Use Authorization

In presenting this thesis in partial fulfillment of the requirements for an advanced degree at Idaho State University, I agree that the Library shall make it freely available for inspection. I further state that permission to download and/or print my thesis for scholarly purposes may be granted by the Dean of the Graduate School, Dean of my academic division, or by the University Librarian. It is understood that any copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Signature _____

Date _____

BACTERIAL POPULATION DIFFERENCES IN *MUS MUSCULUS* CECUM IN RESPONSE TO PROBIOTIC ADMINISTRATION

By

Leasha M. Christensen

A thesis

submitted in partial fulfillment

of the requirements for the degree of

Master of Science in the Department of Biological Sciences

Idaho State University

May 2015

To the Graduate Faculty:

The members of the committee appointed to examine the thesis of LEASHA M. CHRISTENSEN find it satisfactory and recommend that it be accepted.

Peter P. Sheridan, Ph.D.
Advisor

Gene Scalarone, Ph.D.
Committee Member

Cynthia Blanton, Ph.D.
Graduate Faculty Representative

ACKNOWLEDGEMENTS

Appreciation is due to Dr. Sheridan for the guidance and mentoring provided throughout my graduate career. Isolated cecum samples were graciously provided by Dr. Blanton. Dr. Loubin Yang provided invaluable aid in bioinformatics processing. Idaho State University Molecular Research and Core Facility provided the Illumina MiSeq sequencing of paired-end 16S rRNA reads. Funding was provided in part via a Teaching Assistantship scholarship from the Idaho State University Department of Biological Sciences.

TABLE OF CONTENTS

	Page
Figures	vi
Tables	viii
Abstract	ix
Chapter 1. Literature Review	1
I. Microbial Diversity of Gut Microbiome	
A. Normal microbiota of GI tract in mammals	
B. Diversity and Variation of microbiota among intestinal sections	
C. Natural products and the effects thereof on normal flora	
a. Probiotics	
b. Prebiotics	
c. Synbiotics	
II. 16S rRNA and Bacterial Diversity Analysis	
III. Techniques and Statistical Analysis	
A. Culture dependent	
B. Culture independent	
a. T-RFLP	
Processing and analysis	
T-REX	
AMMI	
b. Paired-end 16S rRNA (V3/V4)	
Next-Gen sequencing	
Processing and Analysis	
Mothur	
MG-RAST	
Chapter 2. Terminal Restriction Fragment Length Polymorphism Analysis of Bacterial Population Differences in *Mus musculus* Cecum in Response to Probiotic Administration	84
Chapter 3. Paired End Phylogenetic Analysis of Bacterial Population Differences in *Mus musculus* Cecum in Response to Probiotic Administration	134
Chapter 4. Future Directions	202
REFERENCES	204

FIGURES

Chapter 2

Figure		Page
Figure 1.	AMMI analysis of 5'HEX forward T-RFLP fragments of control diet samples	126
Figure 2.	AMMI analysis of 3' FAM reverse T-RFLP fragments of control diet samples	126
Figure 3.	AMMI analysis of 5' HEX and 3' FAM T-RFLP fragments of control diet samples	127
Figure 4.	AMMI analysis of 5' HEX forward T-RFLP fragments of synbiotic diet samples	127
Figure 5.	AMMI analysis of 3' FAM reverse T-RFLP fragments of synbiotic dies samples	128
Figure 6.	AMMI analysis of 5' HEX and 3'FAM T-RFLP fragments of synbiotic diet samples	128
Figure 7.	AMMI analysis of 5' HEX forward T-RFLP fragments of control and synbiotic diet samples	129
Figure 8.	AMMI analysis of 3' FAM reverse T-RFLP fragments of control and synbiotic diet samples	129
Figure 9.	AMMI analysis of 5' HEX and 3'FAM T-RFLP fragments of control and synbiotic samples	130
Figure 10.	AMMI analysis of 5' HEX forward T-RFLP fragments of representative samples of control and synbiotic diets	130
Figure 11.	AMMI analysis of 3' FAM reverse T-RFLP fragments of representative samples of control and synbiotic diets	131
Figure 12.	AMMI analysis of 5' HEX and 3' FAM T-RFLP fragments of representative samples of control and synbiotic diets	131
Figure 13.	Margalef diversity index of T-RFLP profiles of control and synbiotic samples	132
Figure 14.	Shannon diversity index of T-RFLP profiles of control and synbiotic samples	132

Figure 15. Simpson diversity index of T-RFLP profiles of control and synbiotic samples 133

Chapter 3

Figure 1. Mothur phylogenetic rectangular consensus cladogram 183

Figure 2. MG-RAST control diet sample B rarefaction curve 184

Figure 3. MG-RAST control diet sample H rarefaction curve 184

Figure 4. MG-RAST control diet sample J rarefaction curve 184

Figure 5. MG-RAST synbiotic diet sample C rarefaction curve 185

Figure 6. MG-RAST synbiotic diet sample F rarefaction curve 185

Figure 7. MG-RAST synbiotic diet sample L rarefaction curve 186

Figure 8. MG-RAST control diet composite rarefaction curve 186

Figure 9. MG-RAST synbiotic diet composite rarefaction curve 187

Figure 10. MG-RAST genus maximum level phylogenetic tree of control samples with coloring according to class 188

Figure 11. MG-RAST genus maximum level phylogenetic tree of synbiotic samples with coloring according to class 189

Figure 12. MG-RAST species maximum level phylogenetic tree of control samples with coloring according to class 190

Figure 13. MG-RAST species maximum level phylogenetic tree of synbiotic samples with coloring according to class 191

Figure 14. MG-RAST species maximum level phylogenetic tree of control samples with coloring according to order 192

Figure 15. MG-RAST species maximum level phylogenetic tree of synbiotic samples with coloring according to order 193

Figure 16. Bray-Curtis PCOA analysis of control diet mouse cecum 194

Figure 17. Bray-Curtis PCOA analysis of synbiotic diet mouse cecum 195

Figure 18. Raw abundance counts of control diet mouse cecum samples 196

Figure 19. Raw abundance counts for synbiotic diet mouse cecum 196

Figure 20. MG-RAST normalized abundance counts of control diet mouse cecum 197

Figure 21. MG-RAST normalized abundance counts for synbiotic diet mouse cecum 197

Figure 22. Composite phylogenetic tree with a genus maximum level and coloring according to class for both diets 198

Figure 23. Composite phylogenetic tree with a species maximum level and coloring according to class for both diets 199

Figure 24. Composite phylogenetic tree with a species maximum level and coloring according to order for both diets 200

Figure 25. A normalized bar chart for detected bacteria with corresponding p-values according to diet groupings 201

TABLES

Table Page

Table 1. Normalized average taxonomic abundance values from Mothur analysis of 16S paired-end sequencing 182

BACTERIAL POPULATION DIFFERENCES IN *MUS MUSCULUS* CECUM IN RESPONSE TO PROBIOTIC ADMINISTRATION

Thesis Abstract-Idaho State University (2015)

Within the mammalian gut, the cecum is the site containing both the highest metabolite absorption and abundant populations of microbes that maintain homeostasis and host health. Various disease states, of which there is no standard medical treatment, result from the dysbiotic disruption of the microbial populations in either presence or relative abundance due to various environmental and host factors. Supplementation with probiotics, prebiotics, and synbiotics has shown promise as a therapeutic intervention to combat dysbiosis. Many gut microbiota cannot be isolated by culture-dependent techniques. Utilization of culture independent techniques such as 16S rRNA T-RFLP, corresponding diversity indices, and paired-end analysis of the V3/V4 hypervariable region shows the effective establishment of probiotic strains while maintaining diversity and homeostatic balance within the cecum under the additional supplementation of prebiotics to meet conditions for therapeutic use. Additional research with other combinations of synbiotics is needed to further investigate effectiveness of this treatment option.

I. Microbial Diversity of Gut Microbiome

More than 10^{30} individual species are present within the microbial world. The immense and unexplored microbial universe is roughly 10^9 times greater in number than stars within the universe. Among archaea, viruses, bacteriophage, fungi, and other Eukarya, this environment is composed primarily of bacteria, which attributes more than 1 kg of weight in the human body and over 1100 species. From birth to death, unique communities of microbes are found within and on the human host as well as other vertebrates: be it the skin, G.I. tract, or mucosal surfaces, with the highest populated region being the lower intestine. The human body externally and internally houses 10-100 trillion microbial cells, which outnumbers the amount of human eukaryotic cells by tenfold. Many of these microbes, which provide the host with metabolic and genetic features that are not innate, develop a synbiotic relationship with the host, reside within the confines of the intestine, and are collectively referred to as the "human gut microbiota" and its xenobiotic genome, the "gut microbiome" is often denoted to in literature as being a bodily organ. Resistance to pathogenic colonization is supplied by the host's native microbiota (7, 8, 19, 21, 22, 32, 49, 69, 79). Host health is mediated in part by the 10^{14} of cells and over 1000 species that comprise the gut microbiome, in that it affects a variety of factors such as: immune response regulation (immunomodulation), resistance to pathogens and pathogen colonization, endotoxemia resistance, epithelial development, development of mature mucosal and systemic immunity, regulates motility of the gastrointestinal tract, fortification of intestinal epithelium barrier integrity, stimulating angiogenesis, blood circulation, energy production and homeostasis, nutrition

and nutrient extraction, metabolism and digestion, metabolic profile regulation (ie. Fatty acids), peptide secretion variation, vitamin and micronutrient synthesis and development (8, 21, 36, 41, 50, 54, 57, 66, 84). Amongst other activities, the metabolites produced by this population of microbes mediate the host-microbe interactions (60). There is a high metabolic potential due to the vast population that compromises this microbiome (11, 26).

Many of these microbes cannot be isolated and cultured by traditional culture-dependent techniques, therefore the study of this microbiome is dependent on developing technologies that are culture-independent in nature in order to combat the previously imposed limitations on analysis. Various initiatives in both the US and Europe have been employed to increase the knowledge base through characterization of the microbes and their genomes within the human body for assessment of their impact and role in states of health and disease (Human Microbiome Project and MetaHIT Consortium, respectively) (8). Due to the "Great Plate Count Anomaly," there has been a noted decline in the rate of accession in novel identity of reference sequences. This can be expected by sample size issues for accurate community representation being that two randomly selected samples derived from a lognormal community may have differing compositions thus require thousands of sampling events for all metacommunity representatives to be present in a sampled, new community. This correlates to environments that are physically identical being composed of differing compositions when formed at random from large metacommunities. This may result in a variety of reactions within the environment to variations in present conditions. When a microbial community forms, its composition is dictated by the surrounding environmental structure. There have been observed

differences in the human fecal flora that can be extended to any part of the intestine within a variety of organisms. Diversity differences can result from host selection, or partially from nutritional intake differences within commensal communities, to chaotic dynamics as is seen in bacterial within small biological treatment systems (intestines are stable). Often times, specific environmental factors cannot be traced to the presence, or lack thereof, of a given phylotype. A smaller reservoir of diversity in the metacommunity (less metacommunity diversity to drive change) will yield smaller diversity and possibly greater reproducibility in its daughter communities. Despite having greater diversity stability, there is a potential of having less functional stability due to a lack of functional redundancy. The observed diversity could correlate to the relative stability and reproducibility of a sample. Diversity studies are imperative to analysis of community structure and function, especially in relation to development of pharmaceuticals, probiotics, bioaugmentation, or substrate presence (19).

These co-evolved interactions can be detrimentally affected by the administration of antibiotics and may lead to chronic or acute illnesses. Various health-related disorders can cause perturbations of bacterial communities associated with host health including the following: colorectal cancer, artherogenesis, Crohn's disease, asthma, hay fever, skin allergies, cancer, type II diabetes mellitus, functional diarrhea, inflammatory bowel disease, functional dysbiosis, ulcerative colitis, pseudomembranous colitis, necrotizing enterocolitis, hyperglycemia, toxic megacolon, gastrointestinal traction functional diseases, irritable bowel and metabolic syndromes, antibiotic-associated diarrhea, behavior and psychological illness, Asperger disorder, childhood disintegrative disorder, Rett disorder, pervasive developmental disorder, autism, dysbiosis-related infections such

as *Clostridium difficile* infections, anorexia nervosa and obesity, cardiovascular events and disease, multiple sclerosis, autoimmunity, and atopic disease pertaining to the change in intestinal microbiota communities. All are mitigated by chemical transformations within the gut as a result of the disturbance of the microbial norm (4, 6, 8, 21, 41, 44, 48, 55, 57, 59, 62, 71, 84). Reversing this alteration from the diseased state will alleviate and reduce said pathologies. Many of these diseases show a decrease in bacterial diversity. Various other axes of involvement with the gut microbiota have been confirmed such as gut-brain, gut-lung, and gut-liver (8, 41).

Other notable changes in bacterial composition found in inflammatory bowel diseases include a reduction in overall diversity, change in the ratios of bacteria from the healthy state (*i.e.* an increased ratio of Firmicutes to Bacteroidetes in the obese, and the transplantation of these communities into healthy mice increased the relative fat-mass significantly, and the presence of mucin-degrading *Akkermansia muciniphilia* directly relating to the onset of obesity), adaptations of the Firmicutes present, decreased abundance of *Roseburia hominis and Faecalibacerium prausnitzii* butyrate-producing bacteria, and other shifts in bacterial abundance and distribution. Alterations of gut microbiota in patients suffering from irritable bowel syndrome showed changes in the expression of host genes for amino acid synthesis, inflammatory response, and cell junction integrity. Small intestinal bacterial overgrowth (of most commonly *Escherichia coli*, *Streptococcus*, *Lactobacillus, Bacteroides,* and *Enterococcus* species) has been shown to factor into this disease's etiology due to the increased degradation of carbohydrates which magnify the symptoms of the disease. Alterations in the composition also affects the function and metabolic potential of the gut as is evident in

type 2 diabetes where the dysbiosis from decreased butyrate-producing bacteria and enriched functions for sulfate reduction and oxidative stress resistance are apparent. The third most common cause of cancer mortality, colorectal cancer, has shown links to *E. coli* being 100 times more abundant than in the healthy host, and that tumorigenesis is promoted while selecting for genotoxic bacteria from alterations within the gut microbiota from the healthy norm (84).

The human microbiome is very volatile and can influence and be affected by various aspects of life, health and development; the relative effect that the gut microbiome has on host health is dependent on the interactions between the two. Urbanization, socioeconomic status, dietary changes, food availability, housing and residency, medications, stress, and traveling are a few facets of the "Modern Lifestyle" that can disrupt or change these vital interactions (21, 44, 57). Additional contributing factors include ethnicity, age, gender, body mass index, genetics, sex, genotype, drug use, and birthing mode (11, 19, 57, 84, 86). Country of origin, including the geological, traditional, nutritional, social, and emotional aspects that correlate to it have shown to influence the composition of the gut microbiota under healthy and diseased states (53). Fifty-seven percent of variation in the gut microbiota can be attributed to diet while genetic accounts for a mere 12%. The diversity and social differences among humans make it extremely difficult to determine the molecular origin of these diseases. Regardless of current insights into the human genome and its mutations, no global application has resulted. Pangenomic association studies have allowed for the discovery of candidate genes to aid in identification of genes related to some disease states

resistance and sensitivity. Analysis of epigenic and environmental factors may also aid determining the origin of metabolic diseases (57).

Bidirectional communication between the enteric nervous system (ENS, which is composed of the nerves, hormones, and other related molecules including neuropeptides and cytokines that control the functions of the gastrointestinal tract), or "second brain," and the central nervous system (brain and spinal cord) connecting the brain's cognitive and emotional centers to functions of the peripheral intestine is mediated by the gut-brain axis. The autonomic nervous system (sensory and motor neurons between the CNS and internal organs comprising parasympathetic and sympathetic systems) is highly involuntary and is also affected by the gut-brain axis through changes in homeostasis by gut peristalsis inhibition though a sympathetic system activation. This is the primary focus of the neurogastroenterological field. The gut brain axis facilitates environmental effects (stress, emotions, and hunger) in relation to gut functions. The ENS extends from the esophagus to the anus and contains thousands of ganglia, roughly 400 million neurons (equivalent to the number present in the spinal cord) which are responsible for gastrointestinal motility control, fluid exchange regulation, defense reactions, local blood flow, GI endocrine functions, gastric and pancreatic secretion, and entero-enteric reflexes. The ENS is segregated from luminal content by the intestinal barrier which prevents ion and small solute diffusion across epithelial cells from the lumen with the help of tight junctions and transmembrane proteins. Permeability across the intestinal membrane is also influenced by mucus, IgA, water secretion, chloride, glycocalyx, antimicrobial peptides, and intestinal microbiota. Increase in permeability is associated

with increased inflammation. The gut microbiota affects various aspects of host health including immune response and physiological and metabolic functions (71).

Innate and adaptive immune cells of the gut immune system interact with the intestinal microbiota (including *Bacteroides fragilis, Lactobacilli*, and *Bifidobacteria* amongst others) in order to provide a homeostatic balance within the intestinal ecosystem. The modifications of these systems with age contribute to a state of low-grade inflammation which may potentially progress into pathological conditions of increased severity. When applying to the gut associated lymphoid tissue, epithelial cells in the intestine or enterocytes act as the primary barrier against invasive microbes by either secreting the antimicrobial peptides mucin and/or defensin or utilizing Toll-like receptors to 'sense' pathogens. Additional immunological responses within the gut include specialized intestinal epithelial cells and microfolding cells collecting and transferring microbial antigens to the lamina propria immune cells, such as dendritic cells, which in turn trigger a polarized host response toward T helper 17 cells by acting as antigen presenting cells in the presence of abundant interleukins -6 and -1β and transforming growth factor-β. This particular immune response occurs when suffering from inflammatory bowel disease. The intestinal mucosa also secretes the IgA from B cells in order to counter bacterial adhesion to mucosal surfaces while neutralizing toxins. However, in the gut associated lymphoid tissue of the elderly, the previously described gut immune response can be impaired through the reduction of mucus and α-defensin secretion and the ease at which pathogens enter the mucosal layers to generate an inflammatory response, which is often referred to as "inflamm-ageing", that is low grade and is propagated by increased growth intestinal pathobionts (53).

There is a direct relationship between the gut microbiota composition and gastrointestinal and extradigestive diseases. Various therapeutic approaches have been developed for correction and alleviation of microbial dysbiosis of the gastrointestinal tract. As there is no current accepted treatment option for combating a state of dysbiosis within the gut microbial flora, various options have been utilized without much success or study. These include: antibiotic treatments, nasopharyngeal fecal-oral transplants and bacteriotherapy, and administration of pro-, pre-, and synbiotics.

Antibiotics, although being the most common course of combat against infectious diseases, may cause harm by severely altering the gut microbiota and may result in the development of antibiotic resistance with continued use via phage-encoded genes that confer resistance. At introduction of the antibiotic to the host, both pathogenic and healthy bacteria are reduced. Even absorbable antibiotic can detrimentally affect the gut microbiome due to the systemic diffusion that is widespread throughout the body (8).

Therapeutic antimicrobial application is dependent on the target in question. Overuse of broad spectrum antibiotics to treat unidentified strains causing infections has resulted in wide-spread antibiotic resistance which may lead into other chronic pathologies. *Helicobacter pylori* infection from overuse of antibiotics indicated a relationship to the development of asthma, hay fever, and skin allergies while *C. difficile* overgrowth from the same cause has links to toxic megacolon, pseudomembranous colitis, and antibiotic-associated diarrhea. A mere two week application of cefoperazone had long-term effects of reduced community microbial diversity (84). The severity of the effects of antibiotics are dependent on the mode of action, microbial structure and distribution of antibiotic resistance genes, and the decree of antimicrobial effect- either bactericidal or

bacteriostatic (84). The antibiotic treatments have shown to effect roughly one third of the bacterial taxa within the gut by decreasing richness and diversity as well as community evenness. Community taxonomic composition closely resembled the pretreatment state of the human host within four weeks of treatment application. However, at 6 months, several taxa failed to recover to the similar state. These results contrast to the prior assumption that antibiotic had modest effects on healthy individuals and support the hypothesis of "functional redundancy" in the gut microbiome (38). A metabolomics study of 2000 murine metabolite features in fecal samples showed that a single high dose of streptomycin can cause significant changes in roughly 90% of the features analyzed (41). Additionally, 20% of patients have a recurrent episode of *C. difficile* infection (CDI) after initial antibiotic treatment, and patients having a recurrent episode are 40% more likely to experience another. There is no standardize treatment for recurrent CDI patients; usually patients have to undergo several antibiotic courses (10).

Through 16S rDNA sequencing, it was found that there was a reduction in the diversity of bacterial species within fecal microbiota of patients with initial CDI in relation to healthy patients, with a notable decrease in *Bacteroides* and an increase in *Proteobacteria* (10). The loss in diversity in intestinal microbiota in developed countries can be attributed to the use, or overuse of antibiotics in addition to antibiotic resistance and therefore proliferation of specific bacterial strains over others. Although the affected taxa vary by subject, some are not recovered months after initial treatment and a long term reduction of bacterial diversity usually ensues (23).

A correlation has been made between weight gain or obesity in humans and increased use of antibiotics. Weight gain in malnourished children, neonates, and adults has been

linked to some antibiotics, however the mechanisms of antibiotic weight gain have not yet been characterized. However, such mechanisms many include an exertion of selective pressure on Gram-positive bacteria and gut colonization by *Lactobacillus spp.* Metabolic abnormalities in obese mice treated with antibiotics included reduced glucose intolerance, body weight gain, metabolic endotoxemia, and markers of oxidative stress and inflammation; all of which were shown to reduce gut microbiota diversity (57). The *in vivo* effects of a single course of administration of antibiotics on microbial populations can persist for years (21, 38). Additional research is required in order to examine and characterize intestinal microbial communities at the species and strain level where diversity is the greatest (21, 48, 84).

Antibiotic intervention in infants has shown to have a negative impact on host health by decreasing overall microbial diversity and the proportion of obligate beneficial anaerobes thus resulting in possible asthma, allergic sensitization and rhinitis, obesity, weight gain, and peripheral blood eosinophilia. The mechanism of modulation of microbiota by antibiotics remains unknown but various hypotheses have been proposed, such as altering microbial metabolism (SCFAs associated with cell differentiation, growth, proliferation, and apoptosis) and the ability to transport metabolites, hormones, micronutrients, and other circulating molecules; instigating homeostatic imbalance of the intestine and therefore impacting the integrity of its defenses; causing a decrease in diversity resulting in a distribution to the innate immune response; and T-cell differentiation and activation regulation. Replacing antibiotic use with pre- and probiotics can alleviate many disease states and may aid in preventing necrotizing enterocolitis, atopic eczema, and other infant diseases (11).

The rising administration of fecal microbiota transplants, also referred to as "fecal infusions" and "fecal bacteriotherapy," serves as treatment for specific diseases by the introduction of donor liquid filtrate of fecal samples. The method of administration ranges from nasogastric or nasojejunal tube, retention enema, and colonoscopy to upper endoscopy. Administration through the lower GI has shown higher eradication rates and effectiveness than upper delivery at a rate of 81-86% to 84-93% (6, 8, 39).

The earliest description of this therapeutic method by Ge Hong relates to roughly 1500 years ago, and is followed by Li Shizhen in the 16th century when applied as "transfaunation" to alleviate various gastrointestinal ailments and in veterinary medicine by Fabricius Acquapendente in the 17th century for animals unable to ruminate. Camel stool was known to be consumed by the Bedouins in the Second World War to cure dysentery. In the late 1950s, Colorado surgeon Eiseman utilized fecal enemas to treat pseudomembranous colitis which later resulted in effective *C. difficile* infection treatment by restoring microbiota community composition to a healthy state. The epidemiology of many of these ailments has changed with time in relation to the mortality, severity, and frequency of the disease course requiring adaptations to therapeutic methods currently employed (8).

Fecal bacteriotherapy administration in intestinal microflora for CDI infection has shown to result in a smaller amount of *Firmicutes* and *Bacteroides*. Two weeks after the treatment, *Bacteriodes* dominated the bacterial composition as seen in healthy individuals (39). Microbiota transplantation has been utilized to treat various diseases including: inflammatory bowel disease, irritable bowel and metabolic syndromes, anorexia nervosa and obesity, multiple sclerosis, and autoimmunity (6). The cost of bacteriotherapy is less

than repeated administration of antibiotic courses, hospitalization, loss of work, etc. (39). However, participants of this intervention must face the "yuck" factor and invasiveness of the application in addition to the relative time consuming and costly screening process for potential diseases such as HIV, Hepatitis B and/or C, Viral Hepatitis, and other functional abnormalities from donor to recipient, high risk sexual behaviors, drug usage, tattoo and/ or body piercing presence and number within 6 months, incarceration history, communicable disease exposure and presence, risk for Creutzfeldt-Jakob disease, diarrheal illness, IBD and IBS history, antibiotic use history, history of GI malignancy or polyposis, immunosuppressant chemotherapeutic exposure, consumption of potential allergens, atopic disease, chronic pain syndrome, autoimmune disease, metabolic syndrome, and possible transfer of other pathogens (6, 8, 11, 39, 84).

Eradication effectiveness has shown to be dependent on the relation of the donor to the recipient. Familial relationships resulted in a 93% effectiveness in comparison to an 84% effectiveness rate in non-related individuals. Effectiveness is also dependent on the microbial composition of the donor, and any discrepancies between donor and recipient may lead to a transition back to the dysbiotic state with the necessity of further future treatments. The microbiome of related individuals has increased resemblance in comparison to unrelated individuals, thus indicating genetic background as a molding factor for the gut flora, and host genetic variation via quantitative trait loci detection of genomic markers has correlation to relative taxa abundance of the following: *Bacteroidetes, Firmicutes, Proteobacteria,* and *Actinobacteria* (8, 11, 84).

Normal Microbiota of GI Tract in Mammals

Now considered as "superorganisms" due to the genetic potential contained in the microbial communities and the human genome (encoding greater than 100 times more genes than the human genome), the microbiota endogenous to the human body develops with the host and may adapt the expression of its genes in concordance with its composition pending stress and adaptations within the local environmental conditions. Despite the inherent inter-individual variation in the community structure of the gut microorganisms, functionality over composition is assumed at highest importance due to there is a shared, conserved portion of encoded functions among individuals' that are commonly referred to as the "gut microbiome" which encode functions that are essential but not encoded within the human genome. These encoded functions and subsequent pathways are essential to the host and may be employed to assume correct gut function (84).

The main bacterial phyla comprising the human gastrointestinal microbial community include the Gram-positive Firmicutes, which is the most abundant at 60% and is composed of over 200 genera including: *Mycoplasma, Bacillus, and Clostridium,* and *Actinobacteria* which is representative of roughly 10% of the total population (47, 57). Other abundant phyla include *Proteobacteria, Fusobacteria, Cyanobacteria,* and *Verrucomicrobia* (84). A large proportion being 2-3 orders of magnitude more abundant than facultative anaerobes and aerobic bacteria of the gut microbiota are strict anaerobes. Many of which are gram negative rods in the *Bacteroides* genus which represent roughly 30% of the measured fecal flora. Other dominant groups include: Bifidobacteria,

Eubacteria, Clostridia, Lactobacilli (all gram positive rods), and gram positive cocci. Less abundant groups include the dissimilarity sulfate-reducing bacteria, methanogens, coliforms, and enterococci. The heterogeneity of the gut in both microbial and substrate diversity allows for an immense range in nutritional patterns (saccharolytic, nitrogen utilizing, and hydrogen metabolizing) and metabolic functions. The gut microbiota can be separated into salutary and harmful bacteria where pathogenic effects include putrification and carcinogenesis of the intestine, infection, inflammation, liver and other organ damage, and diarrhea; while beneficial effects can result from immunological stimulation, decrease of distention problems related to gas, digestion and absorption of essential nutrient improvement, vitamin synthesis, and inhibition of growth or dominance of pathogenic bacteria (31). *Lactobacillus spp.* and *Bifidobacteria* have increased abundance where lipids and simple carbohydrates are absorbed in the small intestine. However not all *Lactobacillus* and *Bifidobacterium* are the same or have the same function, as *L. planatarum* and *L. paracasei* are associated with lean individuals while *L. reuteri* has been linked with obese body types. *Lactobacillus spp*. related to weight modulation in lean individuals have developed defense mechanisms for enhanced glycolysis and defense against oxidative stress while strains associated with weight gain contain a limited ability to break down fructose or glucose and might reduce ileal brake effects. A change in gram negative bacteria can change gut permeability and affect metabolic function (57, 84).

Research has shown that there is a higher degree similarity in bacterial community composition at a specific body site between various individuals than various sites within the same individual even with inter-individual variability being noted (57).

Metagenomic analysis of said variability has also shown shared function amongst bacterial groups. Despite large studies such as the Human Microbiome Project, MetaHIT, and other current studies of the gut microbiota, the bacterial component form and function remains to be elucidated in its entirety. However, research including the analysis of the intestinal microbiota of healthy twins, which at the species level have less than fifty percent of shared bacterial taxa, has shown a genetic role in the establishment and regulation of the gut flora due to the influence mitigated by host-specific genome locus and the highly-debated possible enterotype categories for all individuals (each enterotype dominated by a different genera: *Bacteroides*, *Prevotella*, and *Ruminococcus*, respectively) independent of age, gender, ethnicity or body mass index (57).

Fluctuations and disturbances in the gut microbiota occur from birth until death. Sterile pre-birth, the fetal intestinal tract begins to be established through colonization of microbes present during the birthing process as the delivery type determines initial colonization as those born vaginally possess intestinal microbial flora consisting of the maternal vaginal and fecal flora (facultative anaerobic microbial dominance including *Escherichia coli* or enterococci) and an environment selecting for reduced conditions for anaerobic growth while those born by Caesarian section have bacteria characteristic to the skin such as *Staphylococcus* and *Propionibacterium spp.* in addition to having less populated and diverse bacterial counts. Initial and subsequent feeding also affects microbial establishment as breast-fed infants' microbiota is characteristically dominated by fecal flora highly comprised of *Bifidobacterium* and *Ruminococcus* with notably lower rates of colonization by *Escherichia coli, C. difficile, Bacteroides fragilis,* and *Lactobacillus* (roughly 1% of *Enterobacteri*a) than those observed in exclusively formula

fed infants which was less complex and included enterobacterial genera. With the integration of solid food, the bacterial community begins to be similar to the adult counterpart as *Firmicutes* and *Bacteroidetes* begin to predominate. During adulthood, bacterial community variation ceases until stability declines contingent with old age. Temporal longitudinal studies show that the relative stability of microbial community composition is only altered as an adult transiently via external disturbances as previously mentioned (31, 57).

Although there is a correlation between the microbial composition of the gut microbiome and host age (as is evident with the phylogenetic observations of shifts from *Bifidobacterium* and *Lactobacillus* to the *Bacteroidetes* and *Clostridia* genera from infancy to middle age due to the change in dietary intake from a lactate-based metabolism to short-chain fatty acid yield from plant polysaccharide intake and vitamin production from carbohydrate metabolism), the microbiome alterations occurring due to age remain elusive, especially when correlating to any present effects of frailty due to the presence of other confounding variables that are also under alteration with age. Research has shown a decrease in *Clostridia* and increase in *Bacilli* and *Proteobacteria* and associated decrease in short chain fatty acid producers and increase in opportunistic pathogens and facultative anaerobes in the gut microbiomes in centenarians in relation to middle aged. This may correlate to symptoms of frailty-associated inflammation. Past noted trends in composition in the elderly may be linked to other variables that also change with age, however, differences in the functions relative to pharmacology and nutrition of the host are age dependent. *Alistipes* genus of the *Rikenellaceae* family has highest significant overrepresentation according to data generated utilizing 16S rRNA followed by

Ocillibacter and a decrease in abundance of *Eubacteriaceae, Faecalibacterium, and Lactobacillus* in middle aged and elderly test subjects; this trend may be associated with alterations in microbial metabolism and host interactions due to the role of various microbial taxa on molecular and ecological functions within the gastrointestinal tract. However there is at present uncertainty of implication. Other notable trends included the young having more *Lactobacillaceae, Prevotellaceae,* and *Porphyromonadaceae* while the middle and old aged groups have shown to have increased amounts of *Rikenellaceae*, *Lachnospiraceae* (associated with beneficial short chain fatty acid production), *Ruminococcaceae,* and *Clostridiaceae* and a complete lack of the *Akkermansia* (linked to healthy microbiomes) genus (in the elderly). Young and old subjects also have a higher degree of phylogenetic similarity than either did with the middle aged group. Host age and frailty also have an effect on the gut microbiome function profile as an underrepresentation of vitamins cobalamin (B12) and its corresponding biosynthesis protein and biotin (B7) biosynthesis in the large intestine, in addition to bacterial nucleotide excision repair and mutagenesis via SOS genes (increased mutagenesis rates may lead to antibiotic resistance in gut microbiota as well as increased inflammation). Other carbohydrates such as fructooligosaccharides and raffinose, which are commonly used as prebiotics in the elderly for *Lactobacilli* and *Bifidobacteria*, have shown to have underrepresented utilization in the elderly comparative to the other two age groups and could contribute to disturbing the populations of beneficial bacteria populations and thus increasing the risk of opportunistic infection. The decline of host lactase activity with age can be detrimental to host health due to lactate accumulation within the colon which is generally moderated by lactate metabolizing bacteria via lactases and is linked to

inflammatory bowel diseases and ulcerative colitis. Obesity also leads to microbial composition alterations within the gut with an increase in *Firmicutes* and subsequent increase in inflamm-aging propensity (44, 53, 84).

Both host and bacterial nutrients are provided by diet, and changes in microbiota composition in relation to diet occur due to the fact that certain bacterial species are better equipped genetically to utilize metabolic substrates more efficiently than others. Diets comprised of an increase of fat intake yield an increase in the "Gram-negative/Gram-positive index" of microbiota. Humanized germ-free (GF) mice administered a diet rich in sugar and fat and low in plant polysaccharide (western diet) from a diet low in fat and rich in plant polysaccharides have increased *Firmicutes* and decrease in *Bacteroidetes* phylum abundance. Murine studies switching to a high-fat diet have shown a decrease in *Bacteroidetes* and increase in both *Firmicutes* and *Proteobacteria* phyla, while carbohydrate rich diets yield higher populations of *Bacteroidetes*, and calorie restricted diets prevent colonization of *C. coccoides, Lactobacillus spp.,* and *Bifidobacteria spp* which are major butyrate producers that are required for colonocyte homeostasis as butyrate is the preferred source of energy, repair and maintaining cell health in the human digestive system. A decrease in these bacteria have shown to correlate to an increase in the abundance of opportunistic pathogens. Studies have shown that dietary changes can result in notably rapid microbiota community changes within just 24 hours. Low dietary fiber intake has shown to relate in a decrease in *Bacteroidetes* and increase in *Enterobactericeae*. The gut microbiota may coevolve with diet thus allowing a maximization of energy extraction and protection from non-infection intestinal diseases and inflammation. Vegan/vegetarian diets have also been

shown to produce stool with a lower pH and *Bacteroides spp., Bifdobacterium spp., E. coli,* and *Enteroacteriaceae spp.* culturable counts. A decrease in the quantity as well as diversity of *Clostridium* cluster IV, XIV, and XVII is also associated with a vegetarian diet (57, 84).

Increased research has been performed as of late on the interplay between the content of the gut microbial flora and host health. Yet, due to the numerous variables at play, research has not been able to directly utilize humans for study; thus the mouse model and other human flora associated animals have become widely used for direct application to humans as mouse models allow for increased control during analysis as well as a decrease in the number of confounding variables while providing representation to the human microbiome. Recent developments in this research model have improved their use due the existence of inbred, genetically modified lines, ease of breading and transmission of gut floral community, and the comparative similarities that they hold to the model human physiology as they are confirmed to be consistently humanized. In the case that the two microbial environments differ, similar metabolic pathways could take place in both the mammalian and mouse gut to compensate for the metabolites produced. Also, studies have shown that only a subset of the same microbial profile integrated from the donor into the mouse strains, but it was at a different ratio than expressed previously in the host. They suggest that this could be due to the different genetic backgrounds of the hosts. It is to be noted that within these studies, bacterial diversity measures between human host and mouse model were consistent with the exception of measured species richness (19, 26, 44, 86).

Diversity and Variation of Microbiota among Intestinal Sections

The colon, intestine, stomach, esophagus, oral cavity, and mouth comprise the human gastrointestinal tract. The cecum, ascending, transverse, descending, and sigmoid colon are the distinct anatomical regions of the heterogeneous large intestine, which starts at the ileocecal junction and is responsible for electrolyte and water absorption and secretion in addition to waste excretion and storage (31). The large intestine houses a majority of the present 1.5Kg of intestinal microbiota which represents roughly 10^14 organisms. This immense mass of bacteria is often considered a "superorganism," as there is a higher abundance of genes present in the gut microbiome than within the entirety of the human genome. Heterogeneity of the gut flora correlates to individual host variations that occur on a regular basis. The two most prevalent phyla in the human gut microbiome consists of the Gram-positive *Firmicutes* and the Gram-negative *Bacteroidetes*, however it has been shown that both *Proteobacteria* and *Actinobacteria* can also be the most predominant in a portion of individuals (53).

