Developing a Broader Understanding of the Effects of Fecal Microbiota Transplantation on the Health and Microbiome of Donors and Patients with Recurrent Clostridiodes difficile Infection

by
Cailin Harris

A Thesis
presented to
The University of Guelph

In partial fulfillment of requirements
for the degree of
Master of Science
in
Bioinformatics

Guelph, Ontario, Canada
© Cailin Harris, May, 2020
This collection of papers investigates *Clostridiodes difficile* infection (CDI) and its treatment with fecal microbiota transplantation (FMT) from a more holistic perspective. Bacterial composition of FMT donors and patients with CDI at 4 timepoints (pre-FMT, 10 days, 5 weeks, and 13 weeks post-FMT) were analyzed from 16S rRNA sequence data. Trends in microbiome composition pre and post treatment with FMT were observed to establish which bacteria are responsible for curing patients of CDI and restoring quality of life. Bacterial composition was found to adjust over time after treatment to more closely resemble healthy donors. Donors were studied to identify if bacterial diversity remained stable between donors and were found to have no significant difference in Shannon diversity, indicating similarly diverse bacteria were supplied to each patient. By obtaining a more holistic view, we aim to understand the mechanisms of FMT better and produce precision care for patients suffering from CDI.
Acknowledgments

I would like to acknowledge Joyce Rousseau and Scott Weese for their extensive knowledge in both MOTHUR and bacterial sequencing. I would also like to acknowledge Peter Kim for his support and guidance in this project and Zeny Feng for her statistical expertise and input. I would also like to thank Augustine Wigle for her assistance with the laboratory work and Jesslyn Niergarth for her assistance with bioinformatic techniques. Lastly, I would also like to thank the Centre for Public Health and Zoonosis at the University of Guelph for the use of their laboratories.
Contents

Abstract ii
Acknowledgements iii
List of Tables vi
List of Figures vii

1 Introduction 1
1.1 Clostridiodes difficile Infection ................................. 1
1.2 Fecal Microbiota Transplantation .............................. 2
1.3 Engraftment and Augmentation ................................ 3
1.4 Holistic Treatment Approaches and Personalized Medicine  .......... 4
1.5 Research Contributions ........................................ 5
1.6 Thesis Layout .................................................. 5

2 Precision Medicine and Gut Dysbiosis 15

3 Understanding the Changes in Microbiome Composition of Clostridiodes difficile Patients Pre and Post Fecal Microbiota Transplantation 20
3.1 Abstract ...................................................... 20
  3.1.1 Importance .................................................. 21
3.2 Introduction ................................................... 21
3.3 Results ......................................................... 23
3.4 Discussion ...................................................... 27
3.5 Conclusion ...................................................... 31
3.6 Materials and Methods ..................................... 31
  3.6.1 Sample Collection ......................................... 31
  3.6.2 Extraction and Purification of Bacterial DNA ............... 31
  3.6.3 V4 16S rRNA Amplification via Polymerase Chain Reaction .... 32
  3.6.4 Sequencing ................................................... 34
  3.6.5 Sequence Data Processing .................................. 34
  3.6.6 Statistical Analysis ......................................... 34
4 Fecal Microbiota Transplantation Donor Stability Over Time

4.1 Abstract .................................................................................. 46
4.2 Introduction .............................................................................. 47
4.3 Materials and Methods .............................................................. 48
  4.3.1 Sample Collection ................................................................. 48
  4.3.2 Extraction and Purification of Bacterial DNA ....................... 49
  4.3.3 Amplification of the V4 16S rRNA via Polymerase Chain Reaction 49
  4.3.4 Sequencing .......................................................................... 51
  4.3.5 Sequence Data Processing ...................................................... 51
  4.3.6 Visualization Analyses ......................................................... 51
4.4 Results .................................................................................... 52
4.5 Discussion ............................................................................... 54
4.6 Conclusion .............................................................................. 57
4.7 Figures .................................................................................... 57

5 Conclusions and Further Work .................................................. 70

A MOTHUR Code Manuscripts 2 and 3 ....................................... 72
B R Code Manuscript 2 ................................................................. 82
C P-Values and R Code Manuscript 3 .......................................... 96
List of Tables

C.1 Tukey Pairwise Comparisons of Mean Shannon Diversities . . . . . . . . . . 97
# List of Figures

3.1 Shannon Diversity Plots .................................................. 35  
3.2 Phyla Abundances .......................................................... 36  
3.3 Genera Abundances .......................................................... 37  
4.1 Normal Quantile-Quantile Plots ............................................ 58  
4.2 Histograms of Residuals .................................................... 59  
4.3 Donor Phyla Over Time ...................................................... 60  
4.4 Donor Genera Over Time .................................................... 61
Chapter 1

Introduction

1.1 Clostridiodes difficile Infection

Clostridiodes difficile (formerly Clostridium difficile) infection (CDI) is a gastroenteric disease caused by the spore-forming bacillus Clostridiodes difficile.\textsuperscript{1–3} The bacterium typically infects individuals whose gut microbiome (the collection of microbial taxa and genes found in the gut) has been disrupted, often due to intense courses of antibiotics.\textsuperscript{3–8} The bacterium releases toxins into the gut which cause side effects including fever, abdominal pain, and even death.\textsuperscript{1–4,7} The side effects can be very severe and often cause immense disruption in the life of the patient. CDI is often acquired in hospital settings and as such, is one of the leading hospital-acquired infections in North America.\textsuperscript{4} Incidences of the disease have increased since 1993 and it is now the most common cause of antibiotic-associated diarrhea.\textsuperscript{9} In 2011 alone, the disease caused an estimated 453,000 infections in the United States, 29,000 of which resulted in death.\textsuperscript{9,10} In Canada, 37,900 cases were estimated to have occurred in 2012 alone.\textsuperscript{10} As such a prevalent disease, CDI also poses a burden to the Canadian economy due to in-hospital costs, medical costs in the community, lost productivity, and management of CDI relapses. An estimated $281 million is thought to go towards these CDI costs in Canada alone.\textsuperscript{10}
Some of the struggles posed to managing CDI are a result of its spores. Spores are difficult to sterilize properly and tend to be more resistant to typical cleaning procedures.\textsuperscript{11} Along with natural persistence, spores are also more prone to antibiotic resistance.\textsuperscript{11,12} As a result, many patients experiencing CDI become antibiotic resistant and require non-antibiotic forms of treatment.\textsuperscript{12} If CDI worsens or becomes more difficult to treat, it falls into two main categories of infection: refractory CDI or recurrent CDI. Refractory CDI occurs when the patient has persistent or worsening diarrhea along with persistent abdominal pain, fever, or high levels of peripheral white blood cells.\textsuperscript{13,14} If refractory, these symptoms will persist despite rigorous treatment with antibiotics. Recurrent infections are characterized when CDI symptoms return or reappear within 8 weeks of a completed CDI treatment.\textsuperscript{13,14} Often, these recurrences are due to reinfections, exacerbated by alterations and disturbances in the microbiome. Refractory and recurrent infections are quite common and add strain to the patients and the economy. Of the 37,900 patients infected with CDI in Canada in 2012, 21\% resulted in relapses of the disease and 27\% resulted in recurrences.\textsuperscript{10} Furthermore, once a patient has experienced a first recurrence, they are more than 60\% more likely to experience further recurrences.\textsuperscript{13} The more recurrences a patient has of the disease, the more likely it is to become antibiotic resistant.

\section*{1.2 Fecal Microbiota Transplantation}

The current treatment regime for infected patients is a course of broad-spectrum antibiotics.\textsuperscript{3–7} The antibiotics are typically metronidazole or vancomycin and work to reduce levels of \textit{Clostridiodes difficile} in the gut, ultimately eliminating symptoms and curing the patient of the disease.\textsuperscript{4,6,13} Other methods of treating CDI have been studied including exposure to spores produced by nontoxigenic \textit{C. difficile} and development of vaccines or monoclonal antibodies.\textsuperscript{13} Unfortunately, as recurrence rates increase and typical treatments fail, new,
readily available and effective treatments for CDI must be considered and investigated. One such treatment is fecal microbiota transplantation (FMT). FMT involves the transfer of a healthy donor’s stool, into the colon of the diseased patient.\textsuperscript{3,5,6,13} The typical method of transfer is through enema, though other methods of transfer have been of research interest. One such method of transfer is through lyophilization, where the fecal samples from the donors are freeze-dried and encapsulated.\textsuperscript{15} These resulting pills are then orally taken by the patient. Studies have also shown that there is little to no adverse effect to the success rate of FMT if the donor stool used is fresh or frozen.\textsuperscript{13} This indicates that treating infected patients can be done without the immediate need for a healthy donor. Lyophilized samples have also been found to remain stable over time.\textsuperscript{15} For the interest of the manuscripts presented, the donor samples used were either fresh or frozen, not lyophilized. FMT has been found to be incredibly successful in clinical trials with success rates (no recurrence after treatment) typically ranging from 80-90\%.\textsuperscript{13,16,17} Though it is known that the imprinting of the donor microbiome onto the patient is likely the cause of the reduction of symptoms and elimination of the infection, little is known about what mechanisms specifically, make FMT such a successful treatment and which specific bacteria, if any, are responsible for gut restoration.

### 1.3 Engraftment and Augmentation

Two words are typically used when discussing the transfer of bacteria from a donor to a patient with CDI: engraftment and augmentation. Engraftment is defined as newly detected operational taxonomic units (OTUs) present in the donors, but not detectable in the patient until after treatment.\textsuperscript{18} Augmentation is defined as OTUs whose levels increased 10-fold in the patient after treatment.\textsuperscript{18} Another term commonly used is depletion, which is used to describe OTUs whose levels decrease in the patient after treatment. Together, these
terms describe the pattern of bacterial changes that occur in the microbiome after receiving treatment with FMT. They can also be used to establish which bacteria are characteristic of a healthy gut which therefore, may be responsible for resolution of CDI and restoration of the gut microbiome.

1.4 Holistic Treatment Approaches and Personalized Medicine

Monitoring the gut microbiome can be challenging as it varies from person to person and can be affected by diet, environment, sickness, and genetics.\textsuperscript{19,20} It is for this reason that a thorough understanding of all components of FMT is necessary to provide effective treatment for infected patients. Through understanding the patient pain scores, microbiome composition, and donor stability over time, a larger understanding of what makes FMT successful can be established. Through investigating donor stability, we can determine what specifically makes a healthy donor and which donors are therefore ideal for treatment.\textsuperscript{19} This information can also be used in patient-donor matching, where an ideal donor is matched to a patient based on that specific patient’s microbial needs.\textsuperscript{19} Donor stability also infers that the transfer of bacteria should be fairly consistent each time, further helping us understand and identify which bacteria are responsible for resolution. Furthermore, by understanding which bacteria are responsible for a decrease in pain, improvement in quality of life, and restoring the gut microbiome of individuals with CDI, synthetic bacterial cocktails can be developed to target specific microbial deficiencies in the patients. Though FMT is relatively safe with proper donor screening, there is still risk of transferring pathogenic bacteria to an infected individual.\textsuperscript{19,21} By creating synthetic bacterial cocktails, treating patients for microbial deficiencies will be safer, minimizing the risk of pathogenic transfer. Synthetic
treatments are also desirable as they have less stigma associated with them than fecal enemas, making these treatment options more appealing to patients with the disease. Producing bacteria for such treatments is also easy regardless of location, and relatively inexpensive, meaning that treatments such as these can be easily used or supplied in rural or remote areas where properly screened, healthy donors may not be readily available. The combination of investigating all components of FMT in relation to CDI will help to provide precision medicine, targeted for individuals based on their specific needs, fast and effectively to resolve CDI.

1.5 Research Contributions

All of the laboratory procedures and computational analyses presented in this thesis were performed by myself with assistance in laboratory procedures from fellow lab members. This includes all extractions, PCR amplifications, gels, and purifications. These contributions also include following the MOTHUR MiSeq pipeline to analyze sequence data, R statistical analyses, and organizing the data sets for exploration and donor analyses. The three resulting papers from this research are intended to be published with one being successfully done so, one ready for submission, and the third being prepared for submission.

1.6 Thesis Layout

The layout of this thesis is as follows: Chapter 2 contains a publication in Healthcare Management Forum. This chapter describes trends in the phyla composition of patients after treatment with FMT and discusses how these changes are related to an improvement in quality of life for patients with CDI. After summarizing any necessary background information, trends in the phyla composition of patients over time are discussed. Following
this, the implications of these results are discussed as well as their implication in a broader context.

Chapter 3 contains a manuscript that is to be submitted to the journal, *MBio*. This chapter describes the phyla and genera composition of CDI patient microbiomes both pre and post-FMT. After summarizing any necessary background information, the composition of the patient microbiome is explored at both the phyla and genera taxonomic levels at timepoints both pre and post-FMT. In Section 3.4, the bacterial components are explained further as to their role in gut health and the implications this may have on developing personalized treatment.

Chapter 4 contains a manuscript that is to be submitted. This chapter investigates the stability of FMT donor microbiome composition. In Section 4.2, the background information regarding donor composition is discussed, followed by statistical analyses of the diversities of the microbial composition between donors. In Section 4.5, the implications of donor stability are discussed in relation to providing effective treatment as well as this information’s place in the analysis of FMT and its effect on patients with CDI.
References


6 Jonna Jalanka, Eero Mattila, Hanne Jouhten, Jorn Hartman, Willem M de Vos, Perttu Arkkila, and Reetta Satokari. Long-term effects on luminal and mucosal microbiota and


Chapter 2

Precision Medicine and Gut Dysbiosis

This chapter contains a manuscript that has been published in the journal, Healthcare Management Forum.
Precision medicine and gut dysbiosis

Cailin Harris, MSc(c)1; Peter T. Kim, PhD1,2; Dawn Waterhouse, PhD, MBA, CHE2; Zeny Feng, PhD1; Jessym Niergarth, PhD(c)1; and Christine H. Lee, MD2

Abstract

*Clostridioides difficile* Infection (CDI) is a leading cause of healthcare-associated infections in Canada, affecting the gastrointestinal tract which can lead to fever, abdominal pain, and diarrhea. Effective treatment for patients with Recurrent CDI (rCDI) can be achieved by Fecal Microbiota Transplantation (FMT) by introducing the gut micro-organisms of a healthy person (donor) into the bowel of the affected individual. Research has shown that an increase in the specific bacterial phyla post-FMT may be partly responsible for this gut restoration and elimination of disease. Furthermore, in understanding the key bacteria associated with successful FMT, full treatment plans can be developed for the individual needs of the patient by matching an infected individual with a donor possessing ideal microbiota for the specific patient. This development of precision medicine and more systematic adoption of FMT can be the next step toward more rapid resolution of rCDI.

Introduction

*Clostridioides difficile* (*C difficile*) Infection (CDI) is one of the most prevalent hospital-acquired infections and is the primary cause of antibiotic-associated diarrhea.1,2 *Clostridioides difficile* infection is a gastrointestinal disease caused by the anaerobic, spore-forming bacterium *C difficile*. The bacterium releases toxins into the bowel resulting in fever, abdominal pain, diarrhea, and in extreme cases, death.1,3 The infection typically affects those with compromised or altered gut bacteria following antibiotic treatment.2,4 The reduction of enteric bacteria following a course of antibiotics allows *C difficile* to colonize the gut at high levels.3 Challenges arise when patients become unresponsive to typical antibiotic treatments and experience refractory and/or recurring infections.5 These recurrences pose a challenge for healthcare providers, patients, and caregivers due to limited treatment options and compromised quality of life for both the patients and caregivers.6,5

Fecal Microbiota Transplantation (FMT) has shown to be approximately 90% effective in resolving Recurrent CDI (rCDI) by unknown mechanisms.8 However, studies have shown gut composition changes post-FMT, and the replenishment of gut micro-organisms from a donor into the bowel of the affected individual may partly explain its efficacy.3,5,8 This resolution improves the overall health and quality of life following FMT.7

The analyses of the microbiome of FMT donors and the recipient-patients, pre-FMT and post-FMT, may result in discovery of key bacteria associated with successful FMT and may lead to better understanding of the link between FMT and overall health. By obtaining a specific list of the microbes necessary for the restoration of the healthy gut microbiota, future research could lead to the administration of pooled beneficial bacteria, providing easily distributable, safer, and targeted microbial therapies for patients with rCDI. Furthermore, patients may also be matched to ideal donors based on required bacterial composition, employing the use of precision medicine to ensure higher success rates even when single dose of FMT is administered rather than requiring multiple FMTs as seen in some patients.

In this article, we provide an up-to-date literature review of the microbiome of donors and pre- and post-FMT of CDI patients. In addition, the findings from our own studies and microbiome changes to the restoration of quality of life in CDI patients are summarized.

Finally, despite robust evidence coupled with high demand, FMT remains unavailable in most healthcare facilities in Canada. We encourage healthcare leaders to employ the LEADS Framework9 and associated Develop Coalitions and Systems Transformation domains to increase partnership in research and implementation of this cost-effective and improved precision medicine treatment option for people with gut dysbiosis, including rCDI.

