Time Series Analysis of Human Microbiota

Rebecca Cargan
Bryan Featherstone
Bairavi Venkatesh

Microbiome

What is a microbiome?

Biotic and abiotic factors in an environment including the microorganisms, their genomes, and the environment that surrounds them.

Microorgasisms consists of:

•Bacteria

•Archaea

•Eukaryotes

•Viruses

Why Study bacterial compositions?

- Aid in digestion and contribute to the metabolism
- Essential for the development and regulation of the immune system and immune response
- Environmental factors have been shown to alter microbiota composition
- Few studies have looked at the temporal dynamics of microbiomes
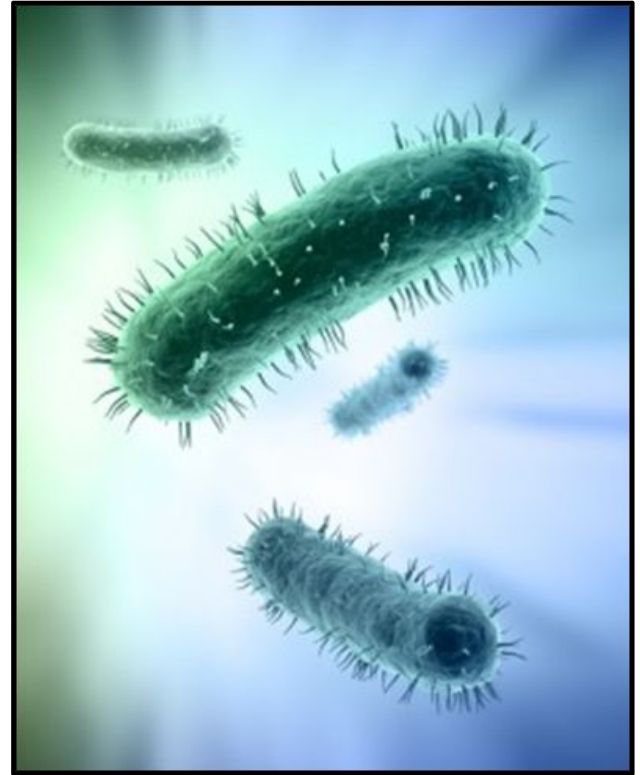
Host lifestyle affects human microbiota on daily timescales

Lawrence A David[1,2,11], Arne C Materna[3], Jonathan Friedman[4], Maria I Campos-Baptista[5], Matthew C Blackburn[6], Allison Perrotta[7], Susan E Erdman[8] and Eric J Alm[4,7,9,10*]

- Compared temporal dynamics of daily human actions to gut and saliva microbiomes.
- 2 individuals over the course of a year
- https://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-7-r89

Findings

- Travel and sickness played the largest role in changing bacterial dynamics.
- Not as many correlations between host lifestyle (~10,000 measurements)and variability in the microbiota. Suggesting future studies shouldn't have to control for host behavior and there's stability between microbes and their host
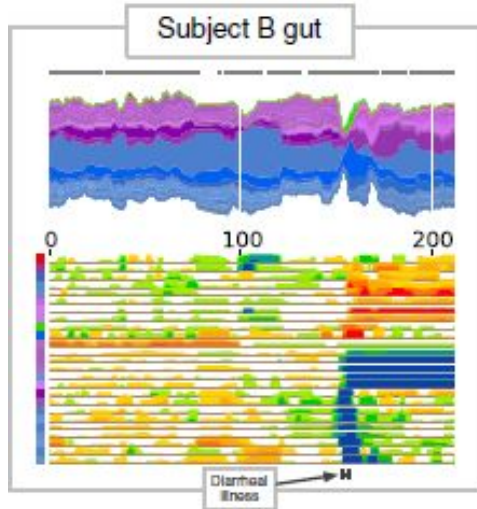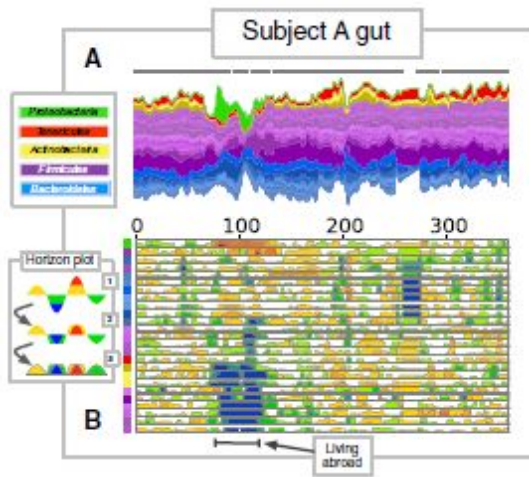
Goals

- Use DADA2 pipeline to look at <u>fine scale variation</u> within microbiota.
- Correlate host lifestyle to changes in bacteria composition using the following analyses
    - Diversity and Ordination Plots
   - Fast Fourier Transform (FFT)
    - ARIMA

DADA 2

- Algorithm correcting for amplicon sequencing error providing fine scale resolution of sequences instead of grouping them into operational taxonomic units (OTUs)
    1. Creates an error model for Illumina sequencing technology
    2. Subjects the error model to your filtered reads
   3. Identify like or unique sequences