Functional activity of the colonic microbiota is directly affected by the composition of the hindgut as the present microbiota fluctuate in function according to the relative availability of substrates, redox potential, oxygen distribution, and pH in the local environment. For example, the proximal colon houses a rich supply of nutrients which allows for an increased rate of growth of local bacterium, decreasing pH due to short chain fatty acid production while the distal colon has a decreased substrate accessibility and therefore a decreased rate of growth and more neutral pH (31). Research has shown a higher proportion of bacterial metabolite production within the cecum and

colon compared to that of the jejunum and ileum. Observed endogenously produced metabolites, pending dietary supplementation, accumulate predominantly in cecal tissue, confining their effect on host physiology at that intestinal ecological level where microbiota is most abundant (26). There are more bacteria in the distal region of the gut with 10^{11}-10^{12} bacteria per gram of colon content in comparison to the upper region which has 10^{3}-10^{4} bacteria/ gram of jejunum material (15, 26, 57, 63, 72). Although high in bacterial population, the potential contribution of the cecum and colon remains poorly studied.

There is an outstanding assumption that under physiological conditions, 95% of dietary lipids are absorbed in the proximal region of the gut (61). This conclusion correlates to past studies that have shown inadequate absorption of an overflow of prebiotic molecules within the jejunum and ileum (20, 26). Additionally, a majority of bacterial metabolites were proportionally lower within the jejunum and ileum (proximal gut) in relation to the cecum and colon (distal gut), which had the highest density of gut microbiota. The increased abundance of these bacterial metabolites within the cecum tissue in comparison to that of the colon also corresponds to the increased capacity of metabolic uptake within that site in relation to its counterpart. This could be due to the comparably thicker mucus layer in the colon which could potentially interfere with intestinal cell fatty acid up-take (82). The bacterial metabolites accumulate and are found within the cecum tissue have shown to exert local ecological effects in host intestinal tissues when rather than direct systemic effects (26). Additional research has postulated that metabolites are produced via bacterial metabolism, not by diet supplementation. For example, PUFA-derived bacterial metabolites are mainly produced in the distal portion of

the gut, however lipid absorption occurs in the small intestine, and fatty acids present in the intestinal lumen are essentially absorbed in the jejunum. The activation of receptors by bacterial metabolism within the gut can affect host metabolism and inflammation, and under pathologic conditions bacterial metabolites could cross the gut epithelium to reach the bloodstream at a higher rate than if under physiological conditions and could possibly reach the lymphatic or portal circulation (26, 84).

Natural Products and the Effects thereof on Normal Flora

The restoration of the gut microbiota from a state of dysbiosis or the prevention or treatment of the resulting disease state can be mediated by dietary intake of the host through the use of prebiotics (nondigestable nutritional components to be utilized by the current microbiota to increase function or presence), probiotics (supplementation with 'beneficial' bacteria), and/or their combined use as synbiotics for direct compositional and metabolomics manipulation of the gut microbiota. Dysbiosis occurs in various degrees of progression where 'acute dysbiosis' is generally regarded as being non-life-threatening and does not require the use of high risk drugs for treatment, while 'chronic dysbiosis' is linked to more severe diseases or functional disease states and can been deemed as life-threatening (41). Microbial ecology analyses include the inherent hypothesis that microbial community structure changes directly affect the function within the community (78).

Fermentation within the gut by gut microbiota allows for the production of a variety of compounds that influence gut physiology and other systemic effects in both a

beneficial and detrimental means (*i.e.* The production of short chain fatty acids, or SCFAs, from complex carbohydrates that have escaped digestion in the upper GI tract [roughly 10-60 g/d of carbohydrate reaches the colon with 8-40 g/d being resistant starch, 8-18 g/d non-starch polysaccharides, 2-10 g/d unabsorbed sugars, and 2-8 g/d oligosaccharides] and protein metabolism for energy). Although relatively stable overtime, the gut microbiota and its metabolic functions can be affected by various physiochemical parameters such as nutrient competition, environmental composition, host condition, bacterial metabolic interactions, and dietary intake and preference which can effect fermentation and other metabolic functions. The small intestine is the area of absorption for most of the sugars and oligosaccharides; sans fructooligosaccharides (oligofructose or inulin), lactose, stachyose, and raffinose which can reach the colon intact along with many other sugar alcohols and food additives. Fermentable carbohydrates such as glycoproteins, mucins, and other polysaccharide derivatives such as chondroitin sulfate can also innately be produced by the host. Endogenously produced substrates are most readily metabolized by *Bifidobacteria*, some *Bacteroides* (polysaccharide utilizers), and *Ruminococci*. The metabolic products from metabolism from one species can later serve as a substrate for another. In addition, saccharolytic species (*Bifidobacterium, Ruminococcus, Eubacterium, Clostridium, and Lactobacillus)* allow for direct growth on polymerized carbohydrates via polyhydrolases and glycosidases yielding SCFAs from the Embden-Meyerhoff Pathway (hydrogen, methane, and carbon dioxide gases removed or excreted). Those that may not be able to directly degrade this substrate can cross-feed on the fragments produced by those who do, and carbohydrate metabolism is regarded as being a cooperative interaction of enzymes from

multiple species. Ethanol, lactate, succinate, and pyruvate are intermediates of carbohydrate metabolism which can be later fermented to SCFA allowing increased energy gain. Proteolytic species can cause accumulation in amines, phenolic compounds and ammonia end products (31, 84).

The International Scientific Association for Probiotics and Prebiotics (ISAAP) has confirmed the relationship between pro- and prebiotics on brain functions, the enteric nervous system and the central nervous system through investigating epithelial cell function, gastrointestinal motility, visceral sensitivity, perception, and behavior in addition to neural, neuroendocrine, neuro-immune, and humoral functions. Behavioral impacts in the developing mammal from initial gut colonization is also apparent, and psychological co-morbidities have been linked to dysbiosis and diseases including inflammation in the bowel. Probiotics can improve intestinal barrier and tight junction development while improving mucin expression, IgA production, epithelial cell apoptosis of the intestine, enteric pathogen colonization inhibition, ion absorption, and upregulation of the immune response. Probiotics have also shown to target intestinal sensory nerves and decrease pain (analgesic effect by increasing the abundance of opioid intestinal mucosa and cannabinoid receptors) perception and prevent stressed-induced hypersensitivity in individuals with dysbiosis. Many of these effects have been observed with heat-killed or conditioned probiotic mediums not only viable cultures. *Bifidobacterium infantis* has demonstrated a decrease in pro-inflammatory cytokines, increase polyunsaturated fatty acids (PUFAs), and exhibited anti-inflammatory effects when used in conjunction with α-linoleic compounds. Increased inflammation leads to increased behavioral anxiety, therefore a decrease in the inflamed state has promise, and

has confirmed preliminary results, in reducing stress, corticosterone anxiety, and depression. Treatment of *L. rhamnosus* probiotics also alters brain neurochemistry by changing GABA mRNA in cortical regions with the addition of a reduction of expression in the locus coeruleus, amygdala, and hippocampus. Patients in a state of stress have alterations in specific bacterial groups, motility within the GI tract, immune function, and hormone and mediator release within the gut (71). The gut microbiota has its own innate antimicrobial effect through the production of ribosomally synthesized peptides bacteriocin from the present bacteria. Ingestion of bacteriocin-producing strains has shown promise therapeutically as a probiotic (84).

Probiotic use also decreases memory dysfunction and may have application in the treatment of neurodevelopmental disorders including autism spectrum disorders (Asperger disorder, childhood disintegrative disorder, Rett disorder, pervasive developmental disorder). GI disturbances and GI system dysfunctions show a direct correlation with the severity of the disorders listed above. A noted increase in diversity in microbiota was present in patients suffering from severe autism when compared to a control. Gut barrier function can be improved by the use of *Streptococcus thermophiles, Lactobacillus bulgaricus, Lactobacillus acidophilus,* and *Bifidobacterium longum* probiotics. Prebiotic supplementation can improve the inflammatory response, decrease putrefactive activity of the gut lumen, prevent GI infections, modulate gut transit, and improve the quality of life by decreasing flatulence and bloating. However, studies of this nature show conflicting results in respect to the identity of the strains undergoing an abundance shift (71).

The use of pre- and probiotics has to overcome a qualitative intrinsic gap as oral administration generally has a three order of magnitude difference from the native 100 trillion bacteria present in the gastrointestinal tract which may result in only temporary colonization in the gut lumen. The mode of entry may also dictate effectiveness in colonization of beneficial bacteria as many Firmicutes are best established through and upper gastrointestinal route while a lower route is more ideal for *Bacteroidetes* due to gastric acid secretion (8).

Probiotics

The human gut microbiota has an immense impact on host health, and thus the direct manipulation of the microbiota to relieve a diseased state and establish a remedial community is of utmost interest. Probiotics are defined by 2001 FAO/WHO as being *"organisms and substances which contribute to intestinal microbial balance" or "a live microbial feed supplement which beneficially affects the host animal by improving its intestinal microbial balance* (8)*."* In order to fit this criteria, which does not include antibiotics, the probiotic needs to be stable and viable and remain as such under storage and during use, survive the intestinal ecosystem, prepared on a large scale, and beneficially effect the host after its integration. Surviving the acidic environment within the gut and then colonizing and becoming active in the colon can be problematic as adherence to the intestinal epithelium may be necessary. Competition of nutrients and ecological sites as well as stress can also cause a decrease in effectiveness in this treatment. Additionally, the probiotic must also remain present after the consumption of

the product initially containing the strain has end. *Lactobacilli* (*Lactobacillus acidophilus, L. casei, L. delbruekii)* and *Bifidobacteria* (*Bifidobacterium adolescentis, B. bifidum, B. longum, B. infatis)*, and *Streptococci* (*Streptococcus salivariius ss. Thermophiles, S. lactis)* are commonly used in probiotic treatments and have shown to alleviate, hepatic encephalopathy, carcinogenesis, diarrhea, colitis, pathogen colonization, constipation, gastroenteritis, immunostimulation, flatulence, gastric acidity, expression variations in microbiome-encoded enzymes in regards to the metabolism of plant polysaccharides, maintain the structure of the gut microbial community, inhibit pathogen invasion via secretion of mucus, improve mucosal integrity, and act on the gut immune system by improving the sensitivity of immune and epithelial cells to microbes within the gut lumen and reduce the permeability of the gut (8, 11, 31, 71). The benefits of the use of these two strains have been summated by the following points: pathogenic growth and survival interference in the gut lumen, mucosal barrier function improvement, gut influence via the systemic immune system and other organs. Application of *Lactobacillus* strains has resulted in a reduction of body weight gain, fat accumulation, plasma insulin, leptin, total-cholesterol, and liver toxicity biomarkers in addition to down-regulation of fatty acid oxidation-related genes and altering gut microbial diversity (84). *Lactobacillus rhamnosos* and *Lactobacillus paracasei* decrease fat storage and a circulating lipoprotein lipase inhibitor to control triglyceride deposition into adipocytes. *Bifidobacteria* is known to reduce the endotoxin levels of the intestine and improve gut barrier function (11). Probiotics are a promising therapeutic supplementation tool for the alleviation and prevention of many diseases and pathologies.

Establishment or increase of presence of beneficial organisms such as *Bifidobacterium* (saccharolytic colonic bacteria comprising 25% of the adult gut microbial population and 95% of the population in newborns) and *Lactobacillus* through introduction and implantation of exogenous bacterium has shown to have health-promoting activities (31). *Bifidobacterium* yields highly acidic end products such as lactate and acetate which may lower the pH of the local environment enough to have an antibacterial effect (bacteriocin-type substance secretion against clostridia, *E. coli, Listeria, Shigella, Salmonella, Vibrio cholerae* and other pathogens) and yielding end products which may inhibit pathogenic bacteria. They are also known to reduce blood ammonia levels, not form aliphatic amines, hydrogen sulfide or nitrites, produce vitamins (B group and folic acid) and digestive enzymes such as casein phosphatase and lysozome, restore intestinal flora to its normal standard during antibiotic usage, lower blood cholesterol, and promote the attack against malignant cells and improve host resistance to pathogens by acting as immunomodulators. Therefore, addition of live cultures of these exogenous bacteria can have a range of desirable health effects upon reaching the colon where they can become active and grow due to the appropriate and selective physiochemical conditions. These bacteria are responsible for the digestion and metabolism of nutrients as well as the energy generation though fermentation of indigestible carbohydrates to SCFAs by the present enzymes. Probiotics have shown to aid in the restoration of the metabolic profile of the bacterial composition within the gut as well as ameliorating infant antibiotic-associated diarrhea, reducing irritable bowel symptoms (through *B. bifidum* supplementation) and small bowel injury due to IBS (*Lactobacillus casei* introduction), treat behavioral alterations due to stress (*Lactobacillus*

rhamnosus) as well as regulate the γ-aminobutyric acid receptor expression. Although the effects of probiotic use may be transient, probiotics have been utilized to shift the presence of intestinal microbial balance to change the overall composition of the gut microbiome (31, 41).

Prebiotics

Prebiotics are defined by 2001 FAO/WHO as being "*a non-digestible food ingredient that beneficially affects the host by selectively stimulating the growth and/or activity of one or a limited number of bacteria in the colon, and thus improves host health* (8)*."* Additional criteria must also be met for a food item to be considered a prebiotic: it cannot be absorbed or hydrolyzed in the upper GI tract, it must serve selectively as a substrate for specific commensal, beneficial bacteria for growth or metabolic activation, alter the microbial composition to that of a healthy state, and cause systemic or luminal effects that benefit host health. Non-digestable foods such as oligosaccharides, polysaccharides, fructooligosaccharides, and other naturally occurring non-digestable carbohydrates (resistant starch, nonstarch polysaccharides [hemicellulose, pectins, gums, plant cell wall polysaccharides]), peptides, proteins, and lipids) have been shown to improve host gut microbiota health by stimulating growth and activity and specific endogenous microbiota, changing the microbial composition of that local, due to the nutrients and metabolites provided. They have been shown to particularly benefit host colonic health, and can directly manipulate metabolism of lipids via products of

fermentation. The chemical structure of these compounds is the cause of lack of digestion and absorption for the use of bacteria (8, 31, 71).

Fructooligosaccharides are a natural non-digestible oligosaccharide that are mainly used as prebiotics. They are short and medium length chains of beta-D fructans which contain a glucose molecule in their initial moiety due to their synthesis in plant cells from fructosyl moiety transfer between sucrose molecules. Mammalian digestive enzymes cannot hydrolyze the glucose-fructose bond or beta-2,1-osidic bond of fructooligosaccharides. The selective anaerobic fermentation of this substrate over other carbohydrates substrate by colonic bacteria (*Bifidobacteria*) also caused a reduced acidity by lactate and acetate production and inhibit the growth of some pathogenic bacteria. Fermentation of carbohydrates to SCFA and lactate in the colon has also shown to increase the absorption of various essential ions including calcium, magnesium, and iron. SCFAs can be used by a number of tissues and organs and also affects endogenous metabolism in addition to lipid and cholesterol metabolism (and decrease VLDL particles in circulation). They have also been shown to decrease total body fat deposition, triglyceridemia, and phospholipidemia (31). Residential status of aged individuals has also shown to have an effect in the bacterial and metabolic composition of the gut microbiota in the elderly. *Firmicute* and SCFAs quantity were higher and *Bacteroidetes* lower in community-dwelling when compared to long-stay facility individuals. Inflammatory molecules correlating to a state of systemic inflammation was also higher in the latter category (53). Prebiotic applications have resulted in reduced inflammation markers in the colon and visceral adipose tissue by initiating the growth of *Bifidobacterium* spp. (in a high fat diet by applying polyphenol-rich extracts) as well as

the inhibition of growth of pathogenic strains and initiating the growth of non-pathogenic strains of *Clostridium (*polyphenol use). Prebiotics have also shown to provide increased stress resistance and anti-inflammatory and anti-allergic responses (fructooligosaccharide supplementation) (25, 26, 41). Prebiotics have shown to have profound effects on the benefit of insulin sensitivity and phytochemical release during fermentation in the colon in the prevention of colorectal cancer through the delivery of fermentation acids to the distal colon. They have also shown beneficial effects in relation to obesity, metabolic diseases, type 2 diabetes, and several cardiovascular diseases. Possible mechanisms of action revolve around the modulation of gut peptides, increasing specific endogenous glucagon-like peptide production, improving the tolerance to glucose, targeting enteroendocrine cell activity, and leptin sensitivity, promoting fermentation in the gut, and regulating the gut inflammatory response (11).

Past studies have shown a restoration of microbiota proportions after high fat treatment via the supplementation of metabolites to decrease weight gain.
Dietary supplementation with chemically produced isomers of conjugated fatty acids in both rodents and humans led to variously notable properties, such as: anti-obesity, anti-atherogenic, and anti-inflammatory properties (33, 58, 87). It is to be noted that the metabolic effects could be isomer-specific (29).

Synbiotics

The combined use of probiotics and prebiotics as synbiotics ("*a mixture of probiotics and prebiotics that beneficially affects the host by improving the survival and*

implantation of live microbial dietary supplements in the gastrointestinal tract, by selectively stimulating the growth and/or by activating the metabolism of one or a limited number of health-promoting bacteria, and thus improving host welfare") shows promise in combating the diseased dysbiosis state by supplementation with functional and health enhancing nutrition by maintenance of the colonic flora in the healthy, balanced state (8). They can also increase the shelf life, viability, and effectiveness, and functional activity of exogenous and endogenous bacteria of one another when used in conjunction than alone. Pharmaceutical application can be made to the following intestinal disease states: colitis from antibiotic use, Crohn's disease and ulcerative colitis among other inflammatory bowel diseases, colorectal cancer, ileocecitis, necrotizing enterocolitis, systemic disorders, septicemia, pancreatitis, and multiple organ failure syndrome. Further use to combat pathogenic bacterial overgrowth, parasite growth, viral infections, burn treatment, stress, and antibiotic therapy effects also applies as these are associated with the translocation of bacteria due to the failure of the intestinal barrier. *Bifidobacterium* has shown to negate bacterial translocation (31).

The production of short chain fatty acids (SCFAs) including butyrate, acetate and propionate contribute to an anti-inflammatory and anti-neoplastic response while protecting the intestinal epithelia through energy supply (to counter atrophy and inflammation in diversion colitis caused by SCFA deficiency) and mucin secretion. The reduction of SCFA may be due to a decrease of specific bacterial strains *Faecalibacterium prausnitzii, Eubacterium hallii* and rectal/*Roseburia* group in the elderly. The impairment of mucin secreting can cause increased colonization of pathogens, such as Gram-negative *Enterobacteriaceae,* in the intestinal mucosa. Release

of lipopolysaccharides or endotoxins by these pathogens can perpetuate the existing inflammatory response. For example, IBD patients having abnormal intestinal microbiota also experience instability in the abundance of dominant species in comparison to their healthy counterparts. There is also decrease of *F. prausnitzii* in Crohn's disease and ulcerative colitis while the abundance of adherent-invasive *E.coli* increases. Colorectal cancer has also shown to relate to an increase of colonization of the bowel by the toxigenic microbes *Helicobacter pylori, Bacteroides fragilis* and *E. coli*. Prebiotics, as components in many fruits, vegetables, and grains, in addition to probiotics and their combined use as synbiotics have shown to enhance the production of mucus, defensins, and IgA by intestinal epithelium and upregulate the immunological response of phagocytic and natural killer cells. In order to benefit from this application in its entirety, increased research into gut mucosal immunity and function in addition to phylogenic identification of microbial community should be pursued. Application of synbiotics has also shown to reduce frequency of the common cold and other winter infections in the elderly while increasing the frequency of micronutrients (vitamins and minerals), thus aiding in overall host health. The relation of gut microbiota to the diseased state as being either the cause or the result remains elusive (31, 53).

The use of synbiotics is best applied to alleviate acute dysbiosis while decreasing development of chronic dysbiosis. However, due to the lack of information pertaining to the composition of the gut microbiota under both healthy and diseased states, supplementation of effective pre- and probiotics has been challenging. Species shifts between the two states of health have remained elusive despite the current knowledge of phyla level changes. This current lack of information further limits the understanding of

microbial community interactions on the supplemented treatments as bacterial spatial distribution within the GI tract is heterogeneous due to environmental differences and different pre- and probiotics may have varying activities at different locations. Within a local environment, the metabolic products of one bacteria can be modified and utilized by another bacterial species so increasing the availability of a molecule in its active form can be mitigated by community level biotransformation reactions. These cooperative interactions directly affect the degree of effectiveness of a prebiotic (as the necessary active form may never reach its target location) or a probiotic (which may contain a strain that does not yield the desired, beneficial effect on microbial composition and function). Further research is needed to investigate and characterize the intestinal communities of microbes to increase the efficiency and effectiveness of synbiotic treatments (31, 41).

II. 16S rRNA and Bacterial Diversity Analysis

Within the three domains of life, a large proportion of the present ribosomal RNA structure is preserved, and designated regions of said structure are specific to at least one of the three domains. The secondary structure of the prokaryotic SSU rRNA consists of various regions that differ to the analogous regions in eukaryotes (some regions are uniformly conserved across the tree of life while others are specific to each domain and differ across the three). Being required for life, all of the phylogenetic domains contain organisms having large insertions of many of the nine major variable regions of SSU rRNA which contain both constant and hypervariable regions and do not participate in horizontal gene transfer events (46). Surveying the distribution of the small subunit

rRNA gene allows for an alternative profiling approach due the genes ubiquity across the domains of life (18S in Eukarya, and 16S in Archaea and Bacteria), the information it provides relevant to phylogenetic identification, relevance to current database compositions (3).

Even prior to the establishment of next generation sequencing as a primary method of analysis, the 16S rRNA gene was the most represented gene present in the GenBank database. However utilization of the 16S rRNA gene results in various limitations including the following: decreased evolutionary rate, lack of correlation to organism function, and the variable copy number. However, there is no other molecular marker, at present, that is found in all organisms while having a low rate of recombination and transfer of horizontal genes, or can differentiate closely related organisms according to the information it includes. There is no universally employed region of the gene currently being analyzed allowing for one or more of the present 9 variable regions within the 16S gene to be targeted and analyzed via specifically designed PCR primers (3, 9, 14, 16, 30, 37, 42, 45, 46, 67, 74-78, 83, 85).

Various biases (microbial population relative abundance misrepresentation in a given sample) and errors (actual sequence misrepresentation as a result of PCR sequencing and amplification) are employed when utilizing the 16S rRNA gene in sequence survey. Whether the relative abundances of the gene being sequenced are equal to the bacterial presence in the sample is affected by the DNA extraction method, DNA purification protocol, selection of PCR primers, PCR cycling conditions, community composition within the sample, and copy number of the 16S gene within the genome as PCR polymerases erroneously result in substitution in 1 of 10^5 to 10^6 bases, and

amplification of heterogeneous templates can result in formation of chimeras and not a sequencing error when incomplete amplification products are present to serve as primers for related fragments at a rate of 5-45%. Sequencing also results in errors due to the homopolymer under-representation at a rate of 0.01-0.02. Undetected chimeras resulted in an increase spurious OTU and phylotypes on the genus level to be identified. Trimming sequences at break points can reduce the overall error rate when identification of breakpoints where're criteria pertaining to quality scores is not met. These biases confound the representation and application of data collected (7).

Due to recent sequencing and molecular-based technological improvements, 16S rRNA phylogenetic analysis have alluded to an exponential increase in the number of unique bacterial species to be that of 10^7 to 10^9. This estimate bypasses the estimate calculated by utilizing culture-based isolation techniques as is evident in the discussion of the great plate count anomaly as only ~1% of total bacterial can be selectively cultured and isolated within a given sample as corresponds to the saying that "everything is everywhere ; but the environment selects." This is apparently due to the differing groups of bacterial species found within geographically different yet ecologically similar environments. It is proposed that as bacteria are a large source of bioactive, natural products, they could be utilized pharmaceutically in future drug therapies. However, it remains unknown if the increase in species diversity will correspond to a like increase in the biosynthetic diversity of secondary metabolites. Therefore, metagenomic and culture independent methods can be utilized to elucidate the potential secondary metabolite source present within the newly-found microbial diversity (67).

Many of the culturable microbes were unfortunately considered "microbial weeds," and did not provide quality or representative results for the diversity being studied. This also leads into the issue that a species being selected for in a controlled environment differing from the conditions where the sample was extracted may not even be a major contributor of the microbial community within its native environment (an extremely minor player of low relational abundance to the whole) (16).

The 16S rRNA gene, though representative of low sequence diversity libraries, is highly conserved, and is considered universal and has greatly increased the rise of complex, novel microbial consortia (27, 30). Species analysis of diversity can be based on the nucleotide sequence identity in mitochondrial DNA within the species. 16S ribosomal RNA coding region of partial nucleotide sequences can discriminate amongst species identity in order to achieve specificity and sensitivity with increased resolution of the results. The nucleotide regions analyzed are highly conserved and many were located in the ribosomal RNA coding region (37). As there are sufficient data available pertaining to the composition and identity of 16S rRNA genes, they can be applied to multiple techniques utilizing next-generation phylogenetic identification of abundant microbes in addition to measures of similarity and or distance, spatial and temporal assessment of microbial diversity changes, or the exploration in treatment effects and disturbances in microbial community composition. Also, microbial ecology analyses include the inherent hypothesis that microbial community structure changes directly affect the function within the community (1, 40, 78).

Alpha and beta diversity measurements, which cannot be obtained in their entirety, can be applied to general changes within a given community and can be effected by sampling (78).

What is Diversity?

Unparalleled taxonomic diversity exists within microbial communities of both the host and environment. Though improved dramatically with current available technologies including the 16S rRNA gene sequencing, analysis and interpretation of the spatial distribution, organization and composition of these communities remains elusive. Alpha (species richness measurement), beta (analysis of diversity comparisons between samples such as Bray-Curtis methods that are based on relative OTU abundance or Unifrac that integrates phylogenetic structure and is separated by being weighted [where phylogenetic structure is based on OTU abundance on a binary scale and showed higher alpha diversity in PE data compared to single direction sequencing reads of the 5' end] and unweighted [which has shown no difference in PE data and single direction sequencing), and gamma diversity analysis progress as facilitated this issue, however current sequencing coverage is lacking for comprehensive analysis of single samples or microbial community comparisons. Rare biosphere taxa include those microorganisms that can live at relative low abundance, have yet to be discovered. Taxonomic diversity has shown to be increasing in lower dominant distributions (3, 74, 78).

Diversity indices of Margalef Species Diversity (species abundance and frequency), Shannon Diversity (entropy or proportional abundance of a given species

amongst a whole), and Simpson Diversity (probability of two randomly selected 'samples' belonging to the same species) can all be elucidated by alpha and beta measures (12).

Species richness is characterized by the number of species within a given community or sample and species evenness is the size of the species population within that same community. Both of which are used as parameters to investigate diversity and structure within a community and are limited by conventional culture-dependent methodology due large fractions (85-99.9%) of microorganisms present in nature being refractory to cultivation. Lack of identical environmental parameters composing the exact environmental structure as is found in nature during cultivation imposes an additional limitation by altering the community structure by the new selective conditions introduced to the environmental sample. This results in an evolved community structure that may not represent the initial structure present during sampling (51).

Host-associated samples have shown to be diverse due to beta diversity analysis of environmental samples as it measures the degree of similarity, or phylogenetic relatedness, between community samples. This type of diversity metric can be used in measuring shifts in the composition and structure of microbial communities across environmental samples. Beta-diversity studies of communities and genus level taxonomic distinctions has insisted that extensive sequence curation and contig formation is unnecessary. However beta-diversity studies, though useful in community comparisons, have limited use to only communities exhibiting clear differences and it does not provide information the details of these differences (9, 43).

In silico digestion can now be accomplished to investigate the efficiency of restriction enzymes in discrimination among sequences and provide information pertaining to the following about an 'ideal enzyme': best estimate of diversity by discriminating between present phylotypes, best phylogenetic group resolution, best paring with primers for a given dataset. However, there are various limitations to the programs that perform this analysis in regards to condition limitations and application and data interpretation. Additionally, this tactic is only done if a specific sequence is in mint and is not ideal for community analysis or overall diversity studies as only a small fraction of bacterial diversity is represented in the databases utilized and generated phylotypes may not correlate to those present within any given database. Primers must be general enough to apply to all bacterial populations of interest yet specific to a given taxonomic group; *i.e*. bacterial 8f is not as universal as though as it has been shown by the probe match tool in the ribosomal database project to amplify at most 76-98% of the bacterial 16S rRNA genes contained on that database (which only include a fraction of the extant bacterial diversity) and is not specific to bacteria alone (as it also matched various archaeal 16S rRNA gene sequences (73).

OTU assignments are made according to the derived genetic distance between sequences. Distribution of sequence abundances among OTUs allows for general estimates of ecological richness, evenness, and diversity (a combination of the two previous) of the community in addition to measurements of the like between communities. Phylogenetic methods investigate differences in communities according to sequence difference. OTU application allows for quantitative measurements to be

collected but is limited due to the increased number of sequences needed to combat the underestimation of richness that is caused by inadequate sampling (74).

Accurate distance based threshold for taxonomic level definitions can be created and therefore consensus-based methods of OTU classification should be employed. This is due to the inability to define bacterial taxonomic levels resulting from the lack of adequate bacterial taxa being cultured or culturable. Many of the present taxonomic outlines and requirements are based on previously cultured organisms causing candidate phyla and non-culturable phyla are lacking in taxonomy identifying to the level of genus or species. No accepted and employed definition of a bacterial species exists, which also increases the difficulty in appropriate taxonomic classification according to phylotype or even define the genera, family, class, order, or phyla of bacteria (77).

There is no biological significance pertaining to alpha diversity when using either the one or each gap calculator. However, when performing analysis according to OTU metrics, an overestimate of similarity between communities resulted when ignoring gaps, and an underestimate resulted when counting each gap. Adjustments in the OTU cut off resulted in a parallel effect in some beta diversity measures; for example, ignoring gaps while increasing the threshold lessens the sequence differences between each other and incorporate additional sequences within an OTU yielding an increased number of OTUs to be shared while increasing the stringency of an OTU resulted in the opposite. However, there was no significant difference between 0.03 and 0.05 OTU cutoffs when using either the one or each gap calculation for distance. A 0.03 OTU threshold (6 differences) used in 200 base reads reduces the application of results when artifacts introduced in PCR, alignment, and sequencing are taken into consideration (76).

Methods of bacterial diversity measurement

The simplest approach of trace analysis is binary comparison of different sample peak presence. Although valid, this approach lacks appropriate quantitative analysis. PCA, or Principal Component Analysis, and other multivariate statistical methods have been used to employ the needed statistical rigor for complex data set analysis. An additional benefit to multivariate analyses is the use of numerous variables that are not constrained to species identification. Shifts in ordinations can be determined by vector distance from the ordinate or from other data points (68).

PCA can be used to show trends in distances in community patterns. However, PCA data is not normally distributed. Additionally, non-linear data from large gradients can cause PCA ordination arching and thus obscure any patterns (68).

More than one group or the influence of various factors are generally integrated into studies of applied microbiology. ANOVA, or analysis of variance, allows analysis of this as well as establishing whether either factor influences the microbial composition on an individual basis or may result from a contribution interaction amongst factors present. For complex data sets, analysis of variance is the most appropriate statistical analysis method. There are multiple variations concerning this analysis and each apply to a particular experimental context. Also, various assumptions concerning the sample set need to be ‘approximately’ true for the ANOVA to be validly applied such as errors being distributed normally among a mean of zero and a standard deviation, and that all variation from group to group is constant in all measured groups. Additionally, individual

treatments are additive instead of multiplicative. Experiments are generally too small to test the validity of these assumptions, which a failure in such can affect both significance and sensitivity. In microbiology applications, quantity is generally estimated and the assumptions may not hold true as there are two common problems: normal distribution is often unlikely among small whole numbers, and heterogeneous variances within different groups may result from a large range of bacterial quantity. If there is doubt in these assumptions, the validity and significance are considered more of approximations. If assumptions are not met, transformations of the data cannot be achieved, making ANOVA impossible (34, 65). Most post hoc tests are available in software packages, however many do not address the ANOVA assumptions. Various different post hoc tests can lead to the same conclusions. The type of test that you use is dependent on the intention of the analysis. However, none of the tests effectively substitute a pre-planned experiment having specific analysis to test in mind (35).

The additive ANOVA model identifies but does not analyze the interaction source. Multiplicative formulations such as AMMI, or Additive Main effects and Multiplicative Interaction, are an alternate to this model and can be used to identify the contribution of genotype vs. environment interaction sensitivity (34, 65, 68).

III. Techniques and Statistical Analysis

Culture Dependent

Microbiology's Golden Age, the early 1900's, had a severe dependency on culture-based isolation and characterization techniques despite the apparent limitations

relative to this methodological practice. Culture-based techniques have led to novel insights into the study of bacterial community structures of culturable organisms within a given environmental sample. However, there is limited application in revealing the complete diversity or phylogenetic assignment of many of the environments being studied. The ex situ presence and growth of organisms, culture dependent methodology utilized, has shown to be highly unrepresentative and lacking in relation to the diversity present within the microbial populations being studied. DNA-based molecular methods have facilitated this and circumvent the limited scope of culture dependent techniques by identifying sequence diversity according to genes present within a sample. With the introduction of the Polymerase Chain Reaction, analysis of microbial communities was facilitated while avoiding culture biases by creating a unique amplified nucleic acid profile or pattern reflecting the microbial community structure of the sample (1, 5, 40, 68).

Traditional microbial ecological investigations integrated clone library generation where every clone consists of a conserved primer PCR amplification product and sequencing of that clone. Many sequence collections are generally necessary for comparison, and intensive sampling of each library is required for accumulation of adequate coverage of a microbial community sample. Due to the log of microbial population in the sample, complete analysis of every cell in the community cannot be performed, and thus analysis relies on statistical integration to generate general diversity measurements (74).

Culture Independent

Biochemical compound analysis from an environmental sample source can be achieved by means of culture-independent methods investigating microbial community composition. Purified, ribosomal nucleic acid or genomic DNA isolates are utilized in said analysis by paring molecular genetic techniques to phylogenetics and integrates species that are related phylogenetically to species that are cultivatable as well as those who have yet to or cannot be isolated by culture-dependent means (1, 5). Many of these molecular techniques are based on the use of PCR amplicon separation by targeted functional or phylogenetic genes, namely the highly conserved region within prokaryotes, the16S rRNA. However, diversity cannot be studied at a level of high resolution within these methods pending the relative simplicity of the community structure due to the limited species designated by observed DNA hybridization rates or sequence visible on a gel. Yet insights into diversity can be obtained in terms of relative abundance of a common species as these rapid, robust methods depict major differences amongst community composition and structure while testing comparative hypotheses. Characterization of complex microbial community samples is an arduous process that limits many microbial ecology studies due to the difficulties of identification and quantification of microorganisms present within any given sample. Culture-dependent methods are restricted due to the limited information collectable from morphological data and the intricacies of isolation. As such, diversity can be more adequately assed by genetically-based techniques due to its application to a wider range of organisms. However, they are not without their inherent biases such as those present during

amplification due to annealing temperature discrepancies as well as poor primer-template homology which skewed relative abundances measures as well as may lead to a loss/lack of amplification in some community members, template concentration, DNA extraction, pipetting, reproducibility, etc. which result in the lack of reflection of TRF abundance and an organisms' abundance. Utilizing a well-homogenized sample and PCR master mix and integrating replication of samples can alleviate some of the associated biases (13).

Construction, screening, and analysis of clone libraries is both time and cost intensive. Other techniques have been developed (such as DNA melting behavior and single-strand conformation) to circumvent clone library use and assess community structure while providing a crude qualitative assessment of species diversity. However, they are limited by the lack of sensitivity of the materials employed in the procedure (ie staining) and to not contribute data relative to phylogenetic assignment or identity of a given microbial community (51).

Most PCR-based techniques utilize 16S rRNA genes from sample DNA for amplification, cloning, and sequencing and can be applied to diversity studies. Cloning is both time and financially exhaustive as several thousand clones must be analyzed from a single sample to definitively characterize a 4000 species/gram soil sample, and the resulting phylogenetic assignment of individual members within the microbial community is not necessary for substantive studies of microbial communities. Other analyses (*i.e.* Amplified Ribosomal DNA Restriction Analysis [ARDRA], Single Stranded Conformation Polymorphism analysis [SSCP], Thermal and Denaturing Gradient Gel Electrophoresis [TGGE and DGGE], Length Heterogeneity analysis [LH}, etc.) are highly dependent on manual visual resolution and may not supply the sufficient

fragment resolution as restriction fragment analysis divides an organisms visual signal thus omitting organisms having an increased amount of restriction sites or provide adequate amounts of data (13, 40).

A common technique amongst those discussed that has proven effective in comparative studies of microbial communities within several environments is that of the TRFLP analysis (5). It is a replacement of RFLP analysis (ARDRA) for comparison of rDNA due to limitations of species richness and evenness estimation as well as identification of specific phylogenetic groups present. The use of fluorescent dye is employed to differentiate between past technologies and circumvent the limitations imposed by said technologies. ARDRA also has limitations as sample complexity increases, resolution decreases (patterns become too complex to measure diversity). TRFLP has a higher resolution (51).