Patient gut composition pre-FMT and post-FMT

Patients who are infected with *C difficile* experience significant alterations to their gut flora (microbiota) due to shifts and changes instigated by the antibiotic(s) in majority of cases.1,3,10 With the distortion of healthy microbiota, *C difficile* is able to colonize, multiply, and produce toxins which are responsible for causing symptoms and signs of the infection.2,5 Of particular notice are the changes in overall bacterial abundances which are shown to be greatly decreased in CDI patients, suggesting a depletion of healthy

1 University of Guelph, Guelph, Ontario, Canada.
2 Vancouver Island Health Authority, Victoria, British Columbia, Canada.

Corresponding author:
Christine H. Lee, Vancouver Island Health Authority, Victoria, British Columbia, Canada.
E-mail: christine.lee@viha.ca
microbes and overall diversity.\textsuperscript{4,10} When observing the bacteria present in the gut of the patient at specific group levels, certain trends have been consistently observed. At the subclass level, Proteobacteria and Bacilli are found in high abundances in rCDI patients when compared to their noninfected counterparts.\textsuperscript{6,8,11-13} Upon further investigation, there are trends within these phyla that show elevated levels of the enterobacteria family along with \textit{Lactobacillus}, \textit{Enterococcus}, \textit{Streptococcus}, and \textit{Veillonella} genera.\textsuperscript{8,11,14} Bacteria that belong to the Firmicutes and Bacteroidetes phyla have shown to be at particularly low levels in CDI patients prior to FMT.\textsuperscript{12,13}

Following successful FMT, there are notable changes in the gut composition of patients.\textsuperscript{3,5-7} As a result of gut microbiota restoration, bacterial levels are converted to more complex and diverse microbiomes which more closely resemble that of the FMT donor.\textsuperscript{2-7,11,12} Although the overall bacterial compositions in cured patients vary, certain commonalities among treated patients have been noted. The following bacteria that belong in the Clostridia class: \textit{Butyrivibrio}, \textit{Eubacterium}, \textit{Anaerostipes}, \textit{Clostridium}, and \textit{Roseburia} genera consistently increase post-treatment.\textsuperscript{6,8,12,14} A distinct increase in \textit{Ruminococcus}, \textit{Bacteroides}, and \textit{Alistipes} has also been seen following successful FMT.\textsuperscript{6,8,13,14} On the contrary, Proteobacteria and Bacilli, especially \textit{Lactobacillus}, decrease following successful treatment. The balance of these key bacteria may be partly responsible for the restoration of health and improved quality of life in cured patients.\textsuperscript{11,12,14}

\textbf{Donor gut composition}

Studies of donor microbiomes have been conducted to determine the role of microbiota and FMT efficacy. Patients who were cured following FMT and their microbiota analyses showed more diverse gut composition similar to the donors compared to pre-FMT and also showed improvements in Health-Related Quality of Life (HRQoL), especially in improved pain scores, compared to their baseline or those who failed FMT.\textsuperscript{4,7,15} In the donor microbiome, the Actinobacteria, Bacteroidetes, Bacteroidia, and Firmicutes were the predominant phyla.\textsuperscript{6,11,13,14} The bacteria at the genus level which were absent in CDI patients initially but present in donors and later found in successfully cured patient were \textit{Alistipes}, \textit{Parabacteroides}, \textit{Prevotella}, \textit{Fecalibacterium}, \textit{Oscillaspira}, \textit{Sporobacter}, \textit{Ruminococcus}, \textit{Subdoligranulum}, \textit{Butyrivibrio}, and \textit{Eubacterium}.\textsuperscript{8}

\textbf{Donor and patient microbiome}

In this section, we describe some of the microbiome changes observed in patients and donors from the randomized controlled clinical trial of FMT for rCDI conducted in Ontario and British Columbia\textsuperscript{16} (see Figure 1). Within the Firmicutes, \textit{Verrucomicrobia}, Bacteria\_unclassified, Actinobacteria, and Bacteroidetes were shown to increase in relative abundances in patients from the very low levels present pre-FMT to significantly higher levels as time following FMT progressed. In the 10-day, 5-week, and 13-week timepoints post-FMT, those levels increased to levels that resemble the healthy donor. The phylum, Proteobacteria, reduced greatly from the high levels found in patients pre-FMT to the low levels found in the FMT donors. Verrucomicrobia and Bacteroidetes increased in patients post-FMT to the levels found in that of the donors. Across the six phyla: Firmicutes, Proteobacteria, \textit{Verrucomicrobia}, Bacteria\_unclassified, Actinobacteria, and Bacteroidetes, whether high or low to start in patients pre-FMT, all levels are found to resemble that of the donors as time progressed post-FMT. These changes in gut composition closely resembled the trends described in patients and donors, specifically the increased level of Firmicutes and Bacteroidetes and the decrease in the level of Proteobacteria.

\textbf{Defining health and HRQoL}

Historically, health was defined as an absence of illness but has since been modernized to represent a holistic and multidimensional concept that encompasses physical, mental, emotional, and social well-being.\textsuperscript{17,18} With the emphasis on the mental, emotional, and social aspects of health, a treatment to be successful must be able to restore the HRQoL of the patients and eliminate the physical illness.\textsuperscript{18,19} Health-related quality of life is defined as an individual’s ability to function and their perceived well-being in the physical, mental, and social domains of life and the impacts health has on that ability.\textsuperscript{17,20} Health-related quality of life can be divided into eight different domains: physical functioning, role limitations due to physical health, emotional well-being, role limitations due to emotional health, social functioning, energy/fatigue, general health perceptions, and bodily pain.\textsuperscript{17} Studies have shown that FMT for CDI has had significant positive impact on the eight domains of health, especially in reducing pain scores as early as day 10 from receiving FMT.\textsuperscript{7,16,21-24}

\textbf{Discussion}

As the leading cause of death due to infectious diarrhea in the United States and Canada, establishing safe and effective treatments for CDI is essential.\textsuperscript{2-4} Fecal microbiota transplantation is one such treatment option, and unlike antibiotic treatment, FMT restores the gut microbiota to a healthy state and prevents future CDI episodes.\textsuperscript{2-4,10,12,14} Although the exact bacteria responsible for this improvement are still being investigated, there is evidence to suggest that the increase in Firmicutes and Bacteroidetes as well as the decrease in Proteobacteria are key in restoring the gut microbiome and improving the overall HRQoL of the patients.\textsuperscript{10-12}

A thorough understanding of exact gut compositions (pre-FMT, post-FMT, and donor) may lead to establishing a true “microbiota-based” therapy for each individual patient. By selecting a donor with specific microbiome for an individual patient may lead to practicing precision and curing patients with one single FMT rather than multiple doses as is needed by some patients. However, further studies are needed to
establish the optimal route of administration, oral capsules, colonoscopy, retention enema and state of FMT, and frozen versus lyophilized (freeze-dried) formulation.

As we gain increased knowledge of the microbiome in general, and for FMT donors and recipients in particular, it is important to maximize the benefit of this research to patients as broadly as possible. We refer to the establishment in the United States, of the Openbiome, through which providers are able to access FMT for their patients. There is currently discussion about establishing a similar network in British Columbia. We point to the alignment here with the Develop Coalitions and Systems Transformation domains of the LEADS framework. By collaborating, by forming networks and consortia, we will be able to mobilize knowledge to bring this effective precision medicine treatment to a larger proportion of the population. We also entreat our health leaders to support innovative solutions such as FMT for various gut disorders associated with disrupted microbiota in addition to CDI by expanding intellectual capital and infrastructure to facilitate FMT.

We ask that leaders employ both Develop Coalitions and Systems Transformation capabilities in order to champion the type of change required to improve care for patients presenting with gut dysbiosis within the health system. This requires leaders to go beyond open and purposeful communication internal to individual facilities. What is required is:

- implementing systematic processes for identification of patients who will benefit from FMT;
- enabling the referral of those patients to providers trained in FMT and the more precise microbiome assessments as outlined in this article;
- encouraging and enabling collaborative research to further our knowledge and understanding of the microbiome and indeed other healthcare areas so as to optimize treatment options; and
- actively pursuing membership/collaboration with networks as outlined above.

By adhering to these recommendations, we believe that health leaders will take great strides in improving the health and quality of life for this and other, patient populations.

ORCID iD
Peter T. Kim, PhD  https://orcid.org/0000-0002-6398-4284

References


Chapter 3

Understanding the Changes in Microbiome Composition of Clostridiodes difficile Patients Pre and Post Fecal Microbiota Transplantation

3.1 Abstract

In recent years, the use of fecal microbiota transplantation (FMT) has been found to be an effective way to treat patients with Clostridiodes difficile infection (CDI). Despite its increase in popularity, the exact mechanism that makes FMT a successful treatment is still unknown. In this paper, we described the changes in the microbiome of 58 patients with recurrent or refractory CDI over 4 different timepoints (n=232): pre-FMT, day 10, week 5, and week 13 post-FMT. Healthy samples from 7 donors (n=193) were used as a
healthy baseline comparison. MOTHUR was used to determine taxonomic classification of the sequence data. Bacterial diversity was found to greatly increase post-FMT with the most noticeable increase at the day 10 timepoint. Our findings indicate that bacteria from the phyla Firmicutes and Actinobacteria may be augmented after treatment with FMT. Specifically, the family *Lachnospiraceae* and order Clostridiales from Firmicutes were noted to increase in relative abundance post-FMT. The phyla Verrucomicrobia and Bacteroidetes are shown to be engrafted into the patient microbiome from the donor microbiome and Proteobacteria is shown to deplete after treatment indicating an association with the infected gut. These results indicate that these bacteria may be key in restoring the gut microbiome of patients with CDI and in congress with health, metabolic, and virome data, can be used to develop donor matching and personalized bacterial treatments for patients.

### 3.1.1 Importance

Information regarding engrafted and augmented bacteria in patients with CDI following FMT treatment can lead to a wealth of understanding regarding the mechanisms of successful FMT. By obtaining a specific list of the microbes necessary for the restoration of the healthy gut microbiota, future research could lead to the development of a synthetic cocktail of bacteria, providing easily distributable, targeted microbial therapy for patients with CDI.

### 3.2 Introduction

*Clostridioides difficile* infection (CDI) is one of the most prevalent causes of infectious diarrhea in developed nations and is the leading cause of death by association.\(^1,2,4,22\) The infection is caused by the anaerobic, spore-forming bacillus *Clostridioides difficile* and results in extreme patient discomfort due to the release of toxins into the gut. Typical symptoms of infection include abdominal pain, diarrhea, and death in certain instances.\(^1-3\) CDI
mainly affects individuals over the age of 65 who have experienced gut microbiome disruption or depletion.\textsuperscript{2,4,22} This disruption is often stemmed from or follows the use of antibiotics which deplete the natural gut flora. The current treatment plan for CDI includes the use of broad-spectrum antibiotics which further act to disrupt the microbiota of the gastrointestinal tract.\textsuperscript{3-6} Upon completion of the antibiotic treatment, 15-30 percent of individuals will experience a recurrence of the disease.\textsuperscript{3,4,22} These recurrences pose a challenge for healthcare providers as patients become non-respondent to typical antibiotic treatment and require new means of treating the disease safely and effectively.\textsuperscript{5,6} In 2012, 37,900 CDI episodes were estimated to have occurred in Canada, 21\% of which were relapses and 27\% of which were recurrences.\textsuperscript{10} The rate of relapses along with the estimated $281 million that went towards CDI related occurrences, demonstrates the need for research into alternative and more effective treatments.\textsuperscript{2,10}

A method of combating CDI and preventing further recurrence is through fecal microbiota transplantation (FMT), a process that introduces the gut microflora of a healthy donor into the colon of the affected individual.\textsuperscript{3,5,6,13} FMT has been found to be a highly effective and successful therapy when compared to typical antibiotic treatments.\textsuperscript{3,5,6,22} Clinical resolution rates have been reported at 80-90\%.\textsuperscript{6,13,16,17} Despite the clinical efficacy of FMT, the mechanism by which FMT restores healthy gut bacteria is still widely unknown.\textsuperscript{16,23}

This study utilized samples from 58 patients at 4 timepoints (pre-FMT, 10 days post-FMT, 5 weeks post-FMT, and 13 weeks post-FMT) and 7 donors at multiple timepoints for a total of 232 and 193 samples, respectively (n=425 samples overall). Patients had experienced refractory or recurrent infections three times before being treated with FMT. Donors underwent a health screening process to ensure the safety of the patient receiving the fecal transplant. The aim of this study was to examine which bacteria have been engrafted or augmented following FMT through analyzing trends in the microbiome of FMT donors, and patients infected with CDI pre-FMT and post-FMT. Information regarding engrafted and
augmented bacteria in patients with CDI following successful FMT treatment can lead to a wealth of understanding regarding the mechanisms of successful FMT. By determining these bacteria, it is possible to identify which specifically are responsible for the restoration of the healthy gut microbiota following successful FMT treatment. The identified bacteria characteristic of the healthy gut microbiome can then be used for personalized medicine/treatments including targeted microbial therapies and strategic treatment plans for infected patients.\textsuperscript{24} With further investigation, this personalized treatment plan can also include donor matching (matching a patient to their ideal donor) depending on the patient’s bacterial needs to provide comprehensive, precise medical care.

### 3.3 Results

**Diversity increased in patients post-FMT.** In a broader attempt to understand changes in the microbiome of patients pre and post-FMT, Shannon’s diversity index was calculated for the phyla and genera. Through evaluating Shannon’s diversity, we confirmed that the diversity of the patients was lowest pre-FMT and increased over the 3 timepoints following treatment (day 10, week 5, and week 13) until the levels returned to similar levels as the donors (Figure 3.1). By week 13 post treatment, the diversity of the patient microbiome greatly increased to levels that much more closely resembled that of the donors which represent our healthy baseline diversity. Though the diversity of the patients had not been completely restored, the noticeable increase was consistent with the elimination of CDI symptoms. The microbiome of the patients was shown to most dramatically change in diversity from pre treatment to 10 days post treatment, after which, the diversity stayed relatively consistent with minor increases over the timepoints following (Figure 3.1).

**Levels of Firmicutes and Bacteroidetes increased in patients post-FMT.** The phylum Firmicutes was found to account for 53.66% of the phyla relative abundances and was
the most abundant phylum in patients and donors. Firmicutes was shown to be distinctly lower in patients pre-FMT than post treatment (Figure 3.2A). Consistent with other phyla, the change in the relative abundance of Firmicutes was most noticeable from pre treatment to 10 days post treatment. After this increase at 10 days post treatment, the levels of Firmicutes steadily increased to levels similar to the donors and were even found at a higher relative abundance than that found in the donors. Although the levels found in patients pre treatment were lower than donor and post treatment levels, Firmicutes were still at relatively high abundances when compared to the other phyla present in the patient microbiome pre-FMT.

Bacteroidetes, which was the sixth most prominent phylum, accounted for 0.75% of overall relative abundances. Levels of Bacteroidetes in patients pre-FMT were found to be very low, but increased in patients post-FMT (Figure 3.2G). These levels of relative abundance were found to stay fairly consistent even at 13 weeks post treatment. Interestingly, there were very low levels of Bacteroidetes in the donors which may indicate that higher levels are not typically indicative of a healthy gut microbiome.

Levels of Proteobacteria decreased in patients post-FMT. Proteobacteria was found to be the second most prominent phylum and accounted for 18.33% of the total relative abundances. The relative abundance levels of Proteobacteria were drastically higher in patients pre-FMT and were found to dramatically decrease at 10 days post treatment (Figure 3.2B). These levels continued to remain low and decreased at 5 weeks and 13 weeks post treatment. The donor levels of Proteobacteria were much lower than those displayed in patients pre-FMT and at 13 weeks post-FMT. Proteobacteria was the only phylum of bacteria that displayed this decrease in abundance in patients.

Nearly non-existent levels of Verrucomicrobia and Actinobacteria were present in patients pre-FMT. Verrucomicrobia was found to be the third most abundant phylum accounting for 14.34% of overall relative abundances. The relative abundance levels were
found to be negligible in patients pre-FMT but then appeared in patients after treatment and continued to rise in abundance at 5 weeks and 13 weeks post treatment (Figure 3.2C). Although found in relatively high abundances in the donors, the levels found in the patients post treatment were found to be higher than the levels found in the donors.

Once established in the gut post treatment, the abundance of Actinobacteria accounted for 5.17% of all relative abundances and was the fourth most prevalent phylum found in the gut microbiome. This phylum followed a similar trend as Verrucomicrobia with incredibly low counts pre-FMT and a steady increase in the timepoints post-FMT until levels more closely resembled the donors (Figure 3.2E).

**Other phyla present in the microbiome appeared to have no distinct trends.** The top 8 phyla that comprised 99.08% of the microbiome of the samples were Firmicutes, Proteobacteria, Verrucomicrobia, Actinobacteria, Fusobacteria, Bacteroidetes, Spirochaetes, and Fibrobacteres, respectively. Fusobacteria, Spirochaetes, Fibrobacteres, and all other phyla relative abundances appeared to have no distinct trends (Figure 3.2F, 3.2H, 3.2I, and 3.2J).

**Levels of unclassified Enterobacteriaceae, Escherichia/ Shigella, and Veillonella decreased in patients post-FMT.** Unclassified members of the *Enterobacteriaceae* family were found to have drastically higher relative abundances in patients pre-FMT than in the donors (Figure 3.3B). The relative abundances of these bacteria dropped as time progressed past the treatment and returned to levels that more closely resembled that of the donor levels. The most drastic drop in abundance occurred from pre-FMT to 10 days post-FMT. The levels then decreased slightly over time but remained fairly consistent at the week 5 and week 13 timepoints.