DADA2 pipeline

- Filtering and Trimming
- Inference
    - Subsample (2 million reads) and dereplicate
    - Learn error rates to create error model
    - Use error model against whole data set
- Sequence Table
    - Make sequence table while removing Chimeras
- Taxa Table
    - Assign taxonomy using Silva Database Version 123
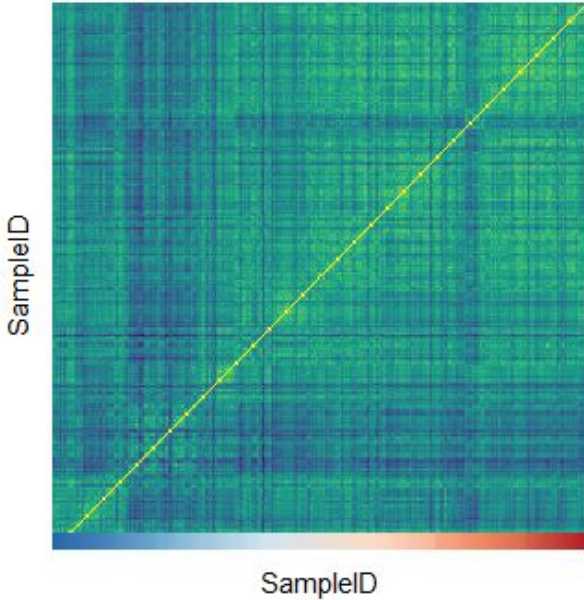
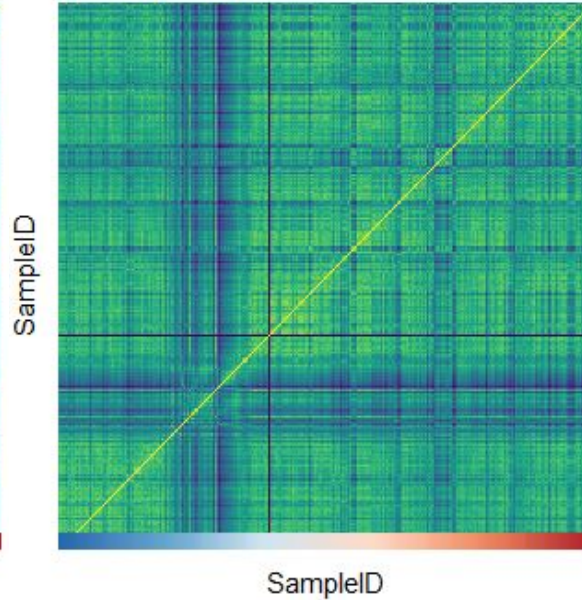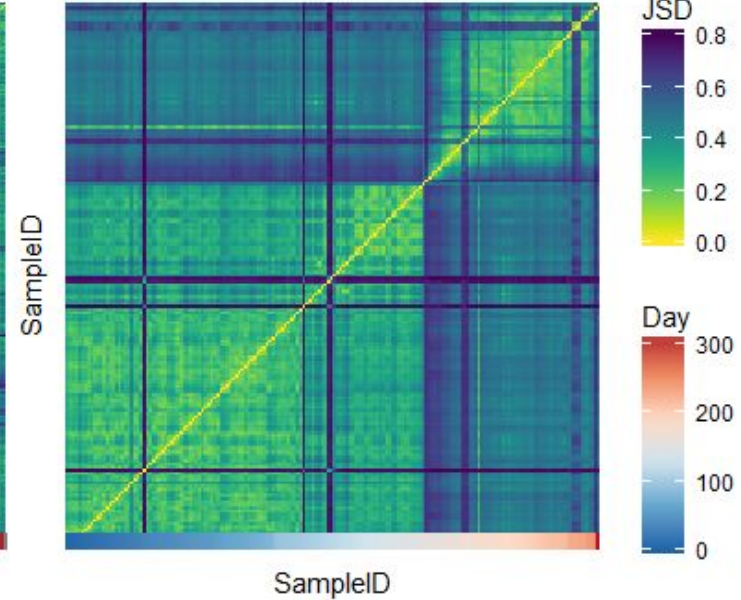# Relative Bacteria Abundance (Order) vs Time
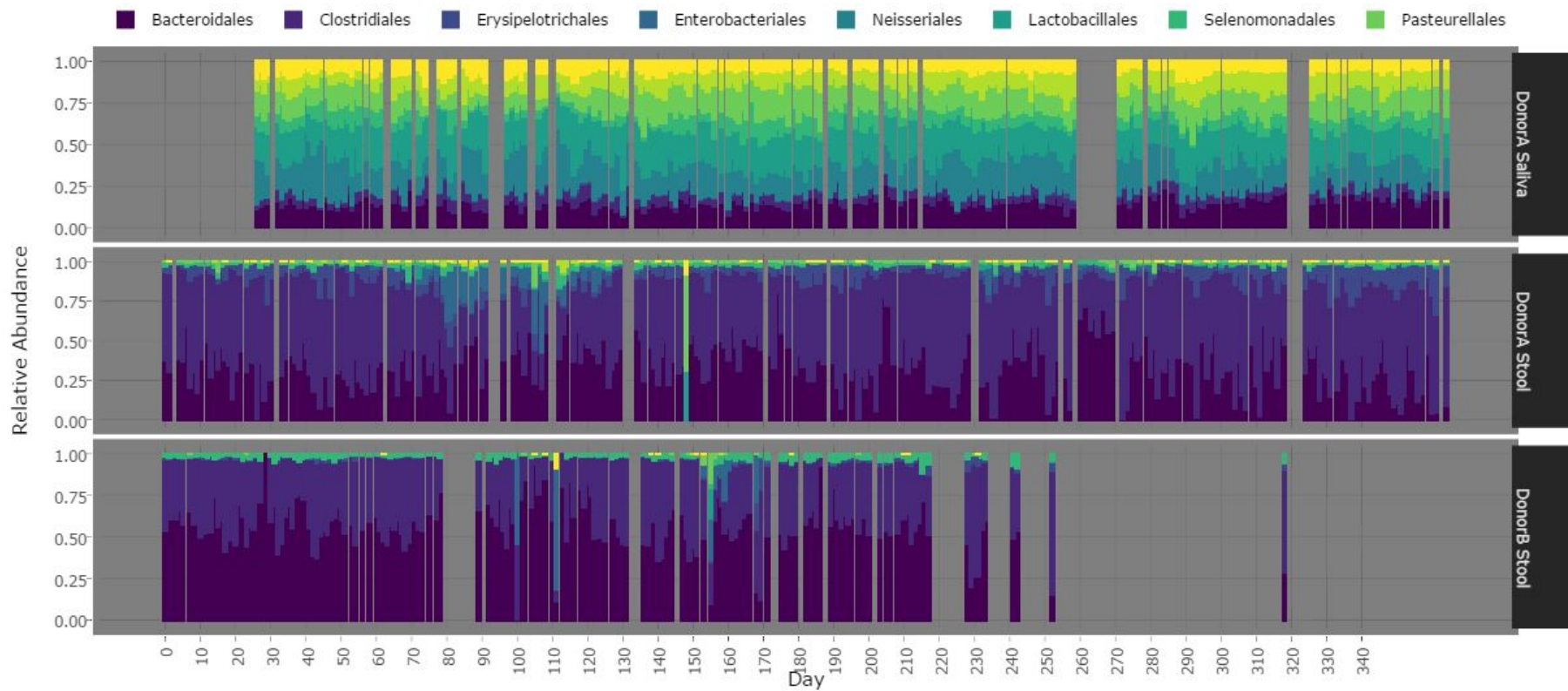
Diversity PCoA plot