TRFLP analysis was introduced nearly two decades ago and quickly became an ideal molecular tool in microbial community analysis due to its nature at being high throughput and having relatively high phylogenetic resolution and application (51, 52, 81). It quickly replaced the culture dependent methodology that was utilized based on extensive characterization and isolation of pure bacterial cultures during the golden age of microbiology circa 1900 due to its inherent limitations. Numerous genetic based approaches based on the production of a pattern or profile of nucleic acids from an amplified sample have been developed and utilized since that era, of which TRFLP analysis stands at the forefront despite its flaws and limitations. The development of TRFLP statistical applications have allowed for adequate noise elimination in TRF profiles as well as retention of information for representative profile interpretation of

microbial community dynamics. PCR-based methodology has proven to yield reproducible data in relation to the abundance of a given amplicon while being in direct proportion to template abundance in a sample of increased diversity (1, 5, 13, 18, 40, 52, 81).

T-RFLP

Rapid, high throughput and robust analysis of microbial community composition, diversity, and structure can be achieved and validated through TRF (Terminal Restriction Fragment) analysis of the 16S rRNA gene. It has been successfully utilized in bacterial community differentiation amongst a wide variety of sample sources and has been compared to denaturing gel electrophoresis and 16S rRNA gene cloning for this purpose. In addition to community differentiation, it can also be applied to analysis of the relative structure and phylotype richness of a community in addition to specific organism identification. This method bypasses the limitations present in cultivation-dependent methods by comparing PCR amplified DNA sequences from environmental samples. Its application has broadened to include analysis of fungal ribosomal genes, 16S rRNA genes, archaeal 16S rRNA genes, functional genes, and physiochemical cycle and metabolic product encoding genes. Small subunit (16S or 18S) rRNA genes from total community DNA is the most commonly targeted gene for amplification and study (1, 5, 13, 18, 40, 51, 52, 73, 81).

T-FRLP (Terminal Restriction Fragment Length Polymorphism) analysis is based on restriction site location variations among generated sequences while the lengths

of fluorescently labeled TRFs are precisely determined on automated DNA sequencers by high-resolution gel electrophoresis and the use of internal size standard in every profile. Detected TRF profiles are quantitatively measured in terms of size in base pairs, fluorescence intensity (peak height), presence, and width (length of fluorescence), etc. Theoretically, the composition of the generated data can be extrapolated from available sequence databases (there was a correlation between a decrease in predicted phylogenetically informative TRFs and TRF bins as the number of available reference sequences decreases) (5, 27, 51).

TRF data is high throughput and easily obtained in comparison to other community profiling methods. It can also confidently have application in community structure analysis of spatial and temporal shifts. Analysis of the data is also greatly facilitated through the use of current fragment analysis and software packages allowing for statistical precision and increased application in microbial community dynamics on a scale previously unobtainable. The use of 16S rRNA amplicons in the production of TRF patterns also have applications related to diversity studies of the community profile. Additionally, the TRF patterns are generated by electrophoresis systems integrated into DNA sequencing platforms and thus allow greater precision and higher resolution that any other current community profiling method available (1, 2, 40, 51).

Many TRFLP methodologies utilize fluorescently labelled universal primers which will anneal to prokaryotic 16S rRNA gene conserved regions for PCR amplification to yield genomic DNA of a microbial isolate or community sample. Only using one fluorescently tagged primer can lead to a sever underestimation of microbial diversity as various bacterial species may have the same terminal restriction fragment

length for a given primer combination; thus resulting in a lack of resolve in comparison to the use of two tagged primers. Multiplexing can be used for communities that contain different taxa. PCR, restriction enzyme digestion, and gel electrophoresis are integrated in the process of TRF pattern generation. A PCR of extracted sample DNA is performed using fluorescently tagged primers that are homologous to the target gene's conserved region of interest before fluorescently labeled amplicons that are specifically designed for a given tetranucleotide recognition sequence present within the amplicon product are subjected to restriction enzyme digestion to produce a pattern of different fragment lengths after which. Purification of digested amplicons is later performed to remove excess salts that were introduced during the reaction processes. Electrophoresis integrating fluorescence detection of labeled terminal restriction fragments (TRFs) in a DNA sequencer is then accomplished. Automated software programs for fragment analysis generates TRF profile peak data based on TRF peak retention time in relation to a DNA size standard (ROX 1000) to generate electropherograms that can later be analyzed via multivariate statistical methods. Most literature supports that the despite the ladder used, accuracy can only be achieved up to 700 bp. From this data, binary measurement of fragment data in addition to profile abundance within samples can be utilized in comparison of differing communities based on distance and similarity measurements (1, 2, 40, 51, 52, 73).

There are PCR biases possible in T-RFLP community analysis including those introduced during amplification that could skew estimates of organism abundance relative to gene copy number. As such, the resulting TRF patterns can only accurately be applied to describe the relative abundance to specific amplicons and not to relative

organism abundance. The appropriate primer size must also be taken into consideration as resulting patterns may misrepresent the actual diversity present if amplicons do not contain the targeted restriction site. The best possible estimation of diversity can be achieved with an optimal amplicon length is between 400 and 700 bp. This will avoid the data loss apparent with long amplicon use (1, 40).

Other biases that may skew data representation include any inconsistencies in loading the same DNA quality as well as quantity across samples for comparisons. Neglect of sample preparation fidelity therefore resulting in a decrease in overall representation as increased amounts of DNA is to be loaded for a TRF pattern with a large number of similarly sized peaks due to the signal being diluted across the peaks. In contrast, less DNA can be used for patterns with few or unevenly distributed peaks in order to avoid saturation of the fluorescence detector. Also, a lack of sample size consistency may result in skewing abundance and presence results of ensuing data in addition to introducing random bias. This is especially applicable to gut microbiome as specific bacterial are present based on location due to the physiological conditions of the system. Unfortunately, these patterns cannot be ascertained in advance and require trial and error or the use of an idea, equal purity DNA quantity range of 50-200 ng digested DNA in 20-50 ul reactions. Again, the use of differing DNA quantities can also lead to errors within the TRF pattern analysis as well as similarity profiles derived on presence/absence data. Inconsistencies among replicate preparation, the disparity of cell lysis, DNA adhesion and presence of extracellular DNA can also introduce error (1, 5, 13, 18, 40, 52, 81).

Manual interpretation and manipulation has shown to reduce overall data representation and introduce additional error as well as bias. Additionally in manual TRF manipulation, sample identity and PCR replication already considered, random lane-to-lane variability is attributed to residual variability and sample variability in small peak detection can be reflected by TRF rarity and not the expected rarity of a given genotype (5).

Peak size discrimination is generally set to 1 base pair during peak alignment. Manual alignment is subjective and should be avoided. Euclidean coefficients are not ideal as data with blocks of double zeros or fragment absences. A 1% threshold during peak filtering can be statistically relevant for community analysis (68). There can be discrepancies in fragment sizes due to the relative error apparent from sequence identity on migration within a polyacrylamide gel. This factor can contribute to an alteration in seemingly phylogenetically specific TRFs and the use of 'binning' for comparative TRF sizes. inconsistencies among replicate preparation, the disparity of cell lysis, DNA adhesion and presence of extracellular DNA Binning is employed to account for TRF drift caused by improperly sized TRFs due to differences in fragment migration and purine content. DNA standardization can decrease replicate TRF profile variation. The use of differing baseline thresholds can decrease reproducibility. DNA standardization can alter branch lengths (18, 27).

Despite the relative simplicity of the TRF procedure, profile reproduction among sample replicates has yet to be achieved in its completeness. Past research has shown that upwards of 85% of all generated TRF peaks can be irreproducible artifacts. Replication is imperative especially if peak representation to a specific phylogenetic assignment is to be

considered. Past TRFLP research has also been based on the mere presence or absence of peak pattern profiles in characterizing the similarity of profiles of a specific sample (the number of peaks to TRF patterns have in common). This contributes to a lack of precise community fingerprint and phylogenetic diversity. However, it may still provide information pertaining to estimated community richness. Complex community structure has also led to incongruences phylotype richness and structure assessment. The inference of phylogenetic composition based on the TRF profile depends on the TRFs phylogenetic resolution, or the similarity of organisms responsible for a specific TRF size, as well as the quality and quantity of comparative reference sequences available. Sequence discrimination by a specific TRF is generally inconclusive and generally yields a skew in sequence distribution for a specific TRF, as extremely few TRFs are specific for a given species or genus. Comparative diversity within a community can still be deduced from phylogenetically relevant TRFs if a larger scale is as well as analysis focusing on a single bacterial division (27, 40).

Various inherent biases are apparent in relation to applications based on DNA and PCR usage that limit TRF pattern interpretation and relevance though the introduction of error and artifact. For example, lack of consistency of fragmentation across samples can skew representation of data, and specificity and bias is also affected by the lack of completeness of ribosomal housekeeping and functional genes in databases for adequate primer design and phylogenetic capture as it is more probable to amplify dominant groups of target sequences than their less abundant counterparts in any given environmental sample. Therefore, in a community having 10^8 cells, only up to 0.1% of populations within the microbial community will be detected. Other sources of error

occur if there is production of single-stranded pseudo-terminal restriction fragments due to incomplete amplification which is cycle number dependent as well as production of secondary structures that are less common for functional genes than 16S rRNA genes due to the difference in complexity. Incomplete digestion due to presence of trace PCR enhancers and additive, template purity and complexity, and PCR salt interference, differential electropherogram migration from resulting fragment length differences, lack of consistency in fragment lengths observed at separation and those sequenced are caused by the differential migration patterns of the fluorophores present on internal ladder sizing standard and the labeled fragments (*i.e.* ROX ladder has an additional twelve carbon atoms than the six carbon FAM label, also FAM and HEX fluorescein dye labeled DNA migrate faster than ROX labeled DNA containing the fluorescein rhodamine dye. Therefore, HEX and FAM generated fragments can be underestimated; there is no current resolution for this problem) also are problematic. Some ladders may result in double peaks after separation be double stranded causing miscalls by software algorithms. Size variation may also be attributed to fragment purine (A/G) content due to the discrepancies in molecular weight. Inconsistencies in gel composition and running conditions, such as temperature and time, may also mitigate migration changes), and data set alignment (peaks are measured in base pair units while the height of each peak is measured in fluorescence units) also result in a lack of representation in the data produced (40, 52, 73).

However this is not unique to this profiling method but also applies to other microbial community analysis methods currently used which also have capabilities of identification of unique profile patterns in relation to the organism responsible for that

result that TRF profiling methods cannot accomplish. 16S rDNA consists of a variable conservation of restriction site positions and thus can lead to a reduction in the resolution achieved by using TRFLP analysis on a species level (unique restriction site position in each species) to that of higher order groups and reduces the TRF profile complexity generated from the community sample limiting the diversity estimates that can be made. The use of a database for phylogenetic assignment of TRF patterns is often nonspecific and imprecise (which is why MiSeq, Mothur, and MG-RAST was performed). Lack of completeness within the database can also lead to incorrect phylogenetic assignment especially applicable for target sequences other than the 16S rRNA gene. The extensive rRNA sequence database has been utilized to determine unique TRF peaks in nearly all reports integrating 16S rDNA TRF patterns with several applying taxonomic assignments to TRF patterns and concluded that often a single TRF can represent several general. Therefore the resolution is lacking and further sequencing and analysis is needed for accurate, high resolution phylogenetic assignment as is apparent in the use of Mothur and MG-RAST software packages. Additionally, it was shown in RDP database analysis that several unrelated organisms can produce the same TRF size (40). Also, despite having developed immensely since its introduction T-RFLP analysis still lacks a consensus in which statistical analyses should be ideally employed (1, 18, 51, 52).

Recent developments in data processing and analysis of TRF profiles has aided in obtaining a reduced-bias, objective analysis of unique, rare, and common profiles which can be applied phylogenetically to aid in diversity studied (1, 5, 13, 18, 40). Various methods have been employed within the literature, however the most ideal method to be utilized is one that reduces the presence of type II errors which concern the lack of

detection of differences in actually different profiles and therefore rejecting the null hypothesis. Actual relationships amongst profiles can be facilitated by the used of analytical replicates and statistical tests in addition to automated software packages that reduce the biases introduced by manual data manipulation while determining significant differences between samples by standardizing data and integrating statistically sensitive algorithms for treatment of outliers, distinguishing 'true peaks', binning, grouping, comparisons, measurements of size, distance and peak area, and noise separation, and other means of differentiating between T-RFLP profiles if derived from the same sample. This is accomplished in spite of errors and innate biases introduced during the procedure (1, 5).

However, there still remains limitations for presumptive assignment of bacterial group identification within a community, and the extent of phylogenetic application is dependent of extent of universalness and specificity of the primers used. No known universal primers can hybridize or amplify all sequences available. The sequences present in the database used to generate said universal primers represent only a portion of the total species diversity present in the natural microbial world and therefore lack complete resolution. Interpreting data should be done cautiously and only as an estimation of diversity as population presence is dependent on rank abundance and those microbial populations that are not dominant numerically are often not represented and therefore species diversity of the environmental sample is vastly underestimated. Also, gene copy number among species and biases introduced though out the procedure can yield skewed, unrepresentative products in relation to rank abundance of the original DNA sample template. Only very general phylogenetic inferences can be drawn and the

degree of phylogenetic information obtainable from this method is dependent on the limited performance of the PCR primer used. Distinct community signatures cannot be produced (51).

Processing and analysis

Capillary or polyacrylamide gel electrophoresis is generally used to determine differences in fragment abundance and length according to a size standard while actual measurements are a mere estimation by algorithm interpolation (abundance is due to fluorescence intensity) and are later assigned to OTU bin categories which may include multiple phylotypes pending primer and enzyme resolution of species complexity and phylogenetic relatedness within the sample. The degree of accuracy in fragment data collection directly relates to its applicability of accurate community composition representation. As noise varies by run, an automated objective procedure is required to determine this baseline threshold (either by peak height or area) by an arbitrarily chosen value of fluorescence units. The use of high threshold noise decreases noise but may lose data related to small reproducible peaks. The use of a threshold also assumes that there is little experimental variation in sample preparation. Run-to-run variability also affects size estimations of the same bacterial phylotype (73).

The operational definition for a species cites a 3% dissimilarity, but is not widely accepted. Methods employing OTU approaches avoid many of the limitations implicit to phylotype analysis due to the lack of bin restrictions since taxonomy outlines are not applied. Sequences can thus be assignment and clustered with equal basis regardless of

reference sequence representation or restrictions issued in outline classifications. OTU assignment is depended on the presence of other sequences in the dataset. However, this methodology assumes that the 16S bacterial rRNA gene evolves at the same rate among all taxonomic affiliations. Analysis and clustering algorithms are also computationally intensive for OTU-based methodology. Mothur utilizes a neighbor joining algorithm that is taxonomy-independent and performs better than deterministic and heuristic methods available. OTUs can represent sequences from multiple lineage assignments due to there being no taxonomic level threshold commonly employed (77).

As TRFLP analysis is increases in popularity, the range of methods for data analysis has also increased. Many of these processes have shown to lack appropriate statistical testing. The Bray-Curtis coefficient can be employed for ideal construction of similarity matrices of TRF profiles (68).

Sequenced data is imputed into Peak Scanner (Applied Biosystems) where defaults are altered to represent the use of ROX 1000 size standard and NPP analysis method. After which, all poor or bad peak signals being weak or noise are removed and remaining data is exported to Microsoft Excel where all peaks that did not represent the fluorophores used were removed from downstream analysis according to electropherogram patterns, or traces, presented within this program that display an array of TRF peak sizes in relation to variable fluorescence intensities produced by the fluorophores used (68).

Remaining data was formatted for diversity analysis via Primer 6 software by using program defaults with the exception of 'samples as rows,' and Shannon, Simpson and Margalef values were generated for each of the imported peaks. After which,

averages and standard deviations of the replicates of each of the sample points were found and plotted in Microsoft Excel based on primer direction and sample site pairing (12).

The same quality Peak Scanner data was also utilized for PCA and AMMI analysis via T-REX after formatting in Microsoft Excel. Default settings were utilized with the exception of settings regarding "using peak area" and setting the clustering threshold at 0.5 when aligning the peaks. Data type was designated as TRF presence. Each site was analyzed by including both reference and village samples while considering directionality of each TRFLP primer individually as well as together. Configured data was utilized to produce PCA graphs (interaction component analysis) for the AMMI study.

Fragment analysis of the TRF data is first achieved via Peak Scanner software and then exported into the proper format for the T-REX database. Raw data files are generated by the size-calling software, peak scanner. The other file required for analysis by TREX software is a labeled file containing data describing each of the samples and correspond to factors within variable design (such as dye/sample peak, sample name, size, height, area, and data point). Samples having more than one fluor were processed by treating the same fluor as a unit of peaks to distinguish between data produced from other fluors.

T-REX

Due to being highly parallel, robust, and automated in nature, use of the microbial fingerprint technique T-RFLP is ideal in developing comprehensive data sets that are

required for adequate study of microbial communities. Past peak recognition algorithms were problematic in discriminating between shoulders and smaller, individual peaks when on the edge of or in close proximity to larger peaks. Additionally, broad peaks were miscalled as multiple sub-peaks having an apex of less than one base pair. These issues have resulted in the use of base binning for approximate peak sizes. Though reducing error, this practice also decreases resolution and introduces the problem of an incorrect assignment into a bin which can result in an alteration in similarity profiles, presence/absence analysis, and individual TRF peak abundance analysis amongst generated patterns. Instituting a similarity coefficient, which is instituted in many databases and current software packages, can correct for this; however dissimilarity and unique peaks cannot be distinguished by this application. To circumvent all of these problems in addition to subjectivity and time consuming factor related to manual alignment, utilization of the raw electrophogram data (as is generated by peakscanner software) in the analysis of programs with more integration algorithms that are more flexible can be performed (TREX). Analysis of peak area surpasses peak height for accurate DNA abundance analysis in an electropherograms due to peak width increasing functionally according to retention time (a standard amount of short DNA will have the same width but a different height compared to a longer fragment of the same amount) (40).

T-RFLP analysis expedited (http://trex.biohpc.org/), or T-REX, is a software package that was utilized in this study for the rapid, flexible, and consolidated processing of raw data into a format matrix for interpretation and analysis via the AMMI, or additive main effects and multiplicative interaction model (17). The use of the AMMI program to

analyze the results of the T-REX program allows for the determination of the primary and secondary drivers of the microbial community samples, sample heterogeneity, and interaction effects as given by pattern signals and noise. The program is a free, web-based molecular-based microbial community fingerprinting technique that allowed for labelling, uploading, and processing raw data while bypassing various barriers that were previously existent in manual T-RFLP analysis such as unwanted variability of error prone steps that are time intensive and subjected researcher bias and variability including difficulty in distinguishing true peaks from "noise," peak alignment across samples, matrix creation and manipulation from raw data of TRFs, and deciding the best multivariate analysis for the given dataset in order to analyze environmental variable effects and treatments on the present microbial community composition (17).

Benefits provided by the use of this program that allow for the specialized and customizable analysis of the applied sample set consist of the following: raw data can be labeled in relation to the attributes of the sample's experimental design and TRFs can be aligned/binned in the same manner, true, active peaks in the corresponding data results can be distinguished from noise via a baseline threshold (that for this study was derived empirically to be 50 relative florescent units, and past applications sub-optimally applied an arbitrary threshold for peak delineation from noise which does not account for any variation of noise within samples according to discrepancies in DNA amount utilized, or sample bias), the data matrix produced via several built-in filtering mechanisms can vary on several complexity measures such as variance distribution among main and interaction effects and sample heterogeneity, and users can manipulate data sets with multiple fluors such as HEX and FAM, for more statistically applicable and representative analysis.

Additionally, within the filtering process, "environments" are established from samples that are replicates or conceptually equivalent groups with designated values, while "replicates" denotes samples attributes in the file are all identical. In aligning T-RFs, the size (base pairs) is determined by utilizing an internal size standard in referencing the TRF, such as ROX1000 (17).

In T-REX, true peaks are those that, assuming a zero mean, exceed the standard deviation generated among all peaks and multiplied by a factor provided that is based on original electropherogram results. TRF alignment is accomplished based on binning according to base pair size in relation to an internal size standard; however, analytical errors may result due to random fluctuations, purine content and fluorophore behavior yielding TRF drifting. There is no standard treatment of these errors. The effect of the alignment used, manual or automated, can only be determined by downstream multivariate analysis (17).

Production of various degrees of data matrices (typically two-way) from tabulated or listed data generated by PEAKSCANNER™ software which produces column specific data per variable; *i.e.* one column representing all TRF sizes and another for all peak heights, which allows for accurate data representation and algorithmic processing including variance distribution among interaction and main effects in addition to heterogeneity within the samples and AMMI analysis of the resulting data matrices. This yields ordination scores and other tables: the first table is a full ANOVA table, the second consists of estimations of interaction sum of squares for patterns and noise for replicated data, the third pertains to percentages of variation from interaction and main effects, and the fourth summarizes interaction signal variation percentage generated by the first

interaction principal component axes. The second axes (IPCA 2) represents secondary drivers of community structure such as site differences. AMMI analysis shows variation contributions according to: TRFS (variability in the means of different TRFs), environments (number of peaks or overall signal strength in TRFLP profiles), and TRF environments (how TRFs respond differently with environments (17).

Most software applications are not equipped with the processing capacity to thoroughly analyze datasets by all data matrix manipulations where the primary data types consists of the following: presence/absence, peak height and area, replicate sample averages, experimental factor examination, spurious TRF deletion. Bay-Curtis or other ordination statistics are used to analyze profiles, and binary matrix assignment can be compared via Jaccard's Correlation. Binary data had the lowest measure of main effects variation and the highest interaction effects in 90% of the analyzed datasets. The lack of processing power can often lead to a lack of representative data including signal and noise differences which can result in obscured ordination patterns. Additionally, most software packages are not exploratory in multivariate analysis and rely heavily on available sequence databases (17, 18).

The web-based free software package, T-Align (http://inismor.ucd.ie/~talign/), which T-REX implements, replaces the time intensive and error prone manual comparison of TRF profiles for identification of microbial community factors and the changes they may produce on community structure. It integrates an automated algorithm for peak alignment and to facilitate in exploratory analysis. The algorithm uses replicate TRF profiles of a given sample to generate a sample profile to remove pseudo TRFs. Later, sample profiles and their relative fluorescence are compared. Those TRF profiles

differing by less than 0.5 bp are considered to be the same and are aligned as such according to the moving average algorithm. Each profile represents only one TRF in the overall alignment, and those profiles that do not contribute a TRF are analyzed again within a greater than 0.5 bp average TRF size. When a new TRF is identified, sa new average will be utilized until all TRFs have been binned. Default alignment rounds to the nearest nucleotide size integer. Peaks are then clustered and grouped with a TRF assignment according to size according to a clustering. Other packages similar to T-Align are expensive and fluctuate in available features. T-Align results in a 2.7% average difference in area percentage of each peak among duplicates while other software can rage up to 7 or 11%. Grouping samples into environments allows for replicate analysis. Rare TRFs generated by the quality control factors within the program can also be omitted (17, 80).

The data produced in T-RFLP analysis can be applied to clone libraries or databases of 16S rRNA sequences for phylogenetic assignment or be interchanged with other techniques that are PCR based. T-RFLP has been applied to analysis of all domains. Bacterial community analysis is generally performed to identify species of a given sample, determine sample diversity, or compare community samples that have been separated according to space and or time. Peak quantity and area of TRF profiles relate to the richness and evenness of a bacterial population. Problems have often risen in relation to profile comparison which is needed for spurious peak elimination as well as species assignment (80).

AMMI

Of the ordination methods that are not those that test a specific hypothesis such as 'two microbial communities being significantly different' or apply collected data to environmental variables *i.e.*canonical correspondence analysis, AMMI, also known as 'doubly-centered PCA', and T-RF centered PCA (not environmentally centered as subtle treatment differences are not consistently captured) are among the most robust methods analyzed by yielding consistent ordinations (1, 18).

AMMI has increased utility as it can be applied to interactions that are non-linear in a complex model while discarding a noise rich residual by decomposing a matrix of residuals produced by fitting the additive main effects according to single decomposition value. Interactions are graphically represented in a biplot. A bilinear AMMI model can be constructed by differential genotypic sensitivity to the most discriminating hypothetical environmental variables that are estimates of the data. A biplot with the first PCA axis (IPCA 1) can be generated by plotting genotypic and environmental PCA scores against their means. If the PCA score is nearly zero, the interaction effect is small. If a genotype and environment have the same sign on the PCA axis, the interaction is positive; if different, it is negative. A biplot can also be employed to describe yield and stability if IPCA 1 comprises a large amount of the interaction SS (34, 65).

Biplot use is most successful when the GxE interaction, or the genotype being under the major environmental effects of genotype by environmental interactions, is concentrated most heavily in the first or the first two PCA axes (IPCA 1 and or 2). Reliable stability conclusions can be elucidated by biplots when the interaction of a

sample is represented by the first or the first two PCA axes (IPCA 1 and or 2) as mainly noise was captured in further interaction principal component axes and thus did not provide any contribution to observations predicting validation of results. It has been noted that the stability increases (according to the Wricke's ecovalence stability measure, $W_{i(AMMI)}$) with increased implementation of PCA axes. The use of biplots to describe the interaction is limited by the dataset (what percentage each IPCA contains separately as well as together). Interactions better represented by a biplot of genotypes and environments of the first two PCA axes by plotting the second PCA scores against their respective first PCA scores. The interaction between the two factors, genotype and environment, is obtained by a projection of either vector on the other. In any given quadrant, the interaction between the two is positive. The end point of the vector from the origin (0,0) determines the stability of a variety or environment. If the vector is in closer proximity to the origin, there will be reduced interaction effects and can thus be labeled as stable. Stability according to this biplot is more precise than a biplot comprised of only the first PCA axis. A significant amount of interaction must be contained within the first two axes for this biplot to be successfully representative (34, 18, 70).

The additive model can still retain structure in the terms representing the interaction and can be modeled by both residual error and multiplicative components, or the number of components chosen so that white noise is represented by the residual. A combination of both of these expressions yields the AMMI model for data organized in a two-way table. Parameters for this model can be projected by use of column and row main effect means and decomposition of the value of the interaction parameter residual

matrix. If the interaction has at least one term, an assumption is made that errors are homogeneous in variance after nonadditive effects have been considered (24). Higher interaction effects are regarded as unstable and the means of the two factors will result in high variability. Stability increases with proximity to the origin. Interaction of a given variable to another is obtained by projecting one variable's vector on to the other and analyzing the length of the vector where the projection occurs. If one variable is in the same quadrant as another, the interaction is positive (34). Even distribution can lead to low interaction (24).

However, the variation in results pending method used provided analysis that is qualitative in nature instead of quantitative as conflicting recommendations for the number of interactions to retain and use vary by dataset (24). The AMMI Stability Value (ASV) can be obtained by integrating results produced by both IPCA 1 and IPCA 2. A large sum of squares for environments correlates increased diversity among environments, and the cause of generated variation among the dataset. If the magnitude of the sum of squares for the genotype to environmental interaction is larger than that of the genotype alone, a significant difference in genotypic response is evident across environments. The percent capture of interaction sum of squares correlates to the degrees of freedom for each IPCA. The combined percentages for each shows the representation of the genotypic environmental interaction. Variability in both main effects and interactions in relation to environments can be shown by scattered data among environmental locations on a biplot. High potential environments are evenly distributed in a given quadrant with minimum interaction effect. Lower potential environments cluster in a given quadrant with high PCA value where lowest yielding environment have

the highest positive scores. Independent of direction, the greater the IPCA score, the more specifically adapted the genotype is to the environment. Higher interaction with environments can also be concluded by high ASV and rank which show erratic, unstable yield across environments. Positive interactions are due to interaction scores having similar signs. A score close to zero and ranked first (least) is ASV value correlates to minimum GEI or stable yield over environments, and negative interaction scores allow increased performance in environments having negative interaction values (70).

The biplot, generated by using genotypic and environmental scores of the first two AMMI components utilized, identifies genotype assignment to a given unique location. The magnitude of the genotype to environment interactions or differential responses dictates varietal ranking across the measured environments. ANOVA can describe the main effects and quantify interactions through an analysis of variance; however, unlike AMMI, this analysis cannot be applied to genotype-environmental interaction identification. When equally studying main effects and interaction, employment of the AMMI model can increase accuracy by incorporating both ANOVA and PCA into a single methodology. Refinement of a given location is made by adjustments composed from information pertaining to other locations within the dataset with the addition of removal of residual or noise variation from the interaction of genotype and environment (70).

The only experimental design requirement is the use of a two-way data matrix to provide a graphical representation in the form of a biplot to simultaneously summarize interaction and main effect data of both environment and genotypic data. Within the AMMI model, ANOVA analysis separates interaction from the additive component of the

data to which PCA is then applied as a multiplicative model for analysis of the interaction generated from the additive ANOVA. PCA scores are then plotted against one another to yield a visual representation of components of the genotype to environment interaction in the resulting biplot. Applying stability statistics to this biplot yields genotypic grouping according to performance across the present environments. Disparity amongst the resulting genotypic responses to the various environments within the genotype to environment interaction can place a limitation on the estimates of accurate yield and the identification of genotypes with high comparative yield. Both biotic and abiotic stressors can cause genotype to environmental interaction, and increased resistance or tolerance of the genotype to these stressors can reduce the interaction if needed (70).

Paired-end 16S rRNA (V3/V4)

Utilization of 16S rDNA sequencing (of two hyper variable regions of the gene, the V3 and V4 regions, respectively) has allowed for circumvention of various hurdles that were present in past analysis (poor resolution and sensitivity as well as an increased financial burden) as well as an in-depth analysis of this "rare biosphere" while maintaining breadth of coverage. The shorter sequences yielded allow diversity indication (21). Paired-end analysis is completed when sequencing both the 5' and 3' ends, which in turn doubles the number of base pair reads on the Illumina platform. The V3 region of the 16S rRNA gene overlaps, while the V4 region does not due to being greater than 200 bp in length. However, both V3 and V4 regions can be utilized to identify OTUs while constructing phylogenetic assignments in tree form. Overlapping

reduces error rate (85). Base pairs in the conserved structures of all the domains and the non-canonical base pairs are present in irregular helices and the V3 region. Secondary structures that are conserved across all domains have shown to form the same three-dimensional structure. Comparative analysis has allowed for the accurate prediction of the secondary structure as well as a number of tertiary structure interactions within many RNAs. Most of the tertiary stricter interactions observed in eukaryotes are also present in prokaryotes. Protein synthesis, and essential function for all life, occurs at the ribosome. Peptidyl transferase and decoding relates to ribosomal RNA. The V4 region at position 588-652 (as present in *E. coli)* forms a compound helix of 55 nucleotides in SSU rRNA within bacteria. There are three base pairs in the pseudoknot in bacterial SSU rRNA at 505-507/524-526, and 3D folding in the 540 region is similar in both prokaryotes and eukaryotes (46).

Paired-end (PE) sequencing is an improvement on the previous 16S rRNA studies; it has 2 PCR steps with different pairs of primers where initially the two primers contain a primer for Illumina specific sequencing, an index, or barcode, sequence, and a gene-specific primer followed by the use of two primers that contain an Illumina adapter and primer sequence for sequencing) although it is limited in that sequencing is based on built in Illumina sequence primers and requires two PCR cycles thus increasing the possibility of artifacts and requiring 20-25 nt for the sequencing index and gene-specific primer (43).

However, an additional method of 16S analysis exists which utilizes only one PCR cycle, primers that contain an index (reverse primer) and Illumina adapter sequence, hairpin formation preventing 10-nt pad, 16S non-complementary 2-nt linker, and a gene-

specific primer. The combination of pad, linker, and primer is used as the sequencing primer for the 5'end, the reverse complement of the pad-linker-primer combination acts as the sequencing primer for the 3' end to sequence the index region, and the combination primer also acts as the 3' sequencing primer to yield an index sequence and 2 250 nt reads after 500 cycles (43).

The assembly of paired-end ~125 base reads advantageously incorporates quality control steps for generation of 16S rRNA gene sequence reads while allowing unprecedented access to reduced abundance microbial DNA. The high number of sequence generation for libraries utilizing the Illumina sequencing methodology allow for high levels of completeness in regards to sampling efficiency in addition to replication abilities. The paired-end analysis technique is advantageous to other high-throughput sequencing techniques in that it reduces the amount of erroneous sequences that are included in downstream analysis (imposing a quality control step) while providing enormous data sets. A smaller quantity of sequences may provide sufficient data for extrapolating underlying patters between highly differing communities, however larger data sets are mandatory for the identification of more acute, subtle responses to environmental factors within less abundant populations for increased sequence coverage of the rare biosphere. The length (two fold coverage), and therefore quality of generated libraries, of the reads is also increasing and index sequence use allows for parallel sequencing of samples. Illumina base calling is improved due to the existing algorithm for optimal cluster identification within the flow cell when there is maximum diversity of nucleotides across the first four bases sequenced in the forward read. It can have issue with low abundance phenotype identification and alpha diversity exact measurements (3).

The V3 and V4 regions of the 16S rRNA gene overlap. A total of 2168 reverse primers with differing indices have been published for the V4 region, and taxonomic groups can be assigned easier to longer sequences when employing a classifier (43).The V3 hypervariable region consists of roughly 200 nucleotides (170-190), has high taxonomic resolution, and consists of a conserved flanking region. It is also compatible with paired-end sequencing which allows for overlapping 3' end sequencing, reduction of sequencing errors, and generation data applicable to pipelines of computational analysis that are publically available. When assembling paired end reads by aligning the 3' ends of both forward and reverse reads, additional quality control is provided. Discarding reads that do not assemble as contigs due to sequencing errors of mismatches between complementary reads, decreases artificial sequence number (3).

When the V3/V4region is sequenced, a 0-7 bp heterogeneity spacer is applied to the index sequence for equal sample proportions to be sequences out of phase to mitigate the set back of 'low sequence diversity' amplicons produced in 16S analysis by allowing the 16S gene amplicons from an equal proportion of samples to be sequenced out of phase thus reducing this issue. Both 250 bp and 300 bp paired end MiSeq protocols for analysis of the low sequence diversity, conserved 16S rRNA gene allow for flexible, cost-efficient sequence options. The V4 hypervariable region is targeted in the most widely used dual-indexing MiSeq paired end approach. There are 9 hypervariable regions flanked by conserved sequence regions in the 16S rRNA gene. A 469 bp region is required to contain both the V3 and V4 regions while maximizing the length of the generated MiSeq reads. This region provides sufficient data for microbial community taxonomic classification of specimens associated with the human microbiome (as is

employed in the human microbiome project). Strict filtering can alleviate error rates that tend to increase at the ends of reads and thus improve the accuracy of taxonomic assignment while avoiding spurious read assembly (30).

Next-Gen sequencing

Next generation sequences has facilitated the interest and understanding of human, animal, and environmental microbial community structure and function in addition to increasing the knowledge of novel pathogens and functionality and effect of a consortia of microbes on a myriad of other ailments and disease states (78). The past employed observational phylogenetic analysis of yet to be cultured novel taxa has shifted to characterization of microbial community taxonomic shifts through experimental techniques. Previously utilized Sanger sequencing has not provided the sufficient coverage for adequate community analysis. 16S gene libraries achieved by Illumina technologies denote the immense increase in sequence number and insights into diversity and composition of community samples proved by extant next-gen sequencing of serial analysis of ribosomal sequence tags (SARST) (previous sequencing capabilities were limited to less than 101 base reads while yielding lack-luster error rates). Illumina is better because it has higher resolution, higher sequence read lengths, and is cheaper than the Roche/454 pyrosequencing platform (Branford, CT, USA). The Illumina platform produces up to 1.5 billion reads per run in comparison to the 1 million reads per run on previously primarily used 454 Pyrosequencing platform at a comparable cost Led by the expanding field of bioinformatics, technological advancement within the last decade has

transitioned the past focus on 16S gene fragment sequencing within the hundreds via clone libraries to next-generation sequencing of millions of fragments through sequencing technologies such as the Illumina MiSeq platform (3, 30, 43, 85).

Capable of generating at least 250 nucleotide paired reads, the MiSeq sequencing platform is adept at multiplexing a large number of samples in addition to sequencing shotgun metagenomes in parallel while maintaining a reduced rate of error, higher sequence coverage, and reduced cost ($1000/lane and $125,000 for the instrument compared to the HiSeq, 300 cycler, $1500/lane and $740,000 instrument cost, and Illumina is 50-12000 fold less expensive per sequenced megabase than pyro and Sanger sequencing) comparable to other available platforms. Also, it employs a chip-based bridge amplification procedure prior to reversible terminator dye nucleotide synthesis mediated sequencing with up to 500 cycles of sequence data where each cycle is split into two individual reads which provides paired reads of the same DNA fragment template. The MiSeq platform has high sequencing throughput as it is capable of utilizing paired 250 nt reads to generate 8.5Gbp (*i.e.* 17 million read pairs). Although the HiSeq has been generally utilized for metagenomics shotgun sequencing, the MiSeq suits 16S rRNA studies due to the increased length of reads it yields in addition to the performance and cost standards. Genomic and metagenomics sequencing on this platform has shown, recently, to be an sufficient complement to 16S rRNA gene sequencing due to its ability of generating high quality reads that can be distributed across a range of samples with ease (43).