*Escherichia/Shigella* and *Veillonella* accounted for 3.71% and 2.95% of the overall relative abundances representing the third and fourth most abundant genera in the patient and donor relative abundances. These genera followed a similar trend to unclassified *Enterobac-*
teriaceae and were found to have higher abundances in the patients pre treatment than the donors which represent the healthy baseline (Figure 3.3F; Figure 3.3H). After treatment, the relative abundance levels of these genera decreased and more closely resembled the relative abundances of the donors. In the case of Veillonella, the abundances dropped to very low levels such that they were almost absent in the sample at 13 weeks post treatment.

Levels of unclassified Lachnospiraceae and unclassified Clostridiales increased in patients post-FMT. Unclassified members of the Lachnospiraceae family were found to be present in very low levels in patients pre-FMT treatment and increased dramatically 10 days post treatment (Figure 3.3D). The relative abundances peaked at 5 weeks post treatment and then reduced at 13 weeks post treatment to levels that more closely resembled the levels of the donors. Even with the decrease 13 weeks post treatment, the relative abundances of unclassified Lachnospiraceae were still higher in the patients than the donors.

The relative abundances of unclassified members of the Clostridiales order followed the same pattern as unclassified Lachnospiraceae and were present in low levels in patients pre-FMT and increased post-FMT until levels resembled that of the donors (Figure 3.3G). Unlike unclassified Lachnospiraceae, the relative abundance levels of the patients did not peak and instead, steadily increased after the initial large increase 10 days post treatment. The relative abundances increased post treatment until they resembled that of the healthy standard of the donor.

Nearly non-existent levels of Akkermansia, Streptococcus, Blautia, and unclassified Ruminococcaceae increased in patients post-FMT. The genus Akkermansia was found to be the most abundant genus accounting for 11.49% of the genera relative abundances. In patients pre-FMT, the levels of Akkermansia were almost zero, however, post treatment, they increased dramatically and continued to increase at 5 weeks and 13 weeks post treatment until the abundances exceeded the levels shown in the healthy donors (Figure 3.3A).
Blautia also followed a similar pattern as Akkermansia and started off at negligible levels and increased significantly after treatment (Figure 3.3I). Blautia was found to represent 2.47% of the total relative abundance of bacteria and was the fifth most abundant genus. Like Akkermansia the relative abundances of Blautia increased until they surpassed the levels of the donors.

Members of the Streptococcus genus and unclassified members of the Ruminococcaceae family followed similar trends as Akkermansia and Blautia. The Streptococcus genus was found to account for 4.21% of the genera relative abundances and was the second most abundant genus. Both Streptococcus and unclassified Ruminococcaceae appeared in negligible levels in patients pre treatment and then increased post treatment until the levels more closely resembled the donors (Figure 3.3C, 3.3E, and 3.3J). Though Streptococcus increased in levels, it did not increase to the same level as the donors. Unclassified Ruminococcaceae returned to levels that much more closely resembled the donors.

3.4 Discussion

The results from this study support the understanding that components of the gut microbiome are replenished after FMT and bacterial composition is restored to relative abundances and diversities that reflect the microbiome of the healthy donors. The composition of the patients at 13 weeks post treatment also reflects the composition of the donors. The most drastic change in the composition of the patient gut microbiome across all phyla and genera was at the day 10 post treatment timepoint which is consistent with the resolution of CDI and any associated symptoms. Due to their absence in patients pre-FMT and their increase in levels after the fact, Verrucomicrobia and Actinobacteria may be indicative of or associated with the healthy gut microbiome.

The bacteria which increased from nearly zero to much more significant relative abun-
dances post treatment reflect the idea of engraftment where the donor’s bacterial ecosystem is imprinting itself onto the patient.\(^\text{18}\) This pattern indicates that the Verrucomicrobia and Actinobacteria phyla may be engrafted onto the patient microbiome. The genera that follow this engraftment pattern are *Akkermansia*, *Streptococcus* and *Blautia* along with unclassified members of the *Ruminococcaceae* family. An increase in *Streptococcus*, *Akkermansia*, and *Ruminococcaceae* after FMT are consistent with what has been seen in other studies.\(^\text{6,16,18,27,28}\)

Patients have shown to have relatively high levels of *Verrucomicrobiaceae*, *Lachnospiraceae*, and *Ruminococcaceae* post-FMT.\(^\text{28}\) *Lachnospiraceae* and *Ruminococcaceae* families are common in healthy gut microbiota and tend to be depleted in patients with CDI.\(^\text{28}\) Many of these bacteria have functions which support immune system responses, cellular processes, and prohibit the growth of *Clostridiodes difficile*.\(^\text{29}\) Engrafted bacteria are provided by the donor, restoring protective gut microbes and are thought to be one of the reasons why FMT is highly effective.\(^\text{29}\)

The bacteria present in patients at low levels which then increased to much more substantial relative abundances are reflective of augmentation, where the addition of the donor microbiome helps to support pre-existing microbes, allowing them to flourish and increase in numbers.\(^\text{18}\) The bacterial phylum that follows this trend is Firmicutes. The majority of bacteria present in patients pre treatment when their relative abundances are low, and after the relative abundances of bacteria have increased post-FMT are found in the Firmicutes phylum. This trend has been found in other studies where there has specifically been an increase in the Clostridia class with an emphasis in the family *Lachnospiraceae*.\(^\text{6,30,31}\) Some of the genera found in these studies include several *Clostridium* species and *Lachnospira*.\(^\text{6,16}\)

Within this study, we too found that unclassified *Lachnospiraceae* were found to be augmented in patients post treatment as well as unclassified Clostridiales. With several studies emphasizing the increase in Firmicutes, an increase in certain types of Clostridia may be a large contributing factor to overall improved health after FMT.\(^\text{27,31}\)
Bacteroidetes is interesting in that it does not follow an engraftment or augmentation pattern. In other studies, an increase in Bacteroidetes is a common theme and is shown to be increased in donors as well as patients post-FMT, specifically, the Bacteroidia class which is especially depleted within Bacteroidetes.\textsuperscript{16,27,30}

Bacteria that decrease in levels post treatment follow a pattern of depletion and may infer that these bacteria are representative of an unhealthy gut microbiome, as they are present at higher levels in the diseased state. Depletion may occur due to competition of resources or due to attack by other microbes or phages present in the donors. Proteobacteria is found in high levels in the patients prior to FMT and unclassified \textit{Enterobacteriaceae} which are a class of Proteobacteria were found to also deplete. The high levels of unclassified \textit{Enterobacteriaceae} in patients pre treatment correspond to the high levels of Enterobacteria and reflect the trend of depletion as time progresses post treatment. Enterobacteria has also been found at elevated levels in patients compared to healthy individuals in similar studies.\textsuperscript{27,30} In other studies, the genera \textit{Escherichia/Shigella} and \textit{Veillonella} also follow a depletion trend and are found at increased levels in patients before treatment.\textsuperscript{16,27}

Many of the bacteria found to increase in the gut post-FMT are bacteria which support the production of short chain fatty acids (SCFA) or support the immune system. SCFAs are important in maintaining gut health as they act as intermediates in processes, are taken up by cells as nutrients, and can alter gene expression.\textsuperscript{32,33} As such, many of them regulate leukocyte functionality which work to reduce inflammation in the gut.\textsuperscript{32} Some of the bacteria found to produce SCFAs are Actinobacteria, \textit{Veillonella}, \textit{Blautia}, \textit{Ruminococcaceae}, \textit{Lachnospiracea}, and Clostridiales.\textsuperscript{34–40} Some bacteria also support a healthy immune system through increasing gene expression of immune responses or induce/support the production of immune cells.\textsuperscript{32} Bacteroidetes species are found to produce polysaccharide A, which induces regulatory T cell growth and cytokine production.\textsuperscript{41} \textit{Akkermansia} has been found to reduce gut barrier disruption by increasing the gene expression of certain immune functions.\textsuperscript{42}
Recent studies have worked to develop and obtain a more holistic view of what improves the gut health of patients after receiving FMT.\textsuperscript{43,44} Studying the virome and the metabolic profile of the patients and donors may also shed some light into understanding what makes FMT so successful.\textsuperscript{28,45} Work has already been done to analyze bacteriophages found in fecal samples of healthy individuals as well as in a patient with CDI.\textsuperscript{28,45} Though the impact of FMT on the gut virome has not yet been extensively studied, present studies have shown a wide variation between patients and donors, likely due to host-specific differences.\textsuperscript{28} The virome has also shown significant changes 6-7 months post-FMT.\textsuperscript{28} The rapid variation seen post-FMT is likely reflective of the adapting ecosystem in the gut during this period.\textsuperscript{28,45} This suggests that the virome changes dynamically with the microbiota of the gut, although its influence on human health is still not fully understood.

Although knowing the phyla potentially responsible for maintaining a healthy gut is useful in understanding the composition of the microbiome of patients with CDI, from that information alone it is difficult to establish which specific bacteria may be associated and utilized in recovery. Furthermore, many of the genera that may be associated with recovery of the healthy gut due to their increase are unclassified. This represents flaws in a few areas of our current understanding of microbes and the microbiome itself. Though many studies have been conducted in recent years regarding the microbiome and which bacteria may be associated with healthy maintenance, many of the studies show conflicting results or statistically insignificant bacterial abundances.\textsuperscript{28,31,46} Discrepancies in the literature may be simply because many of the microbes present in the gut have still been undefined. In relation to this, it is possible that bacterial DNA sequences have been obtained through laboratory practices but are unidentifiable because they are not yet part of current reference databases. This results in one of two potential errors: the sequence is either not matched to the reference database and is ultimately discarded, or the sequence is mismatched and classified incorrectly.
3.5 Conclusion

FMT has proven to be a successful method of treating patients with rCDI although little is known about the exact mechanisms by which this occurs. With many common species of bacteria found between both donor and post-FMT patients, it may be possible to create a targeted mix of bacteria for treatment. Insights into the potential role of short chain fatty acid-producing bacteria in recovery may also aid in the road to targeted treatments. Furthermore, an understanding of changes in the virome of patients suffering from CDI could lead to a greater understanding of the mechanism by which FMT is so successful. By combining microbiome, virome, health, and metabolic data, a more holistic view of patient health and understanding of patient needs can be established for personalized medicine.

3.6 Materials and Methods

3.6.1 Sample Collection

Stool samples were collected from 58 patients with rCDI from a recent clinical trial. Samples were collected prior to FMT treatment as well as 10 days, 5 weeks, and 13 weeks post treatment. This results in 4 samples per patient for a total of 232 samples. Potential FMT donors were screened for pathogens and other diseases before collection of stool samples. A total of 193 donor stool samples were collected from a total of 7 donors and were amalgamated for analyses. The total number of stool samples processed in this research was 425. Samples were maintained at -80°C.

3.6.2 Extraction and Purification of Bacterial DNA

Bacterial DNA was extracted from the stool samples of both patients and donors using the commercially provided kit: E.Z.N.A. Stool DNA kit by Omega Bio-tek. The concentration
of the DNA was then measured using a nanodrop to ensure that the concentrations are equal to or above 25 ng/mL. Concentration values less than this are unable to be sequenced by the Illumina MiSeq system. The resulting DNA was then cleaned and purified using AMPure beads and ethanol. The beads bind the bacterial DNA and the ethanol removes any proteins or eukaryotic DNA. The beads ensure that the resulting DNA is all bacterial and removes protein and eukaryotic DNA in the process. The resulting DNA was stored at -20°C.

3.6.3 V4 16S rRNA Amplification via Polymerase Chain Reaction

PCR amplification of the V4 hypervariable region of the 16S rRNA gene; a highly conserved portion of the bacterial genome necessary for translation, was performed using the forward primer S-D-Bact-0564-a-S-15/P5Adapter: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGAYTGGGYDTAAAGNG, and the reverse primer S-D-Bact-0785-b-A-18/P7Adapter: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTACNVGGGTATC-TAATCC. These primers were chosen because they amplify the V4 region specifically and also contain adapters for the sequencing system we used (Illumina). Following all processing stages, the resulting DNA from each sample was placed into its own well on a 96 well plate. Records of sample locations within each plate were documented.

2.5µL of bacterial DNA was added to 12.5µL of KAPA Ready Mix, 0.5µL of the forward and reverse primers, and 9µL of HyClone HyPureTM Molecular Biology Grade Nuclease-free and RNase-free Water. Mixture was centrifuged to mix. PCR conditions were as follows: denaturation for 3 minutes and 45 seconds at 94°C; annealing for 60 seconds at 53°C; extending for 60 seconds at 72°C; cycle through previous steps a total of 27 times before final extension phase for 10 minutes at 72°C. PCR products were stored at -20°C.

Confirmation of V4 region PCR amplification was performed using gel electrophoresis. 5µL of each PCR product was mixed with 2µL of loading dye and subsequently loaded
into a 1.5% agarose gel (1.5g of UltraPure™ agarose powder dissolved into 100mL of 0.5x TBE buffer and 5µL of RedSafe™). 5µL of GeneDireX® 100bp DNA Ladder was added to the outside wells to provide a scale and control. Gel was run in 0.5x TBE buffer at 150V, 80 milliamps, for 30 minutes. Positive amplification resulted in a bright band present at approximately 400bp.

The resulting DNA was then cleaned and purified using AMPure beads and ethanol. 20µL of AMPure was added to the PCR product and mixed via pipetting before being placed into a well of a 96 well plate. The mixture was left to sit for 5 minutes before being put on a magnetic plate and being left to rest for 2 minutes. The supernatant was removed and 200µL of 80% ethanol was added to the well for a 30 second wash before being removed. The ethanol was repeated and after the ethanol was removed, the plate was left to sit for 10 minutes to dry on the magnetic plate. The plate was then removed from the magnet and 32µL of Tris buffer was added to the well and mixed via pipetting before being left to sit for 2 minutes. The 96 well plate was then placed on the magnet again and after 2 minutes, the supernatant was collected for further use. The resulting DNA from each sample was placed into its own well on a 96 well plate. Records of sample locations within each plate were documented. The resulting DNA was stored at -20°C.

DNA was prepped for Illumina sequencing by a second PCR amplification that utilized primers location specific to a particular well on a 96 well plate. 5µL of purified 16S PCR product was added to 25µL of KAPA Ready Mix, 4µL of the appropriate N and S primers, and 12µL of HyClone HyPure™ Molecular Biology Grade Nuclease-free and RNase-free Water. Mixture was centrifuged to mix. PCR conditions were as follows: denaturation for 3 minutes and 30 seconds at 94°C; annealing for 30 seconds at 55°C; extending for 30 seconds at 72°C; cycle through previous steps a total of 8 times before final extension phase for 10 minutes at 72°C. PCR products were stored at -20°C. Following amplification, products were purified and run in an electrophoresis to confirm amplification as described previously.
3.6.3 Sequencing

Resulting purified and amplified DNA was then sent to the Genomics Facility at the University of Guelph, where it was amplified at 4-6 million reads per 96 well plate using the Illumina MiSeq system. This resulted in approximately 40,000 to 60,000 reads per sample in each well.

3.6.5 Sequence Data Processing

Using the bioinformatics software MOTHUR, the sequences were organized into OTUs based on the variation of sequences in the V4 region based on sequence identity. Commands used to process the data followed the outline specified in the MiSeq SOP28 and were altered to suit our data (Appendix A). The reference database used during bioinformatic analysis and bacterial identification was the Silva database, an online resource for quality checking and aligning ribosomal RNA sequence data. The reference file contains the full-length sequences and taxonomy references of 168,111 bacteria; 4,337 archaea; and 18,213 eukarya.

3.6.6 Statistical Analysis

Shannon diversities were determined to evaluate alpha diversity of the samples both for evenness and richness. The formula used to calculate the Shannon diversities is as follows:

\[- \sum_{j=1}^{S_i} p_j \ln p_j \quad (3.1)\]

where \( S_i \) is the total number of identified bacteria (at a given taxonomic rank) for each sample, \( j \) is the total number of stool samples from 1 to 425, and \( p_j \) is the proportion of bacterial
OTU found for sample $j$. Shannon diversity indices were determined using the diversity function of the vegan package for the statistical software R. Graphical representations of the diversity indices and bacterial phyla and genera were created in R.