Differentiation of bacterial composition by site
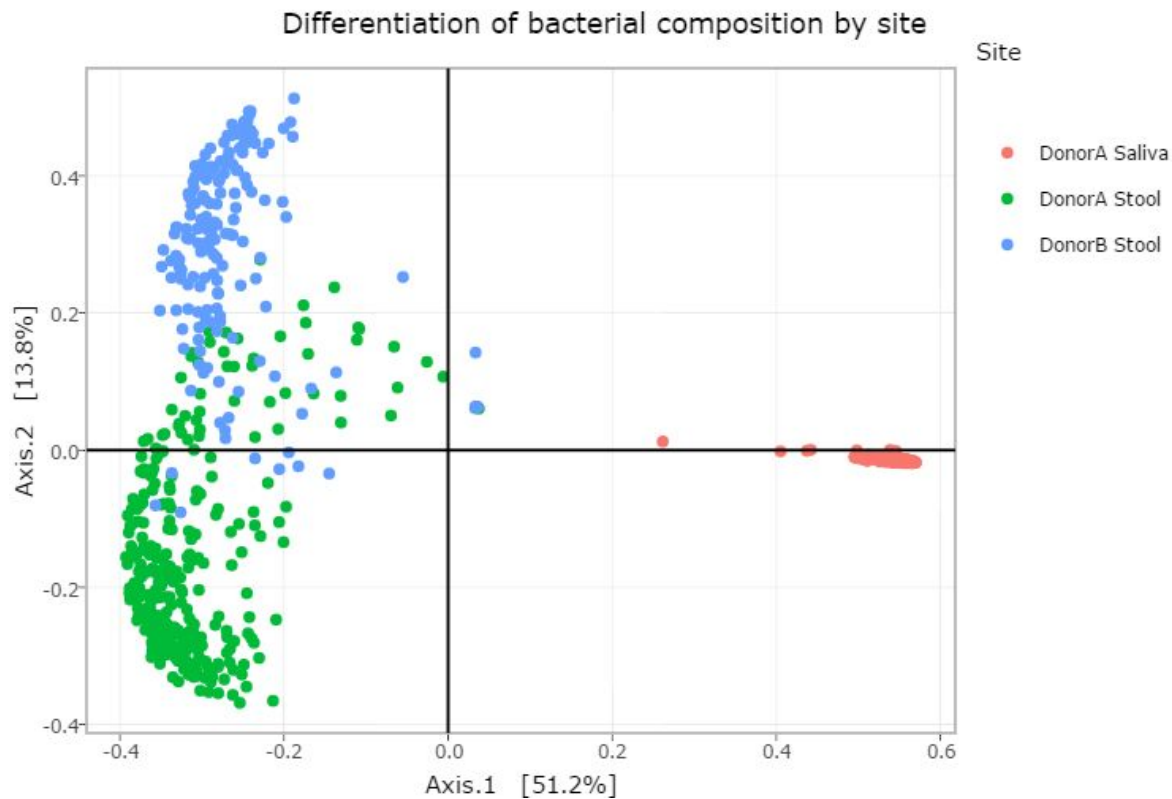
Heatmap using Jensen Shannon Divergence
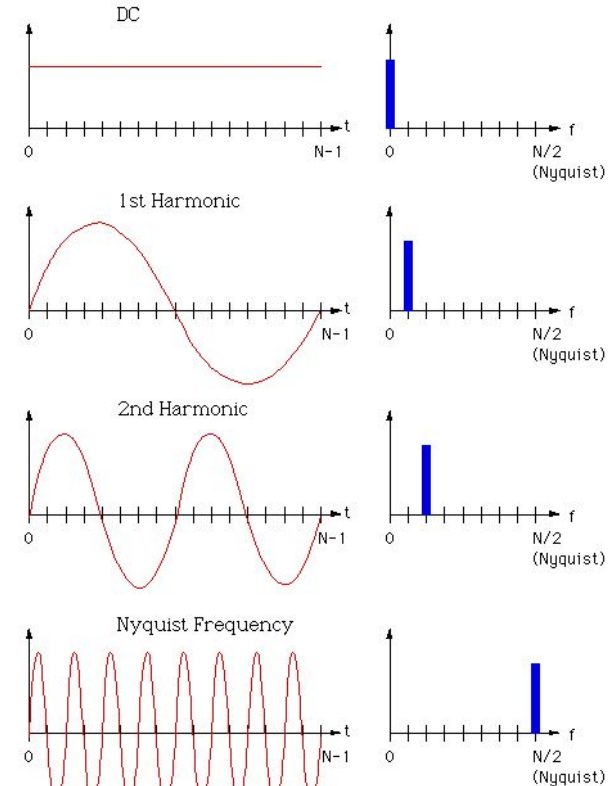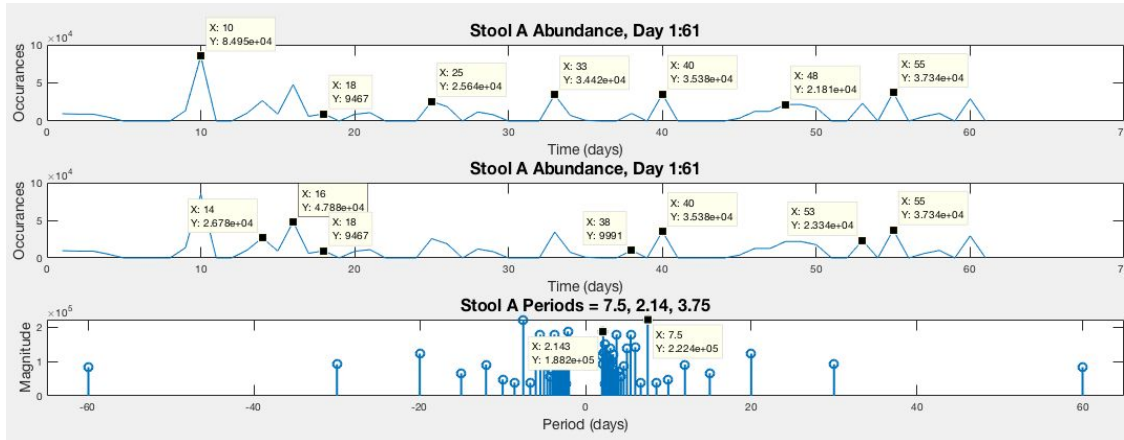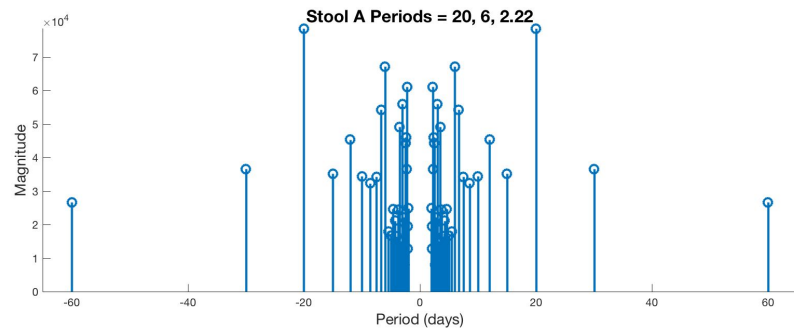
FFT Analysis Background