Despite the increased affordability and high-throughput nature of available next-generation sequencing technologies for analysis and characterization of the composition

of microbial communities, technical limitations inherent to the sequencing platforms utilized provide obstacles for adequate data analysis and processing during culture-independent profiling techniques. This shift in analysis from culture-dependent methodology is due to the realization that the amount of microbes from an environmental or human sample visible during direct staining exceeds by multiple orders of magnitude the number that could be cultured, termed "the great plate count anomaly" (30).

Past pitfalls for utilizing this platform revolves around the lack of sequencing ability when analyzing samples having low genetic diversity as is commonly apparent with 16S rRNA gene amplicon studies. However, technological and methodological updates in platform performance has improved the software for image analysis that was utilized, thus bypassing this issue. Also, only 5-10% of PhiX phage DNA is required in comparison to the past 50% that was used to artificially increase genetic diversity of reads. The platform software has also improved its ability to perform adequate cluster discrimination thereby improving the quality of results of cluster density and data quality is affected by the quantity of DNA loaded into the flow cell. Initial error rates vary according to cluster density. Usage of a pre-clustering step can reduce sequence error and unique sequence number through the brief sorting of sequences according to decreasing abundance prior to comparison of the rare sequences in a sequential manner. If the rare sequence consists of less than a set amount of bases away from a more abundant sequence, it is removed from the data and its abundance added to the more abundant sequence (43).

The low sequence diversity alluded to in regards to the 16S gene applies to the beginning cycles of the MiSeq run where successful cluster identification and pre-

phasing/phasing calibration depend on the composition of targeted amplicon heterogeneous bases in that they require a balanced composition of bases in the initial 12-18 cycles. Due to the inherent nature of the 16S gene, amplicon pools are highly homogenous and need to be sequenced along with a heterogeneous control library (commonly PhiX phage DNA in a 1:1 ratio) to improve the quality of the sequence reads, however there is a possibility of losing half of the reads to a template that is non-targeted (30).

There is a co-evolutionary arms race between new sequencing platforms and new software tools for the acquisition and analysis of data on an emerging, unprecedented scale. In Illumina analysis, a diverse set of environmental samples with an exceedingly large depth of >3million reads/sample can be sequenced, thus allowing for an exceptional level of consistency in respect to taxonomic recovery and the recapture of diversity patterns previously reported in relation to meta-analysis. This provides the possible future applications of analysis of large-scale studies of simultaneous analysis of thousands of samples to provide a survey of the microbial communities and identification of microbial phylotypes at an unparalleled spatial and temporal resolution. The improvement in sequencing technology and well as the programs available for analysis has allowed for a higher degree of relation between data samples to yield clearer, more accurate biological patterns as well as a boom in 16S rRNA gene analysis (9, 14, 83).

Processing and Analysis

Processor time and memory requirements must be minimized by algorithms due to the exponential growth of database content. Algorithms for sequence comparison are generally based on pairwise alignments. The degree of relatedness among sequences according to expectation values derived from local alignments is performed by methods available within databases, such as BLAST. Distance measures according to pairwise identity within the alignment are integrated in both neighbor-joining and UPGMA algorithms. CLUSTAL programs create binary trees via clustering prior to pairwise alignment for tree nodes (28). Despite technical advances in microbial community analysis on spatial, temporal, and taxonomic scales, inconsistencies in data interpretation in regards to genetic diversity of the 16S rRNA gene remain (77). High sample throughput, phylogenetic information pertaining to the species or present phylotypes at varying abundance, and economic affordability is necessary in methodology. RDP Bayesian classifier can be utilized to assign data to taxonomic groups for the generation of taxonomic profiles (3).

Amplicon pools prepared with AMPure XT beads (Beckman Coulter Genomics, Danvers, MA, USA) and Library Quantification Kit for Illumina (Kapa Biosciences, Woburn, MA, USA), and the streamlined Illumina MiSeq platform (San Diego, CA, USA), which is scalable and high-throughput and has application to various environmental samples were utilized for sequencing due to distance values and downstream analyses are impacted significantly by alignment quality (76).

In analysis of the produced data, there are two ways to bin sequences, by phylotype (similarity to reference sequences) or by OTU (similarity to other sequences present in the community sample). Phylotyping allows for a label to be issued on sequence according to relationship of past cultured and characterized microbes. However, innumerable organisms belonging to the same species have differing phenotypes, and some organisms with the same phenotype may belong to differing taxonomic lineage assignments. The phylotype approach is also not advisable for community samples that contain novel sequences that are yet to be identified or contain lineages that are previously unidentified. 6.6% of the reference sequences in the RDP database lack a genus-level assignment. Therefore, classification is limited to the available reference taxonomies as species level assignments are impossible due to taxonomies ending at genus. Phylotype-based methodology is also sensitive to sequencing errors. Additionally, the genus-level cutoff masks the sequencing error (77). OTU based approaches reduces the sequencing error rate from 1.08 to 0.01% (43).

Database-dependent methods are limited due to the lack of representation of rare and novel populations when analyzing the deep coverage existent in many environmental samples (43). Pre-clustering followed by clustering at 3% (equivalent to 97% sequence identity) allows for increased accuracy in OTU characterization, reduction of singleton sequence proportion, and minimally affects the distribution and presence of microbial taxa. Single nucleotide errors also had minimal effect on classification of sequences. However, a 97% sequence identity taxonomic classification standard can show shifts in clusters of higher and reduced dominance (3). However, filtering sequences reduced the genetic diversity between sequences when compared to unfiltered sequences due to the

fact that filtering increases the similarity among sequences and communities. Filters remove sequence information for community population differentiation despite their application to broad scale phylogenetic application at the phylum or kingdom level (76).

Greengenes, one of the many software data packages employed by MG-RAST, performs an alignment that poorly treats variable regions due to a higher prediction of genetic diversity, richness and phylogenetic diversity than produced via the SILVA and RDP-based alignments as employed by Mothur software. However, the RDP alignment does not align variable regions within samples. Data pertaining to variable regions cannot be easily related to full-length genes within the parent genome. A reduction from 1500bp to 200 bp sequence reads affects the application of any phylogenetic observations and OTU assignments made according to the limitation in read size. Most studies assume that partial sequence distances are not significantly different from full-length sequence distances despite the fact that the 16S rRNA gene lacks uniform evolution throughout its length. When employing the proxy species definition specific to full length sequences 3% distance cut off, the variability in evolution within the 16S rRNA gene becomes apparent. Genetic diversity also decreases along the 16S rRNA gene length. Regression coefficients do not adequately explain variation between regions in comparison to the whole gene. Longer reads increase the relation of segmented analysis to the whole gene. The 16S gene is a marker for diversity within a genome and follows a well-determined secondary structure. Sequence alignments also contain various innate biases including that relation by common decent results when same characters are treated as equivalent, and the more shared, the more related. This is treated for by many alignment algorithms by optimally achieving the maximum number of identities. Treatment of gaps in sequence content can

also affect the alignment produced. Phylogenetic analysis of these alignments can be accomplished via maximum likelihood (having a transition-transversion ratio of 2 and focus on variable characters for nucleotide sequence data), maximum parsimony, (assuming evolution is achieved by the most efficient, shortest path, and analyzes variable characters for both nucleotide and protein data), and distance, or neighbor joining, methods (which utilizes both constant and variable characters for establishing the most closely matched sequences, binning those as one, and continuing on throughout the rest of the sequences included in the data for both protein and nucleotide information). Gene analysis via next gen sequencing allows for replicates to be analyzed in addition to increased complexity of experimental designs to be investigated while increasing the breadth and depth of sampling. Pairwise comparisons and multiple sequence alignments are not as beneficial due to the fact that they ignore the secondary structure of the 16S gene (76).

Mothur

Mothur (www.mothur.org/wiki/MiSeq_SOP) utilized an rRNA gene sequence naïve Bayesian classifier to classify each sequence against the Ribosomal Database Project (RDP) 16S rRNA gene training set (version 9). Groups of sequences were generated according to the applicable taxonomic units (OTUs) at a 3% dissimilarity level (43).

Utilization of the simple nearest alignment space termination algorithm allows for quick, flexible generation of a quality alignments independent of the reference alignment

used while integrating the secondary structure of the sequence- thus increasing the possibility that positional homology is conserved between sequences. No difference in alignment quality between full-length sequences and variable region fragments has been shown. Multiple sequence alignments allow for the calculation of pairwise sequence distances and phylogenetic identification for the assignment of sequences to OTUs (78).

Unfortunately, the memory and time requirements for paired-end analysis by this program was extremely extensive compared to other programs and analysis methods integrated in this study. This is due to the scale of the code commands according to space and time for sequence length and quadratically to the third power in time for sequence number. As such as the length of sequences is exceeded by the quantity, the memory required for doubling the number in the alignment increases at minimum of four-fold and the time by eight-fold, realistically limiting alignments to less than 5000 sequences due to the limitations placed on RAM. Pairwise calculations of distances to circumvent this issue and provide sequence homology is even further time intensive. The aligners within the program also do not predict secondary structure of the 16S rRNA molecule making comparisons impossible without further re-alignment steps (78).

MG-RAST

Microbes housed in differing environments and the community dynamics thereof are commonly studied through use of metagenome, or random community genome, analysis by directly sequencing DNA from an environmental sample. As technology advances, the challenge with this methodology has shifted from sample sequence generation to sequence analysis (56). Metagenomic analysis has application for

investigating the community profile of abundant and rare populations, however it cannot be applied to most environmental samples due to limitations of sequencing and computation (3).

Metagenomics may allow insights into all genes, and thus a possible higher understanding of gene function, within a microbial community. 16S rRNA-based analysis allows for insights pertaining the detection of rare phylotypes as well as to past-unexplored biodiversity and ecological characteristics of either individual taxa or while microbial communities in addition to the relationship of trends and patterns observed at the species level to those of the host or environmental parameters evident (9, 42, 45).

MG-RAST serves as a high-throughput pipeline for high performance computing and annotation allowing low cost, next generation means of worldwide metagenomics sequence analysis. The functional sequence assignments of the metagenomics input produced by this software package are automated and are generated from both nucleotide and protein database comparison allowing for functional summaries and comparative phylogenetic analysis. This sequence analysis approach can replace the costly and time and labor intensive DNA plasmid cloning that precedes sanger, or dideoxy chain termination, sequencing (which produces longer reads but includes inherent cloning biases) (56).

The first step in either sequencing technique for analysis is to compare the generated data to known sequences existing on present databases. This computationally intensive task can be extremely time consuming if done manually, however it does provide data for the following analysis: phylogenetic comparison and profiling, functional annotations and metabolic reconstruction, and sequence binning MG-RAST allows for the analysis of raw

environmental sequence data for analysis by first performing a normalization step for unique internal ID generation and processing and prior to data summary generation. Sequence screening by use of BLASTX against the comprehensive non-redundant SEED database (and INSDC databases, rDNA databases [GREENGENES, RDP-II, and the European 16S database], as well as the chloroplast, mitochondrial, and ACLAME database of mobile elements) for protein encoding genes. Matches to existing databases are then computed to derive comparative data (56).

Phylogenetic reconstruction of the metagenomics sample is generated by application of the data provided in the SEED nr database and the ribosomal RNA database. Functional assignments are also generated by application of appropriate database information for the generation of metabolic reconstructions. The software was preemptively designed for later modifications such as the integration of new datasets and algorithms to allow for more extensive and new analysis steps or means of data comparison for any stage within the analysis process. The pipeline is executed in Perl through usage of the SEED framework, NCBI BLAST, SQLite, and Sun Grid Engine. However, the datasets that are currently available are not completely exhaustive and there has been shown to be as few as 10% or as high as 98% lack of sequence matching from a sample to a dataset. Accuracy of annotations is dependent on the quality of the data used. However, unique to this system, samples of differing sequence length can also be analyzed by this means due to the algorithms employed. Assembled sequences are generally longer and allow for increased accuracy in gene function or phylogenetic binning identity than sequences that are initially submitted without first being assembled (56).

CHAPTER 2

TERMINAL RESTRICTION FRAGMENT LENGTH POLYMORPHISM ANALYSIS OF BACTERIAL POPULATION DIFFERENCES IN *MUS MUSCULUS* CECUM IN RESPONSE TO PROBIOTIC ADMINISTRATION

Leasha M. Christensen,* Cynthia Blanton, and Peter P. Sheridan

Department of Biological Sciences, Idaho State University

***Corresponding Author: Leasha M. Christensen**

Email Addresses:

LMC: chrileas@isu.edu, CB: blancynt@isu.edu, PPS: sherpete@isu.edu

ABSTRACT: The cecum aids in maintaining homeostasis and host health by having the highest metabolite absorption and housing the most abundant population of microbes. Various disease states, for which there is no standard medical treatment, result from the disruption of the microbial populations in either presence or relative abundance due to various environmental and host factors. Supplementation with probiotics, prebiotics, and

synbiotics has shown promise as a therapeutic intervention to combat dysbiosis. Many gut microbiota cannot be isolated by culture-dependent techniques. High throughput, culture independent 16S rRNA T-RFLP analysis of cecum samples from mice fed a control or synbiotic diet showed conservation of homeostatic balance in synbiotic-supplemented samples according to Shannon, Simpson, and Margalef indices and insights into possible phylogenetics when further processed by AMMI analysis. Additional research is needed to more directly determine the diversity and phylogenetic effects of synbiotic supplementation on cecum content for the alleviation of dysbiosis among the present microbial populations.

KEYWORDS: dysbiosis, prebiotic, probiotic, synbiotic, T-RFLP, microbial diversity, AMMI

INTRODUCTION

Unique microbial communities are found within and on the human host as well as other vertebrates, such as the skin, gastrointestinal tract, mucosal surfaces, or the most largely populated region being the lower intestine, namely the cecum (11, 36, 53). Many of these microbes, which provide the host with metabolic and genetic features that are not innate and therefore develop a synbiotic relationship with the host, reside within the confines of the intestine and are collectively referred to as the "human gut microbiota," and its genome, the "gut microbiome," is often denoted to in literature as being a bodily organ (14, 15).

Host health is mediated largely by the microbiota existing within the gastrointestinal tract in that it affects various physiological factors such as: resistance to pathogen colonization, immune response regulation, nutrition, metabolism, and

development, and host homeostasis (8, 12, 13, 50). Being co-evolved, these interactions can be detrimentally affected due to administration of antibiotics and alterations of chemical transformations within the gut and may lead to acute and/or chronic illnesses including, though not exhaustively, the following: digestive, bowel, eating and weight disorders, cancer, cardiac events, allergies and asthma, type 2 diabetes, atopic diseases relating to the change in microbial community composition, and neurodevelopmental disorders (3, 10, 14-26, 32, 35, 37, 39, 43-45, 51, 58, 59). The gut microbiota also moderate the programing and control of various physiological functions such as epithelial development, blood circulation, as well as innate and adaptive mechanisms of immunity and energy homeostasis regulation. Therefore, in response to environmental factors that disrupt the host-gut interactions, metabolic diseases may ensue. Various aspects of the 'modern lifestyle' including traveling, dietary changes and restrictions, use of medications, age, as well as urbanization, geographic location, and stress level can contribute a disruption in this essential host-microbe interaction (6, 14-25, 32, 35, 37, 39, 44-46).

Although there is no standard medical treatment, various methods such as fecal-oral transplantations, bacteriotherapy, and antibiotic administration, have been employed to treat severe alterations in gut microbial populations upon the development of a disease to a chronic state (6, 7, 14-16, 42, 44). The lack of treatment options usually result in antibiotic administration, which is often unsuccessful. For instance, 20% of patients suffering from a *Clostridium difficile* infection (CDI) have a recurrent episode after initial antibiotic treatment, and patients having a recurrent episode are 40% likely to experience another (7). Also, metabolomics studies of 2000 murine metabolite features in fecal

samples have shown that a single high dose of streptomycin can cause significant changes in roughly 90% of the features analyzed (25).

The use of fecal-oral transplantations is a highly invasive, extensive process including various donor and recipient screenings for compatibility amongst other factors despite potential use (2, 7, 14, 42). In light of such, the use of probiotics, prebiotics and the combined use as synbiotics as a treatment method shows potential (14-26, 37, 39, 44, 45, 47, 59).

Probiotics are defined as being *"organisms and substances which contribute to intestinal microbial balance" or "a live microbial feed supplement which beneficially affects the host animal by improving its intestinal microbial balance* (6)*."* In order to fit this criteria, which does not include antibiotics, the probiotic needs to be stable and viable and remain as such under storage and during use, survive the intestinal ecosystem, prepared on a large scale, and beneficially effect the host after its integration. Surviving the acidic environment within the gut and then colonizing and becoming active in the colon can be problematic as adherence to the intestinal epithelium may be necessary. Competition of nutrients and ecological sites as well as stress can also cause a decrease in effectiveness in this treatment. Additionally, the probiotic must also remain present after the consumption of the product initially containing the strain has end (14, 25, 32, 35).

Lactobacilli, Bifidobacteria, and *Streptococci* are commonly used in probiotic treatments and have shown to alleviate, hepatic encephalopathy, carcinogenesis, diarrhea, colitis, pathogen colonization, constipation, gastroenteritis, immunostimulation, flatulence, and gastric acidity among other diseased states (25, 35). *L. acidophilus* is the most commonly used and tolerated probiotic, it has been shown to synthesize vitamin K

which is necessary for the conversion of the bone matrix osteocalcin to the active form and may thus aid in improving bone integrity (14, 16-25, 32, 35, 37, 39, 45, 59).

Prebiotics are defined as being "*a non-digestible food ingredient that beneficially affects the host by selectively stimulating the growth and/or activity of one or a limited number of bacteria in the colon, and thus improves host health* (6)*."* Prebiotics must serve selectively as a substrate for specific commensal, beneficial bacteria for growth or metabolic activation while not being absorbed or hydrolyzed prior to reaching its final destination. Its application must also result in microbial composition alterations to achieve a state of health while also providing systemic or luminal effects that benefit host health (14, 25, 32).

Non-digestible foods, or prebiotics, such as oligosaccharides, polysaccharides, fructooligosaccharides, and other naturally occurring non-digestible carbohydrates (resistant starch, nonstarch polysaccharides [hemicellulose, pectins, gums, plant cell wall polysaccharides]), peptides, and lipids (aid in cation absorption [Ca, Fe]) have been shown to improve host gut microbiota health by stimulating growth and activity of specific endogenous microbiota by changing the microbial composition of the local environment. Lack of absorption and digestion of these compounds is due to their chemical structure (25).They have been shown to particularly benefit host colonic health, and can directly manipulate metabolism of lipids via products of fermentation.

The combined use of probiotics and prebiotics is referred to as synbiotics, or "*a mixture of probiotics and prebiotics that beneficially affects the host by improving the survival and implantation of live microbial dietary supplements in the gastrointestinal*

tract, by selectively stimulating the growth and/or by activating the metabolism of one or a limited number of health-promoting bacteria, and thus improving host welfare (6)*."* They have shown promise in combating the diseased dysbiosis state and can also increase the shelf life, viability, and effectiveness, and functional activity of exogenous and endogenous bacteria of one another when used in conjunction than alone. Further use to combat pathogenic bacterial overgrowth, parasite growth, viral infections, burn treatment, stress, and antibiotic therapy effects also applies as these are associated with the translocation of bacteria due to the failure of the intestinal barrier (14, 25, 32, 35).

However, due to the lack of information pertaining to the composition of the gut microbiota under both healthy and diseased states, supplementation of effective pre- and probiotics has been challenging. Species shifts between the two states of health have remained elusive despite the current knowledge of phyla level changes. This current lack of information further limits the understanding of microbial community interactions on the supplemented treatments due to the heterogeneity of bacterial spatial distribution within the GI tract according to environmental differences. This results in differing activities of the components of synbiotic at diffing locales within the gut microbiome. Within a local environment, the metabolic products of one bacteria can be modified and utilized by another bacterial species so increasing the availability of a molecule in its active form can be mitigated by community level biotransformation reactions. These cooperative interactions directly affect the degree of effectiveness of a prebiotic, as the necessary active form may never reach its target location, or a probiotic, which may contain a strain that does not yield the desired, beneficial effect on microbial composition and function. Further research is needed to investigate and characterize the intestinal

communities of microbes to increase the efficiency and effectiveness of synbiotic treatments (25, 35).

Many of these microbes cannot be isolated and cultured by traditional culture-dependent techniques, as characterization of complex microbial community samples is an arduous process that limits many microbial ecology studies due to the difficulties of identification and quantification of microorganisms present within any given sample. Culture-dependent methods are restricted due to the limited information collectable from morphological data and the intricacies of isolation. As such, diversity can be more adequately assed by genetically-based techniques due to its application to a wider range of organisms (30). Therefore, the study of this microbiome is dependent on developing technologies that are culture-independent in nature in order to combat the previously imposed limitations on analysis. Various initiatives in both the US and Europe have been employed to increase the knowledge base through characterization of the microbes and their genomes within the human body for assessment of their impact and role in states of health and disease (Human Microbiome Project and MetaHIT Consortium, respectively) (14, 15, 25, 35).

Although progress has been made, the gut microbiome has yet to be characterized to its entirety. Among archaea, viruses, bacteriophage, fungi, and other Eukarya, this environment is composed primarily of bacteria, which attributes more than 1.5kg of weight and over 1100 species (14, 15, 25, 35). The exponential amount of species being analyzed consisting of 10^{11}-10^{14} bacteria per gram of intestinal content, requires a high-throughput means of DNA analysis due to the lack of accurate representation of diversity from culture-dependent methods of analyses (25, 35, 53).

Terminal restriction length polymorphism (TRFLP) is ideal for utilization as a cultural-independent means of rapid, high throughput, robust and quantitative analysis (statistical as well as molecularly) of the microbial community composition, diversity, and structure present within the mouse cecum tissue culture by means of fluorescent dyes that attach to PCR primers and the resolution achievable in current sequencing technologies (33, 54, 57).

Analysis of the data is also greatly facilitated through the use of current, improved fragment and sequence analysis and software packages allowing for statistical precision and a higher degree of resolution between data samples yielding a clearer, more accurate representation of biological patterns and increased application in microbial community dynamics on a scale previously unobtainable (28, 38). It has been successfully utilized in bacterial community differentiation amongst a wide variety of sample sources. In addition to community differentiation, it can also be applied to analysis of the relative structure and phylotype richness of a community in addition to relative organism identification thus bypassing the limitations present in cultivation-dependent methods. It can also confidently have application in community structure analysis of spatial and temporal shifts. The use of 16S rRNA amplicons in the production of TRF patterns also has application related to diversity studies of the community profile by providing insights pertaining to the detection of rare phylotypes as well as past-unexplored biodiversity and ecological characteristics of either individual taxa or whole microbial communities. This is due to it being present in all organisms, being large enough for informatics purposes (at roughly 1500bp) while containing both conserved and hypervariable loop regions and not being transferred horizontally (4, 9, 38). Additionally, the TRF patterns are generated by

electrophoresis systems integrated into DNA sequencing platforms and thus allowing greater precision and higher resolution that any other current community profiling method available (34, 55, 56, 57).

The most simple approach of trace analysis is binary comparison of different sample peak presence. Although valid, this approach lacks appropriate quantitative analysis. ANOVA, which provides a description of the main effects and quantification of interactions through analysis of variance, allows analysis of this as well as establishing whether either factor influences the microbial composition on an individual basis or may result from a contribution interaction amongst factors present (27, 31, 48). It is integrated in AMMI analysis to separate variation into interactions and main effects (TRF and environmental, or E, variation) and then it later applies PCA to the interactions to create and interaction principal components (IPCs) thus focusing on effects of treatments and environments on resulting TRFs (46). Relative stability conclusions in addition to identification of genotype assignment to a given location while simultaneously summarizing interaction and main effects can be elucidated by biplots and is most successful when the genome-environmental interaction is concentrated most heavily in the first or the first two PCA axes as use of more than two did not contribute to observations or validate results as noise was predominantly captured in subsequent PCAs (27, 31, 46, 49). When equally studying main effects and interaction, employment of the AMMI model can increase accuracy by incorporating both ANOVA and PCA into a single methodology. Within the AMMI model, ANOVA analysis separates interaction from the additive component of the data to which PCA is then applied as a multiplicative

model for analysis of the interaction generated from the additive ANOVA. Both biotic and abiotic stressors can cause genotype to environmental interaction (31).

PCA and other multivariate statistical methods have been used to employ the needed statistical rigor for complex data set analysis. An additional benefit to multivariate analyses is the use of numerous variables that are not constrained to species identification. PCA can be used to show trends in distances in community patterns. However, PCA data is not normally distributed. Additionally, non-linear data from large gradients can cause PCA ordination arcing and thus obscure any patterns, yet AMMI has increased utility as it can be applied to interactions that are non-linear in a complex model while discarding a noise and also aids in variation measures within a dataset including IPC interaction signals (27, 46, 49).

Mouse models allow for increased control during analysis as well as a decrease in the number of confounding variables while providing representation to the human microbiome, and beta diversity results showed that there was no significant difference in the generated means (14, 37, 40). Diversity studies are imperative to analysis of community structure and function, especially in relation to development of pharmaceuticals, probiotics, bioaugmentation, or substrate presence (24). As such, it is proposed that an intensive study of microbial flora within the cecum, the source of highest metabolite absorption (14-23, 44, 45, 52, 53, 59), in response to probiotic administration be performed. The administration of different probiotic diets will lead to a measureable change in the microbial populations within the cecum, where shifts in diversity and frequency of the bacterial lineages will be detected.

METHODS

Sample Preparation:

Three samples each of harvested cecum from male 9 month old mice (Harlan), *Mus musculus*, grown and provided by Dr. Cynthia Blanton, following either a controlled (B, C, F) or synthetic (H, J, L) diet were chosen at random. Based on a powdered form of American Institute of Nutrition (AIN)-93M purified rat diet (Dyets, Inc., Bethlehem, PA), the diets administered to the mouse model were modified to utilize cornstarch in place of sucrose and dextrin in order to reduce the susceptibility of osmotic dehydration of the bacteria studied within the synthetic diet. The isocaloric diets administered were composed based on carbohydrate ingredient manipulation by "assuming energy densities of 4, 0, and 2 kcal/g for cornstarch, cellulose, and fructooligosaccharide, respectively (1)." Nutraceutix, Redmond, WA, provided fructooligosaccharide and lyophilized probiotic cultures ($1x10^{11}$ CFU/g of equal parts *Lactobacillus acidophilus* and *Lactococcus lactis lactis)*. The diets were made fresh three times a week with addition of probiotics immediately prior to feeding each morning for 18 weeks (1).

The harvested diet-specific cecum samples were then sterilely dissected both laterally and vertically, and rehydrated using 6mL 10mM TRIS,pH~8.0, 1% Triton after being stored at -20°C. Samples were then incubated at 80°C for one hour, and a pellet was obtained by a low-speed spin (1000rpm) for five minutes. Residual supernatant was removed (the amount of supernatant produced was not consistent among the samples as the initial dryness of the tissue varied greatly when thawed to room temperature) after centrifugation at 16,000xg for 10 minutes. The cell pellet obtained from half of the initial

cell suspension solution was resuspended by addition of 250uL 10mM TRIS. The second pellet from the remaining original cell suspension was stored at -20°C.

DNA Isolation:

Genomic DNA of the cecum samples was obtained by a mechanical sheer forces protocol utilizing 100µL Lysozyme (100mg/mL) added to the rehydrated cell suspension solution and incubated for 30 minutes at 37ºC. After which, 100µL Proteinase K (10mg/mL in 10mM TRIS) was added, incubated for 30 minutes, and brought up to volume with Herwigs Lysis buffer in a bead-beating tube. Prior to supernatant being drawn off and added to isopropanol, tubes were then bead-beated for 5 minutes, boiled at 80ºC for 10 min, microcentrifuged at 3000xG for 5 minutes, and stored at -20ºC for 12 hours. Samples were then centrifuged at 13000xG for 10 min, and the pellet washed with 200µL 70% ethanol (4ºC) prior to centrifugation at 13000xG for 10 minutes. Ethanol was removed by drying at 37ºC for 60 minutes before DNA was rehydrated in 10mM TRIS, pH ~8.0.

T-RFLP and PCR Amplification:

A bacterial SSU rDNA T-RFLP polymerase amplification with the fluorescently-tagged universal bacterial 1492R (5'-FAM/TTACCTTGTTACGACTT-3') and 8F (5'HEX/AGAGTTTGATCCTGGGCTCAG-3') primers (1mM) and 1U (0.5µL) *Vent* exo (-) (New England Biolabs, Ipswich, MA) was set up and ran in triplicate in 50µl reactions for each of the genomic DNA samples with each reaching containing the following: 1x ThermoPol Buffer (New England Biolabs, Ipsich, MA), 400µM per each deoxynucleotide triphosphate (New England Biolabs, Ipswich, MA), and 1µL genomic DNA template. Individual master mixes and negative controls were used for each sample in a program

consisting of the following steps:10 minute denaturation at 95ºC, followed by 30 cycles of 95ºC/ 1 minute, optimized annealing for 2 minutes at 52.3ºC, and 4 minutes at 72ºC. The final cycle of the previously listed steps is immediately followed by a concluding 10 minute elongation stet at 72ºC. After visualization via low EEO 1% agarose gel stained with ethidium bromide, the TRFLP PCR product was purified by GeneJET PCR Purification kit (Thermo Scientific) and standard protocol.

10uL of the purified TRFLP amplicons were then digested with 5U (0.25uL/reaction) of restriction endonuclease *Taa*I I (Thermo Scientific), 20uL 10X Tango buffer, and 18.75uL nuclease-free water at 65°C for 2 hours. The T-RFLP digests were then purified by precipitating the fragments with 5µL 3M sodium acetate and 100µL 70%, and resuspending the pellet in 10uL nuclease-free water. 1µL purified terminal restriction fragments (T-RFs) were then submitted to the Idaho State University Molecular Core Facility for size determination utilizing GeneScan™ 1000 ROX™ Size Standard (Applied Biosystems) for fluorescently labeled DNA on a 3130 XL Genetic Analyzer (Applied Biosystems, Foster City, CA).

Statistical and Data Analysis:

Raw terminal restriction fragment size and peak area data generated for each sample was analyzed using NPP analysis in PeakScanner Software V.1.0 (Applied Biosystems). Fragments exceeding the threshold value (50 RFU) and designated as representative peaks and not noise in comparison to the size standard were collected and imputed in Primer 6 software (PRIMER-E Ltd, Plymoutoh, United Kingdom) following all default settings with the exception of ‘samples set as rows’ in order to generate respective statistical diversity indices of Margalef Species Diversity (species abundance

and frequency), Shannon Diversity (entropy or proportional abundance of a given species amongst a whole), and Simpson (probability of two randomly selected 'samples' belonging to the same species) Index values for comparison amongst the diets and individual cecum samples.

AMMI:

Additive Main Effects and Multiplicative Interactions (AMMI) analysis investigating microbial community variation and environmental effects on microbial flora within the cecum ecosystem of the mouse intestinal microbiome was performed using T-RFLP peak values and the web based T-REX, or T-RFLP analysis expedited) (http://trex.biohpc.org/) program due to its capabilities in labelling and uploading raw data, processing of molecular-based microbial community fingerprinting techniques, and providing a flexible, rapid, and consolidated analysis specific to T-RFLP data (5). A clustering threshold of 0.5 was selected in addition to analysis according to peak height. All other default settings were utilized. An analysis of variance (ANOVA) was generated and interaction principal component analysis (IPCA) was then performed on resulting ANOVA data through use of the same software to reduce the dimensionality of the multivariate data. This analysis allowed for the visualization of the main effects and interactions of microbial environment and genotypes simultaneously, and IPCA values corresponding to primary and secondary axes were then graphed (Microsoft Excel) for the generation of representative peak variability in relation to the 5' and 3' T-RFLP fragments. Analysis was performed in nine ways: individual analyses was performed on the control diet of each the forward and reverse primer and then both primers together; this was then replicated for the synthetic diet and for both diets together.

RESULTS

Data are presented by means of the AMMI model, which partitions variation measured into main effects and interactions during the analysis of variance through Additive Main Effects and Multiplicative Interactions (ANOVA), and then applies PCA (principal component analysis) to the interactions to create interaction principal components axis (IPCAs) amongst the samples analyzed allowing for simultaneous visualization of microbial environment and genotype main effects and interactions. PCA allows for a reduction of dimensionality in multivariate data by the creation of key variables that characterize the variation within the complete dataset by serving as a composite of various original variables. The new variables generated do not correlate with one another and are utilized without resulting multicolinearity. T-RFLP data was initially analyzed nine different ways according to differing diet and fragment (HEX or FAM) combinations.

Control

The control forward fragment TFR profile depicts the two dimensional interactions and variations produced by analyzing the forward HEX fragment of the control diet fed mice (B, H, J) (Figure 1). A significant total of 74.09% variation has been captured in this analysis thus providing adequate representation of the data. Greater than twice as much variation was generated among the x-axis (IPCA 1) than the y (IPCA 2). The replicates of each of the mice are not all located in proximity of one another and relates to increased variation and interaction amongst sample replicates, as shifts in ordination are determined by relative vector distance from the ordinate or other data points present (48). 66% of the data points are located in quadrants one and two of the

figure. High similarity and overlap is evident in respect to mouse H only. Those points having a position less than or close to 0 on the y-axis have a negative correlation or interaction to the other points and can also be considered more stable due to the relative reduction in interaction effects present on those data points, as the end point of the vector from the origin determines the stability of the genotype in the given environment. Those farther from the origin, as can be determined by vector projection, are more sensitive and adaptable to the environmental parameters applied through the application of the diet. The most ideal genotypic content can be found in quadrant 1, as there is a positive interaction in regards of both IPCA 1 and 2. Quadrant 2 represents a low yielding, stable genotype, quadrant 3 an unstable, low- yielding genotype, and 4 an unstable high-yielding genotype. This data produced a mean square error rate of 0.03571, 36.32% TRF main effects percent variation total, interaction percent variation total composed of 50.6% pattern and 13.08% noise, and an 88.96825 total sum of squares value correlating to increased diversity among environments and thus the cause of generated variation I this dataset (data not shown).

The analysis of the reverse fragment of the control diet analysis represents 285.15% measured variation amongst the replicates of these samples (Figure 2). The cumulative percentage exceeding 100% corresponds to complete capture and complete recovery of all predicted signal. No variation was generated on a two-dimensional basis as there is no variation value represented on the y-axis. Not all replicates share the same location or are in close proximity within the figure. The exception lies with samples J and H. Replicates 1 and 2 of H and 1 and 3 of J share the exact position within the fourth quadrant of the graph. This correlate to decreased variation and a high interaction

amongst these replicates due to a lack of projected vector shared by each of the data points. The second replicate of J is within the same quadrant as the other two replicates of that sample. The location of data points belonging to all of the replicates of J and the first two replicates of H have a positive interaction due their presence in the same quadrant.

Both samples H and B have two replicates within the same quadrant with the remaining replicate in another quadrant. This shows increased, positive interaction between replicates 1 and 2 of each sample independently and decreased interaction with its remaining replicate. Replicates from sample B all have positive interactions with the environment due to their location on the y-axis. The scatter of replicates from sample B correlate to variability in both main effects and interactions. Those samples closer to the origin are less adaptive and less sensitive to the environmental influences wrought on by the diet, and can therefore be characterized as more stable. This data produced a mean square error rate of 0.10741, 48.78%TRF main effects percent variation total, interaction percent variation total composed of 8.81% pattern and 42.41%noise, and a 30.70833 total sum of squares value correlating an overall lack of diversity among environments (data not shown).

Figure 3 shows an interesting ‘mirrored relationship of the sample replicates in an analysis of the control samples for both forward and reverse fragments. In some cases, one direction (either forward or reverse) of the sample replicate is found in the mirrored opposite relative position of the graph as is seen in B2, B3, H1, H2, and H3. J2 is close to following this pattern. Shared data points, represented by stacked replicates, show a positive, increased interaction and decreased variation due to the lack of vector distance between the data points. A total of 69.39% variation of the sample replicates is

represented in the two dimensional analysis above. There is more variation in the x-axis than in the y-axis as is represented by the relative percentages listed. Those points that are closer to the origin are less sensitive and are less adaptable to the environmental effects. Those points having a negative value (contributed by a position having a negative value from either the y or x axis) have a negative interaction and correlation to the environment tested against. Those samples with the most ideal genotype are found in quadrant 1, as their position is relative to a positive value in regards to IPCAs 1 and 2. This data produced a mean square error rate of 0.02673, 21.05% TRF main effects percent variation total, interaction percent variation total composed of 65.78% pattern and 13.17% noise, and a 168.69399 total sum of squares value correlating to increased diversity among environments and variation (data not shown). The scatter of data points across the biplot correlate to variability in both main effects and interactions.