3.7 Figures

Figure 3.1: Diversity of CDI patient microbiome increases after treatment with fecal microbiota transplantation. (A) Boxplot of Shannon diversity indices for donor and patient phyla at timepoints pre-FMT, 10 day, 5 week, and 13 week post-FMT. Higher indices indicate a higher level of bacterial diversity. (B) Boxplot of Shannon diversity indices for donor and patient genera at timepoints pre-FMT, 10 day, 5 week, and 13 week post-FMT. Higher indices indicate a higher level of bacterial diversity.
Figure 3.2: Relative abundances of top 8 bacterial phyla and unclassified bacteria (99.99% of total abundance) in CDI patients change to reflect donor composition after treatment with fecal microbiota transplantation. Boxplots of (A) Firmicutes, (B) Proteobacteria, (C) Verrucomicrobia, (D) unclassified bacteria, (E) Actinobacteria, (F) Fusobacteria, (G) Bacteroidetes, (H) Spirochaetes, (I) Fibrobacteres, and (J) all other phyla relative abundances in donors and patients at timepoints pre-FMT, 10 day, 5 week, and 13 week post-FMT. Higher values indicate higher levels of bacterial abundance.
Figure 3.3: Relative abundances of top 5 bacterial genera and unclassified Enterobacteriaceae, unclassified bacteria, unclassified Lachnospiraceae, unclassified Clostridiales, and unclassified Ruminococcaceae in CDI patients change to reflect donor composition after treatment with fecal microbiota transplantation. Boxplot of (A) Akkermansia, (B) unclassified Enterobacteriaceae, (C) unclassified bacteria, (D) unclassified Lachnospiraceae, (E) Streptococcus, (F) Escherichia/Shigella, (G) unclassified Clostridiales, (H) Veillonella, (I) Blautia, (J) unclassified Ruminococcaceae, and (K) all other genera relative abundances in donors and patients at timepoints pre-FMT, 10 day, 5 week, and 13 week post-FMT. Higher values indicate higher levels of bacterial abundance.
References


6 Jonna Jalanka, Eero Mattila, Hanne Jouhten, Jorn Hartman, Willem M de Vos, Perttu Arkkila, and Reetta Satokari. Long-term effects on luminal and mucosal microbiota and


Chapter 4

Fecal Microbiota Transplantation

Donor Stability Over Time

4.1 Abstract

With Clostridiodes difficile infection (CDI) representing the largest contributor to gastrointestinal disease and death by association, research efforts have gone towards finding new and effective treatments. Due to the propensity for the disease to become non-responsive to antibiotics, fecal microbiota transplantation (FMT) has become a popular method of treatment when traditional antibiotic regimes fail to cure the patient. Many studies have investigated the composition of the patient microbiome both pre and post treatment with FMT to establish which bacteria may be responsible for the restoration of gut health. However, little research has been done to examine donor stability between donors and over time. In this study, we aim to examine the stability of the donors’ microbiomes at the phylum and genus levels both over time and in relation to one another to establish if stable gut composition not only indicates gut health but indicates the restoration of healthy microbial communities to infected patients. Understanding the stability of the donors helps to provide
a holistic view of CDI and FMT and can aid in further understanding the mechanisms that make FMT successful. Through this, sophisticated patient-donor matching can be established and used to match a patient to their ideal donor based on their specific microbial needs. Furthermore, if donor composition is found to remain stable, this may aid in developing standardized screening procedures for FMT donors to maximize the safety of the procedure.

4.2 Introduction

*Clostridiodes difficile* is a spore-forming bacterium which inhabits the gut in high levels after traditional bacterial levels of the microbiome have been depleted, typically after extensive antibiotic use. Infection with this bacterium results in extreme discomfort, abdominal pain, diarrhea and can even cause death in rare circumstances. With *Clostridiodes difficile* infection (CDI) becoming a problematic cause of gastroenteric disease and the leading cause of death by association in industrialized nations, many efforts have gone into developing effective treatments. The standard course of treatment for CDI is a course of broad-spectrum antibiotics which work to diminish the levels of *Clostridiodes difficile* enough to eliminate infection. Unfortunately, reinfection is quite common which can lead to both recurrent (CDI symptoms return within 8 weeks of completed treatment) and refractory (worsening of CDI symptoms) infections. With recurrences occurring in approximately 30% of all cases (in Canada in 2012), establishing new and effective treatment methods is imperative.

In recent years, fecal microbiota transplantation (FMT) has become a popular option for CDI treatment due to its high efficacy rate (80-90% reported in clinical trials). FMT involves the transfer of healthy donor stool to the colon of the infected patient which results in the imprinting of the donor microbiome onto the patient. With the patient’s
microbiome composition returning to levels of bacteria found in healthy individuals, the infection is diminished, and patients are cured of symptoms. With its rise in popularity, understanding the composition of the patient microbiome pre and post-FMT has been an increasing field of study with the aim to develop bacterial cocktails to further improve the effectiveness of treatment and create an ease of use for all types of patients.\textsuperscript{3,24,25} Though many studies have been done to better understand the composition of the microbiome of patients with CDI, little research has been done on the stability of the donor microbiome.

Here we examined the microbiome of 193 samples from 7 different FMT donors across time from when they first donated, to the last. Dates of donation range from July 2012 to December 2015. Using these data, we investigated the stability of the microbiome of each of the 7 FMT donors used for resolution of CDI over time.

Donors were recruited and thoroughly screened for pathogens and diseases before being used in FMT. We examined the Shannon diversities of the overall phyla and genera composition of the donors as well as the Shannon diversities of each donor over the course of their donation dates at both the phylum and genus levels. Through understanding the gut microbiome of donors and their stability, we can establish and test for an ideal FMT donor.\textsuperscript{19}

\section*{4.3 Materials and Methods}

\subsection*{4.3.1 Sample Collection}

FMT donors underwent a screening process for pathogens and other diseases before collection of stool samples. A total of 193 stool samples were collected from a total of 7 donors. Samples were maintained at -80°C.
4.3.2 Extraction and Purification of Bacterial DNA

Bacterial DNA was extracted from the stool samples using the commercially provided kit: E.Z.N.A. Stool DNA kit by Omega Bio-tek. DNA concentration were then measured using a nanodrop. DNA concentrations had to be greater than or equal to 25 ng/mL as concentrations less than this could not be sequenced using the Illumina MiSeq system. The resulting DNA was then cleaned and purified using AMPure and ethanol. The AMPure beads bound the bacterial DNA and the ethanol removed unwanted proteins and eukaryotic DNA. 20µL of AMPure was added to the DNA and mixed via pipetting before being placed into a well of a 96 well plate and left for 5 minutes. The 96 well plate was then placed on a magnetic plate and left to rest for 2 minutes. The supernatant was removed and 200µL of 80% ethanol was added to the well for a 30 second wash and was then removed. Another ethanol wash was performed and then the plate was left to dry for 10 minutes on the magnetic plate. After being removed from the magnetic plate, 32µL of Tris buffer was added to each well and mixed via pipetting before being left to sit for 2 minutes. The 96 well plate was then placed on the magnetic plate for another 2 minutes before the supernatant was collected for further use. Resulting DNA was stored at -20°C.

4.3.3 Amplification of the V4 16S rRNA via Polymerase Chain Reaction

PCR amplification of the V4 hypervariable region of the 16S rRNA gene, was performed using the forward primer S-D-Bact-0564-a-S-15/P5Adapter: TCGTCCGAGAGTCAGAT-GTTATAAGAGACAGAYTGGGYDTAAAGNG, and the reverse primer S-D-Bact-0785-b-A-18/P7Adapter: GTCTCGGAGAGAGATGTGTAAGAGACAGTACNVGGGTATC-TAATCC. These primers contain adapters necessary for Illumina MiSeq and were chosen due to their V4 amplification specificity. The resulting DNA was placed into an individual well
on a 96 well plate.

2.5µL of bacterial DNA was added to 12.5µL of KAPA Ready Mix, 0.5µL of the forward and reverse primers, and 9µL of HyClone HyPureTM Molecular Biology Grade Nuclease-Free and RNase-Free Water. Mixture was centrifuged before being amplified under the following PCR conditions: denaturation for 3 minutes and 45 seconds at 94°C; annealing for 60 seconds at 53°C; extending for 60 seconds at 72°C; cycle through previous steps a total of 27 times before final extension phase for 10 minutes at 72°C. PCR products were stored at -20°C.

Gel electrophoresis was used to confirm V4 region amplification. 5µL of each PCR product was mixed with 2µL of loading dye and loaded into a 1.5% agarose gel (1.5g of UltraPureTM agarose powder dissolved into 100mL of 0.5x TBE buffer and 5µL of RedSafeTM). 5µL of GeneDireX 100bp DNA Ladder was added to the outside wells to provide a scale and control. Gel was run in 0.5x TBE buffer at 150V, 80 milliamps, for 30 minutes. Positive amplification resulted in a bright band present at approximately 400bp. DNA was then cleaned and purified using AMPure beads and ethanol as described in Section 4.3.2. The resulting DNA was stored at -20°C.

A second PCR amplification was then performed on the DNA to prep for Illumina sequencing. Well-specific primers, specific to a particular well on a 96 well plate, are used as follows: 5µL of purified 16S PCR product added to 25µL of KAPA Ready Mix, 4µL of the appropriate N and S primers, and 12µL of HyClone HyPureTM Molecular Biology Grade Nuclease-Free and RNase-Free Water. The resulting mixture was then mixed via centrifugation. PCR conditions were as follows: denaturation for 3 minutes and 30 seconds at 94°C; annealing for 30 seconds at 55°C; extending for 30 seconds at 72°C; cycle through previous steps a total of 8 times before final extension phase for 10 minutes at 72°C. PCR products were stored at -20°C. Following amplification, PCR products were purified as described in section 4.3.2 with the exception of using 50µL of AMPure instead of 20µL. A gel
electrophoresis was then performed as previously described to confirm amplification (Section 4.3.3).

### 4.3.4 Sequencing

Purified and amplified DNA was then sent to the Genomics Facility at the University of Guelph, where it was amplified at 4-6 million reads per 96 well plate using the Illumina MiSeq system. Each sample resulted in approximately 40,000 to 60,000 reads.

### 4.3.5 Sequence Data Processing

The bioinformatics software MOTHUR was used to organize the sequences into OTUs based on the variation of sequences in the V4 region. Commands used to process the data followed the outline specified in the MiSeq SOP28 and were utilized to suit our data (Appendix A). The Silva database was used as a reference in bioinformatic analyses containing full-length sequences and taxonomy references of 168,111 bacteria; 4,337 archaea; and 18,213 eukarya.

### 4.3.6 Visualization Analyses

Shannon diversities were determined to evaluate alpha diversity of the samples both for evenness and richness. The formula used to calculate the Shannon diversities is as follows:

$$-\sum_{j=1}^{S_i} p_j \ln p_j$$  \hspace{1cm} (4.1)

where $S_i$ is the total number of identified bacteria (at a given taxonomic rank) for each sample, $j$ is the total number of stool samples from 1 to 193, and $p_j$ is the proportion of bacterial OTU found in the donor for sample $j$. Shannon diversity indices were determined
using the diversity function of the vegan package for the statistical software R. Graphical representations (boxplots and time series plots) of the diversity indices were created using R.

Analysis of variance (ANOVA) and Tukey multiple pairwise comparisons of mean Shannon diversity indices were also performed in R to obtain p-values. Normal quantile-quantile plots and histograms of residuals for evaluating normality were created using R.

The Bonferroni correction was determined using the following formula:

\[
\text{adjusted } \alpha \text{ value} = \frac{\alpha}{\text{number of pairs}} \quad (4.2)
\]

where for our data, \( \alpha \) is set to 0.05 and there are 21 pairs of donor comparisons. This resulted in an adjusted \( \alpha \) value of 0.00238.

### 4.4 Results

**Summary of donor samples.** From the 7 donors, the following number of samples were collected: 20, 115, 8, 5, 15, 25, and 5 for donors A, B, C, D, E, F, and G, respectively. The number of donations varied as donors only provided samples when they volunteered to do so. Donor A had donations ranging from July 2012 to June 2013. Donor B had samples that ranged from July 2012 to December 2015. Donor C had samples ranging from July 2013 to December 2014. Donor D had samples that ranged from September 2012 to July 2013. Donor E had samples ranging from July 2013 to February 2014. Donor F had samples that ranged from February 2013 to December 2015. Donor G had samples ranging from November 2015 to December 2015.

**Donor bacterial diversity remained consistent across different individuals.** The mean Shannon diversity index was determined for each donor at both the phylum and
Analysis of variance (ANOVA) was performed to test if there was a significant difference in mean Shannon diversities at the phylum level among donors. To check that our data followed a normal distribution, a normal quantile-quantile plot and histogram were made from the model’s residuals. The normal quantile-quantile plot (Figure 4.1A) and the histogram (Figure 4.2A) indicated that our phyla diversities follow a normal distribution. The p-value determined from ANOVA analysis of the donor phyla was 0.5102. Though the p-value indicates that there is no significant evidence to suggest a difference in diversity between donors, Tukey pairwise comparisons were also performed to further investigate if the mean Shannon diversities at the phylum level were the same for each pair of donors. P-values of the donor pairs ranged from 0.5280322 to 1.0 (Table C.1).

The same ANOVA and Tukey procedures were used to analyze the Shannon diversity at the genus level of the donors. Normality checks were performed as done for the phyla diversities. The normal quantile-quantile plot (Figure 4.1B) and histogram (Figure 4.2B) showed that the genera diversities follow a mostly normal distribution with a slight left skew. The skew is very slight and the data mostly follows a normal distribution, satisfying the normality assumption. The p-value of the ANOVA analysis was found to be 0.000252. Tukey pairwise comparisons were performed to test if the mean Shannon diversities at the genus level were the same for each pair of donors. P-values of the donor pairs were found to range from 0.0143822 to 1.0 (Table C.1).

To correct for the increase of observing a rare event due to testing multiple comparisons, we used the Bonferroni correction to establish an adjusted α value which was found to be 0.00238. With this correction, none of our p-values from the Tukey multiple comparisons (for both the phyla and genera) were found to be significant. This means that there is not enough evidence to suggest that there is a difference between the bacterial diversities between each pair of donors. Therefore, these results suggest that both the phyla and genera Shannon diversity indices across donors do not significantly differ.
Donor phyla and genera diversities produced sporadic trends over time. The Shannon diversity index of each time point for each donor was calculated and plotted for the bacterial phyla and genera present in their microbiome (Figure 4.3 and Figure 4.4). Donor G appeared to have a relatively stable diversity over time at both the phylum and genus levels (Figure 4.3G; Figure 4.4G) as did Donor C at the genus level (Figure 4.4C). For all other donors at the phylum and genus levels, the diversities appeared to be very sporadic. In some cases such as Donor B, the diversities remain similar for small periods of time before the diversity changes (Figure 4.3B; Figure 4.4B). In most other cases, when observing graphical trends, the diversities appear to be very random, ranging greatly within each donor over time.

4.5 Discussion

Stability of the donor microbiome is essential in understanding the effects of FMT as it can lead to a further understanding of the bacteria necessary for gut microbiome restoration in patients with CDI. As FMT works to shift the gut microbiome of the patient towards the composition of the donor, donors whose gut composition remain stable over time, are more likely to transfer similar bacteria to patients with each transfer.\textsuperscript{19} The understanding that donor composition remains stable between individuals indicates that healthy donors who undergo proper screening procedures should be providing similarly diverse bacterial compositions to different patients. In this study, we determined that the trends in bacterial diversity change over time. As bacteria are likely to vary due to environmental and genetic factors, this variation is not unexpected.\textsuperscript{19,20} Furthermore, at the genus level, there is a greater likelihood of functional redundancy.\textsuperscript{49} Due to the overlap in the functions and roles of bacteria in the gut, some variability may occur while the gut is functioning healthily. Investigating the functional roles of bacteria present in each donor may help to identify
areas where this functional overlap is occurring and would be recommended for future study.

The stability in the donors also reinforces the importance of proper donor screening for FMT. Safety of FMT is a concern as any pathogenic bacteria present in the donors will be passed to the patients who are more susceptible to infections due to a depletion in microbial composition.\textsuperscript{21,50} If a donor was chosen whose composition was not stable when compared to other donors, this may be indicative of bacterial composition that is unfavorable for treatment. If a donor were found to be not stable across time or in relation to other donors, this may indicate sickness, disease, or other criteria that make that donor unsuitable. Current standards for donor selection for FMT are not regulated or standardized and is often based on criteria set by the doctor implementing treatment.\textsuperscript{21} Through understanding more about donor bacterial composition, these processes can be standardized across medical platforms.\textsuperscript{21}

In this study, we observed that the alpha diversity of the donor microbiome remains stable between donors. Beta and gamma diversities are other measures that can be used to further understand microbial composition. Beta diversity is used to determine the differences in species composition across multiple sites (in our case, donors) and gamma diversity measures the diversity of entire landscape (in our case across all donors).\textsuperscript{51} Though each measure of diversity is important in understanding bacterial landscapes, in this paper we focused on the measures of alpha diversity to investigate the composition of each donor individually. Measuring beta diversity can be determined in three different ways for this data which includes the diversity between samples, the diversity between donors, and the diversity between all samples across all donors. In the future, establishing which measure of beta diversity we wish to look at may assist in understanding more about the donor bacterial composition. Gamma diversity was not evaluated due to the computational difficulty to measure across 193 samples. For future research, investigating other measures of diversity more thoroughly will help to understand any similarities and relationships that the donor bacterial compositions may have with each other.
Characteristics of a healthy gut microbiome and gastrointestinal functionality has been correlated with a diverse and stable bacterial composition.\textsuperscript{52} When observing the results of the donors in our study, the Shannon diversity was found to be stable between donors, supporting that these are healthy donors with proper gastrointestinal functions. If stability is a measure of health, then monitoring the stability of donors is important in obtaining a more thorough understanding of the mechanisms of successful FMT. Some studies have even suggested the idea of super donors, who are found to result in more successful FMT treatments than other donors and ultimately, higher remission rates.\textsuperscript{50} These super donors may be identified by not only bacterial composition, but by stability over time and their overall diversity which is found to be crucial in restoring the patient microbiome.\textsuperscript{50}

The current model used in this paper tested whether the donors differ from each other but did not account for any intercorrelations for each donor. Since the mean of the Shannon diversity indices were being used in the ANOVA and Tukey pairwise comparison models for each donor, it is possible that there may be variation in the Shannon diversity at different donation timepoints for individual donors which are overlooked when the diversities are averaged. Looking at the averages may mask any variability in each donor and provide a less accurate depiction of each donor’s diversity when compared to others. In the future, utilizing a model which accounts for these relationships may help to develop a better understanding of each individual donor’s microbial composition. When observing the time plots of each donor, we do not see any distinct trends and they appeared to be quite random. Due to the variability in the number of samples per donor, with many of them being small with under 20 samples, it may be difficult to fairly evaluate trends over time. The time ranges of the donors also vary greatly with some donating in 2012 and others not providing their first donation until 2015. This variability in time lapses may also make comparing time-based trends for each donor difficult. However, if an appropriate model is determined, testing this randomness in the future would help to establish a greater understanding of the stability of
the donor microbiome over time.