$$F(x) = \sum_{n=0}^{N-1} f(n)e^{-j2\pi(x\frac{n}{N})}$$

- Periodicity: How often you can expect a pattern to repeat itself
- Periodicities are not mutually exclusive!

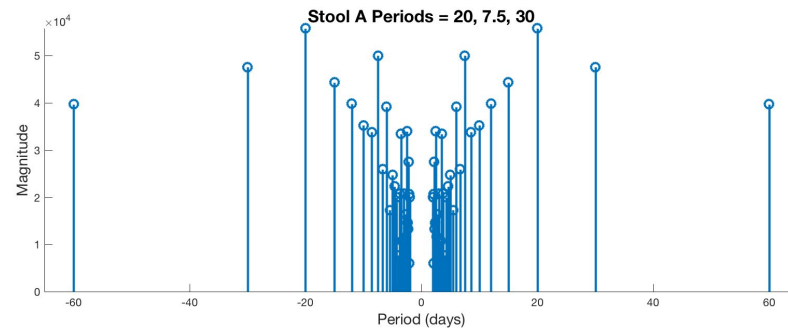http://paulbourke.net/miscellaneous/dft/
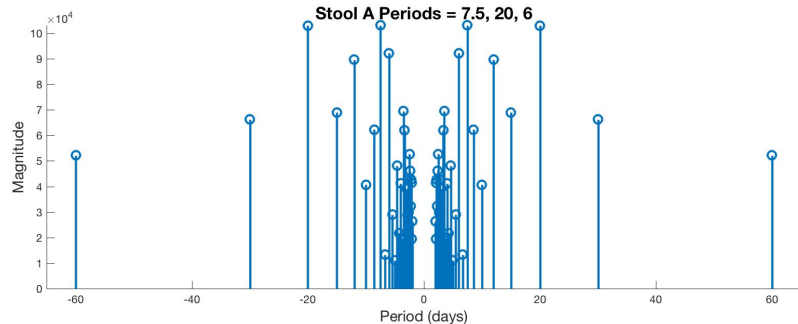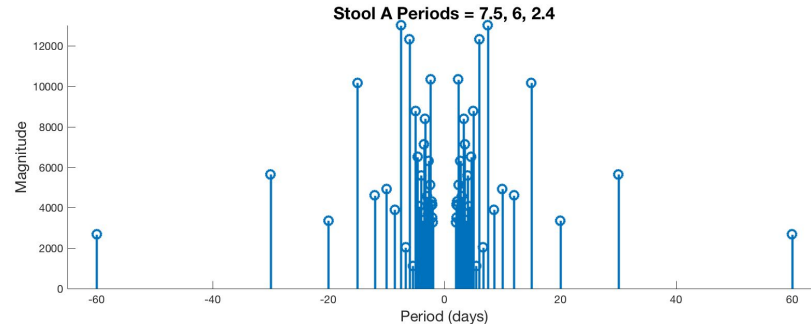
FFT Analysis: Lachnospiraceae Family

Seq 13. Lachnospiraceae: *Blautia*

Seq 18. Lachnospiraceae: *Fusicatenibacter*

Seq 9. Lachnospiraceae: *Anaerostipes*
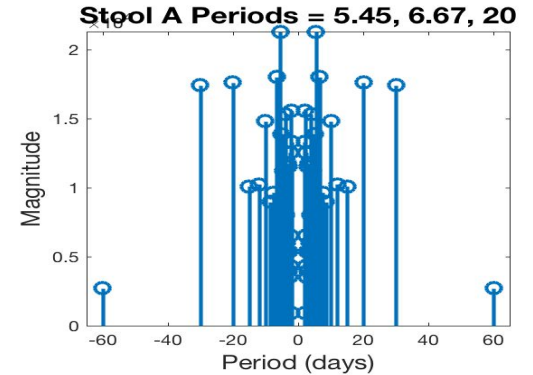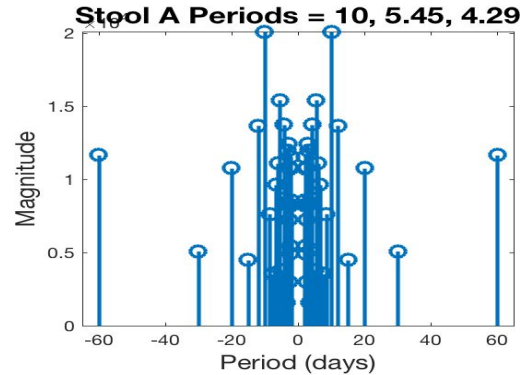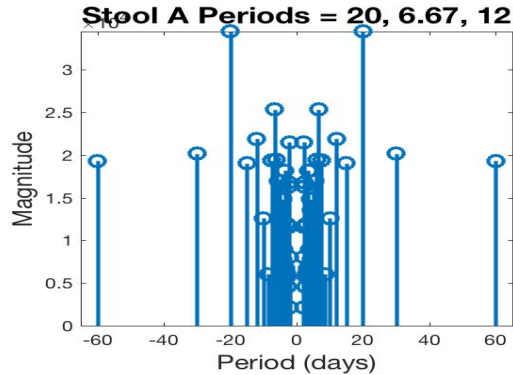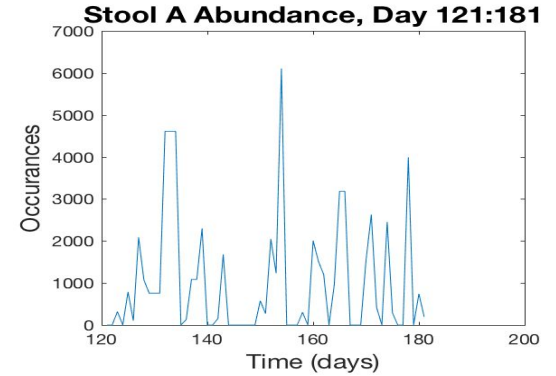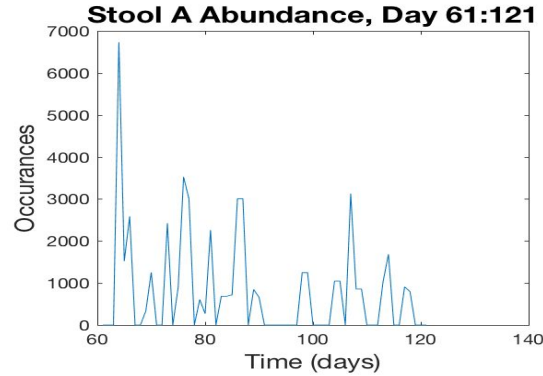