Synbiotic

The two-dimensional analysis of the synbiotic forward fragment captures a total of 59.56% variation (Figure 4). There is slightly more variation correlated to the x-axis (IPCA 1) than the y-axis (IPCA 2) as is represented by the listed percent variations on the axis legends. There is no overlap of data points within this figure. However, there seems to be increased localization of data points amongst the origin thus demonstrating increased stability of variety and environment due to the reduced interaction effects. Sample replicates do not share close proximity with the exception of the slight closeness within the F sample replicates. The samples are spread out showing increased variation and less interaction and stability among sample replicates. Increased dimensional analysis could include a higher degree of variation. Those points having a negative value

(contributed by a position having a negative value from either the y or x axis) have a negative interaction and correlation to the environment tested against. Replicates in the same quadrant share a positive interaction comparable to one another. Those samples with the most ideal genotype are found in quadrant 1, as their position is relative to a positive value in regards to both axis. Replicate L1 is the most unstable as its vector from the origin is the largest and its overall value is negative relative to its position in quadrant 4. This data produced a 26.57%TRF main effects percent variation total, a 70.58% interaction percent variation total, and 42043682.93 total sum of squares value correlating to immense diversity among environments and variation (data not shown).

Data corresponding to the reverse 3' FAM fragments yielded by T-REX analysis of the synbiotic mice is shown in Figure 5. A total of 89.92% is captured by both IPCA 1 and 2 values. Increased variation can be found in IPCA 1 compared to IPCA 2 as is evident in the value listed on the axis. The most ideal genotypes can be found in quadrant 1 as there is an overall positive value for the point relative to being greater than 0 in relation to both IPCA 1 and 2 values. Clustering of data points as is evident in both quadrant 1 and 2 refers to increased interaction and decreased variability of those data points. Data having a negative value in respect to axis location, have a negative interaction and correlation to the other points and environmental conditions imposed from the diet. Those points farthest from the origin are more sensitive and adaptable to environmental effects due to the projected vector from the origin, and points located about or near the origin have increased stability. L1 is the most isolated of all the rest of the data points. Quadrant 1 houses the most replicates. However, 7 of the 9 data points are either on or very near to the origin or have a close to 0 value in terms of IPCA 2

value. The lack of shared location corresponds to increased variation amongst samples. This data 42.27% TRF main effects percent variation total, 42.27% interaction percent variation total, and a 4715170.35069 total sum of squares value correlating to increased variation and diversity among environments (data not shown).

A 54.63% total variation is captured from analysis of IPCA 1 and 2 of the forward and reverse TRF profiles of synbiotic mice (Figure 6). However, this does not contribute to a significant amount of interaction capture within the biplot to be considered fully representative of the dataset. More variation can be found in IPCA 1 than 2 as is relative to the value listed on the axis. The most ideal genotypes can be found in quadrant 1 (5 replicates present) as there is an overall positive value for the point relative to the origin in relation to bother IPCA 1 and 2 values. Those that have a negative value, from either of the axis, have a negative interaction and correlation to the other points and environmental conditions imposed from the diet. Those points farthest from the origin are more sensitive and adaptable to environmental effects. However, many of the replicate data points are found among the origin or across the x-axis, correlating to increased stability. The clustering of data points, when regarded in terms of ordination shifts according to vector distance from the ordinate or from other data points yields decreased interaction across data within the same general location. L1f has the most negative interaction relative to the rest of the data points. The reverse fragments are more clustered than the forward, and with the exception of l3f, there is separation of the fragments on the graph. The forward fragments are found further from the origin than the reverse. Scattered data across the biplot corresponds to variability in both interactions and main effects to the environment. This data produced a 12.77% TRF and 3.72% environmental

main effects percent variation total, 83.52% interaction percent variation total, and 51663738.15435 total sum of squares value correlating to increased diversity among environments and thus the cause of generated variation within this dataset (data not shown).

Combined Diet

A lack of significant interaction capture to be considered fully representative of the dataset applies to this analysis of the forward fragment of both of the diets combined as only 56.41% variation is captured by both IPCA axis values (Figure 7). IPCA 1 has the most variation and also has increased replicate clustering about its axis. The clustering of data points about the origin relates to increased stability, and less adaptability of the replicates within the effects of their given environmental changes imposed by the diet or other factors within the analysis. The diets do not separate from one another or follow any distinct patterns. All replicates of a given sample are not found within the same area. Quadrant 1 houses no replicates, and therefore there is no positive interaction or correlation amongst the replicates present. C3 is the largest outlier, having the largest ordinate vector to the origin or other data samples, as it is isolated and found furthest from all other samples. Some of the replicates (H1 and 2 as well as F1 and 3) share a location on the graph. The overlap or shared location of data points corresponds to decreased variability and increased interaction amongst those of the data points. This data analysis yielded a mean square error rate of 0.02857, 33.09%TRF main effects percent variation total, interaction percent variation total composed of 54.13% pattern and 12.78%noise, and a 172.21587 total sum of squares value correlating to the diversity among environments and variation (data not shown).

A complete capture in variation as is evident in an IPCA value of 958.57 % in the first axis was generated in when investigating the reverse fragment of the combined analysis of both the control and synbiotic diets (Figure 8). 14 of the 18 replicates share data point, or location on the graph. This correlates to an increase in interaction and lack of overall variation amongst all of the samples sharing this pattern. Only 2 of all of the replicates measured are found in quadrant 1. These 2 data points have a positive correlation and variability. The rest of the data points have negative variation due to their location given a negative value contributed by IPCA 1, 2, or both. None of the replicates share the same location for a given sample. Those replicates furthest from the origin having the largest ordinate vector are more sensitive and adaptable to environmental conditions such as alterations caused by diet supplementation. Data points closer to the origin are more stable to environmental effects caused by diet supplementation than those farther away with a larger vector. This data analysis yielded a mean square error rate of 0.089027, 58.95% TRF main effects percent variation total, interaction percent variation total composed of 2.1% pattern and 38.94% noise, and a 62.35985 total sum of squares value correlating to the diversity among environments and variation (data not shown).

A total of 63.5% variation in represented in the two dimensional graphic analysis of both the forward and reverse fragments of both diets combined (Figure 9). This value does not contribute to adequate representation of the sample summary by the analysis provided. There is more variation about the x-axis (IPCA 1) than the y (IPCA 2) as correlates to relative percentage value listed. Stacked replicates designations show shared data point position, increased interaction, and decreased variation amongst replicates included. 27 of the data points share data share the same location with at least one other

data point. 14 of the 27 are reverse replicates. Replicate data points having a shared locale interestingly have the same fragment designation, either forward or reverse. There are no shared data points between both of the fragment types. Additionally, each quadrant only contains one fragment designation. The graph is split with forward fragments on the right side (quadrants 1 and 4) and reverse fragments on the left side (quadrants 2 and 3). The forward fragments are clustered more around the x-axis than the reverse. Increased proximity could correlated to decreased variation and increased interaction between the sample data points. Those data points located in quadrant 1 have a positive interaction and correlation to variability. Data points closer to the origin are less adaptable and sensitive to environmental factors and are therefore more stable due to the projected ordinate vector from the origin. This data analysis yielded a mean square error rate of 0.02158, 22.77% TRF main effects percent variation total, interaction percent variation total composed of 64.89% pattern and 12.34% noise, and a 341.95604 total sum of squares value correlating to the diversity among environments and variation (data not shown).

Representative Diet Samples

The two dimensional interactions and variations produced by analyzing the forward HEX fragment of the "best" representative control (H) and synbiotic (L) diet as determined by Peak Scanner absorbance plots of TRFLP data are graphically represented in Figure 10. None of the data points included in this figure are clustered about the origin. This represents increased sensitivity and adaptation to environmental conditions and therefore decreased stability of the variety or environment due to an increase in interaction effects. However, each representative sample produced two of its three

replicates within the same general area. The higher interaction effects are unstable causing the measured variability in the produced means. The closeness, as measured from a projected vector connecting the data points, correlates to decreased variation and increased interaction. Those with a positive correlation to variation, provided by a positive value from IPCA 1, 2, or both, can be found in Quadrants 1 and 2 on this graph. Quadrant 1, which houses a shared data point for H replicates 1 and 2, has the most positive correlative values. L2 in quadrant 3 has the most negative correlative value due to negative value contribution from both IPCA 1 and 2. Neither sample has all of its replicates clustered about a given data point. There is more variation in IPCA 1 than IPCA2, and a total of 138.23% is represented by both axis, which cumulatively captures all interaction within this dataset. This data analysis yielded a mean square error rate of 0.06667, 40.32%TRF main effects percent variation total, interaction percent variation total composed of 30.79% pattern and 28.9%noise, and a 32.15 total sum of squares value correlating to the diversity among environments and variation (data not shown). The relatively small sum of squares for the environments of this dataset coincide with decreased diversity among the environments and the general lack of generated variation among the dataset.

A depiction of the reverse 3' FAM fragments yielded by T-REX of the control (H) and synbiotic (L) mice is shown in Figure 11. Each sample has triplicate forward and reverse data points. The number after the initial sample letter designation depicts the replicate number. There is an overall negative variation among both IPCA 1 and 2 as is evident with the cumulative -147.7 % variation from the axis totals. The negative values listed by the axis correlate to the difference in sign between the genotype and

environment and can allow increased performance in environments having interaction values that are negative. IPCA 1 has more negative variation than IPCA 2. Four of the 6 replicates share the same data point due to a lack of variation and increased positive interaction amongst those samples. For either sample, all of the replicates are not grouped together. Quadrant 1 houses the most positive correlational replicates and quadrant 3 the most negative. Samples furthest from the origin, L1, have increased sensitivity and adaptive properties to environmental changes, such as diet, as can be measured by the relative small length a projected vector from the origin to the data point. This data analysis yielded a mean square error rate of 0.14286, 69.51%TRF main effects percent variation total, interaction percent variation total composed of -20.64% pattern (a negative interaction) and 51.14% noise, and an 18.38095total sum of squares value correlating to the diversity among environments and variation (data not shown). The relatively small sum of squares for the environments of this dataset, which is the smallest measured value of all of the datasets, is due to decreased diversity among the environments and the general lack of generated variation.

Figure 12 depicts the 5' forward HEX and reverse 3' FAM fragments yielded by T-REX of the control (H) and synbiotic (L) mice "best" sample designated by data collected by Peak Scanner. Each sample has triplicate forward and reverse data points. The number after the initial sample letter designation depicts the replicate number and the concluding f or r forward and reverse, respectively. Complete variation capture among the first two PCA values yielded a total of 90374.3% variation is represented by both IPCA 1 and 2 with increased variation in IPCA 1. Out of the 12 replicates present, half share a common location, or data point, on the graph. This corresponds to decreased

variation and increased interaction amongst those samples as would be evident from a projected vector connecting the data points. The lack of clustering about the origin is representative of increased sensitivity and adaption of the samples to their environment in addition to an increased lack of stability to interactions and measured multivariate effects. Of the 12 samples, 10 have at least one positive contributing value from either IPCA 1 or 2. The most positive correlational values in terms of variation are found in quadrant 1 as both IPCA 1 and 2 supply positive values, thereby contributing positive variation effects. Quadrant 3 is the most negative due to negative supplied values. Reverse fragments cluster more amongst each other and away from forward fragments. . This data analysis yielded a mean square error rate of 0.04545, 24.56% TRF main effects percent variation total, interaction percent variation total composed of 47.54%pattern and 27.89% noise, and a 63.64091 total sum of squares value correlating to the diversity among environments and variation (data not shown). This is relatively small sum of squares comparable to all other datasets and contributes to decreased diversity among the environments and the general lack of generated variation among the samples measured.

Diversity Indices

Diversity indices representing the average Margalef species richness (d) of TRF profiles generated by cecum content of control (B, H, J) and synbiotic (C, F, L) diet fed mice is represented in Figure 13. This diversity index measures species abundance and frequency. A larger standard deviation is represented by the relative size of the error bars for the control mice in relation to the synbiotic. This correlate to a lesser degree of consistency amongst the individual sample replicates. In contrast, the small standard deviation evident in sample C, a synbiotic diet fed mouse, reflects replicated results with

less variation, and thus the value derived being more reflective of the present species richness within the sample. There is a consistency of measured species richness among the control samples as is evident by their relative position on the graph. However, analysis of the data points correlating to the symbiotic diet shows a slight decrease in diversity comparative to the control samples, there is a noticeable spike in the average measured species richness within the synbiotic sample F, and sample L produced the lowest value. Overall trend analysis correlates to a conservation of species richness during the administration of probiotics. Therefore, one can conclude that community composition is not inversely disturbed and function and homeostatic balance is maintained as is contrary to antibiotic administration, and there would be a decreased chance of dysbiosis occurring as a side effect to this diet treatment.

Amongst the control samples, B has a greater measured standard deviation than the other two mice which both share a relatively small error bar size when analyzing the Shannon diversity measure for entropy or abundance of a species amongst a whole community sample (Figure 14). The relatively small standard deviation error bar size of H and J compared to B correlates to a higher degree of representation of the entropy or proportional abundance of a given species amongst a whole as is reflected by this index value. The control diet has the highest degree of variation as sample B also has the highest standard deviation and lowest Shannon diversity index value of all the mice while H yielded the highest Shannon value. The synbiotic mice contained the sample point with the least standard deviation, L. The other mice within this category had the second (F) and third (C) highest standard deviations amongst all of the mice analyzed. The measured Shannon diversity within this category was fairly consistent. When comparing all of the

samples together, the measured average Shannon diversity value for the replicates of each sample were similar to one another as differences were based on the hundredth. This correlates to a conservation of Shannon diversity after then administration of synbiotics when compared to the normal, control diet composition. The consistency across samples alludes to a maintenance and lack of disturbance of gut intestinal community composition and therefore activity upon synbiotic treatment.

When analyzing the Simpson diversity Index (1-ʎ'), the probability of two randomly selected samples being composed of the same species identification, of both diet categories of cecum content, the control diet contained the top two highest standard deviations (B, J) amongst all of the samples, but it also contained the second smallest measured (H). Mouse H also had the highest diversity value. However analysis across samples showed a low degree of variation in the average diversity value obtained. Mice fed the synbiotic diet had small standard deviations with the exception of F. The lower standard deviation could correlate to a higher degree of representation to the actual community structure in relation to the species identity measured by this diversity index. The lack of measurable difference in the Simpson values obtained related to a conservation of homeostatic presence and balance after the administration of the synbiotic diet in comparison to the controlled diet. This correlates to a decreased chance of ensuing dysbiosis from the diet treatment and thus validates the treatment option when compared to the negative effects often resulting from antibiotic or other invasive treatments.

DISCUSSION

Culture-based techniques have led to novel insights into the study of bacterial community structures of culturable organisms within a given environmental sample. However, there is limited application in revealing the complete diversity or phylogenetic assignment of many of the environments being studied. DNA-based molecular methods have facilitated this and circumvent the limited scope of culture dependent techniques by identifying sequence diversity according to genes present within a sample (49). Species richness is characterized by the number of species within a given community or sample and species evenness refers the size of the species population within that same community. Both are used as parameters to investigate diversity and structure within a community and can be qualitatively estimated based on the unique frequency of occurrence of ribotypes detected during TRFLP despite being limited by conventional culture-dependent methodology due large fractions (85-99.9%) of microorganisms present in nature being refractory to cultivation. Lack of identical environmental parameters composing the exact environmental structure as is found in nature during cultivation imposes an additional limitation by altering the community structure by the new selective conditions introduced to the environmental sample. This results in an evolved community structure that may not represent the initial structure present during sampling. Molecular approaches utilizing isolated total community DNA as a template for study, such as TRFLP which couples PCR and rRNA-based phylogeny, avoid these limitations by obviating the requirement for cultivation while providing useful insights into the sample identity by pairing results with database integration (29-31, 33, 34, 40, 46, 55-57).

Construction, screening, and analysis of clone libraries is both time and cost intensive. Other techniques have been developed (such as DNA melting behavior and single-strand conformation) to circumvent clone library use and assess community structure while providing a crude qualitative assessment of species diversity. However, they are limited by the lack of sensitivity of the materials employed in the procedure (ie staining) and do not yield data relative to phylogenetic assignment or identity of a given microbial community. There is also a limitation for presumptive assignment of bacterial group identification within a community (29-31, 33, 34, 40, 46, 55-57).

However, the extent of phylogenetic application is dependent of extent of universalness and specificity of the primers used. Currently, no known universal primers are present that can hybridize or amplify all sequences available. And if utilizing cloning for phylogenetic identification, the sequences present in the database used to generate said universal primers represent only a portion of the total species diversity present in the natural microbial world and therefore lack complete resolution (33). T-RFLP analysis therefore stands as a robust, high throughput, automated culture independent means of analysis that bypasses many of the limitations imposed by other techniques employed.

The TRF profiles produces had no consistent patterns when analyzing either the forward or the reverse fragments across samples within a given diet or between diets. The shifts in ordinations, applied as the measured distance of projected vectors from the ordinate or from other data points, produced a range of variability, stability, and interaction amongst samples from a given environmental factor (such as diet) as is evident from the data point locations from the origin of the figures previously mentioned. An increased distance from the origin correlated to decreased stability where the means

of the measured factors contribute in increased measured variability. Those data points that shared a given quadrant had a positive interaction. The PCA scores or (IPCA%) generated on either axis of the figures represent the interaction intensity of the genotype to the environment; the smaller the score, the less interaction effect, and similarity in sign between these two factors correlates to a positive interaction shown while dissimilarity, the inverse. Stability according to the biplot utilized in this study is more precise than a biplot comprised of only the first PCA axis, though a significant amount of interaction must be contained within the first two axes to be successfully representative. Being relatively high despite ranging among the datasets, a large sum of squares for environments as generally reported in this analysis correlates increased diversity among environments, and the cause of generated variation among the dataset. Variability in both main effects and interactions in relation to environments can be shown by scattered data among environmental locations on a biplot, as was evident in many of the analysis performed, and those high potential environments are evenly distributed in a given quadrant with minimum interaction effect. Lower potential environments clustered in a given quadrant with high IPCA value where lowest yielding environments had the highest positive IPCA scores (31, 40).

Qualitative diversity measurements of the TRF profiles produced from the cecum content of each of the representative diets generated an overall trend corresponding species richness conservation during the administration of probiotics (Figure 13). This supports the conclusion that community composition is not negatively affected. Also that function and homeostatic balance is maintained, leading to a decreased probability of inducing a diseased state of dysbiosis as a result of probiotic administration which is

contrary to typical antibiotic administration. Also, Shannon diversity values for all samples were comparable due to differences observed according to the hundredth of the value observed (Figure 14). This correlates to a conservation of Shannon diversity between the diets. This consistency also correlates to a preservation and lack of disturbance of gut intestinal community composition and activity upon synbiotic treatment. A reduction in the probability dysbiosis and conservation of homeostatic balance from the diet treatment in addition to validation of the synbiotic treatment option when compared to the negative effects often resulting from antibiotic or other invasive treatments is shown by the lack of significant difference amongst measured Simpson diversity values and low generated standard deviation when compared to the control diet (Figure 15). These findings could also correlate to a higher degree of representation to the actual community structure in relation to the species identity

Analysis based on peak area was accomplished due to relative peak height investigation resulting in the presence of the following error: deletion of the smallest peaks, often believed to be within the range of observed noise, causing variation in the level of effect it had on dendogram error rate and thus correlating that those peaks could represent frequently occurring, important TRFs in terms of distinguishing between samples (54). However, fragment is not technically representative of an OTU as multiple organisms can produce identical TRF profiles (46). There has also been difficulty reported in designating accurate identity to each TRF in complex gene profiles. Each peak in a complex microbial community sample is generally representative of multiple TRFs of the same size produced by multiple species (56).

Despite its use and application, interpreting data generated during T-RFLP analysis should be done cautiously and only as an estimation of diversity as population presence is dependent on rank abundance and those microbial populations that are not dominant numerically are often not represented and therefore species diversity of the environmental sample is vastly underestimated. Also gene copy number among species and biases introduced though out the procedure can yield skewed, unrepresentative products in relation to rank abundance of the original DNA sample template. Only very general phylogenetic inferences can be drawn and the degree of phylogenetic information obtainable from this method is dependent on the limited performance of the PCR primer used. Distinct community signatures cannot be produced (29-31, 33, 34, 40, 46, 54-57). Also, the profile reproduction among sample replicates has yet to be achieved in its completeness, thus contributing to a lack of precise community fingerprint and phylogenetic diversity analysis while still providing information pertaining to estimated community richness. Complex community structure has also lead to incongruences phylotype richness and structure assessment. The inference of phylogenetic composition based on the TRF profile depends on the TRFs phylogenetic resolution (the similarity of organisms responsible for a specific TRF size) and well as the quality and quantity of comparative reference sequences available. Sequence discrimination by a specific TRF is generally inconclusive and generally yields a skew in sequence distribution for a specific TRF, as extremely few TRFs are specific for a given species or genus. Comparative diversity within a community can still be deduced from phylogenetically relevant TRFs. Additionally, there can be discrepancies in fragment sizes due to the relative error apparent from sequence identity on migration within a polyacrylamide gel. This factor

can contribute to an alteration in seemingly phylogenetically specific TRFs and the use of 'binning' for comparative TRF sizes.

Various inherent biases are also apparent in relation to applications based on DNA and PCR usage that limit TRF pattern interpretation and relevance though the introduction of error and artifact including those present during sample preparation and DNA isolation, amplification, digestion, electropherogram migration and inconsistencies in gel composition and running conditions, and data set alignment. However this is not unique to this profiling method but also applies to other microbial community analysis methods currently used (56, 57). Further metagenomic studies are necessary to elucidate more exact measures of diversity and phylogenetic assignments for the activity and effectiveness of synbiotics as treatment options to combat dysbiosis and preserve homeostatic balance within he host as is mitigated by the gut microbial flora.

Acknowledgements

MRCF

References

1. Abdo, Z., Schüette, U. M. E., Bent, S. J., Williams, C. J., Forney, L. J., and P. Joyce. (2006) Statistical methods for characterizing diversity of microbial communities by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes. *Environmental Microbiology.* 8(**5**):929-938.
2. Blackwood, C.B., Marsh, T., Kim, S-H., and E. A. Paul. (2003) Terminal Restriction Fragment Length Polymorphism Data Analysis for Quantitative Comparison of Microbial Communities. *Appl Environ Microbiol*. 69(**2**):926-932.
3. Blanton, C. Personal Communication. 2014.

4. Borody, T. J., and A. Khoruts. (2012) Fecal microbiota transplantation and emerging applications. *Nat Rev Gastroenterol Hepatol*. 9:88-96.
5. Bäckhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., et al. (2004) The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci.* 101:15718-15723.
6. Cammarota, G., Ianiro, G., Bibbò, S., and A. Gasbarrini. (2014) Gut microbiota modulation: probiotics, antibiotics or fecal microbiota transplantation? *Internal and Emergency Medicine*. 9:365-373.
7. Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., and R. Knight (2010). Global Patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS*. 108:4516-4522.
8. Chen, J., He, X., and J. Huang. (2014) Diet effects in gut microbiome and obesity. *Journal of Food and Science*. Doi: 10.1111/1750-3841.12397
9. Clement, B. G., Kehl, L. E., Debord, K. L., and C. L. Kitts. (1998) Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities. *Journal of Microbiological Methods*. 31:135-142.
10. Costello, E.K., et al. (2009) Bacterial community variation in human body habitats across space and time. *Science*. 326:1694-1697.
11. Cotter, P. D. (2011) Small intestine microbiota. *Curr Opin Gastroenterol*. 27:99-105 .

12. Culman, S. W., Bukowski, R., Gauch, H. G., Cadillo-Quiroz, H., and D. H. Buckley. 2009. T-REX: Software for the Processing and Analysis of T-RFLP data. *BMC Bioinformatics*. 10:171.
13. Culman, S. W., Gauch, H. G., Blackwood, C. B., and J. E. Thies. (2008) Analysis of T-RFLP data using analysis of variance and ordination methods: A comparative study. *Journal of Microbiological Methods*. 75:55-63.
14. Cunningham-Rundles, S., Ahrné, S., Johann-Lang, R., Abuav, R., Dunn-Navarra, A., Grassey, C., Bengmark, S., and J. S. Cervia. (2011) Effect of Probiotic Bacteria on Microbial Host Defense, Growth, and Immune Function in Immunodeficiency Virus Type-1 Infection. *Nutrients*. 3:1042-1070.
15. Curtis, T. P., and W. T. Sloan. (2004) Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. Ecology and Industrial Microbiology. 7:221-226.
16. de Wit, N., Derrien, M., Bosch-Vermeulen, H., Oosterink, E., Keshtkar, S., et al. (2012) Saturated fat stimulates obesity and hepatic steatosis and affects gut microbiota composition by an enhanced overflow of dietary fat to the distal intestine. *Am J Physiol Gastrointest Liver Physiol*. 303:589-599.
17. de Wit, N.de Wit, N. J. W., Bosch-Vermeulen, H., de Groot, P. J., Hooiveld, G. J. E. J., Grootte Bromhaar, M. M., Jansen, J., Müller, M., and R. van der Meer. (2008) The role of the small intestine in the development of dietary fat-induced obesity and insulin resistance in C57BL/6J mice. *BMC Medical Genomics*.1:14. Doi:10.1186/1755-8794-1-14

18. Dethlefsen, L., Huse, S., Sogin, M. L., and D. A. Relman. (2008) The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biol*. 6(**11**): 2383-2400.
19. Dethlefsen, L., McFall-Ngai, M., and D. A. Relman. (2007) an ecological and evolutionary perspective on human-microbe mutalism and disease. *Nature*. 449:811-818.
20. Di Bella. (2013) Fecal microbiota transplantation: the state of the art. *Infectious Disease Reports*. 13(**5**):43-45.
21. Druart, C., Neyrinck, A. M, Dewulf, E. M., De Backer, F. C., Possemiers, S., et al. (2013) Implication of fermentable carbohydrates targeting the gut microbiota on conjugated linoleic acid production in high-fat-fed mice. *Br J Nutr*: 1-14.
22. Druart, C., Neyrinck, A. M., Vlaeminck, B., Fievez, V., Cani, P. D., Delzenne, N.M. (2014) Role of the Lower and Upper Intestine in the Production and Absorption of Gut Microbiota-Derived PUFA Metabolites. *PLoS ONE*. 9(**1**): e87560. Doi: 10.1371/journal.pone.0087560
23. Dunbar, J., Ticknor, L. O., and C. R. Kuske. (2001) Phylogenetic Specificity and Reproducibility and New Method for Analysis of Terminal Restriction Fragment Profiles of 16S rRNA. *Applied and Environmental Microbiology.* 67(**1**):190-197.
24. Fritz, j. V., Desai, m. S, Shah, P., Schneider, J. G., and P. Wilmes. (2013) From meta-omics to causality: experimental models for human microbiome research. *Microbiome*. 1:14.

25. Langille, G. I., M., Meehan, C. J., Koenig, J. E., Dhanani, A. S., Rose, R. A., Howlett, S. E., and R. G. Beiko. (2014) Microbial shifts in the aging mouse gut. *Microbiome*. 2:50.

26. Hayashi, H., Takahashi, R., Nishi, T., Sakamoto, M., and Y. Benno. (2005) Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *J Med Microbiol*. 54:1093-1101.

27. Hilton, A., and R. A. Armstrong. (2006). Statnote 6: one-way analysis of variance (ANOVA). *Microbiologist*. 33-37.

28. Hooper, L. V. (2004) Bacterial contributions to mammalian gut development. *Trends Microbiol*. 12:129-134.

29. Khoruts, A., Dicksved, J., Jansson, J. K., and M. J. Sadowsky. (2010) Changes in the composition of the human fecal microbiome after bacteriotherapy for recurrent Clostridium difficile-associated diarrhea. *J Clin Gastroenterol.* 44:354-360.

30. Klemashevich, C., Wu, C., Howsmon, D., Howsmon, D., Alaniz, R. C., Lee, K., and A. Jayaraman. (2014). Rational identification of diet-derived postbiotics for improving intestinal microbiota function. *Current Opinion in Biotechnology*. 26:85-90.

31. Konstantinidis, K. T., and J. M. Tiedje. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA.* 102:2567-2572.

32. Kostic, A. D., Howitt, M. R., and W. S. Garrett. (2013) Exploring host-microbiota interactions in animal models and humans. *Genes Dev.* 27:701-718

33. Lazarevic, V., et al. (2009) Metagenomic study of the oral microbiota by illumine high-throughput sequencing. *J Microbiol Methods.* 79:266-271.

34. Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., et al. (2005) Obesity alters gut microbial ecology. *Proc Natl Acad Sci USA.* 102:11070-11075.

35. Ley, R. E., Peterson, D. A., and J. I. Gordon. (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*. 124:837-848.

36. Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., et al. (2008) Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci USA*. 105:2117-2122.

37. Liu, W-T., Marsh, T.L., Cheng, H., and L. J. Forney. (1997) Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Genes Encoding 16S rRNA. *Appl Environ Microbiol.* 63(**11**):4516-4522.

38. Magrone, T., and E. Jirillo. (2013) The interaction between gut microbiota and age-related changes in immune function and inflammation. *Immunity & Ageing*. 10:31.

39. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O., and D. L. Kasper. (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*. 122:107-118.

40. McGarr, S. E., Ridlon, J. M., and P. B. Hylemon. (2005) Diet, anaerobic bacterial metabolism, and colon cancer: a review of the literature. *J Clin Gastroenterol.* 39:98-109.

41. Mujico, J. R., Baccan, G. C., Gheorghe, A., Diaz, L. E., and A. Marcos. (2013) Changes in gut microbiota due to supplemented fatty acids in diet-induced obese mice. *Br J Nutr*: 1-10.

42. Nicholson, J. K., Holmes, E., and I. D. Wilson. (2005) Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol.* 3:431-438.

43. Niot, I., Poirier, H., Tran, T. T., and P. Besnard. (2009) Intestinal absorption of long-chain fatty acids: evidence and uncertainties. *Prog Lipid Res*. 48:101-115

44. Noverr, M. C., and G. B. Huffnagle. (2005) The 'microflora hypothesis' of allergic responses. *Clin Exp Allergy.* 35:1511-1520.

45. O'Hara, A. M., and F. Shanahan. (2006) The gut flora as a forgotten organ. *EMBO Rep*. 7:688-693.

46. O'Shea, E. F., Cotter, P. D., Stanton, C., Ross, R. P., and C. Hill. (2012) Production of bioactive substances by intestinal bacteria as a basis for explaining probiotic mechanisms: bacteriocins and conjugated linoleic acid. *Int J Food Microbiol*. 152: 189-205.

47. Osborn, A.M., Moore, E.R.B., and K. N. Timmis. (2000) An evaluation of terminal restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ Microbiol.* 2(**1**):39-50.

48. Raju, B. M. K. (2002) A study on AMMI model and its biplots. *Journal of the Indian Society of Agricultural Statistics*. 55(**3**):297-322.

49. Rankoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and R. Medzhitov. (2004) Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell*. 118:229-241

50. Rees, G. N., Baldwin, D. S., Watson, G. O., Perryman, S., and D. L. Nielsen. (2004) Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: application in multivariate statistics. *Antonie van Leeuwenhoek*. 86:339-347

51. Sadeghi, S. M., Samizadeh, H. Amiri, E., and M. Ashouri. (2011) Additive main effects and multiplicative interactions (AMMI) analysis of dry leaf yield in tobacco hybrids across environments. *African Journal of Biotechnology*. 10(**21**):4358-4364.

52. Saulnier, D. M., Ringel, Y., Heyman, M. B., Foster, J. A., Bercik, P., Shulman, R. J., Versalovic, J., Verdu, E. F., Dinan, T. G., Hecht, G., and F. Guarner. (2013) The intestinal microbiome, probiotics and prebiotics in neurogastroenterology. *Gut Microbes*. 1(**4**):17-27.

53. Saulnier, D. M., Molenaar, D., de Vos, W. M., Gibson, G. R., and S. Kolida. (2007). Identification of prebiotic fructooligosaccharide metabolism in Lactobacillus plantarum WCFS1 through microarrays. *Applied and environmental microbiology*: 73(**6**), 1753-1765.

54. Sartor, R. B. (2008) Microbial influences in inflammatory bowel diseases. Gastroenterlogy. 134:577-594.

55. Schütte, U. M. E., Abdo, Z., Bent, S. J., Shyu, C., Williams, C. J., Pierson, J. D., and L. J. Forney. (2008) Advances in the use of terminal restriction fragment

length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Applied Microbiology and Biotechnology*. 80:365-380.

56. Stres, B. (2006) The first decade of terminal restriction fragment length polymorphism (T-RFLP) in microbial ecology. *Acta agriculturae Slovenica*. 88:65-73
57. Walsh, C. J., Guinane, C. M., O'Toole, P. W., and P. D. Cotter. (2004) Beneficial modulation of the gut microbiota. *FEBS lett*. 588**(22)**:4120-4130.
58. Zhang, X., Zeng, B., Liu, Z., Liao, Z., Li, W., Wei, H., and X. Fang. (2014) Comparative diversity of gut microbiota in two different human flora-associated Mouse strains. *Current Microbiology*. 69:365-373.
59. Zoetendal, E. G., Raes, J., van den Bogert, B., Arumagam, M., Booijink, C. C., et al. (2012) The human small intestinal microbiota is driven by rapid intake and conversion of simple carbohydrates. *ISME J.* 6:1415-1426.

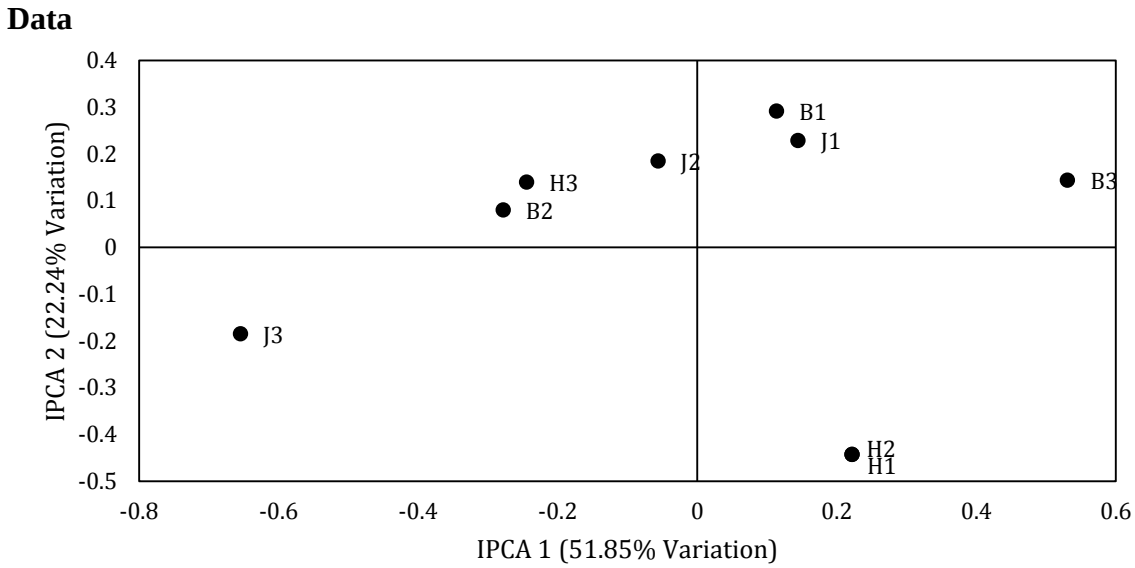

Figure 1. 16S bacterial amplicon T-RFLP analysis from the molecular-based microbial community fingerprinting technique, T-REX (T-RFLP analysis Expedited), on the forward 5' HEX fragment of triplicate mouse cecum samples harvested from the cecum content of control (B, H, J represented by closed circles) diet fed mice. The letter for each data point represents the sample and the following the number the replicate number. Stacked point designations represent data having the save value and coordinate.

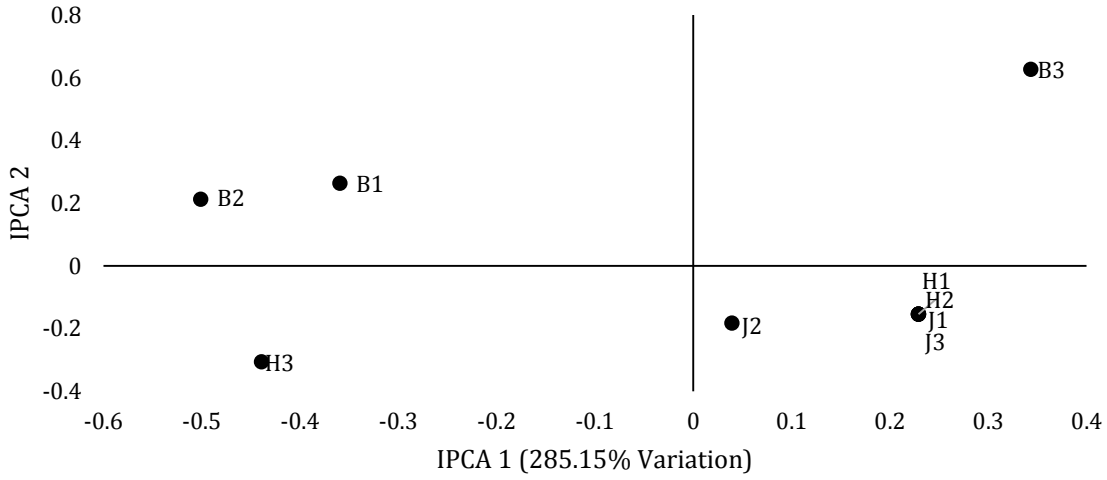

Figure 2. 16S bacterial amplicon T-RFLP analysis on triplicate mouse cecum samples harvested from the cecum content of control (B, H, J represented by closed circles) diet fed mice where the letter represents the sample and number the replicate. Data is presented by means of the AMMI model, which organizes variations and interactions amongst data by means of ANOVA and PCA. The figure above depicts the reverse 3' FAM fragments yielded by T-REX of the control mice. Stacked point designations represent data having the same value and coordinate.