4.6 Conclusion

Our findings suggest that microbial diversity remains stable across FMT donors, providing CDI patients with roughly equivalent microbial diversity through treatment regardless of donor. Understanding the donor stability is not only important for resolution of CDI, but for other gastrointestinal diseases as well. FMT has been used to treat irritable bowel syndrome, ulcerative colitis, inflammatory bowel disease, and other gastrointestinal disorders and so regulating screening processing and increasing understanding can help to maximize the efficiency of using FMT as a legitimate treatment option.\textsuperscript{19,50,53} Through understanding diversity and stability of donors over time, we can evaluate gut health of donor individuals and contribute to the overall understanding of FMT to develop more personalized medicine for those affected with CDI and other gastroenteric diseases. By implementing patient-donor matching, we can further increase the success rate of FMT and ideally match a patient based on their unique bacterial needs.\textsuperscript{19}

4.7 Figures
Figure 4.1: Normal quantile-quantile plots of ANOVA residuals show that the samples follow a normal distribution. (A) ANOVA residuals for the phyla Shannon diversities. (B) ANOVA residuals for the genera Shannon diversities.
Figure 4.2: Histograms of ANOVA residuals demonstrate that the samples follow a normal distribution with the mean centering around zero. (A) ANOVA residuals for the phyla Shannon diversities. (B) ANOVA residuals for the genera Shannon diversities.
Figure 4.3: Shannon diversity indices of microbial phyla across time from first donor sample to last donor sample. Higher indices indicate a higher level of bacterial diversity. (A) Shannon diversities for Donor A. (B) Shannon diversities for Donor B. (C) Shannon diversities for Donor C. (D) Shannon diversities for Donor D. (E) Shannon diversities for Donor E. (F) Shannon diversities for Donor F. (G) Shannon diversities for Donor G.
Figure 4.4: Shannon diversity indices of microbial genera across time from first donor sample to last donor sample. Higher indices indicate a higher level of bacterial diversity. (A) Shannon diversities for Donor A. (B) Shannon diversities for Donor B. (C) Shannon diversities for Donor C. (D) Shannon diversities for Donor D. (E) Shannon diversities for Donor E. (F) Shannon diversities for Donor F. (G) Shannon diversities for Donor G.
References


6 Jonna Jalanka, Eero Mattila, Hanne Jouhten, Jorn Hartman, Willem M de Vos, Perttu Arkkila, and Reetta Satokari. Long-term effects on luminal and mucosal microbiota and


To conclude, the manuscripts presented in this study work to view and analyze CDI and FMT from a more holistic perspective to gain a broader and more thorough understanding of what makes FMT successful. These collective works have demonstrated that donors provide stable and diverse microbiomes to patients with CDI. It is this diverse microbiome that is then imprinted onto the patient whereby some bacteria become augmented in the patient and some are engrafted. This engraftment and augmentation is ultimately what restores bacterial composition and relative abundance levels in CDI patients to those that more closely resemble healthy donors. Finally, it is this change in the microbiota of patients that works to relieve their symptoms, alleviating pain and improving their overall quality of life. By understanding all aspects of FMT treatment and the microbiome, we can work to produce and provide precision medicine which will work to relieve patients of disease quicker and more effectively. Furthermore, personalized medicine works to make treating patients with CDI a safe, cost effective, and easy procedure which is not only beneficial for patients and their health, but for the economy as well.

In the future, we aim to look at not only the microbiome of patients and donors, but the virome and metabolome as well. Currently, little is known about the exact effects of the gut
on health and studying the microbiome is a small piece of the overall story. By looking at bacteriophages in the gut and metabolites involved in FMT success, we can obtain a greater understanding of the disease and FMT to provide the most effective care possible for those suffering from CDI.
Appendix A

MOTHUR Code Manuscripts 2 and 3
Output files:

stability.files

nano makecontigs.sh
make.contigs(file=stability.files, processors=12)
chmod u=rwx makecontigs.sh
nohup ./mothur makecontigs.sh

Output files:

stability.trim.contigs.fasta
stability.scrap.contigs.fasta
stability.contigs.report
stability.contigs.groups

nano summary1.sh
summary.seqs(fasta=stability.trim.contigs.fasta)
chmod u=rwx summary1.sh
nohup ./mothur summary1.sh

It took 253 secs to summarize 54133127 sequences.

Output File Names:

stability.trim.contigs.summary

nano screenseqs1.sh
screen.seqs(fasta=stability.trim.contigs.fasta,
group=stability.contigs.groups, maxambig=0, minlength=239,
maxlength=243)
chmod u=rwx screenseqs1.sh
nohup ./mothur screenseqs1.sh

Output File Names:

stability.trim.contigs.good.fasta
stability.trim.contigs.bad.accnos
stability.contigs.good.groups

nano summary2.sh
summary.seqs(fasta=stability.trim.contigs.good.fasta)
chmod u=rwx summary2.sh
nohup ./mothur summary2.sh

Using 24 processors.

<table>
<thead>
<tr>
<th></th>
<th>Start</th>
<th>End</th>
<th>NBases</th>
<th>Ambigs</th>
<th>Polymer</th>
<th>NumSeqs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum:</td>
<td>1</td>
<td>239</td>
<td>239</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2.5%-tile:</td>
<td>1</td>
<td>239</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td>792788</td>
</tr>
<tr>
<td>25%-tile:</td>
<td>1</td>
<td>239</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td>7927874</td>
</tr>
<tr>
<td>Median:</td>
<td>1</td>
<td>240</td>
<td>240</td>
<td>0</td>
<td>5</td>
<td>15855748</td>
</tr>
<tr>
<td>75%-tile:</td>
<td>1</td>
<td>240</td>
<td>240</td>
<td>0</td>
<td>5</td>
<td>23783622</td>
</tr>
<tr>
<td>97.5%-tile:</td>
<td>1</td>
<td>243</td>
<td>243</td>
<td>0</td>
<td>6</td>
<td>30918708</td>
</tr>
<tr>
<td>Maximum:</td>
<td>1</td>
<td>243</td>
<td>243</td>
<td>0</td>
<td>13</td>
<td>31711495</td>
</tr>
<tr>
<td>Mean:</td>
<td>1</td>
<td>239</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td>31711495</td>
</tr>
</tbody>
</table>
# of Seqs:      | 31711495 |

It took 148 secs to summarize 31711495 sequences.

Output File Names:
stability.trim.contigs.good.summary

nano uniqueseqs.sh
unique.seqs(fasta=stability.trim.contigs.good.fasta)
chmod u=rwx uniqueseqs.sh
nohup ./mothur uniqueseqs.sh

Output File Names:
stability.trim.contigs.good.names
stability.trim.contigs.good.unique.fasta

nano countseqs.sh
count.seqs(name=stability.trim.contigs.good.names, group=stability.contigs.good.groups)
chmod u=rwx countseqs.sh
nohup ./mothur countseqs.sh

Output File Names:
stability.trim.contigs.good.count_table

nano summary3.sh
summary.seqs(fasta=stability.trim.contigs.good.unique.fasta, count=stability.trim.contigs.good.count_table)
chmod u=rwx summary3.sh
nohup ./mothur summary3.sh
Using 24 processors.

<table>
<thead>
<tr>
<th></th>
<th>Start</th>
<th>End</th>
<th>NBases</th>
<th>Ambigs</th>
<th>Polymer</th>
<th>NumSeqs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum:</td>
<td>1</td>
<td>239</td>
<td>239</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2.5%-tile:</td>
<td>1</td>
<td>239</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td>792788</td>
</tr>
<tr>
<td>25%-tile:</td>
<td>1</td>
<td>239</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td>7927874</td>
</tr>
<tr>
<td>Median:</td>
<td>1</td>
<td>240</td>
<td>240</td>
<td>0</td>
<td>5</td>
<td>15855748</td>
</tr>
<tr>
<td>75%-tile:</td>
<td>1</td>
<td>240</td>
<td>240</td>
<td>0</td>
<td>5</td>
<td>23783622</td>
</tr>
<tr>
<td>97.5%-tile:</td>
<td>1</td>
<td>240</td>
<td>240</td>
<td>0</td>
<td>6</td>
<td>30918708</td>
</tr>
<tr>
<td>Maximum:</td>
<td>1</td>
<td>243</td>
<td>243</td>
<td>0</td>
<td>13</td>
<td>31711495</td>
</tr>
<tr>
<td>Mean:</td>
<td>1</td>
<td>239</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td># of unique seqs:</td>
<td>8694877</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total # of seqs:</td>
<td>31711495</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It took 350 secs to summarize 31711495 sequences.

Output File Names:
- stability.trim.contigs.good.unique.summary
  
  ```
  nano alignseqs.sh  
  align.seqs(fasta=stability.trim.contigs.good.unique.fasta,  
             reference=silva.v4.fasta)  
  chmod u=rwx alignseqs.sh  
  nohup ./mothur alignseqs.sh
  ```

Output File Names:
- stability.trim.contigs.good.unique.align
- stability.trim.contigs.good.unique.align.report
- stability.trim.contigs.good.unique.flip.accnos
  
  ```
  nano summary4.sh  
  summary.seqs(fasta=stability.trim.contigs.good.unique.align,  
               count=stability.trim.contigs.good.count_table)  
  chmod u=rwx summary4.sh  
  nohup ./mothur summary4.sh
  ```

Using 24 processors.

<table>
<thead>
<tr>
<th></th>
<th>Start</th>
<th>End</th>
<th>NBases</th>
<th>Ambigs</th>
<th>Polymer</th>
<th>NumSeqs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum:</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2.5%-tile:</td>
<td>3082</td>
<td>13400</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td>792788</td>
</tr>
<tr>
<td>25%-tile:</td>
<td>3082</td>
<td>13400</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td>7927874</td>
</tr>
<tr>
<td>Median:</td>
<td>3082</td>
<td>13400</td>
<td>240</td>
<td>0</td>
<td>5</td>
<td>15855748</td>
</tr>
<tr>
<td>75%-tile:</td>
<td>3082</td>
<td>13400</td>
<td>240</td>
<td>0</td>
<td>5</td>
<td>23783622</td>
</tr>
<tr>
<td>97.5%-tile:</td>
<td>3082</td>
<td>13400</td>
<td>240</td>
<td>0</td>
<td>6</td>
<td>30918708</td>
</tr>
<tr>
<td>Maximum:</td>
<td>13425</td>
<td>13425</td>
<td>243</td>
<td>0</td>
<td>13</td>
<td>31711495</td>
</tr>
<tr>
<td>Mean:</td>
<td>3082</td>
<td>13399</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td># of unique seqs:</td>
<td>8694877</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>total # of seqs:</td>
<td>31711495</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It took 6063 secs to summarize 31711495 sequences.

Output File Names:
stability.trim.contigs.good.unique.summary

nano screenseqs2.sh
screen.seqs(fasta=stability.trim.contigs.good.unique.align, count=stability.trim.contigs.good.count_table, summary=stability.trim.contigs.good.unique.summary, start=3082, end=13400, maxhomop=8)
chmod u=rwx screenseqs2.sh
nohup ./mothur screenseqs2.sh

Output File Names:
stability.trim.contigs.good.unique.good.summary
stability.trim.contigs.good.unique.good.align
stability.trim.contigs.good.unique.bad.accnos
stability.trim.contigs.good.good.count_table

nano filterseqs.sh
filter.seqs(fasta=stability.trim.contigs.good.unique.good.align, vertical=T, trump=., processors=12)
chmod u=rwx filterseqs.sh
nohup ./mothur filterseqs.sh

Output File Names:
stability.filter
stability.trim.contigs.good.unique.good.filter.fasta

nano uniqueseqs2.sh
unique.seqs(fasta=stability.trim.contigs.good.unique.good.filter.fasta, count=stability.trim.contigs.good.good.count_table)
chmod u=rwx uniqueseqs2.sh
nohup ./mothur uniqueseqs2.sh

nano precluster.sh
pre.cluster(fasta=stability.trim.contigs.good.unique.good.filter.unique.fasta, count=stability.trim.contigs.good.unique.good.filter.unique.count_table, diffs=3)
chmod u=rwx precluster.sh
nohup ./mothur precluster.sh

Output File Names:
stability.trim.contigs.good.unique.good.filter.unique.precluster.fasta
stability.trim.contigs.good.unique.good.filter.unique.precluster.count_table
stability.trim.contigs.good.unique.good.filter.unique.precluster.010213A.map
stability.trim.contigs.good.unique.good.filter.unique.precluster.05071
3E.map
stability.trim.contigs.good.unique.good.filter.unique.precluster.05071
5B.map
stability.trim.contigs.good.unique.good.filter.unique.precluster.05091
2D.map
stability.trim.contigs.good.unique.good.filter.unique.precluster.05111
2B.map
stability.trim.contigs.good.unique.good.filter.unique.precluster.05121
3B.map
stability.trim.contigs.good.unique.good.filter.unique.precluster.06011
4B1.map
...

nano chimeravsearch.sh
chimera.vsearch(fasta=stability.trim.contigs.good.unique.good.filter.
unique.precluster.fasta,
count=stability.trim.contigs.good.unique.good.filter.unique.preclus
ter.count_table, dereplicate=t)
chmod u=rwx chimeravsearch.sh
nohup ./mothur chimeravsearch.sh

Output File Names:
stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.
vsearch.pick.count_table
stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.
vsearch.chimeras
stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.
vsearch.accnos

nano removeseqs.sh
remove.seqs(fasta=stability.trim.contigs.good.unique.good.filter.
unique.precluster.fasta,
accnos=stability.trim.contigs.good.unique.good.filter.unique.preclus
ter.denovo.vsearch.accnos)
chmod u=rwx removeseqs.sh

Output File Names:
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.
fasta

nano summary5.sh
summary.seqs(fasta=stability.trim.contigs.good.unique.good.filter.
unique.precluster.pick.fasta,
count=stability.trim.contigs.good.unique.good.filter.unique.preclus
ter.denovo.vsearch.pick.count_table)
chmod u=rwx summary5.sh
nohup ./mothur summary5.sh

Using 24 processors.
<table>
<thead>
<tr>
<th>Start</th>
<th>End</th>
<th>NBases</th>
<th>Ambigs</th>
<th>Polymer</th>
<th>NumSeqs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum:</td>
<td>1</td>
<td>513</td>
<td>232</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>2.5%-tile:</td>
<td>1</td>
<td>513</td>
<td>239</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>25%-tile:</td>
<td>1</td>
<td>513</td>
<td>239</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Median:</td>
<td>1</td>
<td>513</td>
<td>240</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>75%-tile:</td>
<td>1</td>
<td>513</td>
<td>240</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>97.5%-tile:</td>
<td>1</td>
<td>513</td>
<td>240</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Maximum:</td>
<td>2</td>
<td>513</td>
<td>243</td>
<td>0</td>
<td>8</td>
</tr>
</tbody>
</table>

Mean: 1 513 239 0 4

# of unique seqs: 3077564
Total # of seqs: 31044224

It took 168 secs to summarize 31044224 sequences.

Output File Names:

stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.summary

nano classifyseqs.sh
classify.seqs(fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta,
count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.count_table,
reference=trainset9_032012.pds.fasta,
taxonomy=trainset9_032012.pds.tax, cutoff=51, processors=12)
chmod u=rwx classifyseqs.sh
nohup ./mothur classifyseqs.sh

Output File Names:

stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.taxonomy
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.tax.summary

nano removelineage.sh
remove.lineage(fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.fasta,
count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.count_table,
taxonomy=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.taxonomy, taxon=Chloroplast-Mitochondria-unknown-Archaea-Eukaryota)
chmod u=rwx removelineage.sh

Output File Names:

stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.taxonomy
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.accnos

79
stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.pick.count_table
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta

nano summary6.sh
summary.seqs(fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta,
count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.pick.count_table)
chmod u=rwx summary6.sh
nohup ./mothur summary6.sh

Using 24 processors.