Seq 37. Lachnospiraceae: *Coprococcus*

FFT Analysis : Lachnospiraceae Blautia (24) First 6 Mo.
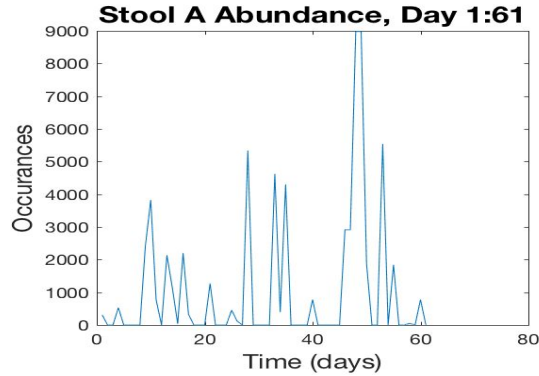
FFT Analysis: Lachnospiraceae Blautia (24) Last 6 Mo.
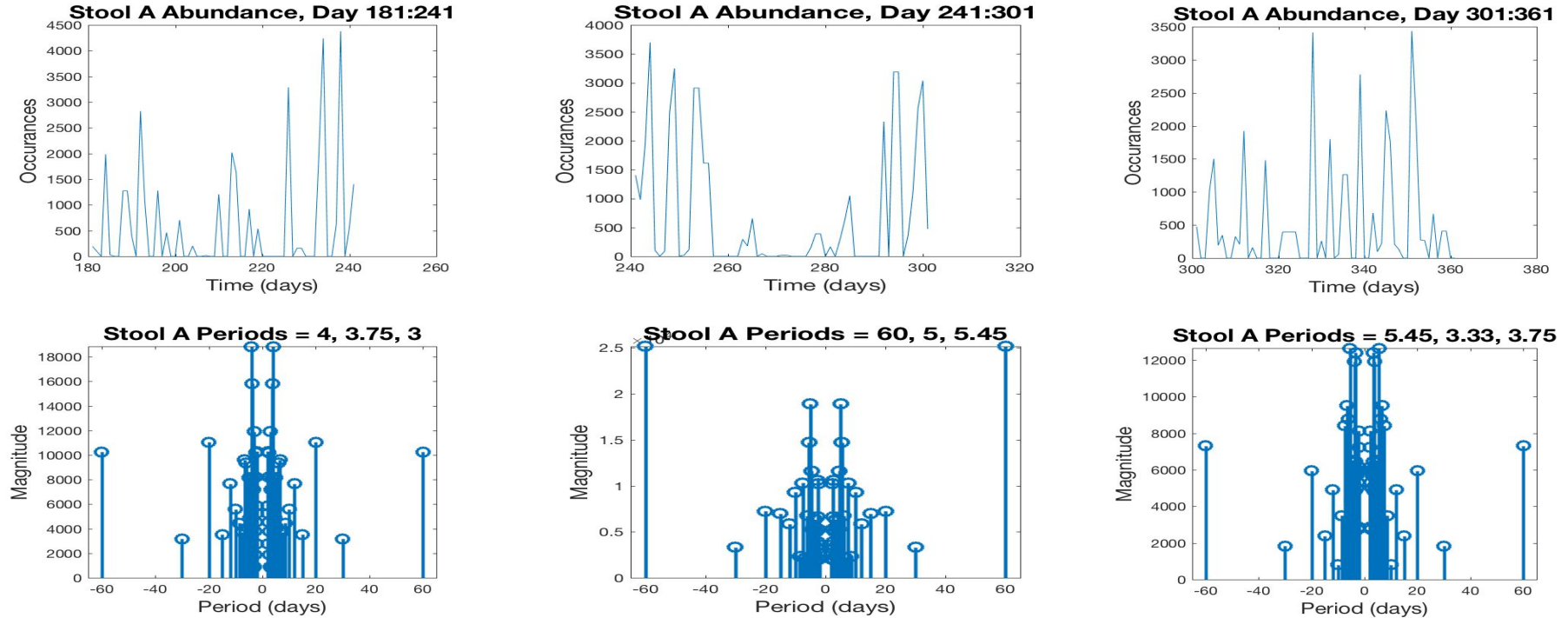
"Host-specificity and host-preference patterns of organisms within this genus are driven by host physiology more than dietary habits."