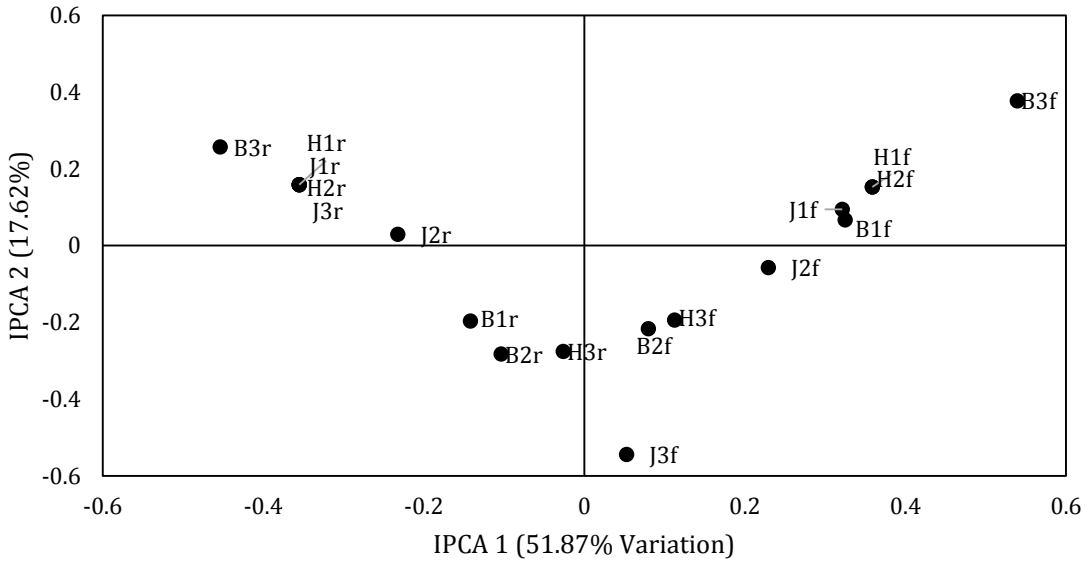

Figure 3. 16S bacterial amplicon T-RFLP analysis on triplicate mouse cecum samples harvested from the cecum content of control (B, H, J represented by closed circles) diet fed mice. Data is presented by means of the AMMI model, which organizes variations and interactions amongst data by means of ANOVA and PCA. The figure above depicts the forward 5' forward HEX and reverse 3' FAM fragments yielded by T-REX of the control mice. Each sample has triplicate forward and reverse data points. The number after the initial sample letter designation depicts the replicate number and the concluding f or r forward and reverse, respectively. Stacked point designations represent data having the same value and coordinate.

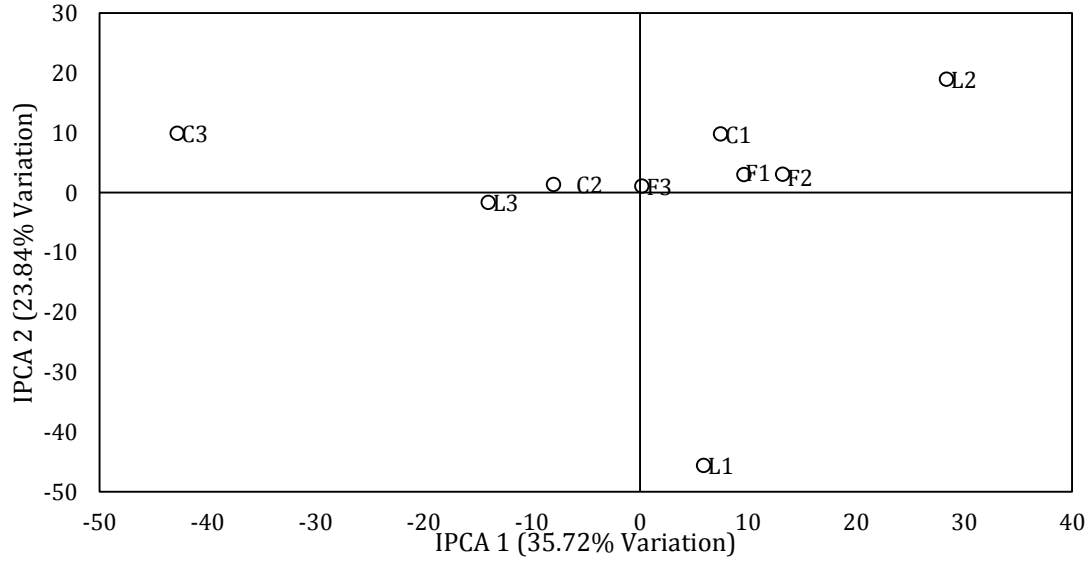

Figure 4. 16S bacterial amplicon T-RFLP analysis on triplicate mouse cecum samples harvested from the cecum content of synbiotic (C, F, L) diet fed mice where the letter represents the sample and number the replicate. Data is presented by means of the AMMI model, which organizes variations and interactions amongst data by means of ANOVA and PCA. The figure above depicts the forward 5' HEX fragments yielded by T-REX analysis of the synbiotic mice which are designated as open circles.

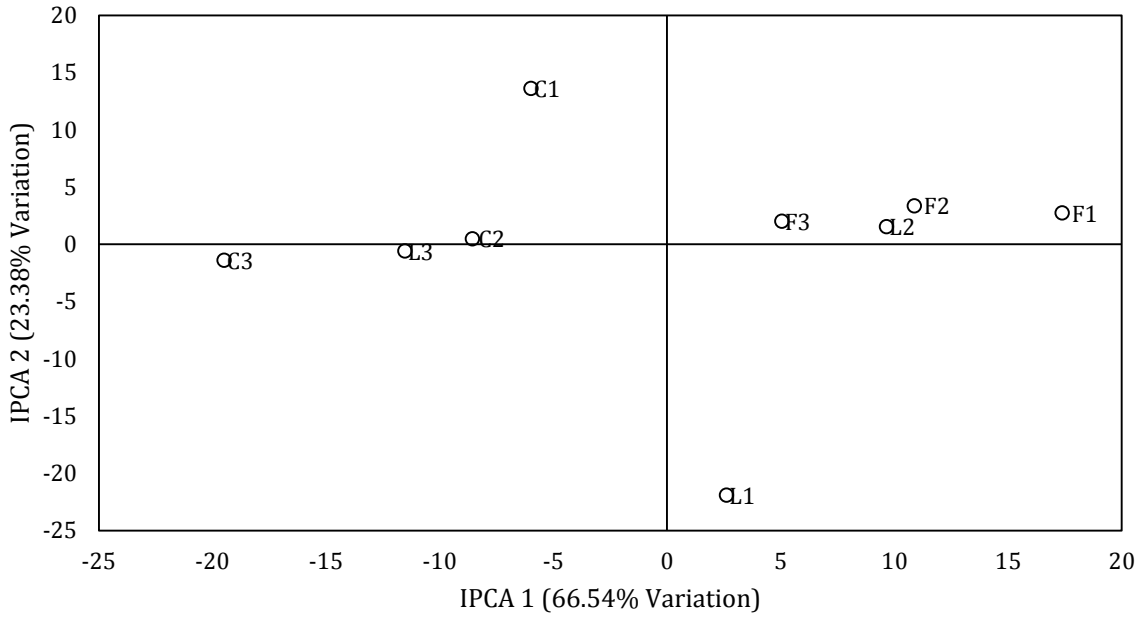

Figure 5. 16S bacterial amplicon T-RFLP analysis on triplicate mouse cecum samples harvested from the cecum content of synbiotic (C, F, L designated as open circles) diet fed mice of the reverse 3' FAM fragment where the letter of each data point corresponds to the sample and number the replicate. Data is presented by means of the AMMI model, which organizes variations and interactions amongst data by means of ANOVA and PCA.

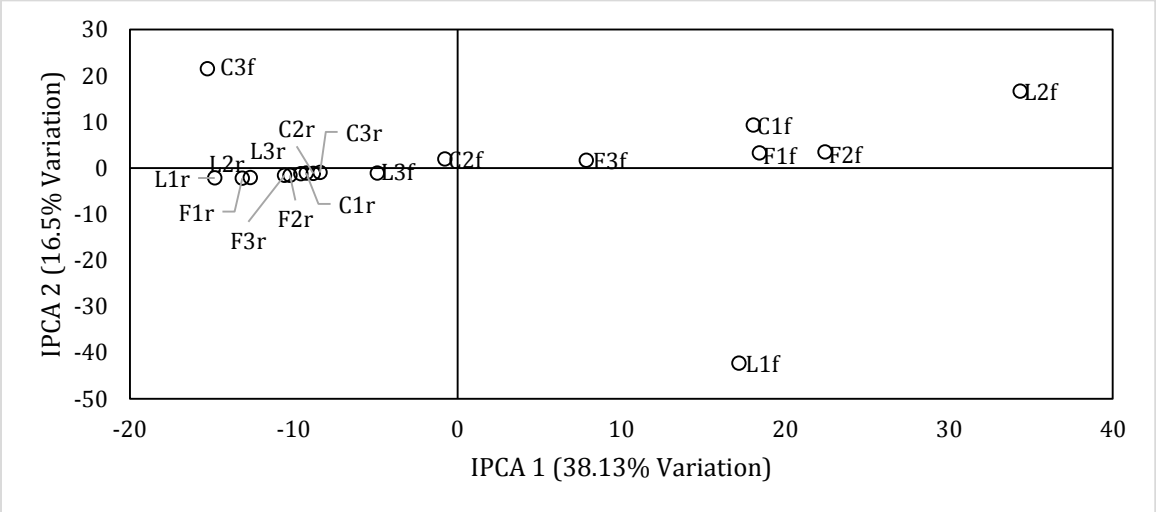

Figure 6. 16S bacterial amplicon T-RFLP analysis on triplicate mouse cecum samples harvested from the cecum content of synbiotic (C, F, L which are designated as open circles) diet fed mice. Data is presented according to AMMI analysis. The figure above depicts the forward 5' HEX and reverse 3' FAM fragments yielded by T-REX analysis of the synbiotic mice. Each sample has triplicate forward and reverse data points. The number after the initial sample letter designation depicts the replicate number and the concluding f or r forward and reverse, respectively. Stacked point designations represent data having the same value and coordinate.

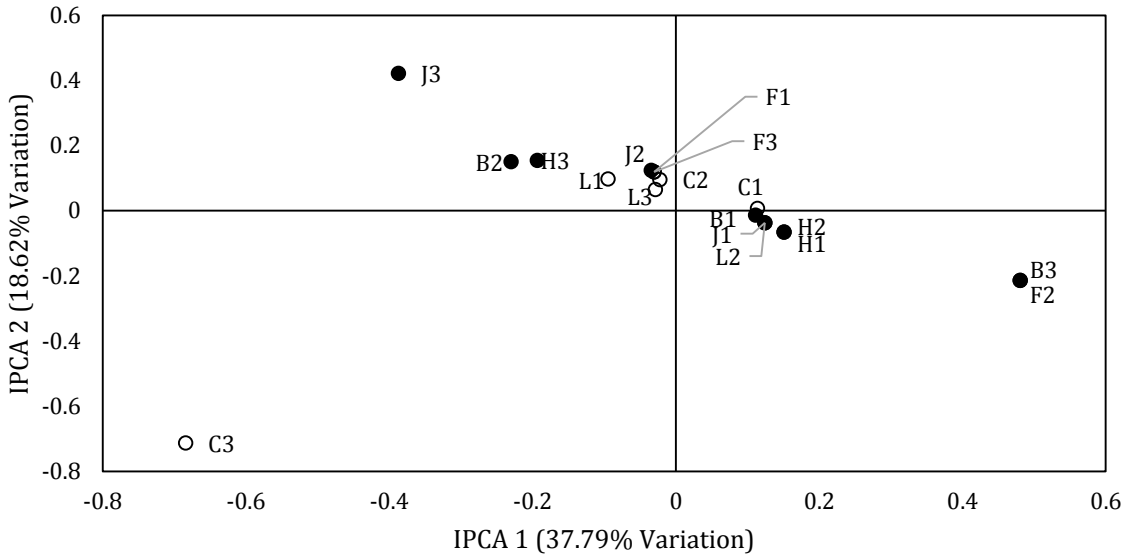

Figure 7. 16S bacterial amplicon T-RFLP analysis on triplicate mouse cecum samples harvested from the cecum content of control (B, H, J designated by closed circles) and synbiotic (C, F, L represented by open circles) diet fed mice. Data is presented by means of the AMMI model, which organizes variations and interactions amongst data by means of ANOVA and PCA. This is a graphic representation of the 5' forward HEX yielded by T-REX of the control and synbiotic mice. Stacked point designations represent data having the same value and coordinate. The letter in the label of each data point corresponds to the sample and the number the replicate.

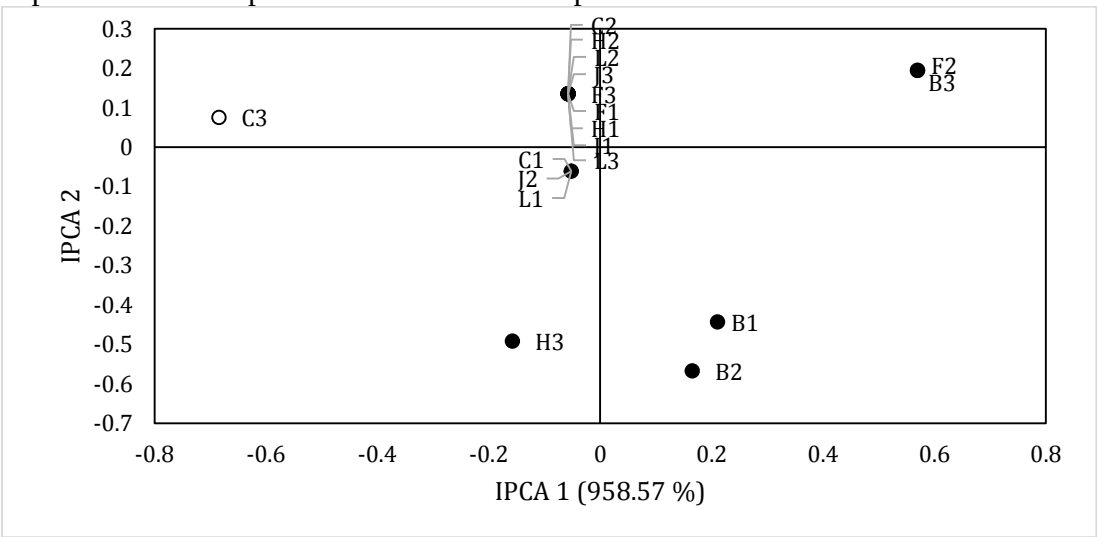

Figure 8. 16S bacterial amplicon T-RFLP analysis on triplicate mouse cecum samples harvested from the cecum content of control (B, H, J represented by a closed circle) and synbiotic (C, F, L designated as an open circle) diet fed mice. Data is presented by means of the AMMI model, which organizes variations and interactions amongst data by means of ANOVA and PCA. The figure above depicts the reverse 3' FAM fragments yielded by T-REX of the control and synbiotic mice. Stacked point designations represent data having the same value and coordinate. Data labels correspond to the sample (letter) and replicate (number).

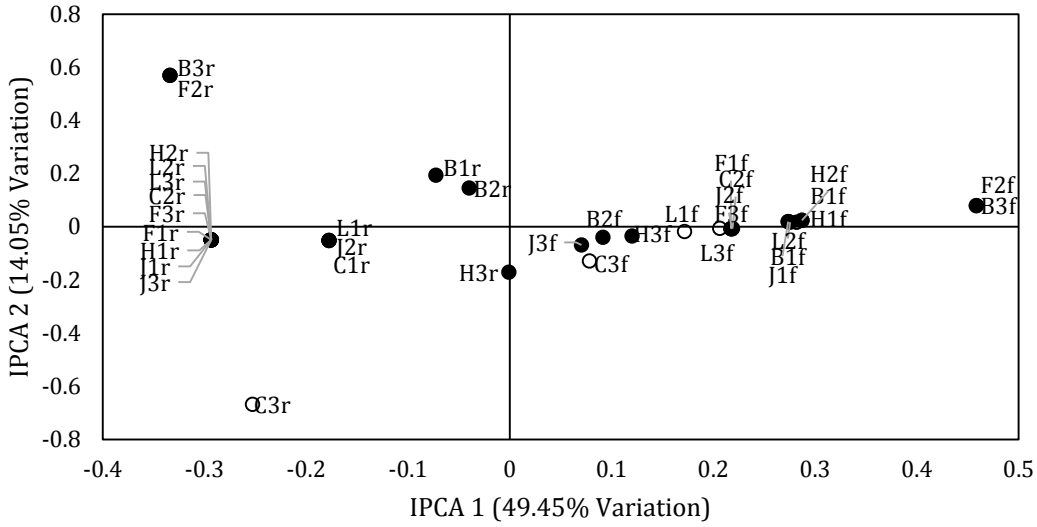

Figure 9. 16S bacterial amplicon T-RFLP analysis on triplicate mouse cecum samples harvested from the cecum content of control (B, H, J represented by a closed circle) and synbiotic (C, F, L designated as an open circle) diet fed mice. Data is presented by means of the AMMI model, which organizes variations and interactions amongst data by means of ANOVA and PCA. The figure above depicts the 5' forward HEX and reverse 3' FAM fragments yielded by T-REX of the control and synbiotic mice. Each sample has triplicate forward and reverse data points. The number after the initial sample letter designation depicts the replicate number, and the concluding f or r forward and reverse, respectively. Stacked point designations represent data having the same value and coordinate.

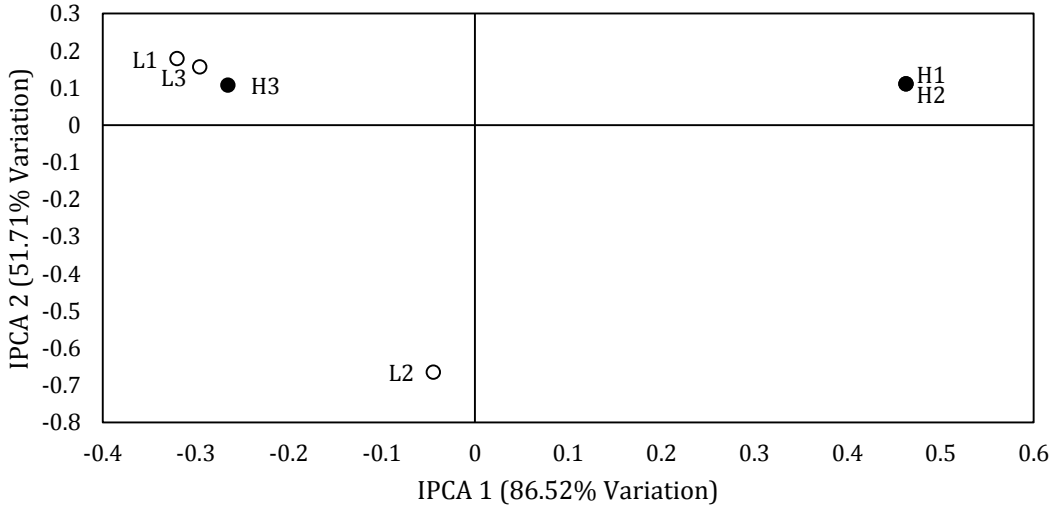

Figure 10. 16S bacterial amplicon T-RFLP analysis from the molecular-based microbial community fingerprinting technique, T-REX (T-RFLP analysis Expedited), on mouse cecum samples harvested from the cecum content of control (H, designated by a closed circle) and synbiotic (L, represented by an open circle) diet fed mice. Data is presented by means of the AMMI model. The figure above depicts the two

dimensional interactions and variations produced by analyzing the forward 5' HEX fragment of the "best" representative control (H) and synbiotic (L) diet as determined by Peak Scanner absorbance plots of TRFLP data. Stacked point designations represent data having the same value and coordinate.

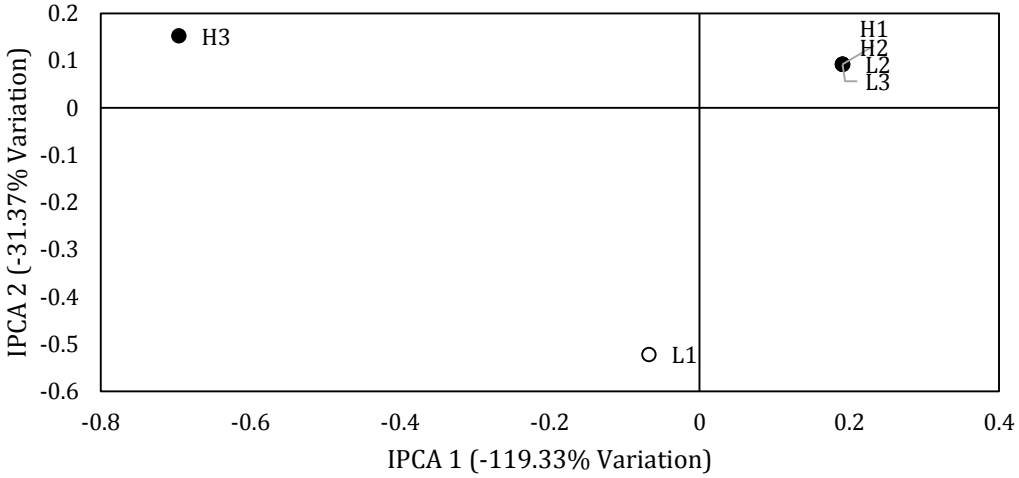

Figure 11. 16S bacterial amplicon T-RFLP analysis on triplicate mouse cecum samples harvested from the cecum content of control (H) and synbiotic (L) diet fed mice. Data is presented by means of the AMMI model, which organizes variations and interactions amongst data by means of ANOVA and PCA. The figure above depicts the reverse 3' FAM fragments yielded by T-REX of the control (H) and synbiotic (L) mice. Each sample has triplicate forward and reverse data points. The number after the initial sample letter designation depicts the replicate number. Stacked point designations represent data having the same value and coordinate.

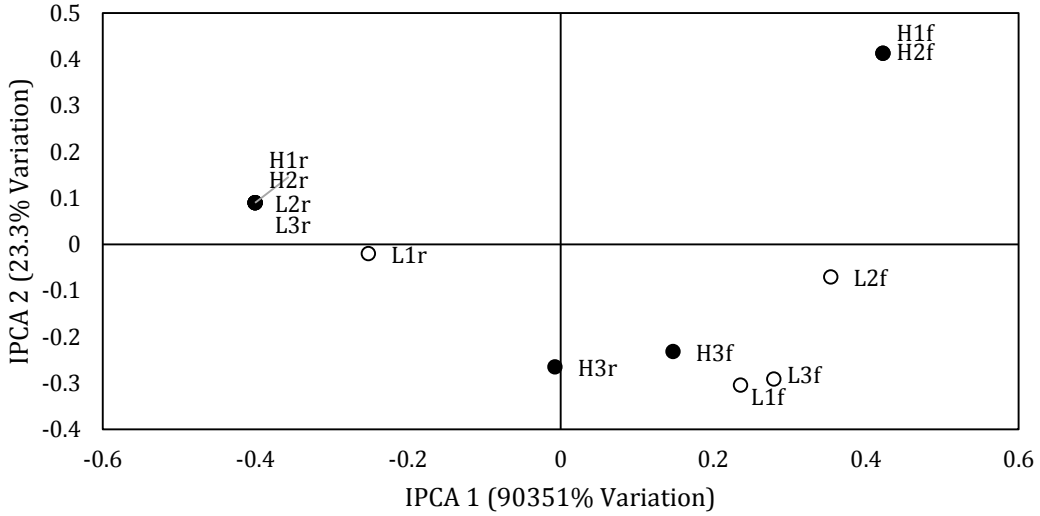

Figure 12. 16S bacterial amplicon T-RFLP analysis on triplicate mouse cecum samples harvested from the cecum content of control (H, closed circle) and synbiotic (L, open circle) diet fed mice. Data is presented by means of the AMMI model, which organizes variations and interactions amongst data by means of ANOVA and PCA.

The figure above depicts the 5' forward HEX and reverse 3' FAM fragments yielded by T-REX of the most representative sample of each the control and synbiotic mice. Each sample has triplicate forward and reverse data points. The number after the initial sample letter designation depicts the replicate number and the concluding f or r forward and reverse, respectively. Stacked point designations represent data having the same value and coordinate.

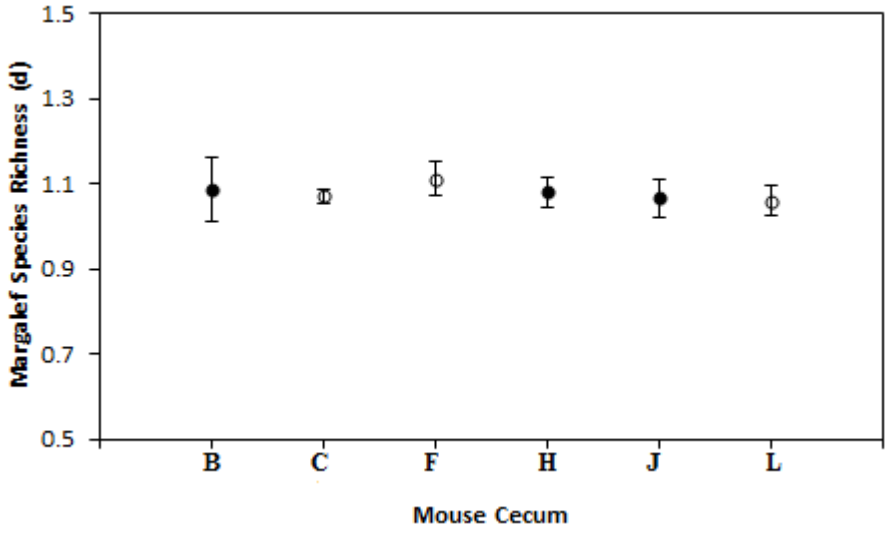

Figure 13. Diversity indices representing the average Margalef species richness (d) values produced by Primer 6 software during terminal restriction fragment length polymorphism analysis comparing cecum content of control (B, H, J designated as closed circles) and synbiotic (C, F, L represented as open circles) diet fed mice. This diversity index measures species abundance and frequency.

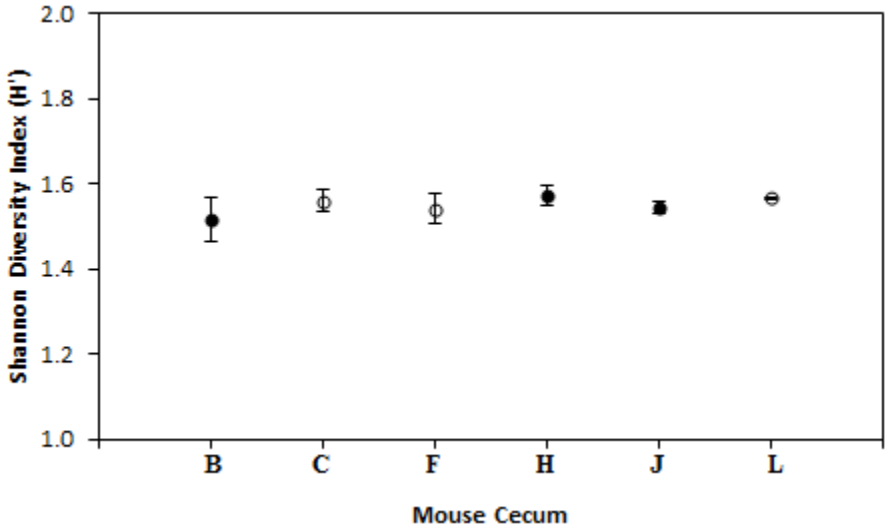

Figure 14. Average Shannon diversity indices (H') values of sample replicates produced by Primer 6 software during T-RFLP analysis of mouse cecum tissue samples harvested from mice fed either a controlled (samples B, H, J designated as closed circles) or synbiotic diet (samples C, F, L represented as open circles). Error bars were generated from obtaining standard deviations of the averaged replicates.

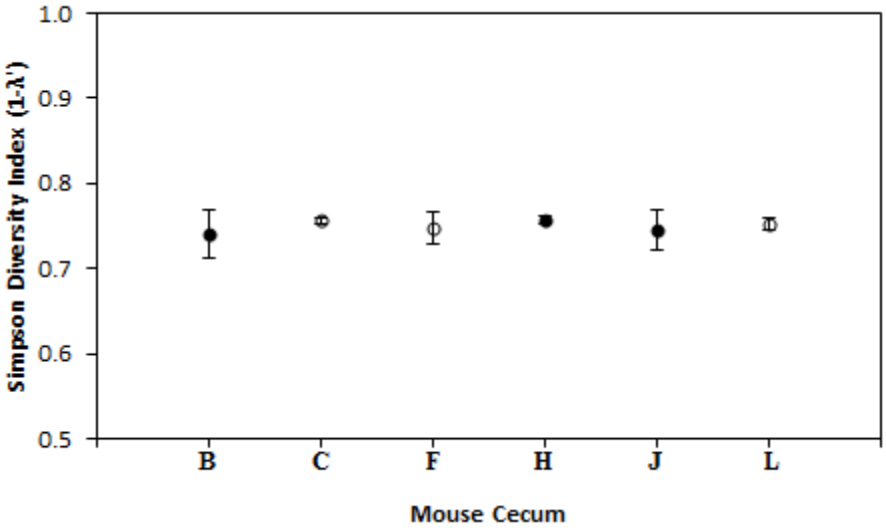

Figure 15. Primer 6 generated average values depicting the probability of two randomly selected samples belonging to the same species, or the Simpson Index (1-ʎ'), from T-RFLP data of mouse cecum tissue samples collected from mice fed either a controlled (samples B, H, J designated as closed circles) or synbiotic diet (samples C, F, L represented as open circles).

CHAPTER 3

PAIRED END PHYLOGENETIC ANALYSIS OF BACTERIAL POPULATION DIFFERENCES IN *MUS MUSCULUS* CECUM IN RESPONSE TO PROBIOTIC ADMINISTRATION

Leasha M. Christensen,* Cynthia Blanton, and Peter P. Sheridan

Department of Biological Sciences, Idaho State University

***Corresponding Author: Leasha M. Christensen**

Email Addresses:

LMC: chrileas@isu.edu, CB: blancynt@isu.edu, PPS: sherpete@isu.edu

ABSTRACT: The cecum has the highest metabolite absorption and houses the most abundant population of microbes which are responsible for maintaining homeostasis and host health. Disruption of the microbial populations in either presence or relative abundance due to various environmental and host factors can result in various disease state that do not have a standard form of medical treatment. Supplementation with probiotics, prebiotics, and synbiotics has shown promise as a therapeutic intervention to combat dysbiosis. Many gut microbiota cannot be isolated by culture-dependent techniques. Culture independent, high resolution paired end16S rRNA genomic analysis of the V3/V4 hypervariable region shows the establishment of probiotic strains maintaining diversity within the cecum under the additional supplementation of prebiotics to meet conditions for therapeutic use in combating dysbiosis and maintaining host health and homeostatic balance as is mitigated by the gut microbial flora.

KEYWORDS: paired end, prebiotic, probiotic, synbiotic, diversity, 16S rRNA, next-gen

INTRODUCTION

Microbial communities abound in a range of environments including that of the human host which contains upwards of 100 trillion microbial cells, outnumbering human eukaryotic cells tenfold. The most densely populated environment is located in the lower intestine of the gastrointestinal tract, specifically the cecum, where the human gut microbiota and its genome is known collectively as the ‘gut microbiome.’ A variety of exogenous metabolic and genetic features bestowed from microbe to host are the result of a co-evolved, synbiotic relationship between these two entities (7, 18, 21, 38, 63). For example, the gut microbiome is largely responsible for the mediation of host health through means of moderating host homeostasis, development and regulation of the

immune response, contributing to pathogen colonization resistance, and actively participating in nutrition and metabolism (27, 39, 42, 51). Epithelial development, regulation of energy homeostasis, blood circulation, and adaptive and innate immunity mechanisms are also mitigated by the host gut microbiome. Disruptions in the normal activity of host interactions of this microbial community including adaptations to the various aspects of the 'modern lifestyle' including travel, diet, age, geographic locale, stress, and use of medications can therefore results in a diseased state (1, 7, 13, 15, 17, 20, 24-26, 31, 45, 47, 49, 50, 53-55, 63, 65, 66).

Due to the nature of these interactions, pharmaceutical interventions to further moderate host health that inadvertently affect the microbial community composition of the gut microbiota can have detrimental effects. Lack of treatment can result in the development of many illnesses both acute and chronic with varying prognoses including the following: asthma, pseudomembranous colitis, functional diarrhea, as well as additional implications such as cancer, obesity and other eating disorders, cardiac events, inflammatory bowel disease, chronic low-grade inflammation, inflammatory bowel syndrome, Crohn's disease, ulcerative colitis, type 2 diabetes, asthma, hay fever, skin allergies, antibiotic-associated diarrhea, pseudomembranous colitis, toxic megacolon, atopic diseases relating to the change in microbial community composition, and neurodevelopmental disorders including autism spectrum disorders (Asperger disorder, childhood disintegrative disorder, Rett disorder, pervasive developmental disorder). All of these disease states are mitigated by chemical transformations within the gut as a result of the disturbance of the microbial norm (1, 6, 7, 9, 13, 15, 20, 25, 26, 31, 37, 41, 43, 45-50, 53-55, 63, 65, 66).

A number of interventions have been integrated into the management of the dysbiotic state in the gut microbiome as there is currently no standard or agreed-upon method for treatment due to the exhibited broad range of effectiveness. Such methodologies include antibiotic administration, bacteriotherapy, and fecal-oral transplantation (7, 17, 19, 30, 53, 63, 65). Antibiotics are generally administered due to this lack of effective treatment and are often unsuccessful, resulting in recurrent infections and altering the microbial community composition at a rate of roughly 90% (19, 31). Promising, yet highly invasive and requiring an extensive screening process, fecal-oral transplantations are also employed but with varying success rates (5, 7, 19, 30). Utilization of the emerging pre-, pro-, post-, and synbiotics has shown increasing potential for correcting the dysbiotic state and maintaining host health (1, 7, 13-15, 20, 25, 26, 31, 41, 45, 47, 49, 50, 53-55, 63, 65, 66).

Probiotics are classified as "*organisms and substances which contribute to intestinal microbial balance"* or *"a live microbial feed supplement which beneficially affects the host animal by improving its intestinal microbial balance"* and must meet specific guidelines by remaining stable and viable under storage and use, be prepared on a large scale, and benefit host health after introduction in addition to surviving and remaining present in the integrated ecosystem after the product used for initial introduction has been depleted (7, 31, 54). Commonly used probiotics that show effectiveness in various pathologies include *Lactobacilli* (*Lactobacillus acidophilus, L. casei, L. delbruekii) Bifidobacteria* (*Bifidobacterium adolescentis, B. bifidum, B. longum, B. infatis),* and *Streptococci* (*Streptococcus salivariius ss. Thermophiles, S. lactis)* (31, 41). However, *L. acidophilus* is the most prevalently utilized and tolerated probiotic, and

it has been shown to synthesize vitamin K which is necessary for the conversion of the bone matrix osteocalcin to the active form and may thus aid in improving bone integrity (1, 7, 13, 15, 20, 25, 26, 31, 41, 45, 47, 49, 50, 53-55, 65, 66).

Prebiotics are "a *non-digestible food ingredient that beneficially affects the host by selectively stimulating the growth and/or activity of one or a limited number of bacteria in the colon, and thus improves host health* (7)*."* Like probiotics, additional criteria must be satisfied. A prebiotic cannot be absorbed or hydrolyzed in the upper GI tract, and it must serve selectively as a substrate for specific commensal, beneficial bacteria for growth or metabolic activation. It also must alter the microbial composition to that of a healthy state, and cause systemic or luminal effects that benefit host health (7, 31, 54). Compounds meeting these specifications include non-digestible foods such as oligosaccharides, polysaccharides, fructooligosaccharides, and other naturally occurring non-digestible carbohydrates, peptides, proteins, and lipids. Host health is improved through the stimulation of growth and activity of specific endogenous microbiota by altering the microbial composition in a given locale due to supplementation of nutrients and metabolites (31).