<table>
<thead>
<tr>
<th></th>
<th>Start</th>
<th>End</th>
<th>NBases</th>
<th>Ambigs</th>
<th>Polymer</th>
<th>NumSeqs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimum</td>
<td>1</td>
<td>513</td>
<td>232</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>2.5%-tile</td>
<td>1</td>
<td>513</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td>770361</td>
</tr>
<tr>
<td>25%-tile</td>
<td>1</td>
<td>513</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td>7703601</td>
</tr>
<tr>
<td>Median</td>
<td>1</td>
<td>513</td>
<td>240</td>
<td>0</td>
<td>5</td>
<td>15407201</td>
</tr>
<tr>
<td>75%-tile</td>
<td>1</td>
<td>513</td>
<td>240</td>
<td>0</td>
<td>5</td>
<td>23110801</td>
</tr>
<tr>
<td>97.5%-tile</td>
<td>1</td>
<td>513</td>
<td>243</td>
<td>0</td>
<td>6</td>
<td>30044041</td>
</tr>
<tr>
<td>Maximum</td>
<td>2</td>
<td>513</td>
<td>243</td>
<td>0</td>
<td>8</td>
<td>30814400</td>
</tr>
<tr>
<td>Mean</td>
<td>1</td>
<td>513</td>
<td>239</td>
<td>0</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

# of unique seqs: 3042597
total # of seqs: 30814400

It took 169 secs to summarize 30814400 sequences.

Output File Names:
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.summary

nano removegroups.sh
remove.groups(count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.pick.count_table,
          fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta,
          taxonomy=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.taxonomy,
groups=MC_S95-MC_S96-MCp4_S96-MCp5_S96)
chmod u=rwx removegroups.sh
remove.groups(count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.pick.count_table,
          fasta=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta,
          taxonomy=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.taxonomy,
groups=MC-MCp4-MCp5)
nohup ./mothur removegroups.sh
Output File names:
stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.pick.pick.count_table
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pick.fasta
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.pick.taxonomy

nano taxsummary.sh
summary.tax(taxonomy=stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.pick.taxonomy,
count=stability.trim.contigs.good.unique.good.filter.unique.precluster.denovo.vsearch.pick.pick.pick.count_table)
chmod u=rwx taxsummary.sh
nohup ./mothur taxsummary.sh

Output File Names:
stability.trim.contigs.good.unique.good.filter.unique.precluster.pick.pds.wang.pick.pick.tax.summary
Appendix B

R Code Manuscript 2
R Code for Plotting Figure 1

#### Diversity Plots Code for Phyla ####

#### Load packages ----

```r
library(vegan)
#vegan package is for the diversity command
```

#### Load data ----

```r
phyla <- read.csv(file.choose())
#the file being used for the phyla is phyla_counts.csv
#in this file, patients are indicated by rows and phyla are columns
```

#### Shannon diversity ----

```r
pshannon <- list(diversity(phyla[1:58,2:23], MARGIN = 1),
                 diversity(phyla[59:116,2:23], MARGIN = 1),
                 diversity(phyla[117:174,2:23], MARGIN = 1),
                 diversity(phyla[175:232,2:23], MARGIN = 1),
                 diversity(phyla[233:425,2:23], MARGIN = 1))
```

#this extracts each timepoint from the data and determines the
diversity for each individual timepoint, then converts all of that
information into a list format

#for extracting columns and rows: [1,2] where 1 is the row and 2 is
the column

#MARGIN indicates if it is pulling data (calculating diversity) for
rows or columns (1=rows, 2=columns)

```r
names(pshannon) <- c("Pre", "Day 10", "Week 5", "Week 13", "Donor")
#this changes the names in our list to the timepoint names
```

```r
boxplot(pshannon, main = "Shannon Diversity Index")
```

#### Diversity Plots Code for Genera ####

#### Load data ----

```r
genera <- read.csv(file.choose())
#the file being used for the genera is genera_counts.csv
#in this file, patients are indicated by rows and phyla are columns
```

#### Shannon diversity ----

```r
gshannon <- list(diversity(genera[1:58,-1], MARGIN = 1),
                 diversity(genera[59:116,-1], MARGIN = 1),
                 diversity(genera[117:174,-1], MARGIN = 1),
                 diversity(genera[175:232,-1], MARGIN = 1),
                 diversity(genera[233:425,-1], MARGIN = 1))
```

#this extracts each timepoint from the data and determines the
diversity for each individual timepoint, then converts all of that
information into a list format

#for extracting columns and rows: [1,2] where 1 is the row and 2 is
the column

#MARGIN indicates if it is pulling data (calculating diversity) for
rows or columns (1=rows, 2=columns)

```r
names(gshannon) <- c("Pre", "Day 10", "Week 5", "Week 13", "Donor")
#this changes the names in our list to the timepoint names
```
diversity(genera[233:425,-1], MARGIN = 1))

#this extracts each timepoint from the data and determines the diversity for each individual timepoint, then converts all of that information into a list format
#-1 indicates to take all columns except for the first one

names(gshannon) <- c("Pre", "Day 10", "Week 5", "Week 13", "Donor")
#this changes the names in our list to the timepoint names

boxplot(gshannon, main = "Shannon Diversity Index")

### Plotting diversities side by side ###

tiff("Shannon Diversities.tiff", units="in", width=9, height=5, res=300)

par(mfrow=c(1,2))

boxplot(pshannon,
  main = "Phyla Shannon Diversity Indices",
  cex.main = 0.9,
  col = c("firebrick2", "gold", "olivedrab3", "lightskyblue3", "plum3"),
  las = 2,
  ylab = "Diversity Index",
  xlab = "Timepoint",
  cex.axis = 0.75,
  cex.lab = 0.9)

mtext("A", adj=-0.1, padj=1.0)

boxplot(gshannon,
  main = "Genera Shannon Diversity Indices",
  cex.main = 0.9,
  col = c("firebrick2", "gold", "olivedrab3", "lightskyblue3", "plum3"),
  las = 2,
  ylab = "Diversity Index",
  xlab = "Timepoint",
  cex.axis = 0.75,
  cex.lab = 0.9)

mtext("B", adj=-0.1, padj=1.0)

dev.off()

#If saving as a png, use 700, by 500 margins

R Code for Plotting Figure 2

###Loading the datasets ----
Firmicutes <- read.csv(file.choose())
Proteobacteria <- read.csv(file.choose())
Verrucomicrobia <- read.csv(file.choose())
Bacteria_unclassified <- read.csv(file.choose())
Actinobacteria <- read.csv(file.choose())
Fusobacteria <- read.csv(file.choose())
Bacteroidetes <- read.csv(file.choose())
Spirochaetes <- read.csv(file.choose())
Fibrobacteres <- read.csv(file.choose())
Other_phyla <- read.csv(file.choose())

###For creating a plot with the 9 most relative abundances ----

#Set up save file and save naming convention
pdf("Phyla_Abundances_Barplots.pdf", paper="A4r", width = 50, height = 20)

#Set our plot layout
#Test to visualize where the plots will go
layout.matrix <- matrix(c(1, 4, 7, 2, 5, 8, 3, 6, 9, 0, 10, 0),
nrow = 3,
ncol = 4)

layout.matrix

layout(mat = layout.matrix,
       heights = c(2, 2, 2, 2),
       widths = c(2, 2, 2, 2))

par(mar=c(4.25, 4.25, 2.5, 2.5))
#This first part of the par function is needed for layout

#Firmicutes plot
Firmicutesplot <- boxplot(Firmicutes$Pre,
                          Firmicutes$X10.Days,
                          Firmicutes$X5.Weeks,
                          Firmicutes$X13.Weeks,
                          Firmicutes$Donor,
                          main = "Firmicutes",
                          xlab = "Timepoint",
                          names = c("Pre", "10 Days", "5 Weeks",
                                    "13 Weeks", "Donor"),
                          col = c("firebrick2", "gold", "olivedrab3",
                                  "lightskyblue3", "plum3"),
                          ylab = "Relative Abundance",
                          cex.main = 1.5,
                          cex.lab = 1.5,
                          cex.axis = 0.8)

mtext("A", adj=-0.2, padj=-1.0)

85
# Proteobacteria plot
Proteobacteriaplot <- boxplot(Proteobacteria$Pre, 
                               Proteobacteria$X10.Days, 
                               Proteobacteria$X5.Weeks, 
                               Proteobacteria$X13.Weeks, 
                               Proteobacteria$Donor, 
                               main = "Proteobacteria", 
                               xlab = "Timepoint", 
                               names = c("Pre", "10 Days", "5 Weeks", 
                                         "13 Weeks", "Donor"), 
                               col = c("firebrick2", "gold", 
                               "olivedrab3", 
                               "lightskyblue3", "plum3"), 
                               ylab = "Relative Abundance", 
                               cex.main = 1.5, 
                               cex.lab = 1.5, 
                               cex.axis = 0.8)

mtext("B", adj=-0.2, padj=-1.0)

# Verrucomicrobiplot
Verrucomicrobiaplot <- boxplot(Verrucomicrobia$Pre, 
                                 Verrucomicrobia$X10.Days, 
                                 Verrucomicrobia$X5.Weeks, 
                                 Verrucomicrobia$X13.Weeks, 
                                 Verrucomicrobia$Donor, 
                                 main = "Verrucomicrobia", 
                                 xlab = "Timepoint", 
                                 names = c("Pre", "10 Days", "5 Weeks", 
                                           "13 Weeks", "Donor"), 
                                 col = c("firebrick2", "gold", 
                                 "olivedrab3", 
                                 "lightskyblue3", "plum3"), 
                                 ylab = "Relative Abundance", 
                                 cex.main = 1.5, 
                                 cex.lab = 1.5, 
                                 cex.axis = 0.8)

mtext("C", adj=-0.2, padj=-1.0)

# Bacteria_unclassified plot
Bacteria_unclassifiedplot <- boxplot(Bacteria_unclassified$Pre, 
                                       Bacteria_unclassified$X10.Days, 
                                       Bacteria_unclassified$X5.Weeks, 
                                       Bacteria_unclassified$X13.Weeks, 
                                       Bacteria_unclassified$Donor, 
                                       main = "Bacteria_unclassified", 
                                       xlab = "Timepoint", 
                                      
                                      }
names = c("Pre", "10 Days", "5 Weeks", "13 Weeks", "Donor"),
col = c("firebrick2", "gold", "olivedrab3", "lightskyblue3", "plum3"),
ylab = "Relative Abundance",
cex.main = 1.5,
cex.lab = 1.5,
cex.axis = 0.8)

mtext("D", adj=-0.2, padj=-1.0)

#Actinobacteria plot
Actinobacteriaplot <- boxplot(Actinobacteria$Pre,
Actinobacteria$X10.Days,
Actinobacteria$X5.Weeks,
Actinobacteria$X13.Weeks,
Actinobacteria$Donor,
main = "Actinobacteria",
xlab = "Timepoint",
names = c("Pre", "10 Days", "5 Weeks", "13 Weeks", "Donor"),
col = c("firebrick2", "gold", "olivedrab3", "lightskyblue3", "plum3"),
ylab = "Relative Abundance",
cex.main = 1.5,
cex.lab = 1.5,
cex.axis = 0.8)

mtext("E", adj=-0.2, padj=-1.0)

#Fusobacteria plot
Fusobacteriaplot <- boxplot(Fusobacteria$Pre,
Fusobacteria$X10.Days,
Fusobacteria$X5.Weeks,
Fusobacteria$X13.Weeks,
Fusobacteria$Donor,
main = "Fusobacteria",
xlab = "Timepoint",
names = c("Pre", "10 Days", "5 Weeks", "13 Weeks", "Donor"),
col = c("firebrick2", "gold", "olivedrab3", "lightskyblue3", "plum3"),
ylab = "Relative Abundance",
cex.main = 1.5,
cex.lab = 1.5,
Bacteroidetes plot
Bacteroidetesplot <- boxplot(Bacteroidetes$Pre,
   Bacteroidetes$X10.Days,
   Bacteroidetes$X5.Weeks,
   Bacteroidetes$X13.Weeks,
   Bacteroidetes$Donor,
   main = "Bacteroidetes",
   xlab = "Timepoint",
   names = c("Pre", "10 Days", "5 Weeks",
             "13 Weeks", "Donor"),
   col = c("firebrick2", "gold",
           "olivedrab3",
           "lightskyblue3", "plum3"),
   ylab = "Relative Abundance",
   cex.main = 1.5,
   cex.lab = 1.5,
   cex.axis = 0.8)

Spirochaetes plot
Spirochaetesplot <- boxplot(Spirochaetes$Pre,
   Spirochaetes$X10.Days,
   Spirochaetes$X5.Weeks,
   Spirochaetes$X13.Weeks,
   Spirochaetes$Donor,
   main = "Spirochaetes",
   xlab = "Timepoint",
   names = c("Pre", "10 Days", "5 Weeks",
             "13 Weeks", "Donor"),
   col = c("firebrick2", "gold",
           "olivedrab3",
           "lightskyblue3", "plum3"),
   ylab = "Relative Abundance",
   cex.main = 1.5,
   cex.lab = 1.5,
   cex.axis = 0.8)

Fibrobacteres plot
Fibrobacteresplot <- boxplot(Fibrobacteres$Pre,
   Fibrobacteres$X10.Days,
   Fibrobacteres$X5.Weeks,
   Fibrobacteres$X13.Weeks,
   Fibrobacteres$Donor,
main = "Fibrobacteres",
xlab = "Timepoint",
names = c("Pre", "10 Days", "5 Weeks",
         "13 Weeks", "Donor"),
col = c("firebrick2", "gold",
       "olivedrab3",
       "lightskyblue3", "plum3"),
ylab = "Relative Abundance",
cex.main = 1.5,
cex.lab = 1.5,
cex.axis = 0.8)

mtext("I", adj=-0.2, padj=-1.0)

# Other phyla plot
OtherPhylaPlot <- boxplot(Other_phyla$Pre,
                         Other_phyla$X10.Days,
                         Other_phyla$X5.Weeks,
                         Other_phyla$X13.Weeks,
                         Other_phyla$Donor,
                         main = "Other Phyla",
xlab = "Timepoint",
names = c("Pre", "10 Days", "5 Weeks",
         "13 Weeks", "Donor"),
col = c("firebrick2", "gold", "olivedrab3",
       "lightskyblue3", "plum3"),
ylab = "Relative Abundance",
cex.main = 1.5,
cex.lab = 1.5,
cex.axis = 0.8)

mtext("J", adj=-0.2, padj=-1.0)

dev.off()

# If saving as a png, use 1200 by 850 margins

# Reset mar
par(mar=c(1,1,1,1))

R Code for Plotting Figure 3

### Loading the datasets ----

Akkermansia <- read.csv(file.choose())
Enterobacteriaceae_unclassified <- read.csv(file.choose())
Bacteria_unclassified <- read.csv(file.choose())
Lachnospiraceae_unclassified <- read.csv(file.choose())
Streptococcus <- read.csv(file.choose())
Escherichia_Shigella <- read.csv(file.choose())
Clostridiales_unclassified <- read.csv(file.choose())
Veillonella <- read.csv(file.choose())
Blautia <- read.csv(file.choose())
Ruminococcaceae_unclassified <- read.csv(file.choose())
Other_genera <- read.csv(file.choose())

# Plotting the top 9 abundances ----

#pdf("Genus_Abundances_Barplots.pdf", paper="A4r", width = 11, height = 8)

layout.matrix <- matrix(c(1, 5, 8, 2, 6, 9, 3, 7, 10, 4, 0, 11),
nrow = 3,
ncol = 4)
layout.matrix

layout(mat = layout.matrix,
heights = c(2, 2, 2, 2),
widths = c(2, 2, 2, 2))

par(mar=c(4.25, 4.25, 2.5, 2.5))

# Akkermansia plot
Akkermansiaplot <- boxplot(Akkermansia$Pre,
Akkermansia$X10.Days,
Akkermansia$X5.Weeks,
Akkermansia$X13.Weeks,
Akkermansia$Donor,
main = "Akkermansia",
names = c("Pre", "10 Days", "5 Weeks", "13 Weeks", "Donor"),
col = c("firebrick2", "gold", "olivedrab3", "lightskyblue3", "plum3"),
xlab = "Timepoint",
ylab = "Relative Abundance",
cex.main = 1.5,
cex.lab = 1.5,
cex.axis = 0.8)

mtext("A", adj=-0.125, padj=-1)

# Enterobacteriaceae_unclassified plot
Entero_unclassplot <- boxplot(Enterobacteriaceae_unclassified$Pre,
                            Enterobacteriaceae_unclassified$X10.Days,
Enterobacteriaceae_unclassified$X5.Weeks,
Enterobacteriaceae_unclassified$X13.Weeks,
                            Enterobacteriaceae_unclassified$Donor,
main = "Enterobacteriaceae_unclassified",
names = c("Pre", "10 Days", "5 Weeks", "13 Weeks", "Donor"),
col = c("firebrick2", "gold",
"olivedrab3",
"lightskyblue3", "plum3"),
xlab = "Timepoint",
ylab = "Relative Abundance",
cex.main = 1.5,
cex.lab = 1.5,
cex.axis = 0.8)
mtext("B", adj=-0.2, padj=-1)