FFT Analysis: Ruminococcaceae Ruminococcaceae
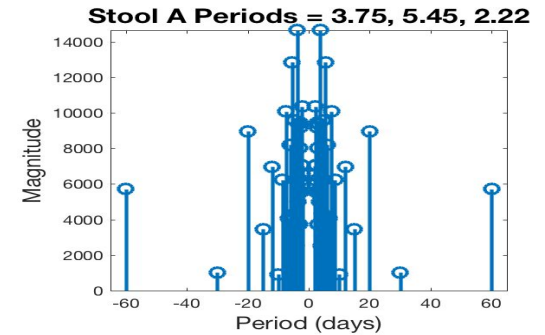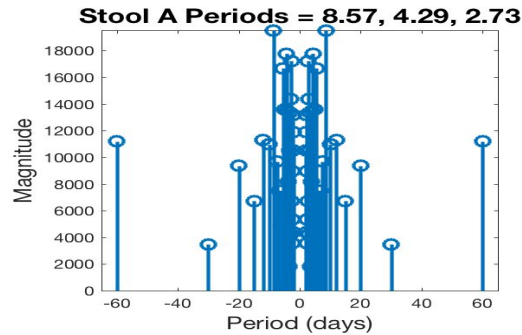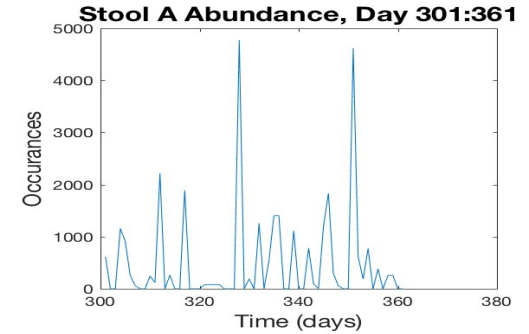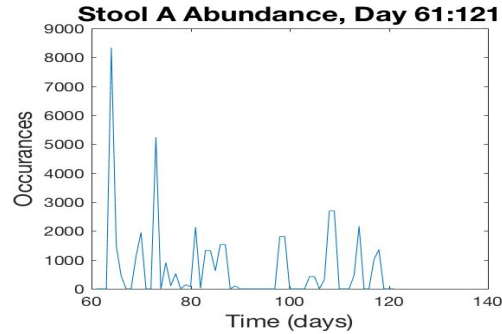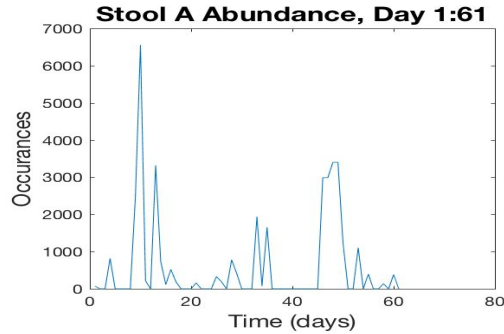
Increased levels of Ruminococcaceae have been associated with diets high in resistant starch.

ARIMA - Background

- ARIMA = AutoRegressive Integrated Moving Average
- "AR" = Models the relationship between an observation and its prior points
- "MA" = Regression errors are linear combos of errors that occurred in the past
- "I" = Data values are replaced with the difference between their current and previous values
- Goal ➜ Fit a model to time series data in order to better understand or predict future points in the data.
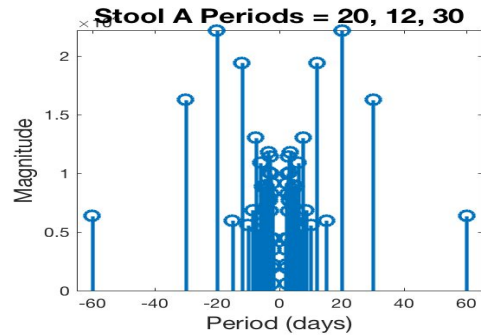- Multiplicative ARIMA Model ➜ Addresses seasonality and period trends

Forecast and 95% Forecast Interval
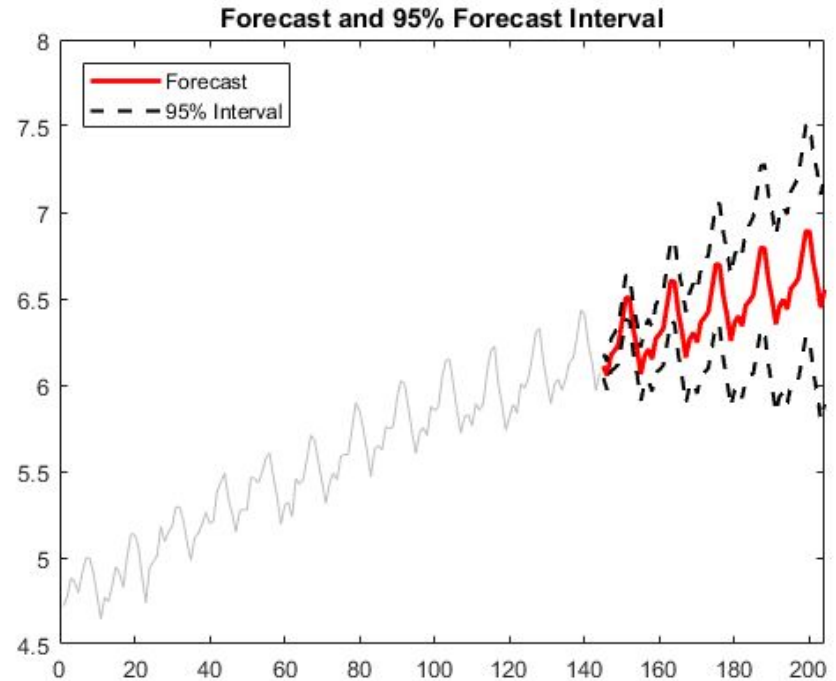
ARIMA - Preprocessing the Data

- DADA2 output yielded relative abundances for 1557 sequences
- However, ~560 sequences had no relative abundance measures across all samples.
- Analysis narrowed to only 994 sequences.
- Clustering was attempted to group together those sequences that showed similar abundance levels across the time duration of experiment.
- Preprocessing Steps:
  - Normalization ➜ (ΣRelative Abundance)/Max Abundance level
  - Euclidean distance map created
  - Distance map was analyzed to choose threshold for sequence similarity
  - 1 representative sample was chosen from cluster to perform ARIMA.