The term 'synbiotics' includes the combined use pro- and prebiotics and is defined as "*a mixture of probiotics and prebiotics that beneficially affects the host by improving the survival and implantation of live microbial dietary supplements in the gastrointestinal tract, by selectively stimulating the growth and/or by activating the metabolism of one or a limited number of health-promoting bacteria, and thus improving host welfare* (7)*."* Synbiotics have shown to combat the diseased dysbiosis state via supplementation with functional and health enhancing nutrition by maintenance of the

colonic flora in the healthy, balanced state. The combined usage of pre- and probiotics has shown to increase the effectiveness of each individual component (7, 31, 41, 54).

The current lack of information concerning the phylogenetic diversity and composition of the gut microbial flora both under a state of health and disease has introduced various challenges for the use of pre-, pro-, and synbiotics in treating dysbiosis. The exact identity of species shifts between the two states of host health has remained elusive despite the limited knowledge of phyla level changes which impedes understanding of microbial community interactions during treatments as spatial distribution within the GI tract is heterogeneous due to environmental differences, and different pre- and probiotics may have varying activities and cooperative interactions with other bacteria at different locations. Further research is needed to investigate and characterize the intestinal communities of microbes to increase the efficiency and effectiveness of synbiotic treatments (31, 41).

Due to the uncultivable nature of the gut microflora, the lack of accurate representation of diversity and inherent biases and limitations present in culture-dependent methods of analysis, and the exponential amount of species being analyzed (on the order of 10^{11}-10^{14} bacteria per gram of intestinal content, more than 1.5kg total weight, and representing over 1100 species), a high-throughput means of DNA analysis was necessary (10, 21, 52). The Human Microbiome Project and MetaHIT Consortium have both utilized emerging culture-independent technologies for the characterization of microbes and their genomes within the human body to assess its composition on various degrees of host health and disease states (7, 21, 31, 41, 63). Mouse models have been employed to decrease biases and confounding variables while providing representation to

the human microbiome (7, 25, 33, 63). Diversity studies are crucial for the development of pharmaceuticals and therapeutic interventions to alleviate a diseased host state (15). Being that the cecum is the source of highest metabolite absorption, analysis of its content is beneficial to determine the effectiveness of synbiotics (1, 7, 13, 16, 20, 21, 26, 34, 45, 47, 49, 50, 53, 55, 63, 65, 66).

Analysis of the gut microbiome is facilitated through the use of current, improved technologies, sequence analysis, and software packages by providing increased statistical precision, resolution between samples, and accurate representation of biological patterns and applications in microbial community dynamics on a scale previously unobtainable (12, 35). In addition to community differentiation, relative structure, phylotype richness, organism identification, and identification of spatial and temporal shifts, species analysis of diversity can be based on nucleotide sequence identity in mitochondrial DNA within a species while achieving specificity and sensitivity with increased resolution and while avoiding the limitations inherent to cultivation-dependent techniques (29). The 16S rRNA amplicons commonly utilized in community profiles and diversity can aid in the detection of rare phylotypes, ecological characteristics, and taxonomic identification due to it being present in all organisms, being large enough for informatics purposes (at roughly 1500bp) while containing both conserved (the most highly conserved structural element in rRNA) and hypervariable loop regions, not being transferred horizontally, and being considered universal- thus greatly increased the rise of complex, novel microbial consortia (8, 23, 32, 35). Microbial community structure comparisons of hundreds of samples can be accomplished via 16S rRNA high-throughput gene sequencing. The Illumina platform produces up to 1.5 billion reads per run in comparison to the 1 million

reads per run on previously primarily used 454 Pyrosequencing platform at a comparable cost. This platform is also 50-12000 fold less expensive per sequenced megabase than pyro and Sanger sequencing (2, 34, 64).

The V3/V4 loop region of the bacterial 16S ribosome can be analyzed through utilization of paired-end sequencing and annotation (36). The paired-end analysis technique is advantageous to other high-throughput sequencing techniques in that it reduces the amount of erroneous sequences that are included in downstream analysis (imposing a quality control step) while providing enormous data sets. The length at two fold coverage has increased quality of generated libraries. Paired-end analysis is completed when sequencing both the 5' and 3' ends, doubling the number of base pair reads on the Illumina platform. Both V3 and V4 regions can be utilized to identify OTUs (thus reducing the observed sequencing error rate from 1.08 to 0.01%) while constructing phylogenetic assignments in tree form. The overlapping present in these regions also allows for a reduction in generated error (34, 64). The V3 hypervariable region consists of roughly 200 nucleotides and has high taxonomic resolution and a conserved flanking region. It is also compatible with paired-end sequencing which allows for overlapping 3' end sequencing, reduction of sequencing errors, and generation data applicable to pipelines of computational analysis that are publically available (2). The MiSeq Illumina platform also provides a cost efficient means of paired-end analysis comparative to its HiSeq counterpart (23, 34). In addition, utilizing the V3/V4 and a 7bp heterogeneity spacer reduces the lower diversity measurements that were previously plaguing MiSeq data. There are 9 hypervariable regions flanked by conserved sequence regions in the 16S rRNA gene. A 469bp region is required to contain both the V3 and V4 regions while

maximizing the length of the generated MiSeq reads. This region provides sufficient data for microbial community taxonomic classification of specimens associated with the human microbiome (as is employed in the human microbiome project). Strict filtering can alleviate error rates that tend to increase at the ends of reads and thus improve the accuracy of taxonomic assignment while avoiding spurious read assembly (60).

Advanced tools for analysis facilitate the detection of integral components of host homeostatic balance through mediation and integration of pre, pro, and synbiotics. Further research, specifically metabolomics, is required to identify the metabolic capabilities and activities of influential gut microbiota for sustaining human host health.

MATERIALS AND METHODS

Sample Preparation:

Harvested cecum from male 9 month old mice (Harlan), *Mus musculus* following either a controlled (B, C, F) or synthetic (H, J, L) diet were chosen at random. The diets, based on a powdered form of American Institute of Nutrition (AIN)-93M purified rat diet (Dyets, Inc., Bethlehem, PA), were administered to the mouse model and were modified to utilize cornstarch in place of sucrose and dextrin in order to reduce the susceptibility of osmotic dehydration of the bacteria studied within the synthetic diet. These isocaloric diets were developed according to carbohydrate ingredient manipulation by assuming energy densities of 4, 0, and 2 kcal/g for cornstarch, cellulose, and fructooligosaccharide, respectively. Fructooligosaccharide and lyophilized probiotic cultures (Nutraceutix, Redmond, WA) were composed of $1x10^{11}$ CFU/g of equal parts *Lactobacillus acidophilus* and *Lactococcus lactis lactis*. The diets were made fresh three times a week with addition of probiotics immediately prior to feeding each morning for 18 weeks (4).

The triplicate diet-specific cecum samples were sterilely dissected both laterally and vertically and rehydrated using 6mL 10mM TRIS,pH~8.0, 1% Triton after being stored at -20°C. Samples were then incubated at 80°C for one hour, and centrifuged (1000rpm) for five minutes to achieve a pellet. Residual supernatant was removed after centrifugation at 16,000xg for 10 minutes. The cell pellet obtained was resuspended in 250uL 10mM TRIS.

DNA Isolation:

Genomic DNA of the cecum samples was obtained by a mechanical sheer forces protocol utilizing 100µL Lysozyme (100mg/mL) added to the rehydrated cell suspension solution and incubated for 30 minutes at 37ºC. After which, 100µL Proteinase K (10mg/mL in 10mM TRIS) was added, incubated for 30 minutes, and brought up to volume with Herwigs Lysis buffer in a bead-beating tube. Prior to supernatant being drawn off and added to isopropanol, tubes were then bead-beated for 5 minutes, boiled at 80ºC for 10 min, microcentrifuged at 3000xG for 5 minutes, and stored at -20ºC for 12 hours. Samples were then microcentrifuged at 13000xG for 10 min, and the pellet washed with 200µL 70% ethanol (4ºC) prior to centrifugation at 13000xG for 10 minutes. Ethanol was removed by drying at 37ºC for 60 minutes before DNA was rehydrated in 10mM TRIS, pH ~8.0.

Paired End and PCR Amplification:

A gradient bacterial paired end PCR (200µM/L 16S paired-end bacterial designed primers [forward primer: $^{5'}$-TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG CCT ACG GGN GGC WGC AG-$^{3'}$ and reverse primer: 5'-GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA GGA CTA CHV GGG TAT CTA ATC C-3'] ,

Vent polymerase, 40-60°C annealing temperature with a four minute extension time) was run on an *Escherichia coli* K12 (+) control which yielded an optimized temperature of 41.5C for annealing. A 16S PCR was then run on the genomic DNA samples in triplicate in 50 µL reactions for each of the genomic DNA samples with each reaction containing the following: the same 16S paired-end primers (Integrated DNA Technologies), *Vent* polymerase (New England Biolabs, Ipswich, MA), 1x ThermoPol Buffer (New England Biolabs, Ipsich, MA), 400µM per each deoxynucleotide triphosphate (New England Biolabs, Ipswich, MA), and 1µL genomic DNA template. Individual master mixes and negative controls were used for each sample in a program consisting of the following steps:10 minute denaturation at 95ºC, followed by 30 cycles of 95ºC/ 1 minute, optimized annealing for 2 minutes at 41.5ºC, and 4 minutes at 72ºC. The final cycle of the previously listed steps was immediately followed by a concluding 10 minute elongation step at 72ºC. After a low EEO 1% agarose "check" gel verified the results, the best of each of the samples (dictated by present DNA concentration as verified by Image J software) was Nano Dropped (NanoDrop ND1000 Spectrophotometer, Thermo Scientific, Wilmington, DE) in triplicate to achieve a proper DNA concentration, and submitted to Idaho State University Molecular Research Core Facility (MCRF) for flow cell sequencing using the Illumina 16S Metagenomic Sequencing Library Protocol (Illumina), the Illumina Nextera XT Index Kit for PCR indexing, and the Illumina MiSeq Reagent Kit v.3 600 cycle chip and the MiSeq Software Suite for sequencing on the Illumina MiSeq 2 Instrument allowing automated generation of DNA clonal clusters via bridge amplification and analysis through utilization of a reversible dye terminator.

Annotation and Phylogenetic Analysis:

MOTHUR

The FastQ-formatted forward and reverse files of the 16S hypervariable gene reads of paired-end tags (PET) produced by sequencing and uploaded to the ISU Galaxy server were assembled by first making a stability file in .txt format and annotated according to the Mothur 1.33.3 MiSeq SOP default settings with the following exceptions: sequences shorter than 35nt and longer than 600nt (with a limit set at 500) having homopolymers longer than 8nt were omitted from further analysis in addition to reads with ambiguous base calls or incorrect primer sequences, a 'pre.cluster' command was utilized to denoise and identify OTUs by applying a pseudo-single linkage algorithm to remove sequences subjected to pyrosequencing errors (28, 62), putative chimeras were identified and removed utilizing the Chimera Uchime algorithm, 'pcr.seqs' command was run across the entire SILVA v.4 database, clustering was accomplished via the 'cluster.split' command, taxonomic classification of each identified OTU was established by setting the distance matrix cutoff at .2 to avoid clustering below 80% similarity and a species cut off set at .03, and a final cut off was set at .1 for genus level identification at 95% and species at 97% according to the RDP (Ribosomal Database Project) database (34, 59).

Normalized average and standard deviations of taxonomic abundance values of significant phylogenetic identifications were obtained from the final file produced by the Mothur of 16S paired-end sequencing of control and synbiotic diet samples. An increase or decrease in abundance was determined from synbiotic values relative to their control counterpart to show the effects of the synbiotic treatment on the diversity and

phylogenetic profile of the gut microbial flora (Table 1). Significant values, as represented in the previously discussed table, are reflected on Figure 1, where an increase in abundance in synbiotic samples comparative to control samples is depicted by an closed diamond, and decrease an open diamond.

A 1000 boot strap value maximum likelihood phylogenetic tree having a 50% majority rule of significance was then generated in Bio Edit software from collected genus level strain assignments produced by Mothur analysis of the 16S paired-end genomic DNA (PET sequencing) of the diet specific mouse cecum samples (Figure 1). Significant strain determined according to a T-test algorithm (data not shown) with all singletons, sequences having a threshold occurrence value less than 10, sequence lengths less than 1400 base or greater than 2300 bases of 16S paired-end sequencing according to type strains present on the Ribosomal Data Base (http://rdp.cme.msu.edu) removed were then truncated on the V3/V4 loop region of the ribosome. 16S was utilized to insure proper directionality (8-1492) on the correct strand while also keeping track of duplicates and the occurrence (frequency and distribution) of the nucleotide sequence. Maximum likelihood, and not the optimized model option available in Model Test software, with default values was utilized due to the large expanse of 14 billion years of evolution of which the database covers. *Thermus aquaticus*, the organism most closely related to the 'in group' without being a member of such, was used as the outgroup and imputed first due to formatting requirements within the program. Clustal-X (sequence alignment), PAUP (analysis), and Model Test (selection of evolutionary model) software were utilized for the multiple alignment with pairwise comparisons initially generated between each followed by group comparison to the closest group.

MG-RAST

The FastQ-formatted forward and reverse files produced by sequencing and uploaded to the ISU Galaxy server were subjected to the following commands on Galaxy prior to submission to Mg-RAST: "FASTQ Groomer" with input quality scores set to "Sanger" while all other options set to default, "NGS: QC and manipulation" in the FASTX-Toolkit for FASTQ data, "Rename sequences" to "numeric," converting the file to 'fasta' format, and each sample's forward and reverse file joined via "Joiner" with the appropriate sequence of file order left to right (3, 11). Once sequences were uploaded into MG-RAST (metagenomics.anl.gov), phylogenetic trees restricted to bacteria having maximum levels of order, genus, and species were generated in addition to analyses of rarefaction, PCOA, and heatmap. Comparisons were also made within diets and between synbiotic and control diets. MG-RAST utilizes M5NR e-values based on MD5 checksums of sequences and annotations provided by a non-redundant data and base developed at Argonne National Laboratory that separate sequence from annotation data (sequence and potential species identifiers in addition to annotations) collected from public databases (protein data base sources: GO, IMG, KEGG, NCBI [RefSeq & GenBank], SEED, UniProt, eggnog, and PATRIC; ribosomal database sources: RDP, Silva, and greengenes). This allowed for the use of higher level functional groups from annotation sets to compare sequence sets and functional hierarchies. Protein coding and ribosomal gene prediction in MG-RAST was accomplished utilizing the protein FragGeneScan and ribosomal gene BLAT similarity searches. This in turn in was utilized to depict the alpha diversity, or description of species, within a given sample. BLAT was also utilized to identify homologous sequences in the M5NR database. Alignments were

made according to sequence similarity comparisons, and amino acid alignments utilizing FragGeneScan to collect ORF predictions prior to BLAT integration for translated amino acid sequence identification within the M5NR database. Annotation within MG-RAST were based on putative gene function collected from the public databases previously discussed (44).

RESULTS

Control Diet Cecum Composition Comparisons

Analysis via MG-RAST of control diet sample 'B,' uploaded on 9/24/14, produced a total of 567 identified ribosomal RNAs (from the predicted 758 predicted rRNA features) from 24,579 (4.7%) of the 521,139 sequences (205,384,411bps) that passed quality control filtering algorithms (pre-quality control bp count at 231,268,701 bp). 608 sequences failed and were removed from downstream analysis. The uploaded mean sequence length was 443+/- 14bp and uploaded mean GC content 54+/- 2% which compares to 394+/-3% mean sequence length post quality control and GC percent post quality control of 54+/-3%. SILVA SSU database generated the highest identification of annotated ribosomal RNA genes (24418), followed by RDP (20096), Greengenes (19671), and SILVA LSU (1) all of which had an e-value raised to -30 and less. 15,945,780 sequences are present in the M5NR protein database which include all unique sequences from applied protein databases, and 309,342 sequences in the M5RNA ribosomal database that contains unique sequences from the utilized ribosomal RNA databases. 97.1% of the detected sequences were of the bacterial domain.

A total number of 316,307 sequences containing 189,417,901 bp, and having an average length of 589 +/- 7bp was initially uploaded for MG-RAST analysis of control

diet sample ‘H’ on 9/30/2014. 444 sequences, or 0.1%, failed to pass the imposed quality control pipeline imposed in this software. Post quality control mean sequence length was 599bp and GC % 56+/-3%. Of those sequences that passed quality control, only 10,766 sequences (3.4%) contained ribosomal RNA genes with 719 alignment identified rRNA features from the initially predicted 1,205. The number of features identifies were annotated by the datasets employed in this sample analysis including protein databases, protein databases containing functional hierarchy information, and rRNA databases. Different databases yield varying results due to the completeness of annotated data contained within the database. SILVA SSU database generated the highest identification of annotated ribosomal RNA genes (9758), followed by RDP (7394), and Greengenes (7183) with an e-value raised to -30 and less. 88.4% of detected sequences belonged to the bacterial domain.

MG-RAST analysis of control diet mouse cecum sample J (9/30/2014) resulted in a total number of 774,847 sequences totaling 463,639,316bp with an average length of 598bp+/-13bp and a mean GC content of 56+/-3%. 0.3% of the sequences (2,307 sequences) failed to pass the quality control. The post quality control sequences (772,540 and 462,750,694bp) had a mean sequence length of 598 +/- 1bp and a mean GC content of 56+/-3%. 35,277 sequences (4.6%) of the sequences that passed quality control contained rRNA genes with 1,194 rRNA features identified from the predicted 2,992. SILVA SSU database generated the highest identification of annotated ribosomal RNA genes (30887), followed by RDP (27684), Greengenes (23560), and SILVA LSU (1) with an e-value raised to -30 and less. 98.1% of the detected sequences belonged to bacteria.

Phylum level taxonomic hits distribution denotes the most abundant phyla in mouse cecum sample B as the following: *Firmicutes*, 30217 (68.3%); *Verrucomicrobia*, 4990 (11.3%); *Actinobacteria*, 3674 (8.3%); *Proteobacteria*, 2321 (5.2%); *Tenericutes*, 795; Unclassified (derived from bacteria), 474; *Bacteroidetes*, 327; *Chordata*, 73; *Cyanobacteria*, 69; *Synerigistetes*, 62; and Unclassified sequences, 57; where listed abundance of annotations were derived on a log scale and is representative of richness and evenness of that taxonomic level within the given sample. Phylum level assignments in sample H also lead with Firmicutes being the most abundant (13704 at 64.4%) while *Actinobacteria* (3799, 17.8%) were the following most prevalent assignments. Both samples contained 'main players' composed of many the same phylum but at different relative abundances (sample H also contained: *Tenericutes*, 460; *Cyanobacteria*, 302; *Proteobacteria*, 289; *Bacteroidetes*, 91; Unclassified (derived from bacteria), 77; *Verrucomicrobia*, 75; and unclassified (derived from other sequences), 41). Sample J also followed the same patterns previously expressed and was composed of the following 'main players' on a phylum level: *Firmicutes* 64.3%, 33263; *Verrucomicrobia* 9.8%, 5071; *Actinobacteria* 9.3%, 4834; *Bacteroidetes* 9.2%; 4775; *Proteobacteria*, 1270; *Tenericutes*, 756; Unclassified (derived from Bacteria), 463; *Cyanobacteria*, 292; and Unclassified (derived from unclassified sequences), 246 (data not shown).

Class level taxonomic abundance between control diet mouse cecum samples B, H, and J also were composed of many the same assignments. The most prevalent class amongst the group was *Clostridia* (B: 26421, 59.7%; H: 7059, 33.2%; and J: 22672, 43.8%). Class assignments of *Clostridia* dominated the most in both B and J while H was closely followed by *Bacilli* at 28.9% (6150). Bacilli was high in the other two samples

but not as such comparative proportions (8512 or 16.4% in sample J and 3283 or 7.4% in B). *Actinobacteria* (3674 [8.3%], 3799 [17.8%], 4834 [9.3]) and *Mollicutes* (795, 460, 756) were also relatively abundant in class assignments for the control diet B, H and J samples, respectively (Figures 10 and 12).

Taxonomic abundance according to order level distinctions among the control mice B, H, and J was led by the most abundant *Clostridiales* (26421, 59.7%; 7056, 33.1%; 22667, 43.8%). Other predominant order identifications included *Lactobacillales*, *Bifidobacteriales*, *Bacillales*, and *Actinomycetales* (data not shown).

The lack of grouping exhibited by the three control diet samples in PCOA analysis demonstrates dissimilarity of taxonomic or functional abundance profiles amongst the control diet samples. The reduced dimensionality of this graphic depiction to limited variables simplifies the data contained within these samples. Neither axis demonstrates a high r^2 value for the dataset (Figure 16). Raw abundance counts relative to each control diet sample is illustrated in Figure 18 while normalized values for the control samples is depicted in figure 20. Distribution of these abundance counts is denoted as a positive integer between 0 and 1 (a uniform scaling that has no impact on the value differences within a single sample or between samples) for the number of times a particular taxon has been detected is explicative of comparisons of normality of the relative distributions of abundance. As raw abundance value distribution characteristically varies among samples (Figure 18), there is still enough similarity amongst the samples as is evident by the clustering at the bottom of each of the samples. The normalization of the values (Figure 20) by a common variable allowing for data of different scales to be compared and generated for the samples, reduced the variation among the sample distributions of

abundance for a more concise analysis of the five number summary (the minimum, first quartile or 25% coverage, median, third quartile or 75% coverage, and maximum) of abundance values for each sample. Each of the samples had a median line at roughly 0.3, a first quartile of ~0.1, and a third quartile nearing 0.6. Some variation was evident in the minimum and maximum values, as J had the lowest, nearing zero, followed by B, and H had the highest maximum value being the closest to 1. J had the lowest maximum value.

Rarefaction curves generated for control diet fed mouse cecum samples reflected annotated species richness where the annotation number total was derived as a function of the original number of sampled sequences. Species abundance is reflected by the initial curve evident in Figures 2, 3, and 4 for the control samples as the data generating the curve is calculated by the observed abundance of species within the utilized datasets subsample annotations. Each of the figures demonstrated a sharp initial slope indicative of novel species within the initiation of the analysis. The rounding off or eventual plateau observed from decreased slope steepness with increasing number of reads correlates to adequate sampling within each of the control samples where additional sampling is unlikely to result in a significant amount of detected novel species assignments. The general trend observed within this graphic analysis is a sharp initial rise followed by a plateau at an asymptote as demonstrative of decreasing detection of new species per unit of collected individuals. Figure 8 depicts this trend by superimposing all rarefaction curves for samples of the control diet. J had the largest slope, as is represented its higher alpha diversity value followed by B and H. However, sample H did have a larger alpha diversity than B but also contained less reads than B.

The alpha diversity, or relative organism diversity with a single number in a sample, of control diet cecum sample B is 36.412 species within the range of two standard deviations and is estimated from species level annotation distribution which is a measure of species richness. Shannon diversity is reflected by the average weighted abundance of the logarithmic value of relative abundance of the generated annotated species within this sample as collected by annotation source databases utilized in MG-RAST. Mouse B had the lowest observed alpha diversity among the control diet cecum samples followed sample H (44.185 species) and the highest value of 55.725 species in sample J (data not shown).

K-mer profile of the k-mer rank abundance of 15-mer coverage according to sequence size for sample B yielded a decrease of optimal coverage of 1202604 after 67 sequences from the initial generated plateau until 1.2e^6 sequences when no detectable coverage was measured (data not shown). This correlates to a decrease in abundance of high-coverage, repetitive sequences as analysis progressed within this sample set. In comparison, sample H yielded a decrease of optimal coverage of 442413 after 25 sequences from the initial generated plateau until 8.89e^6 sequences when no detectable coverage was measured. Sample J also produced a decrease of optimal coverage from a value of 1202604 after the pattern exhibited by mouse H (22).

Synbiotic Diet Cecum Composition Comparisons

A total number of 664,839 sequences containing 398,224,212 bp, and having an average length of 598 +/- 5bp was initially uploaded for MG-RAST analysis of synbiotic diet sample ‘C’ on 9/26/2014. 417 sequences, or 0.1%, failed to pass the imposed quality control pipeline. Post quality control mean sequence length was 599bp and GC % 55+/-

3%. Of those sequences that passed quality control, only 20,139 sequences (3.0%) contained ribosomal RNA genes with 1,145 alignment identified rRNA features from the initially predicted 2,502. SILVA SSU database generated the highest identification of annotated ribosomal RNA genes (17820), followed by RDP (17710), and Greengenes (16170) with an e-value raised to -30 and less. 94.6% of detected sequences belonged to the bacterial domain.

MG-RAST analysis of synbiotic sample 'F' on 9/30/14 contained 1,079,526 sequences having 646,611,472bp and an average length of 598+/-5bp from which 650 sequences, or 0.1% failed quality control filtering pipelines within the software package. 66,505 (6.2%) of the post quality control sequences which had a mean sequence length of 599bp and GC% of 54 +/-3% contained ribosomal RNA genes with 2,291 of the predicted 4,322 aligned rRNA features being identified. SILVA SSU database generated the most identified ribosomal RNA genes (61126) with RDP (60122), Greengenes (59102), and SILVA LSU (1) following. 98.5% of the detected sequences were of the bacterial domain.

Analysis of the synbiotic diet sample 'L' on 9/30/2014 was performed on 926,558 sequences totaling 554,892,258 bp having an average length of 598+/- 6bp. 721 sequences (0.1%) failed to pass the quality control. Post quality control sequences had a mean sequence length of 599bp and a mean GC percent content of 56+/-3% with 2,335 rRNA features identified of the initially predicted 3,573. Source hits distribution was highest with the SILVA SSU database (24964) followed by RDP (23587), and Greengenes (22592) databases. 96.4% of the detected sequences were of the bacterial domain.

Phylum level taxonomic hits distribution for individual synbiotic diet cecum samples result in the most abundant classification, measuring at 74.6% (30249), of the phyla assignments generated for synbiotic mouse C belonged to *Firmicutes*. *Actinobacteria* (11.1%, 4488), *Verrucomicrobia* (1665), *Proteobacteria* (748), *Tenericutes* (366), and *Cyanobacteria* (420) then followed as the most abundant identifications for this sample. Samples F and L also resulted with many of the same phyla being the most prevalent, respectively: *Firmicutes* (69.6%, 67777; 66.1%, 37812), *Actinobacteria* (13.2%, 12891; 22.9%, 13106), *Verrucomicrobia* (12.3%, 11981), *Proteobacteria*, (1669, 1151), *Cyanobacteria* (567, 863), and *Tenericutes* (410, 1765). Many of the following phyla assignments also correlated, just at differing abundances per sample (data not shown).

Class level taxonomic abundance between synbiotic diet mouse cecum samples C, F, and L also were composed of many the same assignments. All lead with *Bacilli* (37.6%, 15263; 46.7%, 45447; 41.3%, 23625) and had many of the same following class assignments at relatively high abundance including: *Clostridia* (29.8%, 12103; 21.9%, 21328; 17.55%, 9993), *Actinobacteria* (29.8%, 4488; 13.2%, 12891; 17.5%; 13106), *Erysipelotrichi* (7.1%, 2866; 299; 7%, 4008), *Verrucomicrobiae* (1665; 12.3%, 11981; 79), and *Gammaproteobacteria* (540; 848; 675) (Figures 11 and 13).

Order level distinctions of taxonomic abundance among the synbiotic mice C, F, and L were led by *Lactobacillales* (30.5%, 12374; 41.5%, 40376; 34.4%, 19702) and followed by *Clostridiales* (29.8%, 12093; 21.9%, 21324; 17.5%, 9986), *Bacillales* (7.1%, 2889; 5.2%, 5071; 6.9%, 3923), and *Bifidobacteriales* (6.5%, 2639; 9.3%, 9018; 18.2%, 10390) as the most prevalent distinctions. Comparative to assignments based on the level

of class, order distinctions are less in agreement across the three samples of the synbiotic diet samples (data not shown).

PCOA analysis of the synbiotic samples (Figure 17) demonstrates reduced dimensionality and dissimilarity of taxonomic or functional abundance profiles through the lack of clustering in the position of the samples on the figure. Comparative to the results of the control data previously discussed, neither axis contained a high r^2 value for the dataset (PCO1, 0.61386; PCO2, .38614). Raw abundance counts relative to each control diet sample is illustrated in Figure 19 while normalized values for the control samples is depicted in Figure 21. Distribution of these abundance counts depicts normality of the taxon detection amount as explained previously. There was variation among the raw abundance count distribution despite the similarity exhibited among the samples (Figure 19). Each sample had two data points distributed higher than the rest of the data and mean numbers which were clustered at the bottom of the graphic. However, sample F, listed as the first sample on the figure, had higher placement of the top two data points comparative to the other two samples. The remaining data points were similar across samples. Normalized values (Figure 21) generated a median line similar to the control diet at roughly 0.3. However, the first quartile of the synbiotic diet samples was higher at roughly 0.2, but the third quartile was similar to the control diet by being located at roughly 0.6 for each of the samples. Although all of the samples had a minimum value less than 1, C was the highest, and F was the lowest. C also produced the highest maximum value of the diet samples, nearing 1, while both F and L were closer at roughly 0.9.

Synbiotic diet fed mouse cecum samples rarefaction curves (figures 5, 6, and 7 for samples C, F, and L, respectively) depicting the species richness of the annotations within the samples is shown by the initial steep with high relative slope and curve produced in each of the figures. Each the curves produced a plateau with increasing read number to show adequate sampling were detection of novel species assignments would be unlikely with integration of additional samples in the analysis. Figure 9 depicts this trend by superimposing all rarefaction curves for samples of the synbiotic diet. F had the largest slope and most reads followed by C and L. However, sample L contained more reads than that of sample C.

Measured alpha diversity also differed amongst the samples. Although it had the largest slope, sample F had the lowest alpha diversity of the synbiotic diet at a value of 25.40 species, while C had the highest generated alpha diversity measurement at 43.36 followed by L at 38.17.

K-mer rank abundance of 15-mer coverage according to sequence size yielded a profile in sample C depicting a decrease of optimal coverage of 1202604 right before the plateau formed by the initial 25 sequences until 8.89e^6 sequences when no detectable coverage was measured (data not shown). The level of coverage of rare sequences followed by high-coverage, repetitive sequences is shown by this decreasing value across sequences. Sample F yielded a decrease of optimal coverage of 3269017 after 25 (further than sample C) sequences from the initial generated plateau until 2.2e^7 sequences when no detectable coverage was measured. Sample L also produced a decrease of optimal coverage at the same value of C but at the same point in sequencing as F.

Cecum Composition Comparison between Diets

The normalized average of taxonomic abundance values of significant phylogenetic identifications of control and synbiotic diet samples showed an increase in strain abundance in all taxons listed in Table 1 with the exception of *Nocardioides* genus, *Clostridia* class, *Clostridiales* order, and *Lachnospiraceae* family while comparing synbiotic treatment results to control diet results. The measured increase in abundance was determined from synbiotic sample values relative to the control (Table 1). Not all significant genus level assignment abundances were included in the Mothur-based tree. Included within the table were higher level taxonomic classifications which showed significant abundances that may not have included the genus level assignments such as *Sphingomonas* or higher levels of taxonomic classification beyond the designation included in the figure.

A 1000 bootstrap 16S paired-end genus level phylogenetic rectangular consensus cladogram according to a 50% majority rule of significant singletons as derived from a T-test (data not shown) and containing sequences with a less than 10 threshold occurrence value, length between 1400 and 2300 bases was generated from data produced by the Mothur software package for type strains present in the Ribosomal Data Base (Figure 1). Firmicutes and *Actinobacteria* were among the most prevalent phylogenetic assignments, followed by *Gammaproteobacteria, Alphaproteobacteria, Bacteriodetes*, and *Betaproteobacteria,* Increased statistical significance is integrated by use of 100 bootstraps for resampling. The shorter the branch lengths, the increased relation among the designations, as is evident among the *proteobacteria*. Both of the supplemented probiotic genus were detected and are present among the *Firmicutes*. Other commonly

utilized strains of probiotics, as previously discussed, are also present. As this figure is representative of the type strains present in RDP, exact phylogenetic identity is unknown. However, this figure is representative of the immense diversity present within the cecum. Figure 1 is limited compared to the phylogenetic analyses generated by MG-RAST in that assignments are not depicted according to individual sample or diet but collectively according to the integration of all samples from both analyzed diets. The Mothur-based tree contained more strain abundance increases when comparing synbiotic to the control diet, however not all designations included on the table appeared on the figure.

Comparative MG-RAST analysis of all of the samples from both the control diet mice B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3) and synbiotic C (MG-RAST accession 4581845.3), F (4582170.3), and L (4582211.3) diet was performed on 11/19/2014 and finished on 2/10/2015. The 16S paired-end phylogenetic trees of the cecum microbiome of the samples yielded the following color designations for the samples as is depicted in the center of the figure: B, silver; C, blue; F, red; H, gold; J, purple; and L, green (Figures 22-24). Data for each of the figures was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases). Each of the figures also contained stacked bar leaf weights but differed in the maximum phylogenetic level and taxonomic coloring assignments (figure 22 contained a genus maximum and class coloring, Figure 23 species maximum and class coloring, and Figure 24 species maximum and order coloring). Figure 22 showed the highest abundance of class level identification among *Clostridia* followed by *Gammaproteobacteria, Bacilli, and Actinobacteria*. Many of the samples amongst both

diets contained the same genus level identifications. It is notable that all of the strains contained the *Lactobacillus* probiotic which was added to the synbiotic diet. Unlike Figure 1, *Lactococcus* was not detected in this figure. *Bifidobacterium*, a common probiotic utilized, was also present in all of the samples. Figure 23 provides species level classification in concordance to the trends depicted in Figure 22. However, the supplemented probiotics utilized in the synbiotic diet were not detected through analysis utilizing MG-RAST. Many of the commonly utilized strains of probiotics are present, however. The most prevalent order, as depicted by Figure 24, is that of *Clostridiales* followed by *Bacillilales* and *Lactobacillilaes*- all of which are present in each of the 6 analyzed samples. Analysis of these figures shows that there is not a designated prevalence of genus or species according to diet, as assignments are not limited to the three samples composing the diet, but include at least one sample of the opposing diet. There are phylogenetic identifications listed that are specific to a given sample, however, as is depicted by the singular bar coloring found at the center of the figure.

A normalized bar chart comparing the detected bacteria according to groups designated by diet with corresponding p-values is provided in Figure 25. The data was compared to M5NR with a 1e-5 e value maximum, minimum identity at 60%, and minimum alignment length of 15bp according to RNA database hits. The smallest p value, which indicates the most distinct difference between the diet groups, was found in *Actinobacteria* with a value of 0.104, and the highest p value in *Spirochaetes* with a value of 1. Probiotic-containing *Firmicutes* was the most abundant and had a p value of 0.1165, followed by *Actinobacteria* and *Cyanobacteria* (0.1628).

DISCUSSION AND CONCLUSION

The exact identity of species shifts between the two states of host health has remained elusive despite the limited knowledge of phyla level changes. This impedes understanding of microbial community interactions during treatments of dysbiosis due to the heterogeneity of the GI tract as a result of differences within microenvironments and spatial distribution of flora and metabolites. This can cause variations of effectiveness and activity of pre- and probiotics by location within the gut microbiome as the metabolic products of one bacteria can be modified and utilized by another bacterial species, increasing the availability of a molecule in its active form can be mitigated by community level biotransformation reactions. These cooperative interactions directly affect the degree of effectiveness of a prebiotic, as the necessary active form may never reach its target location, or a probiotic, which may contain a strain that does not yield the desired, beneficial effect on microbial composition and function (31, 41).

Next-gen high resolution, automated DNA based technologies have replaced time and labor extensive traditional microbial ecological investigations that integrate clone library generation where every clone consists of a conserved primer PCR amplification product and sequence of that clone. As many sequence collections are generally necessary for comparison, intensive sampling of each library is required for accumulation of adequate coverage of a microbial community sample. Due to the log of microbial population in the sample, complete analysis of every cell in the community cannot be performed, and thus analysis relies on statistical integration to generate general diversity measurements (60). Next generation sequencing has facilitated the interest and understanding of human, animal, and environmental microbial community structure and

function in addition to increasing the knowledge of novel pathogens and functionality and effect of a consortia of microbes on a myriad of other ailments and disease states. The structure-function hypothesis, when applied in a biomedical standpoint, consists of the concept that structural and stability alterations within microbial communities can results in changes in the health and disease states of the human host. This is generally applied across the Human Microbiome Project (National Institute of Health and MetaHit funded in the US and European Commission) while investigating the effects of microbial diversity on health and disease (61). Diversity analysis integrating alpha (species richness), and beta (between sample comparisons), though not obtainable in their entirety, can aid in determining these elusive ecological trends and assignments while integrating a quality threshold that reduces potential biases from filtering (61, 64).

OTU assignments were utilized in this study according to the derived genetic distance between sequences. Distribution of sequence abundances among OTUs allowed for general estimates of ecological richness, evenness, and diversity of the community in addition to measurements of the like between communities of differing diet supplementation. Phylogenetic methods employed investigated differences in communities according to sequence difference. The application of the OTU approach allowed for quantitative measurements to be collected (60). However, it is noted that employing the OTU definition can result in an overestimation in similarity amongst community comparisons (58). Past research has claimed that accurate distance based threshold for taxonomic level definitions can be created through consensus-based methods of OTU classification. This is due to the inability to define bacterial taxonomic levels resulting from the lack of adequate bacterial taxa being cultured or culturable (60).

