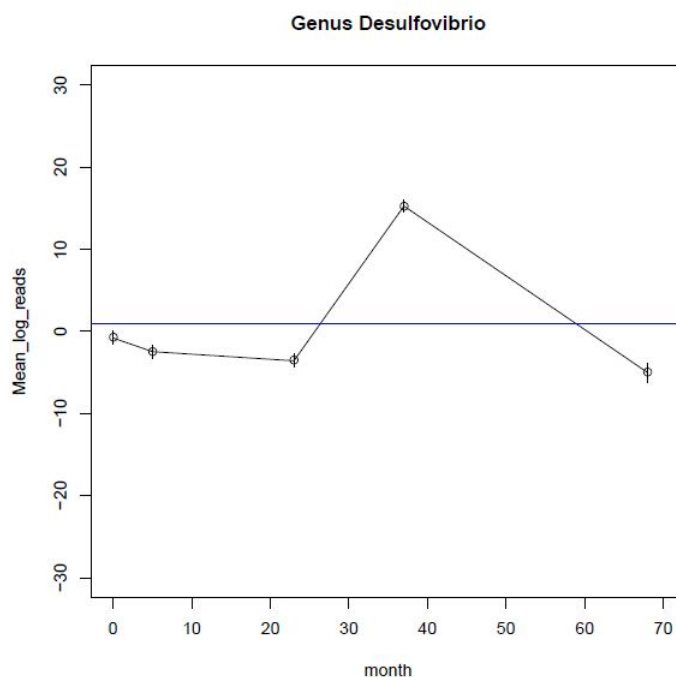
After screening the data we have got 129 genera who have more than 50 counts. Then we run one way ANOVA model to find which genera are showing significant change over time. We have found 123 genera are significantly changing over time.

The insignificant genera are following:

"Treponema"	"SHD-231"
"Acinetobacter"	"Desulfomicrobium"
"Polynucleobacter"	"Yersinia"

To see the generation trend of the genus we have plotted their estimated means(after normalization) with respect to time and have found some of them showing interesting trend of growing during the time interval.

Figure 2: Time Profile of Desulfovibrio



Vertical bars represent 95% confidence intervals.

Note that a genus whose average level stays constant over time will not be flagged by this method. We check for homogeneity of variance by using Levene's test. Among 123 genus only 61 genera p-values are greater than significance level 0.05.

We calculate confidence intervals for parameters for each of the genus by "Wald" method.

See the Appendix(plotofgenus.pdf) and Appendix(confidenceInterval.pdf) for all genera profiles and confidence Intervals.

## References

- [1] Jonathon Tylor *Introduction to regression and analysis of variance*. <http://statweb.stanford.edu/~jtylo/courses/stats203/notes/anova.fixed.pdf>

- [2] Clarridge, J. E. [*Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases*]. Clinical Microbiology Reviews, 17(4), 840–862, 2004.
- [3] Rebecca A. Reiss, Peter Guerra, Oleg Makhnin [*Metagenome phylogenetic profiling of microbial community evolution in a tetrachloroethene-contaminated aquifer responding to enhanced reductive dechlorination protocols*]. Stand Genomic Sci. 2016; 11: 88.