#Bacteria_unclassified plot
Bacteria_unclassifiedplot <- boxplot(Bacteria_unclassified$Pre,
Bacteria_unclassified$X10.Days,
Bacteria_unclassified$X5.Weeks,
Bacteria_unclassified$X13.Weeks,
Bacteria_unclassified$Donor,
main = "Bacteria_unclassified",
names = c("Pre", "10 Days", "5 Weeks", "13 Weeks", "Donor"),
col = c("firebrick2", "gold",
"olivedrab3",
"lightskyblue3",
"plum3"),
xlab = "Timepoint",
ylab = "Relative Abundance",
cex.main = 1.5,
cex.lab = 1.5,
cex.axis = 0.8)
mtext("C", adj=-0.125, padj=-1)

#Lachnospiraceae plot
Lachnospiraceae_unclassifiedplot <- boxplot(Lachnospiraceae_unclassified$Pre,
Lachnospiraceae_unclassified$X10.Days,
Lachnospiraceae_unclassified$X5.Weeks,
Lachnospiraceae_unclassified$X13.Weeks,
Lachnospiraceae_unclassified$Donor,
main = "Lachnospiraceae_unclassified",
names = c("Pre", "10 Days", "5 Weeks", "13 Weeks", "Donor"),
col = c("firebrick2", "gold", "olivedrab3", "lightskyblue3", "plum3"),
xlab = "Timepoint",
ylab = "Relative Abundance",
cex.main = 1.5,
cex.lab = 1.5,
cex.axis = 0.8)
mtext("D", adj=-0.2, padj=-1)

# Streptococcus plot
Streptococcusplot <- boxplot(Streptococcus$Pre,
Streptococcus$X10.Days,
Streptococcus$X5.Weeks,
Streptococcus$X13.Weeks,
Streptococcus$Donor,
main = "Streptococcus",
names = c("Pre", "10 Days", "5 Weeks", "13 Weeks", "Donor"),
col = c("firebrick2", "gold", "olivedrab3", "lightskyblue3", "plum3"),
xlab = "Timepoint",
ylab = "Relative Abundance",
cex.main = 1.5,
cex.lab = 1.5,
cex.axis = 0.8)
mtext("E", adj=-0.125, padj=-1)

# Escherichia_Shigella plot
Escherichia_Shigellaplot <- boxplot(Escherichia_Shigella$Pre,
Escherichia_Shigella$X10.Days,
Escherichia_Shigella$X5.Weeks,
Escherichia_Shigella$X13.Weeks,
Escherichia_Shigella$Donor,
main = "Escherichia_Shigella",
names = c("Pre", "10 Days", "5 Weeks", "13 Weeks", "Donor"),
col = c("firebrick2", "gold", "olivedrab3", "lightskyblue3", "plum3"),
xlab = "Timepoint",
ylab = "Relative Abundance","}

92
cex.main = 1.5,
cex.lab = 1.5,
cex.axis = 0.8)

mtext("F", adj=-0.125, padj=-1)

#Clostridiales_unclassified plot
Clostridiales_unclassifiedplot <-
  boxplot(Clostridiales_unclassified$Pre,
          Clostridiales_unclassified$X10.Days,
          Clostridiales_unclassified$X5.Weeks,
          Clostridiales_unclassified$X13.Weeks,
          Clostridiales_unclassified$Donor,
          main = "Clostridiales_unclassified",
          names = c("Pre", "10 Days",
                    "5 Weeks",
                    "Donor"),
          col = c("firebrick2",
                  "olivedrab3",
                  "lightskyblue3",
                  "plum3"),
          xlab = "Timepoint",
          ylab = "Relative Abundance",
          cex.main = 1.5,
          cex.lab = 1.5,
          cex.axis = 0.8)

mtext("G", adj=-0.125, padj=-1)

#Viellonella plot
Veillonellaplot <- boxplot(Veillonella$Pre,
                            Veillonella$X10.Days,
                            Veillonella$X5.Weeks,
                            Veillonella$X13.Weeks,
                            Veillonella$Donor,
                            main = "Veillonella",
                            names = c("Pre", "10 Days", "5 Weeks",
                                       "13 Weeks", "Donor"),
                            col = c("firebrick2", "gold", "olivedrab3",
                                    "lightskyblue3", "plum3"),
                            xlab = "Timepoint",
                            ylab = "Relative Abundance",
                            cex.main = 1.5,
                            cex.lab = 1.5,
                            cex.axis = 0.8)
cex.lab = 1.5,
cex.axis = 0.8)

mtext("H", adj=-0.125, padj=-1)

# Blautia plot
Blautiaplot <- boxplot(Blautia$Pre,
                        Blautia$X10.Days,
                        Blautia$X5.Weeks,
                        Blautia$X13.Weeks,
                        Blautia$Donor,
                        main = "Blautia",
                        names = c("Pre", "10 Days", "5 Weeks",
                                  "13 Weeks", "Donor"),
                        col = c("firebrick2", "gold", "olivedrab3",
                                "lightskyblue3", "plum3"),
                        xlab = "Timepoint",
                        ylab = "Relative Abundance",
                        cex.main = 1.5,
                        cex.lab = 1.5,
                        cex.axis = 0.8)

mtext("I", adj=-0.125, padj=-1)

# Ruminococcaceae unclassified plot
Ruminococcaceae_Unclassplot <- boxplot(Ruminococcaceae_unclassified$Pre,
                                         Ruminococcaceae_unclassified$X10.Days,
                                         Ruminococcaceae_unclassified$X5.Weeks,
                                         Ruminococcaceae_unclassified$X13.Weeks,
                                         Ruminococcaceae_unclassified$Donor,
                                         main = "Ruminococcaceae Unclassified",
                                         names = c("Pre", "10 Days", "5 Weeks",
                                                   "13 Weeks", "Donor"),
                                         col = c("firebrick2", "gold", "olivedrab3",
                                                 "lightskyblue3", "plum3"),
                                         xlab = "Timepoint",
                                         ylab = "Relative Abundance",
                                         cex.main = 1.5,
                                         cex.lab = 1.5,
                                         cex.axis = 0.8)
mtext("J", adj=-0.2, padj=-1)

#Other Genera plot
Other_generaplot <- boxplot(Other_genera$Pre,
                      Other_genera$X10.Days,
                      Other_genera$X5.Weeks,
                      Other_genera$X13.Weeks,
                      Other_genera$Donor,
                      main = "Other Genera",
                      names = c("Pre", "10 Days", "5 Weeks",
                                "13 Weeks", "Donor"),
                      col = c("firebrick2", "gold",
                                "olivedrab3",
                                "lightskyblue3", "plum3"),
                      xlab = "Timepoint",
                      ylab = "Relative Abundance",
                      cex.main = 1.5,
                      cex.lab = 1.5,
                      cex.axis = 0.8)

mtext("K", adj=-0.125, padj=-1)

dev.off()

#If saving as a png, use 1200 and 850 margins

#Reset mar
par(mar=c(1,1,1,1))
Appendix C

P-Values and R Code Manuscript 3
Table C.1: The resulting p-values from the Tukey pairwise comparisons of donor pairs. The second column contains p-values resulting from comparing mean Shannon diversities at the phylum level. The third column contains p-values resulting from comparing mean Shannon diversities at the genus level.

<table>
<thead>
<tr>
<th>Donor Pairing</th>
<th>Phyla</th>
<th>Genera</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-A</td>
<td>0.9999814</td>
<td>0.0143822</td>
</tr>
<tr>
<td>C-A</td>
<td>0.9966299</td>
<td>0.1469995</td>
</tr>
<tr>
<td>D-A</td>
<td>0.8214370</td>
<td>0.6144732</td>
</tr>
<tr>
<td>E-A</td>
<td>0.9868671</td>
<td>0.0052037</td>
</tr>
<tr>
<td>F-A</td>
<td>0.9992095</td>
<td>0.9985825</td>
</tr>
<tr>
<td>G-A</td>
<td>0.9899778</td>
<td>0.0864298</td>
</tr>
<tr>
<td>C-B</td>
<td>0.9786760</td>
<td>0.9940613</td>
</tr>
<tr>
<td>D-B</td>
<td>0.8222551</td>
<td>1.0000000</td>
</tr>
<tr>
<td>E-B</td>
<td>0.8994895</td>
<td>0.6792159</td>
</tr>
<tr>
<td>F-B</td>
<td>0.9998090</td>
<td>0.0440729</td>
</tr>
<tr>
<td>G-B</td>
<td>0.9657894</td>
<td>0.8795409</td>
</tr>
<tr>
<td>D-C</td>
<td>0.6583701</td>
<td>0.9997626</td>
</tr>
<tr>
<td>E-C</td>
<td>1.0000000</td>
<td>0.9993833</td>
</tr>
<tr>
<td>F-C</td>
<td>0.9604324</td>
<td>0.2759497</td>
</tr>
<tr>
<td>G-C</td>
<td>0.9999937</td>
<td>0.9977844</td>
</tr>
<tr>
<td>E-D</td>
<td>0.5280322</td>
<td>0.9849948</td>
</tr>
<tr>
<td>F-D</td>
<td>0.9263823</td>
<td>0.7893644</td>
</tr>
<tr>
<td>G-D</td>
<td>0.6434084</td>
<td>0.9795855</td>
</tr>
<tr>
<td>F-E</td>
<td>0.8768121</td>
<td>0.0145275</td>
</tr>
<tr>
<td>G-E</td>
<td>0.9999944</td>
<td>0.9999840</td>
</tr>
<tr>
<td>G-F</td>
<td>0.9441048</td>
<td>0.1599911</td>
</tr>
</tbody>
</table>
### Determining Shannon Diversities for Phyla ###

#### Load packages ----

```r
library(vegan)
#vegan package is for the diversity command
```

#### Load data ----

```r
phyla <- read.csv(file.choose())
#the file being used for the phyla is donor_phylacounts.csv
#in this file, patients are indicated by rows and phyla are columns
```

#### Shannon diversity ----

```r
pshannon <- list(diversity(phyla[1:20,2:23], MARGIN = 1),
                  diversity(phyla[21:135,2:23], MARGIN = 1),
                  diversity(phyla[136:143,2:23], MARGIN = 1),
                  diversity(phyla[144:148,2:23], MARGIN = 1),
                  diversity(phyla[149:163,2:23], MARGIN = 1),
                  diversity(phyla[164:188,2:23], MARGIN = 1),
                  diversity(phyla[189:193,2:23], MARGIN = 1))
```

#this extracts each timepoint from the data and determines the diversity for each individual timepoint, then converts all of that information into a list format
#for extracting columns and rows: [1,2] where 1 is the row and 2 is the column
#MARGIN indicates if it is pulling data (calculating diversity) for rows or columns (1=rows, 2=columns)

```r
names(pshannon) <- c("A", "B", "C", "D", "E", "F", "G")
```

#this changes the names in our list to the timepoint names

### Determining Shannon Diversities for Genera ###

#### Load data ----

```r
genera <- read.csv(file.choose())
#the file being used for the genera is Donor_generacounts.csv
#in this file, patients are indicated by rows and phyla are columns
```

#### Shannon diversity ----

```r
gshannon <- list(diversity(genera[1:20,-1], MARGIN = 1),
                 diversity(genera[21:135,-1], MARGIN = 1),
                 diversity(genera[136:143,-1], MARGIN = 1),
                 diversity(genera[144:148,-1], MARGIN = 1),
                 diversity(genera[149:163,-1], MARGIN = 1),
                 diversity(genera[164:188,-1], MARGIN = 1),
                 diversity(genera[189:193,-1], MARGIN = 1))
```
diversity(genera[149:163,-1], MARGIN = 1),
diversity(genera[164:188,-1], MARGIN = 1),
diversity(genera[189:193,-1], MARGIN = 1))

#this extracts each timepoint from the data and determines the
diversity for each individual timepoint, then converts all of that
information into a list format
#-1 indicates to take all columns except for the first one

names(gshannon) <- c("A", "B", "C", "D", "E", "F", "G")
#this changes the names in our list to the timepoint names

###Performing ANOVA on phyla and genera ----

####phyla ----

pAdf <- as.data.frame(pshannon$A)
names(pAdf) <- c("diversity")
pAdf$label="A"

pBdf <- as.data.frame(pshannon$B)
names(pBdf) <- c("diversity")
pBdf$label="B"

pCdf <- as.data.frame(pshannon$C)
names(pCdf) <- c("diversity")
pCdf$label="C"

pDdf <- as.data.frame(pshannon$D)
names(pDdf) <- c("diversity")
pDdf$label="D"

pEdf <- as.data.frame(pshannon$E)
names(pEdf) <- c("diversity")
pEdf$label="E"

pFdf <- as.data.frame(pshannon$F)
names(pFdf) <- c("diversity")
pFdf$label="F"

pGdf <- as.data.frame(pshannon$G)
names(pGdf) <- c("diversity")
pGdf$label="G"

aovphylla <- rbind(pAdf, pBdf, pCdf, pDdf, pEdf, pFdf, pGdf)

model <- aov(diversity~label, data = aovphylla)
summary(model)

Analysis of Variance Table

99
Response: diversity

<table>
<thead>
<tr>
<th></th>
<th>label</th>
<th>Df</th>
<th>Sum Sq</th>
<th>Mean Sq</th>
<th>F value</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>label</td>
<td></td>
<td>6</td>
<td>0.486</td>
<td>0.081007</td>
<td>0.8807</td>
<td>0.5102</td>
</tr>
<tr>
<td>Residuals</td>
<td></td>
<td>186</td>
<td>17.109</td>
<td>0.091985</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#check to see if the residuals follow a normal distribution
par(mfrow=c(2,2))
plot(model)

#perform Tukey multiple comparisons if normal
TukeyHSD(model)

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = diversity ~ label, data = aovphyla)

$label

<table>
<thead>
<tr>
<th></th>
<th>diff</th>
<th>lwr</th>
<th>upr</th>
<th>p adj</th>
</tr>
</thead>
<tbody>
<tr>
<td>B-A</td>
<td>0.018047534</td>
<td>-0.2009846</td>
<td>0.2370796</td>
<td>0.9999814</td>
</tr>
<tr>
<td>C-A</td>
<td>-0.07636124</td>
<td>-0.4548381</td>
<td>0.3015659</td>
<td>0.9966299</td>
</tr>
<tr>
<td>D-A</td>
<td>0.206453417</td>
<td>-0.2455845</td>
<td>0.6584913</td>
<td>0.8214370</td>
</tr>
<tr>
<td>E-A</td>
<td>-0.080521753</td>
<td>-0.3893224</td>
<td>0.2282789</td>
<td>0.9868671</td>
</tr>
<tr>
<td>F-A</td>
<td>0.042480075</td>
<td>-0.2287426</td>
<td>0.3137028</td>
<td>0.9992095</td>
</tr>
<tr>
<td>G-A</td>
<td>-0.111979461</td>
<td>-0.5640173</td>
<td>0.3400584</td>
<td>0.989778</td>
</tr>
<tr>
<td>C-B</td>
<td>-0.094638659</td>
<td>-0.4252537</td>
<td>0.2358646</td>
<td>0.9786760</td>
</tr>
<tr>
<td>D-B</td>
<td>0.188405883</td>
<td>-0.2246050</td>
<td>0.6014168</td>
<td>0.8222551</td>
</tr>
<tr>
<td>E-B</td>
<td>-0.098569287</td>
<td>-0.3467580</td>
<td>0.1496194</td>
<td>0.8994895</td>
</tr>
<tr>
<td>F-B</td>
<td>0.024432541</td>
<td>-0.1750707</td>
<td>0.2239358</td>
<td>0.9998090</td>
</tr>
<tr>
<td>G-B</td>
<td>-0.130026995</td>
<td>-0.5430379</td>
<td>0.2829839</td>
<td>0.9657894</td>
</tr>
<tr>
<td>D-C</td>
<td>0.283089541</td>
<td>-0.2323129</td>
<td>0.7984920</td>
<td>0.6583701</td>
</tr>
<tr>
<td>E-C</td>
<td>-0.003885629</td>
<td>-0.3996877</td>
<td>0.3919165</td>
<td>1.0000000</td>
</tr>
<tr>
<td>F-C</td>
<td>0.119116199</td>
<td>-0.2481211</td>
<td>0.4863535</td>
<td>0.9604324</td>
</tr>
<tr>
<td>G-C</td>
<td>-0.035343337</td>
<td>-0.5507458</td>
<td>0.4800591</td>
<td>0.9999937</td>
</tr>
<tr>
<td>E-D</td>
<td>-0.286975170</td>
<td>-0.7538379</td>
<td>0.1798875</td>
<td>0.5280322</td>
</tr>
<tr>
<td>F-D</td>
<td>-0.16393342</td>
<td>-0.6068782</td>
<td>0.2789315</td>
<td>0.9263823</td>
</tr>
<tr>
<td>G-D</td>
<td>-0.318432878</td>
<td>-0.8902206</td>
<td>0.2533548</td>
<td>0.6434084</td>
</tr>
<tr>
<td>F-E</td>
<td>0.123001828</td>
<td>-0.1722681</td>
<td>0.4182717</td>
<td>0.8768121</td>
</tr>
<tr>
<td>G-E</td>
<td>-0.031457708</td>
<td>-0.4983204</td>
<td>0.4354050</td>
<td>0.9999944</td>
</tr>
<tr>
<td>G-F</td>
<td>-0.154459536</td>
<td>-0.5973644</td>
<td>0.2884453</td>
<td>0.9441048</td>
</tr>
</tbody>
</table>