Threshold Selection for Sequence Similarity
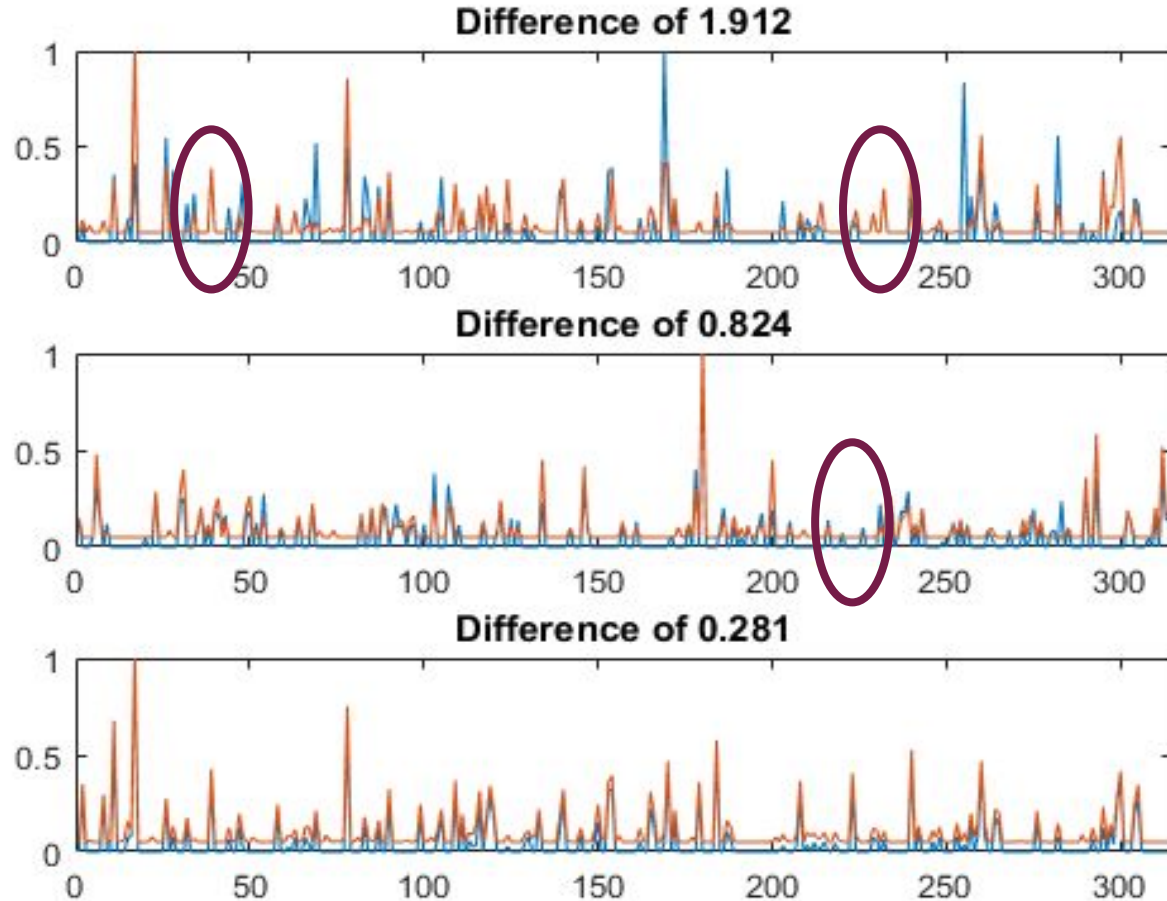
Similar Sequences
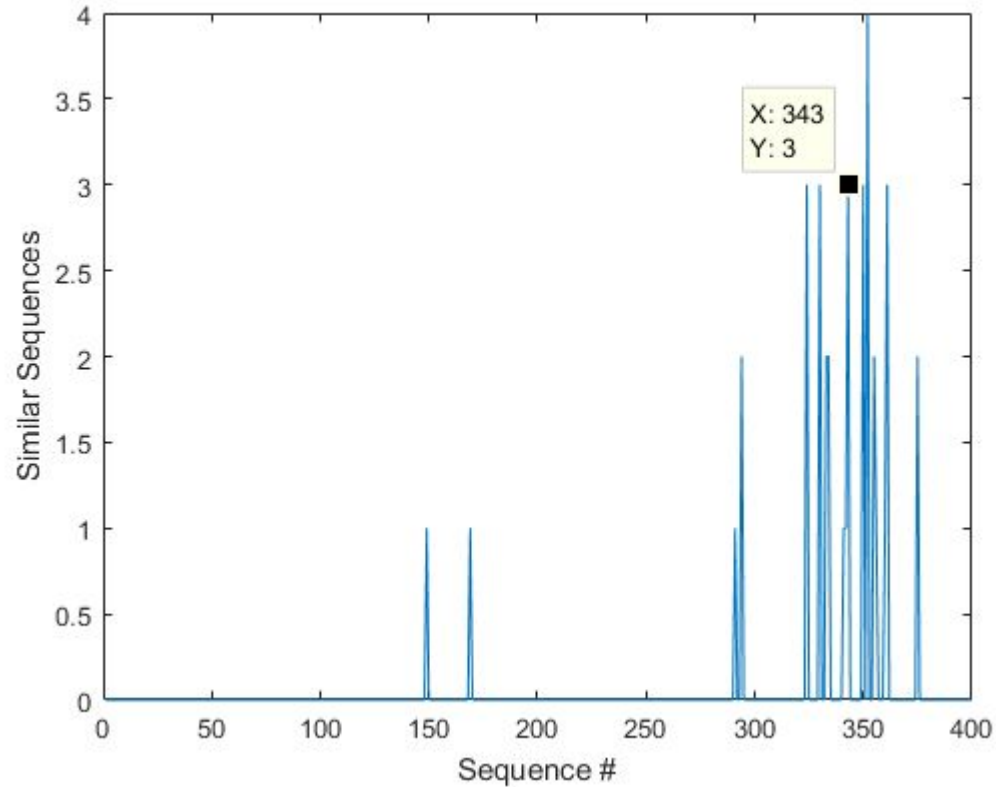
Sequence 343 has 3 other similarities