###genera ----

gAdf <- as.data.frame(gshannon$A)
names(gAdf) <- c("diversity")
gAdf$label="A"

gBdf <- as.data.frame(gshannon$B)
names(gBdf) <- c("diversity")
gBdf$label="B"

gCdf <- as.data.frame(gshannon$C)
names(gCdf) <- c("diversity")
gCdf$label="C"

gDdf <- as.data.frame(gshannon$D)
names(gDdf) <- c("diversity")
gDdf$label="D"

gEdf <- as.data.frame(gshannon$E)
names(gEdf) <- c("diversity")
gEdf$label="E"

gFdf <- as.data.frame(gshannon$F)
names(gFdf) <- c("diversity")
gFdf$label="F"

gGdf <- as.data.frame(gshannon$G)
names(gGdf) <- c("diversity")
gGdf$label="G"

aovgenera <- rbind(gAdf, gBdf, gCdf, gDdf, gEdf, gFdf, gGdf)

modelg <- aov(diversity~label, data = aovgenera)

summary(modelg)

Df Sum Sq Mean Sq F value Pr(>F)
label 6 4.69 0.7811 4.543 0.000252 ***
Residuals 186 31.98 0.1719

#check to see if the residuals follow a normal distribution
par(mfrow=c(2,2))
plot(modelg)

#perform Tukey multiple comparisons if normal
TukeyHSD(modelg)

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = diversity ~ label, data = aovgenera)

$label
  diff   lwr   upr   p adj
B-A  0.34124107 0.04179097 0.64069116 0.0143822
C-A  0.44279561 -0.07426388 0.95985511 0.1469995
D-A  0.35321017 -0.26479413 0.97121447 0.6144732
E-A  0.52395108 0.10177388 0.94612840 0.0864298
F-A  0.06435939 -0.30644319 0.43516197 0.9985825
G-A  0.57520671 -0.04279759 1.19321101 0.0864298
R Code for Figure 3 and Phyla Median Indices

### Donor A phyla timepoint plots ----

```R
# load necessary packages
library(vegan)

# load necessary data files
```

---

```R
# determine the model residuals
eij = residuals(model)
geij = residuals(modelg)

# plot histograms of the residuals
par(mfrow = c(1, 2))

hist(eij, main = "Donor Phyla Diversity Residuals", cex.main = 0.75, xlab = "Residuals", ylab = "Frequency", xlim = c(-1, 1), las = 2)
mtext("A", padj = -1.2, adj = -0.3)

hist(eij, main = "Donor Genera Diversity Residuals", cex.main = 0.75, xlab = "Residuals", ylab = "Frequency", las = 2)
mtext("B", padj = -1.2, adj = -0.3)
```

---

```
R Code for Figure 3 and Phyla Median Indices

### Donor A phyla timepoint plots ----

# load necessary packages
library(vegan)

# load necessary data files
```
DonorA <- read.csv(file.choose())
DonorB <- read.csv(file.choose())
DonorC <- read.csv(file.choose())
DonorD <- read.csv(file.choose())
DonorE <- read.csv(file.choose())
DonorF <- read.csv(file.choose())
DonorG <- read.csv(file.choose())

# calculate Shannon diversities for each donor
Ashannon <- diversity(DonorA[,-(1:2)], MARGIN = 1)
Bshannon <- diversity(DonorB[,-(1:2)], MARGIN = 1)
Cshannon <- diversity(DonorC[,-(1:2)], MARGIN = 1)
Dshannon <- diversity(DonorD[,-(1:2)], MARGIN = 1)
Eshannon <- diversity(DonorE[,-(1:2)], MARGIN = 1)
Fshannon <- diversity(DonorF[,-(1:2)], MARGIN = 1)
Gshannon <- diversity(DonorG[,-(1:2)], MARGIN = 1)

# create dates for each donor
Adate <- as.Date(DonorA$Date)
Bdate <- as.Date(DonorB$Date)
Cdate <- as.Date(DonorC$Date)
Ddate <- as.Date(DonorD$Date)
Edate <- as.Date(DonorE$Date)
Fdate <- as.Date(DonorF$Date)
Gdate <- as.Date(DonorG$Date)

# plot the donor diversities over time
par(mar=c(3,3,3,3))
par(mfrow=c(4,2))

#### Donor A time plot ----

# create a vector of dates for the x axis
adate <- seq(as.Date("2012-07-10"), by = "week", length.out = 51)

plot(Adate, Ashannon, type = "l",
     main = "Donor A Phyla Shannon Diversities over Time",
     cex.main = 0.75,
     ylim = c(0,1.8),
     xlim = c(min(adate), max(adate)),
     xaxt = "n")

axis.Date(1,
          at=seq(min(adate),max(adate), by = "weeks"),
          format = "%d-%m-%y",
          las=2,
          cex.axis = 0.5)
points(Adate, Ashannon, pch = 19)
mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("A", padj = -1.2, adj = -0.1)

### Donor B time plot ----

bdate <- seq(as.Date("2012-07-20"), by = "month", length.out = 43)

plot(Bdate, Bshannon, type = "l",
     main = "Donor B Phyla Shannon Diversities over Time",
     cex.main = 0.75,
     cex.lab = 0.2,
     ylim = c(0,1.8),
     xlim = c(min(bdate), max(bdate)),
     xaxt = "n")

axis.Date(1,
          at=seq(min(bdate),max(bdate), by = "months"),
          format = "%m-%y",
          las=2,
          cex.axis = 0.8)

points(Bdate, Bshannon, pch = 19)

mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("B", padj = -1.2, adj = -0.1)

### Donor C time plot ----

cdate <- seq(as.Date("2013-07-19"), by = "month", length.out = 19)

plot(Cdate, Cshannon, type = "l",
     main = "Donor C Phyla Shannon Diversities over Time",
     cex.main = 0.75,
     ylim = c(0,1.8),
     xlim = c(min(cdate), max(cdate)),
     xaxt = "n")

axis.Date(1,
          at=seq(min(cdate),max(cdate), by = "months"),
          format = "%m-%y",
          las=2,
          cex.axis = 0.8)

points(Cdate, Cshannon, pch = 19)

mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("C", padj = -1.2, adj = -0.1)

### Donor D time plot ----
ddate <- seq(as.Date("2012-09-04"), by = "week", length.out = 47)

plot(Ddate, Dshannon, type = "l",
     main = "Donor D Phyla Shannon Diversities over Time",
     cex.main = 0.75,
     ylim = c(0,1.8),
     xlim = c(min(ddate), max(ddate)),
     xaxt = "n")

axis.Date(1,
          at=seq(min(ddate),max(ddate), by = "weeks"),
          format = "%d-%m-%y",
          las=2,
          cex.axis = 0.65)

points(Ddate, Dshannon, pch = 19)
mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("D", padj = -1.2, adj = -0.1)

### Donor E time plot ----

edate <- seq(as.Date("2013-07-05"), by = "week", length.out = 34)

plot(Edate, Eshannon, type = "l",
     main = "Donor E Phyla Shannon Diversities over Time",
     cex.main = 0.75,
     ylim = c(0,1.8),
     xlim = c(min(edate), max(edate)),
     xaxt = "n")

axis.Date(1,
          at=seq(min(edate),max(edate), by = "weeks"),
          format = "%d-%m-%y",
          las=2,
          cex.axis = 0.65)

points(Edate, Eshannon, pch = 19)
mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("E", padj = -1.2, adj = -0.1)

### Donor F time plot ----

fdate <- seq(as.Date("2013-02-01"), by = "month", length.out = 37)

plot(Fdate, Fshannon, type = "l",
     main = "Donor F Phyla Shannon Diversities over Time",
     cex.main = 0.75,
ylim = c(0,1.8),
xlim = c(min(fdate), max(fdate)),
xaxt = "n")

axis.Date(1,
at=seq(min(fdate),max(fdate), by = "months"),
format = "%m-%y",
las=2,
cex.axis = 0.8)
points(Fdate, Fshannon, pch = 19)
mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("F", padj = -1.2, adj = -0.1)

#### Donor G time plot ----

gdate <- seq(as.Date("2015-11-03"), by = "day", length.out = 32)

plot(Gdate, Gshannon, type = "l",
     main = "Donor G Phyla Shannon Diversities over Time",
     cex.main = 0.75,
     ylim = c(0,1.8),
     xlim = c(min(gdate), max(gdate)),
     xaxt = "n")

axis.Date(1,
at=seq(min(gdate),max(gdate), by = "days"),
format = "%d-%m-%y",
las=2,
cex.axis = 0.55)
points(Gdate, Gshannon, pch = 19)
mtext("Timepoint", padj = 19.9, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("G", padj = -1.2, adj = -0.1)

# if exporting as png, save with 700 by 700 dimensions

dev.off()

#### Determine medians of donors ----

Avector <- unlist(Ashannon)
Anumeric <- as.numeric(Ashannon)
median(Anumeric)
# 0.9037963

Bvector <- unlist(Bshannon)
Bnumeric <- as.numeric(Bshannon)

106
median(Bnumeric)
#1.051711

Cvector <- unlist(Cshannon)
Cnumeric <- as.numeric(Cshannon)
median(Cnumeric)
#0.8297564

Dvector <- unlist(Dshannon)
Dnumeric <- as.numeric(Dshannon)
median(Dnumeric)
#1.3596

Evector <- unlist(Eshannon)
Enumeric <- as.numeric(Eshannon)
median(Enumeric)
#0.8356209

Fvector <- unlist(Fshannon)
Fnumeric <- as.numeric(Fshannon)
median(Fnumeric)
#1.041363

Gvector <- unlist(Gshannon)
Gnumeric <- as.numeric(Gshannon)
median(Gnumeric)
#0.9214965

R Code for Figure 4 and Genera Median Indices

####Donor Genera timepoint plots ----

#load necessary packages
library(vegan)

#load necessary data files
DonorA <- read.csv(file.choose())
DonorB <- read.csv(file.choose())
DonorC <- read.csv(file.choose())
DonorD <- read.csv(file.choose())
DonorE <- read.csv(file.choose())
DonorF <- read.csv(file.choose())
DonorG <- read.csv(file.choose())

#calculate shannon diversities for each donor
Ashannon <- diversity(DonorA[, -(1:2)], MARGIN = 1)
Bshannon <- diversity(DonorB[, -(1:2)], MARGIN = 1)
Cshannon <- diversity(DonorC[, -(1:2)], MARGIN = 1)
Dshannon <- diversity(DonorD[, -(1:2)], MARGIN = 1)
Eshannon <- diversity(DonorE[, -(1:2)], MARGIN = 1)
Fshannon <- diversity(DonorF[,-(1:2)], MARGIN = 1)
Gshannon <- diversity(DonorG[,-(1:2)], MARGIN = 1)

# create date frames for each donor
Adate <- as.Date(DonorA$Date)
Bdate <- as.Date(DonorB$Date)
Cdate <- as.Date(DonorC$Date)
Ddate <- as.Date(DonorD$Date)
Edate <- as.Date(DonorE$Date)
Fdate <- as.Date(DonorF$Date)
Gdate <- as.Date(DonorG$Date)

# plot the donor diversities over time
par(mar=c(3,3,3,3))
par(mfrow=c(4,2))

#### Donor A time plot ----
# create a vector of dates for the x axis
adate <- seq(as.Date("2012-07-10"), by = "week", length.out = 51)
plot(Adate, Ashannon, type = "l",
     main = "Donor A Genera Shannon Diversities over Time",
     cex.main = 0.75,
     ylim = c(0.5,3.5),
     xlim = c(min(adate), max(adate)),
     xaxt = "n")
axis.Date(1,
         at=seq(min(adate),max(adate), by = "weeks"),
         format = "%d-%m-%y",
         las=2,
         cex.axis = 0.5)
points(Adate, Ashannon, pch = 19)

mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("A", padj = -1.2, adj = -0.1)

#### Donor B time plot ----

bdate <- seq(as.Date("2012-07-20"), by = "month", length.out = 43)
plot(Bdate, Bshannon, type = "l",
     main = "Donor B Genera Shannon Diversities over Time",
     cex.main = 0.75,
     cex.lab = 0.2,
     ylim = c(0.5,3.5),
     xlim = c(min(bdate), max(bdate)),
     xaxt = "n")
axis.Date(1,
         at=seq(min(bdate),max(bdate), by = "months"),
         format = "%d-%m-%Y",
         las=2,
         cex.axis = 0.5)
points(Bdate, Bshannon, pch = 19)

mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("B", padj = -1.2, adj = -0.1)
xaxt = "n")

axis.Date(1,
          at=seq(min(bdate),max(bdate), by = "months"),
          format = "%m-%y",
          las=2,
          cex.axis = 0.8)

points(Bdate, Bshannon, pch = 19)

mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)

mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)

mtext("B", padj = -1.2, adj = -0.1)

#### Donor C time plot ----

cdate <- seq(as.Date("2013-07-19"), by = "month", length.out = 19)

plot(Cdate, Cshannon, type = "l",
     main = "Donor C Genera Shannon Diversities over Time",
     cex.main = 0.75,
     ylim = c(0.5,3.5),
     xlim = c(min(cdate), max(cdate)),
     xaxt = "n")

axis.Date(1,
          at=seq(min(cdate),max(cdate), by = "months"),
          format = "%m-%y",
          las=2,
          cex.axis = 0.8)

points(Cdate, Cshannon, pch = 19)

mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)

mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)

mtext("C", padj = -1.2, adj = -0.1)

#### Donor D time plot ----

ddate <- seq(as.Date("2012-09-04"), by = "week", length.out = 47)

plot(Ddate, Dshannon, type = "l",
     main = "Donor D Genera Shannon Diversities over Time",
     cex.main = 0.75,
     ylim = c(0.5,3.5),
     xlim = c(min(ddate), max(ddate)),
     xaxt = "n")

axis.Date(1,
          at=seq(min(ddate),max(ddate), by = "weeks"),
          format = "%d-%m-%y",
          las=2,
cex.axis = 0.65)
points(Ddate, Dshannon, pch = 19)

mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("D", padj = -1.2, adj = -0.1)

#### Donor E time plot ----
edate <- seq(as.Date("2013-07-05"), by = "week", length.out = 34)

plot(Edate, Eshannon, type = "l",
    main = "Donor E Genera Shannon Diversities over Time",
    cex.main = 0.75,
    ylim = c(0.5,3.5),
    xlim = c(min(edate), max(edate)),
    xaxt = "n")

axis.Date(1,
           at=seq(min(edate),max(edate), by = "weeks"),
           format = "%d-%m-%y",
           las=2,
           cex.axis = 0.65)
points(Edate, Eshannon, pch = 19)

mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("E", padj = -1.2, adj = -0.1)

#### Donor F time plot ----
fdate <- seq(as.Date("2013-02-01"), by = "month", length.out = 37)

plot(Fdate, Fshannon, type = "l",
    main = "Donor F Genera Shannon Diversities over Time",
    cex.main = 0.75,
    ylim = c(0.5,3.5),
    xlim = c(min(fdate), max(fdate)),
    xaxt = "n")

axis.Date(1,
           at=seq(min(fdate),max(fdate), by = "months"),
           format = "%m-%y",
           las=2,
           cex.axis = 0.8)
points(Fdate, Fshannon, pch = 19)

mtext("Timepoint", padj = 21, adj = 0.5, cex = 0.5)
mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)
mtext("F", padj = -1.2, adj = -0.1)
### Donor G time plot ----

gdate <- seq(as.Date("2015-11-03"), by = "day", length.out = 32)

plot(Gdate, Gshannon, type = "l",
     main = "Donor G Genera Shannon Diversities over Time",
     cex.main = 0.75,
     ylim = c(0.5, 3.5),
     xlim = c(min(gdate), max(gdate)),
     xaxt = "n")

axis.Date(1,
          at = seq(min(gdate), max(gdate), by = "days"),
          format = "%d-%m-%y",
          las = 2,
          cex.axis = 0.55)

points(Gdate, Gshannon, pch = 19)

mtext("Timepoint", padj = 19.9, adj = 0.5, cex = 0.5)

mtext("Shannon Diversity", padj = -24, adj = 1.3, cex = 0.5, las = 2)

mtext("G", padj = -1.2, adj = -0.1)

# if exporting as png, save with 700 by 700 dimensions

dev.off()

### Determine medians of donors ----

Avector <- unlist(Ashannon)
Anumeric <- as.numeric(Ashannon)
median(Anumeric)
# 2.328748

Bvector <- unlist(Bshannon)
Bnumeric <- as.numeric(Bshannon)
median(Bnumeric)
# 2.634837

Cvector <- unlist(Cshannon)
Cnumeric <- as.numeric(Cshannon)
median(Cnumeric)
# 2.638935

Dvector <- unlist(Dshannon)
Dnumeric <- as.numeric(Dshannon)
median(Dnumeric)
# 2.55145

Evector <- unlist(Eshannon)
Enumeric <- as.numeric(Eshannon)
median(Enumeric)
# 2.857225

Fvector <- unlist(Fshannon)
Fnumeric <- as.numeric(Fshannon)
median(Fnumeric)
# 2.396452

Gvector <- unlist(Gshannon)
Gnumeric <- as.numeric(Gshannon)
median(Gnumeric)
# 2.937808