Analysis of 1 Clustered Group - *Clostridiales* Order
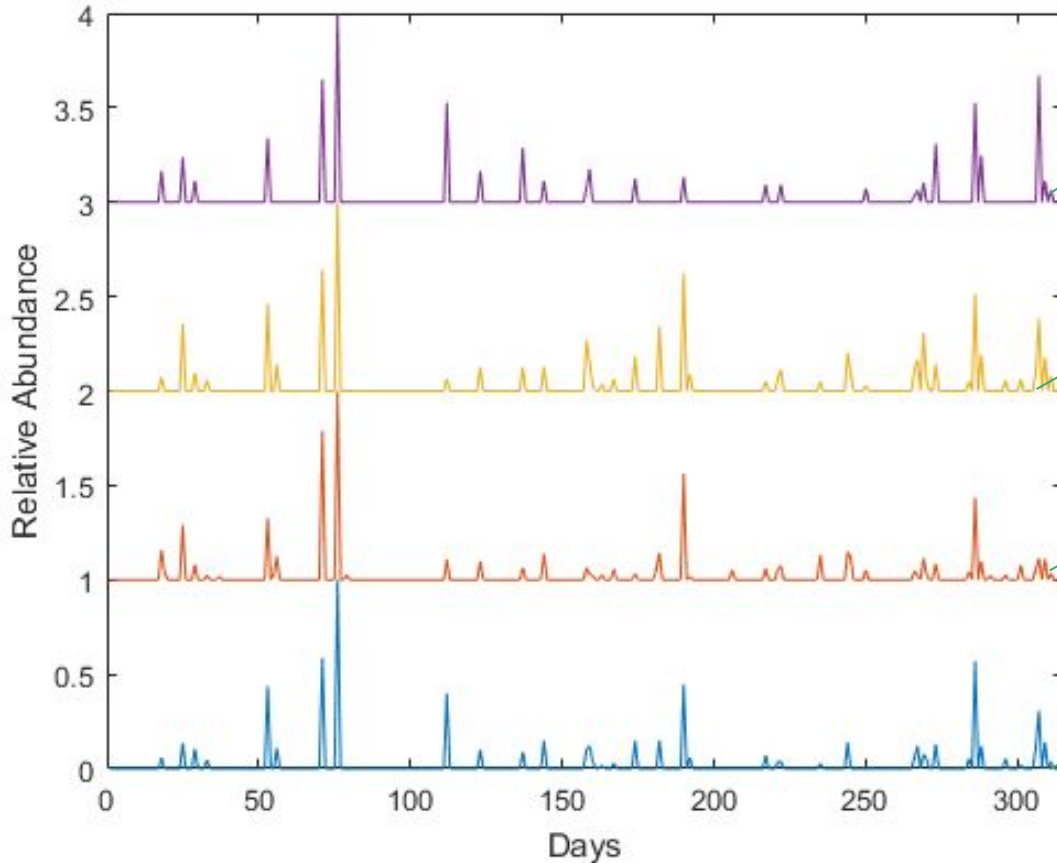
Ruminococcaceae, Ruminococcaceae_UCG-002

Family_XIII, [Eubacterium]_nodatum

NA, NA

Family_XIII, Family_XIII_AD3011

ARIMA Forecasting

- **B** = Subject went abroad
- **A** and **C** = "normal" data
- Given there is more data in Region C, use it as training set
- How closely can the model approximate the "normal" data in Region A?

ARIMA Prediction of Clostridiales Cluster

Discussion

- Bacterial composition were more similar within sites than between sites. However, there was greater variation within the gut composition compared to saliva.
- When comparing within sites, JSD index suggests different time periods when the gut compositions were more similar. We can attribute these changes to travel, sickness, and one unknown.
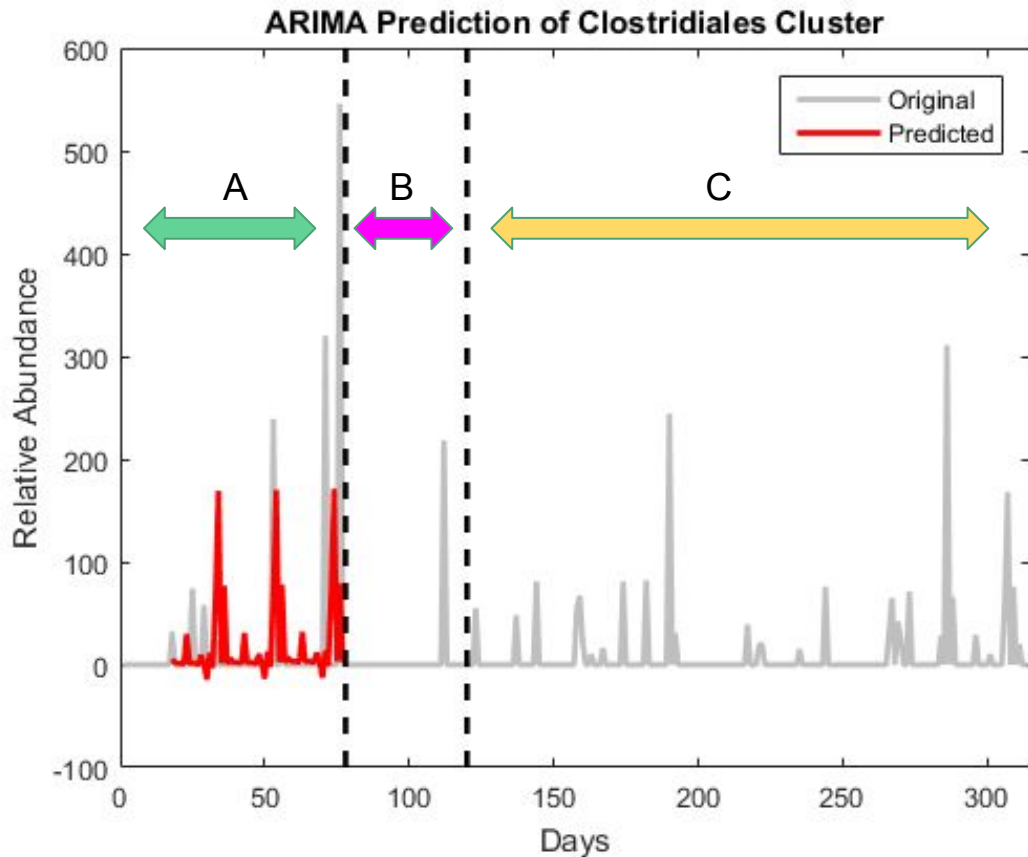- FFT Analysis could be useful for identifying which sequences belong to which families within a certain time span, but vary in usefulness over time based on the particular bacteria.
- A multiplicative ARIMA model can be used to model the relative abundances in the *Clostridiales* cluster in Donor A.

References

Amato, Katherine R., et al. "Habitat degradation impacts black howler monkey (Alouatta pigra) gastrointestinal microbiomes." *The ISME journal* 7.7 (2013): 1344-1353.

Callahan, Benjamin J., et al. "DADA2: high-resolution sample inference from Illumina amplicon data." *Nature methods* (2016).

Eren AM, Sogin ML, Morrison HG, Vineis JH, Fisher JC, Newton RJ, et al. "A single genus in the gut microbiome reflects host preference and specificity." ISME J. 2015;9:90–100.

Gomez, Andres, et al. "Gut microbiome composition and metabolomic profiles of wild western lowland gorillas (Gorilla gorilla gorilla) reflect host ecology." *Molecular ecology* 24.10 (2015): 2551-2565.

Grover, Madhusudan. "Role of gut pathogens in development of irritable bowel syndrome." *Indian Journal of Medical Research* 139.1 (2014): 11.

Lagier, Jean-Christophe et al. "Human Gut Microbiota: Repertoire and Variations." *Frontiers in Cellular and Infection Microbiology* 2 (2012): 136. *PMC*.