Many of the present taxonomic outlines and requirements are based on previously cultured organisms causing candidate phyla and non-culturable phyla that are lacking in taxonomy identifying to the level of genus or species. However, there is currently no accepted and employed definition of a bacterial species which increases the difficulty in appropriate taxonomic classification according to phylotype or even the genera, family, class, order, or phyla of bacteria. The operational definition for a species cites a 3% dissimilarity, but it is not widely accepted. Also, preclustering followed by clustering at 3% (equivalent to 97% sequence identity which shows shifts in clusters of higher and reduced dominance) as employed in this study, allowed for increased accuracy in OTU characterization in addition to providing a reduction of singleton sequence proportion and minimally affecting the distribution and presence of microbial taxa. Single nucleotide errors also had minimal effect on classification of sequences. Methods employing OTU approaches avoid many of the limitations implicit to phylotype analysis due to the lack of bin restrictions since taxonomy outlines are not applied. Sequences can thus be assignment and clustered with equal basis regardless of reference sequence representation or restrictions issued in outline classifications. OTU assignment is depended on the presence of other sequences in the dataset. However, this methodology assumes that the 16S bacterial rRNA gene evolves at the same rate among all taxonomic affiliations which is disputed (61, 64).

Despite the varying sequences produced in each of the samples within the two diets, sequences had similar length (roughly 600bp) and contained comparable GC % content (between 50 and 60%) after the quality control pipelines imposed during analysis via MG-RAST. SILVA and RDP databases produced the most hits among all samples.

Phylogenetic analysis results produced by the extensive filtering and quality control pipelines employed by both Mothur and MG-RAST were comparable but varied to some degree. Data produced from Mothur analysis was later graphically represented based on type strains collected on the Ribosomal Data Base of statistically relevant sequences based on OTU assignments due to the computational, memory, and time limitations imposed by the Mothur (57) software- a methodology not applied to the same sequencing files imputed into MG-RAST, as previously discussed. It is also due to these limitations that individual phylogenetic assignments could not be designated to the separate samples or between the diets studied. Mothur identified *Firmicutes* and *Actinobacteria* as the most prevalent phylogenetic assignments, followed by *Gammaproteobacteria, Alphaproteobacteria, Bacteriodetes,* and *Betaproteobacteria*. Also shorter branch lengths and therefore increased relationship between designations was most evident among *Proteobacteria*. Additionally, both of the genus contained in the synbiotic diet supplementation were detected among the generated *Firmicutes*. Other commonly utilized strains of probiotics were also present on the Mothur-generated phylogenetic analysis of identity among all of the samples. However, as Mothur-generated results were representative of the type strains present in RDP, exact identity is unknown, limiting the data compared to the phylogenetic analyses generated by MG-RAST in that assignments are not depicted according to individual sample or diet but collectively according to the integration of all samples from both analyzed diets. Yet, immense diversity present within the cecum is still apparent from the results generated.

The normalized average of taxonomic abundance values of significant phylogenetic identifications of control and synbiotic diet samples showed an increase in

strain abundance in all taxons listed in Table 1 with the exception of *Nocardioides* genus, *Clostridia* class, *Clostridiales* order, and *Lachnospiraceae* family when comparing synbiotic sample values relative to the control (Table 1). Indigenous gut flora determine *C. difficile* colonization and infection. *Clostridia* bacterium have shown to induce various degrees of dysbiosis on the gut microbiome, detrimentally affecting host heath as previously discussed. A reduction of abundance, but not complete removal of presence, conserves microbial diversity within the gut microbial community while decreasing the propensity of disease development. *Nocardioides* has not shown any involvement in human pathology. Dominant colonization with *Lachnospiraceae* bacterial family is common in the diseased mammalian host, and has shown to have an inverse relationship with *Clostridia.* The significant increase in abundance of the strains included in the probiotic supplementation within the synbiotic diet contribute to the effectiveness of the probiotic by achieving the requirements previously discussed. Not all significant genus level assignment abundances were included in the Mothur-based tree. Included within the table were higher level taxonomic classifications which showed significant abundances that may not have included the genus level assignments such as *Sphingomonas* or higher levels of taxonomic classification beyond the designation included in the figure.

Application of the Mothur data can be made to the results generated to MG-RAST as comparative analysis of all of the samples from both the control and synbiotic diets (as capable of the MG-RAST software) (Figures 22-24) showed the highest abundance of class level identification among *Clostridia* followed by *Gammaproteobacteria, Bacilli, Actinobacteria*, similar genus level identifications, and the presence of probiotic *Lactobacillus* which was added to the synbiotic diet. Unlike Mothur results, *Lactococcus*

was not detected. However, the common probiotic *Bifidobacterium* was identified in all of the samples. The supplemented probiotics utilized in the synbiotic diet were not detected through analysis utilizing MG-RAST by species level classifications; yet, like Mothur results, many of the commonly utilized strains of probiotics were identified. The most abundant order in all of the samples was identified as *Clostridiales* followed by *Bacillilales* and *Lactobacillilaes*. However, individual analysis of these figures shows that there is not a designated prevalence of genus or species according to diet, as assignments are not limited to the three samples composing the diet, but include at least one sample of the opposing diet. The smallest p value indicating the most distinct difference between the diet groups was found in *Actinobacteria* with a value of 0.104, and the highest p value in *Spirochaetes* with a value of 1. Probiotic-containing *Firmicutes* was the most abundant and had a p value of 0.1165, followed by *Actinobacteria* and *Cyanobacteria* (0.1628) (Figure 25). The distance between branches within the trees correlate to the relation of the identifications included.

Individual MG-RAST mitigated analyses of both the control and synbiotic diet samples identified *Firmicutes* as containing the highest number of phylum level taxonomic hits with the same phyla assignments following in abundance though not always in the same order among samples. Class level taxonomic abundance between control diet mouse cecum samples also were composed of many the same assignments as *Clostridia* dominated and less prevalent assignments were among the same groupings of Class. Despite class level taxonomic abundance between synbiotic diet mouse cecum samples containing of many the same assignments within that diet, the most abundant was not the same as the control diet samples as all lead with *Bacilli*. Abundance

according to order level distinctions within the control yielded *Clostridiales* as the most abundant compared to synbiotic samples which identified *Lactobacillales* as the most abundant. This is notable as it correlates to the supplemented strain of probiotic within the diet, and supports the effectiveness of the synbiotic treatment combination as successfully establishing and selectively cultivating the probiotic strain through the presence of the prebiotic compounds contained in that diet (data not shown).

PCOA analysis of the control and synbiotic samples demonstrates reduced dimensionality and dissimilarity of taxonomic or functional abundance profiles through the lack of clustering in the position of the samples (Figures 16, 17). Comparative to the results of the control data, neither axis contained a high r^2 value for the dataset. The distribution of raw abundance and normalized depicts normality of the taxon detection amount as previously mentioned. Variation among the raw abundance count distribution was evident in both diets due to clustering patterns despite the similarity exhibited among the samples. Normalized values generated a median line at 0.3 for both diets. However, the first quartile of the synbiotic diet samples was higher at roughly 0.2 compared to the control diet at ~0.1, and the third quartile location was roughly 0.6 for both diets. Although all of the samples had a minimum value less than 0.1, there was variation amongst the samples of each of the diets. This trend was also evident in analysis of the maximum value, where all sample locations were less than 1, but varied to some degree among diets. Transformation of data to achieve a normal Gaussian distribution via Normalization resulted in the reduction of variation or biases introduced throughout sample preparation and analysis that are not under experimental control in addition to decreasing their impact on interpretation of results (Figures 20 and 21). This also improve

the results exhibited in other comparative analyses (PCA, heatmap, etc) that assume a normal distribution of data for applicable, statistically relevant data interpretation. Means of data normalization are discussed in further detail on the MG-RAST V 3.0 database. The comparable results evident between the diets is represented of the stability of the gut microbiomes in both of the diets. This could represent a general homeostatic balance required for functionality in a healthy host. The variations observed correlate to the individual differences in the profiles analyzed which could be attributed to the difference in taxonomic or functional abundances produced by the supplementation of probiotic strains in the synbiotic diet as previously discussed.

Rarefaction curves reflecting annotated species richness where the annotation number total was derived as a function of the original number of sampled sequences produced an initial steep slope representing species for each of the diets. This sharp initial slope indicates of novel species within the initiation of the analysis of both of the samples. Both diets also yielded an eventual plateau at an asymptote from decreased slope steepness with increasing number of reads correlating to adequate sampling among diets where additional sampling is unlikely to result in a significant amount of detected novel species assignments. Each of the samples within each diet produced variation in the steepness of the initial slope with varying read amounts. The slope of the samples did not correlate to alpha diversity levels among diets as is evident in the variation among samples. The control diet generated a higher average alpha diversity among samples at 44.441 compared to the synbiotics 35.643 species. This could be due to the difference in the number of available reads between samples, however as this measure is dependent on that variable.

The k-mer rank abundance profile of 15-mer coverage according to sequence size yielded a decrease of optimal coverage at less than 70 sequences for all samples contained in both diets from the initial generated plateau until no detectable coverage was measured (data not shown). Redundancy, or repetitiveness, of sequences within each sample according to occurrence of distinct 15 bp patterns is representative of rare, or low coverage sequences being initially depicted, followed by high-coverage, repetitive sequences (the decrease on coverage previously discussed). The rank abundance plot is a function of abundance rank with the most abundant sequences being first listed. This represents a decrease in abundance of high-coverage, repetitive sequences as analysis progressed within this sample set. Use of k-mer applications allows for the identification of the closest sequence template for the generated data pertaining to a given sample. These results could represent taxonomic diversity increasing in lower dominant distributions (2, 22, 57, 59).

As technology advances, the challenge with phylogenetic analysis methodology has shifted from sample sequence generation to sequence analysis. MG-RAST serves as a high-throughput pipeline for high performance computing and annotation allowing a low cost, next generation means of worldwide metagenomic sequence analysis. The functional sequence assignments of the metagenomic input produced by this software package are automated and generated from both nucleotide and protein database comparison allowing for functional summaries and comparative phylogenetic analysis (59, 64). Biological databases are growing exponentially, and algorithms that minimize processor time and memory requirements are becoming increasingly important, which was problematic while utilizing Mothur. Analysis and clustering algorithms are also

computationally intensive for OTU-based methodology. Mothur utilizes a neighbor joining algorithm that is taxonomy-independent and performs better than most deterministic and heuristic methods available. However, the generated OTUs can represent sequences from multiple lineage assignments due to there being no taxonomic level threshold commonly employed. The genetic distance between full 16S gene sequences in a given taxonomic assignment were continuous in each hierarchy level (60).

Even prior to the establishment of next generation sequencing as a primary method of analysis, the 16S rRNA gene was the most represented gene present in the GenBank database. However, various biases such as a microbial population relative abundance misrepresentation in a given sample and errors including actual sequence misrepresentation as a result of PCR sequencing and amplification can be present when utilizing the 16S rRNA gene in sequence survey. The DNA extraction method, DNA purification protocol, selection of PCR primers, PCR cycling conditions (PCR polymerases erroneously result in substitution in 1 of 10^5 to 10^6 bases, and amplification of heterogeneous templates can result in formation of chimeras (not a sequencing error) when incomplete amplification products are present to serve as primers for related fragments at a rate of 5-45%. Sequencing also results in errors due to the homopolymer under-representation at a rate of 0.01-0.02), community composition within the sample, and copy number of the 16S gene within the genome can effect whether the relative abundances of the gene being sequenced are equal to the bacterial presence in the sample. These biases confound the representation and application of data collected. Additionally, microbial ecology analyses include the inherent hypothesis that

microbial community structure changes directly affect the function within the community (56, 61).

Most studies assume that partial sequence distances are not significantly different from full-length sequence distances despite the fact that the 16S rRNA gene lacks uniform evolution throughout its length. When employing the proxy species definition specific to full length sequences 3% distance cut off, the variability in evolution within the 16S rRNA gene becomes apparent. Genetic diversity also decreases along the 16S rRNA gene length. Regression coefficients do not adequately explain variation between regions in comparison to the whole gene. Longer reads increase the relation of segmented analysis to the whole gene. The 16S gene is a marker for diversity within a genome and follows a well-determined secondary structure. The analysis of this gene via next gen sequencing allows for replicates to be analyzed in addition to increased complexity of experimental designs to be investigated while increasing the breadth and depth of sampling. Technical limitations are based on conserved PCR primer availability, fragment length, and gene quality generation while analytical limitations are dependent on accurate sequence classification and genetic diversity within a region availability. This necessitates the use of only a select region of a gene to be studied. Differing regions will be selected until analysis becomes standardized (58). However, the datasets that are currently available for comparison are not completely exhaustive as there has been shown to be as few as 10% or as high as 98% lack of sequence matching from a sample to a dataset. The accuracy of annotations is dependent on the quality of the data used (44). This also applies to beta-diversity studies, though useful in community comparisons, have limited use to only communities exhibiting clear differences and does not provide

information pertaining to the details of these differences due to database-dependent methods that are limited according to the lack of representation of rare and novel populations when analyzing the deep coverage existent in many environmental samples (34, 40).

The results generated in this study conclude that there was a measurable phylogenetic difference in microbial community composition between the two diets administered, thus supporting the use of probiotics as an effective means of establishing homeostatic balance of beneficial bacteria within the cecum content of the host in addition to having an application in regards to pharmaceutical intervention for correcting a dysbiotic state in the gut microbial flora of a diseased host over the alternative, invasive and possible harmful choices of intervention. However, it is noted that competition of nutrients and ecological sites as well as stress can also cause a decrease in effectiveness in this treatment (7, 31, 54). Further research is needed to investigate and characterize the intestinal communities of microbes, their metabolic activity, and functionality to increase the efficiency and effectiveness of synbiotic treatments.

Acknowledgements

MRCF

References

1. de Wit, N., Derrien, M., Bosch-Vermeulen, H., Oosterink, E., Keshtkar, S., et al. (2012) Saturated fat stimulates obesity and hepatic steatosis and affects gut microbiota composition by an enhanced overflow of dietary fat to the distal intestine. *Am J Physiol Gastrointest Liver Physiol*. 303:G589-599.
2. Bartram, A. K., Lynch, M. D., Stearns, J. C., Moreno-Hagelsieb, G., and J. D. Neufeld. (2011) Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Applied and Environmental Microbiology*. **77**(11)3846-3852. DOI: 10.1128/AEM.02772-10.
3. Blankenberg, D., Gordon, A., Von Kuster, G., Coraor, N., Taylor, J., and A. Nekrutenko; Galaxy Team. (2010) Manipulation of FASTQ data with Galaxy. *Bioinformatics*. **26**(14):1783-5.
4. Blanton, C. Personal Communication. 2014.
5. Borody, T. J., Khoruts, A. (2012) Fecal microbiota transplantation and emerging applications. *Nat Rev Gastroaenterol Hepatol*. 9:88-96.
6. Bäckhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., et al. (2004) The gut microbiota as an environmental factor tht regulates fat storage. *Proc Natl Acad Sci* USA 101:15718-15723.
7. Cammarota, G., Ianiro, G., Bibbò, S., and A. Gasbarrini. (2014) Gut microbiota modulation: probiotics, antibiotics or fecal microbiota transplantation? *Internal and Emergency Medicine*. 9:365-373.

8. Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., and R. Knight. (2010) Global Patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS*. 108: 4516-4522.
9. Chen, J., He, X., and J. Huang. (2014) Diet effects in gut microbiome and obesity. *Journal of Food and Science*. Doi: 10.1111/1750-3841.12397
10. Clement, B. G., Kehl, L. E., Debord, K. L., and C. L. Kitts. (1998) Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities. *Journal of Microbiological Methods*. 31:135-142.
11. Cock, P.J., Fields, C.J., Goto, N., Heuer, M.L., and P. M. Rice. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*. 2009 Dec 16.
12. Costello, E.K., et al. (2009) Bacterial community variation in human body habitats across space and time. *Science.* 326:1694-1697.
13. Cotter, P. D., (2011) Small intestine microbiota. *Curr Opin Gastroenterol*. 27:99-105.
14. Cunningham-Rundles, S., Ahrné, S., Johann-Lang, R., Abuav, R., Dunn-Navarra, A., Grassey, C., Bengmark, S., and J. S. Cervia. (2011) Effect of Probiotic Bacteria on Microbial Host Defense, Growth, and Immune Function in Immunodeficiency Virus Type-1 Infection. *Nutrients.* 3:1042-1070.
15. Curtis, T. P., and W. T. Sloan. (2004) Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Ecology and Industrial Microbiology*. 7:221-226.

16. de Wit, N., Derrien, M., Bosch-Vermeulen, H., Oosterink, E., Keshtkar, S., et al. (2012) Saturated fat stimulates obesity and hepatic steatosis and affects gut microbiota composition by an enhanced overflow of dietary fat to the distal intestine. *Am J Physiol Gastrointest Liver Physiol*. 303:G589-599.
17. Dethlefsen, L., Huse, S., Sogin, M. L., and D. A. Relman. (2008) The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biol*. **6**(11): 2383-2400.
18. Dethlefsen, L., McFall-Ngai, M., and D. A. Relman. (2007) an ecological and evolutionary perspective on human-microbe mutalism and disease. *Nature*. 449:811-818.
19. Di Bella (2013) Fecal microbiota transplantation: the state of the art. *Infectious Disease Reports*. (**5**)13:43-45.
20. Druart, C., Neyrinck, A. M, Dewulf, E. M., De Backer, F. C., Possemiers, S., et al. (2013) Implication of fermentable carbohydrates targeting the gut microbiota on conjugated linoleic acid production in high-fat-fed mice. *Br J Nutr*: 1-14.
21. Druart, C., Neyrinck, A. M., Vlaeminck, B., Fievez, V., Cani, P. D., and N. M. Delzenne. (2014) Role of the Lower and Upper Intestine in the Production and Absorption of Gut Microbiota-Derived PUFA Metabolites. *PLoS ONE*. 9(**1**): e87560. Doi: 10.1371/journal.pone.0087560 pgs1-5
22. Edgar, R. C. (2004) Local homology recognition and distance measures in linear time using compressed amino acid alphabets. *Nucleic Acids Res.* 32:380–385.

23. Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Grotman, R. M., and J. Ravel. (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*. (**2**)6.

24. Fritz, j. V., Desai, m. S, Shah, P., Schneider, J. G., and P. Wilmes. (2013) From meta-omics to causality: experimental models for human microbiome research. *Microbiome*. 1:14.

25. GI Langille, M., Meehan, C. J., Koenig, J. E., Dhanani, A. S., Rose, R. A., Howlett, S. E., and R. G. Beiko. (2014) Microbial shifts in the aging mouse gut. *Microbiome.* 2:50.

26. Hayashi, H., Takahashi, R., Nishi, T., Sakamoto, M., and Y. Benno. (2005) Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *J Med Microbiol*. 54:1093-1101.

27. Hooper, L. V. (2004) Bacterial contributions to mammalian gut development. *Trends Microbiol*. 12:129-134.

28. Huse, S. M., Welch, D. M., Morrison, H. G., and M. L. Sogin. (2010) Ironing out the wrinkles in the rare biosphere through improved OUT clustering. *Environmental Microbiology*. 12(**7**):1889-1898.

29. Imaizumi, K., Akutsu, T., Miyasaka, S., and M. Yoshino. (2007) Development of species identification tests targeting the 16S ribosomal RNA coding region in mitochondrial DNA. *International Journal of Legal Medicine*. 121:184-191.

30. Khoruts, A., Dicksved, J., Jansson, J. K., and M. J. Sadowsky. (2010) Changes in the composition of the human fecal microbiome after bacteriotherapy for

recurrent Clostridium difficile-associated diarrhea. *J Clin Gastroenterol*. 44:354-360.

31. Klemashevich, C., Wu, C., Howsmon, D., Howsmon, D., Alaniz, R. C., Lee, K., and A. Jayaraman. (2014) Rational identification of diet-derived postbiotics for improving intestinal microbiota function. *Current Opinion in Biotechnology*. 26:85-90.
32. Konstantinidis, K. T., and J. M. Tiedje. (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci* USA 102:2567-2572.
33. Kostic, A. D., Howitt, M. R., and W. S. Garrett. (2013) Exploring host-microbiota interactions in animal models and humans. *Genes Dev.* 27:701-718.
34. Kozich, J.J., Westcott, S.L., Baxter, N.T., Highlander, S.K., and P. D. Schloss. (2013) Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*. 79(**17**):5112-20.
35. Lazarevic, V., et al. (2009) Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. *J Microbiol Methods.* 79:266-271.
36. Lee, J. C., and R. R. Gutell. (2012) A comparison of the crystal structure of eukaryotic and bacterial SSU ribosomal RNAs reveals common structural features in the hypervariable regions. *PLoS ONE*. (**7**)5:1-6.
37. Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., et al. (2005) Obesity alters gut microbial ecology. *Proc Natl Acad Sci* USA 102: 11070-11075.

38. Ley, R. E., Peterson, D. A., and J. I. Gordon. (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*. 124:837-848.

39. Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., et al. (2008) Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci* USA. 105:2117-2122.

40. Liu, W.T., Marsh, T.L., Cheng, H., and L. J. Forney. (1997) Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Genes Encoding 16S rRNA. *Appl Environ Microbiol*. 63(**11**):4516-4522.

41. Magrone, T., and E. Jirillo. (2013) The interaction between gut microbiota and age-related changes in immune function and inflammation. *Immunity & Ageing*. 10:31.

42. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O., and D. L. Kasper. (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*. 122:107-118.

43. McGarr, S. E., Ridlon, J. M., and P. B. Hylemon. (2005) Diet, anaerobic bacterial metabolism, and colon cancer: a review of the literature. *J Clin Gastroenterol.* 39: 98-109.

44. Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., and R. A. Edwards. (2008) The metagenomics RAST server- a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 9:386.

45. Mujico, J. R., Baccan, G. C., Gheorghe, A., Diaz, L. E., and A. Marcos. (2013) Changes in gut microbiota due to supplemented fatty acids in diet-induced obese mice. *Br J Nutr*. 1-10.

46. Nicholson, J. K., Holmes, E., and I. D. Wilson. (2005) Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol*. 3:431-438.

47. Niot, I., Poirier, H., Tran, T. T., and P. Besnard. (2009) Intestinal absorption of long-chain fatty acids: evidence and uncertainties. *Prog Lipid Res*. 48:101-115

48. Noverr, M. C., and G. B. Huffnagle. (2005) The 'microflora hypothesis' of allergic responses. *Clin Exp Allergy*. 35:1511-1520.

49. O'Hara, A. M., and F. Shanahan. (2006) The gut flora as a forgotten organ. *EMBO Rep.* 7:688-693.

50. O'Shea, E. F., Cotter, P. D., Stanton, C., Ross, R. P., and C. Hill. (2012) Production of bioactive substances by intestinal bacteria as a basis for explaining probiotic mechanisms: bacteriocins and conjugated linoleic acid. *Int J Food Microbiol*. 152: 189-205.

51. Rankoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and R. Medzhitov. (2004) Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell*. 118:229-241

52. Reddy, B. V. B., Kallifidas, D., Kim, J. H., Harlop-Powers, Z., Feng, Z., and S. F. Brady. (2012) Natural Product Biosynthetic Gene Diversity in Geographically Distinct Soil Microbiomes. *Applied and Environmental Microbiology*. 78 (**10**);3744-3752. Doi:10.1128/AEM.00102-12.

53. Salnier, D. M., Ringel, Y., Heyman, M. B., Foster, J. A., Bercik, P., Shulman, R. J., Versalovic, J., Verdu, E. F., Dinan, T. G., Hecht, G., and F. Guarner. (2013) The intestinal microbiome, probiotics and prebiotics in neurogastroenterology. *Gut Microbes*. 1(**4**): 17-27.

54. Saulnier, D. M., Santos, F., Roos, S., Mistretta, T., Spinler, J. K., Molenaar, D., Teusink, B., and J. James Versalovic (2011) Exploring metabolic pathway reconstruction and genome-wide expression profiling in *Lactobacillus reuteri* to define functional probiotic profiles. *PLoS One*. 6(4)1:e18783.

55. Sartor, R. B. (2008) Microbial influences in inflammatory bowel diseases. *Gastroenterology*. 134:577-594.

56. Schloss, P. D. (2008) Evaluating different approaches that test whether microbial communities have the same structure. *The International Society for Microbial Ecology Journal*. 2: 265-275.

57. Schloss, P. D. (2009) A high-throughput DNA sequence aligner for microbial ecological studies. *PLoS ONE*. (4)12: e8230.

58. Schloss, P. D. (2010) The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Computational Biology*. 7(**6**):e1000844.

59. Schloss, P. D. et al. (2009) Introducing mother: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*. 75: 7537-7541.

60. Schloss, P. D., and S. L. Westcott. (2011) Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Applied and Environmental Microbiology*. 10(**77**):3219-3226.

61. Schloss, P. D., Gevers, D., and S. L. Westcott. (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE*. 6(**12**):e27310.

62. Tripathi, B. M., et al. (2014). Spatial Effects on Soil Bacterial Communities in Malaysian Tropical Forests. *Environmental Microbiology*. DOI:10.1007/s00248-014-0404-7.

63. Walsh, C. J., Guinane, C. M., O'Toole, P. W., and P. D. Cotter. (2004) Beneficial modulation of the gut microbiota. *FEBS lett*. 588(**22**):4120-4130.

64. Werner, J. J., Zhou, D., Caporaso, J. G., Knight, R., and L. T. Angenent. (2012) Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. *The International Society for Microbial Ecology Journal*. 6:1273-1276.

65. Zhang, X., Zeng, B., Liu, Z., Liao, Z., Li, W., Wei, H., Fang, X. (2014) Comparative diversity of gut microbiota in two different human flora-associated Mouse strains. *Current Microbiology*. 69:365-373.

66. Zoetendal, E. G., Raes, J., van den Bogert, B., Arumagam, M., Booijink, C. C., et al. (2012) The human small intestinal microbiota is driven by rapid intake and conversion of simple carbohydrates. *ISME J*. 6:1415-1426.

DATA

Table 1. Normalized average taxonomic abundance values from Mothur analysis of 16S paired-end sequencing of control and synbiotic diet samples. An increase or decrease in abundance is determined from synbiotic values relative to their control counterpart.

Taxanomic Level	Taxon	Control Average	Control Standard Deviation	Synbiotic Average	Synbiotic Standard Deviation	Increase	Decrease
6	Unclassified Acidobacteria	0	0	1.300861804	2.253158738	*	
6	Unclassified Acidimicrobiales	0	0	0.525978471	0.911021435	*	
6	Kocuria	0	0	0.983764235	0.975747665	*	
5	Nocardiaceae	0	0	1.051956941	1.822042869	*	
6	Rhodococcus	0	0	1.051956941	1.822042869	*	
6	Aeromicrobium	0	0	0.333333333	0.577350269	*	
6	Nocardioides	1.474009324	1.609876591	0	0		*
4	Nitriliruptorales	0	0	1.826840274	1.963162868	*	
5	Nitriliruptoraceae	0	0	1.826840274	1.963162868	*	
6	Nitriliruptor	0	0	1.826840274	1.963162868	*	
4	Solirubrobacterales	0	0	1.176409372	1.035762123	*	
5	Conexibacteraceae	0	0	1.176409372	1.035762123	*	
6	Conexibacter	0	0	1.176409372	1.035762123	*	
5	Prevotellaceae	0	0	1.176409372	1.035762123	*	
6	Prevotella	0	0	1.176409372	1.035762123	*	
5	Sphingobacteriaceae	0	0	0.666666667	1.154700538	*	
6	Sphingobacterium	0	0	0.666666667	1.154700538	*	
6	Chelatococcus	0	0	1.176409372	1.035762123	*	
6	Agromonas	0	0	0.333333333	0.577350269	*	
4	Sphingomonadales	3.43529961	0.249407594	6.038970823	1.917194353	*	
5	Sphingomonadaceae	3.43529961	0.249407594	5.388539921	1.224108691	*	
6	Acinetobacter	0	0	0.666666667	1.154700538	*	
6	Pseudomonas	0.980645143	0.882758007	3.228366313	1.405243312	*	
6	Moraxella	0	0	0.525978471	0.911021435	*	
3	Bacilli	7693.249296	4193.816257	22078.15274	2393.512187	*	
4	Lactobacillales	7666.847041	4177.209249	22034.67544	2372.19706	*	
5	Aerococcaceae	0	0	0.666666667	1.154700538	*	
6	Aerococcus	0	0	0.666666667	1.154700538	*	
5	Enterococcaceae	3.85089896	5.257604443	20.68721768	12.64405135	*	
6	Enterococcus	3.85089896	5.257604443	19.70345345	11.66832969	*	
5	Lactobacillaceae	7326.48491	4182.193139	13659.47212	785.1739086	*	
6	Lactobacillus	6926.249723	4327.384756	13085.72426	792.5078663	*	
5	Streptococcaceae	156.6942946	57.45560938	8014.39676	2483.549268	*	
6	Lactococcus	17.5148527	6.451849537	7943.217961	2490.62685	*	
6	unclassified	2.294081843	2.002301437	12.21862984	2.983466826	*	
5	unclassified	172.436285	85.70878406	328.5511499	78.23330943	*	
6	unclassified	172.436285	85.70878406	328.5511499	78.23330943	*	
3	Clostridia	24984.25364	4590.072698	10322.34422	2991.153003		*
4	Clostridiales	24956.43237	4591.263597	10299.22301	2992.641808		*
5	Eubacteriaceae	11.61933553	2.677567758	35.00220898	4.604352731	*	
6	Anaerofustis	10.63869039	3.291712254	34.66887564	4.53162293	*	
5	Lachnospiraceae	17512.64269	5257.16041	7585.27055	2141.076544		*
6	Moryella	0	0	1.051956941	1.822042869	*	
6	Faecalibacterium	0	0	0.666666667	1.154700538	*	

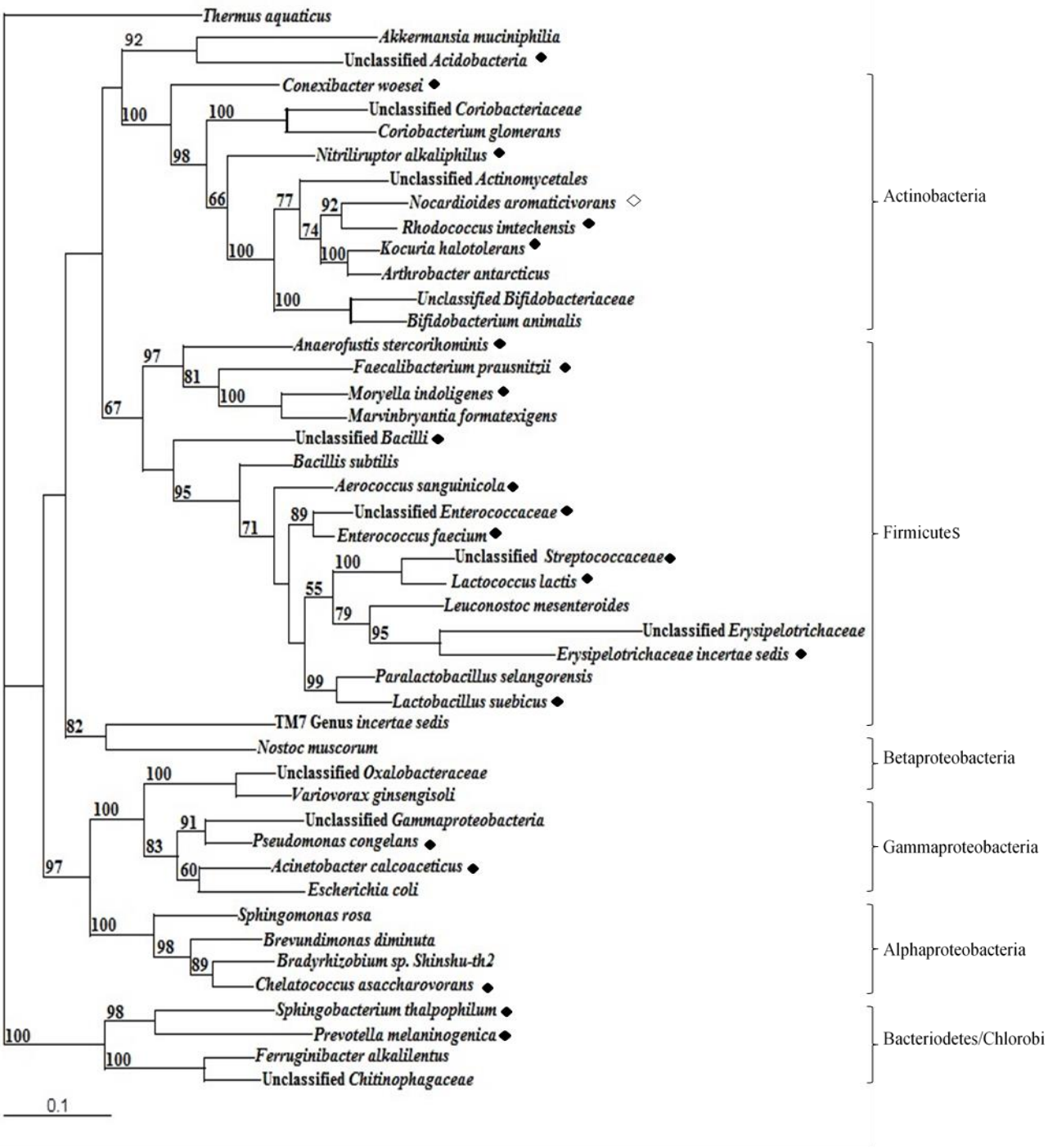

Figure 1. Maximum likelihood 1000 boot strap value phylogenetic rectangular consensus cladogram with 50% majority rule of significant (according to a T-test algorithm [data not shown] with all singletons, sequences having a threshold occurrence value less than 10, sequence lengths less than 1400 base or greater than 2300 bases removed) genus level strains reported by Mothur analysis of 16S paired-end sequencing according to type strains present on the Ribosomal Data Base. Significant taxa increase from normalized averages as shown in Table 1 is depicted by a closed diamond and decrease by an open diamond.

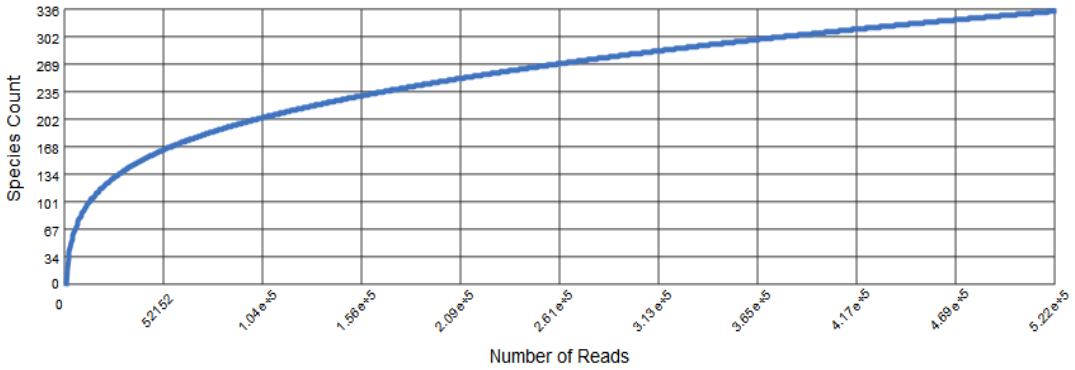

Figure 2. Rarefaction curve generated by MG-RAST of control diet cecum sample B pertaining to annotated richness of species. The distinct species annotation total number is a function of the original sampled sequence number. Curve data is calculated from generated species abundance, and curves are representative of the complete dataset's subsample annotations of the average number of differing species.

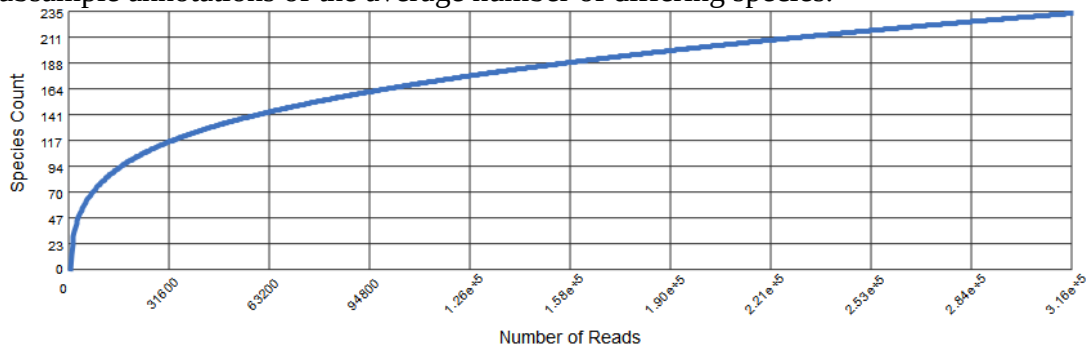

Figure 3. Rarefaction curve of control diet cecum sample H generated by MG-RAST pertaining to annotated richness of species. The distinct species annotation total number is a function of the original sampled sequence number. Curve data is calculated from generated species abundance, and curves are representative of the complete dataset's subsample annotations of the average number of differing species.

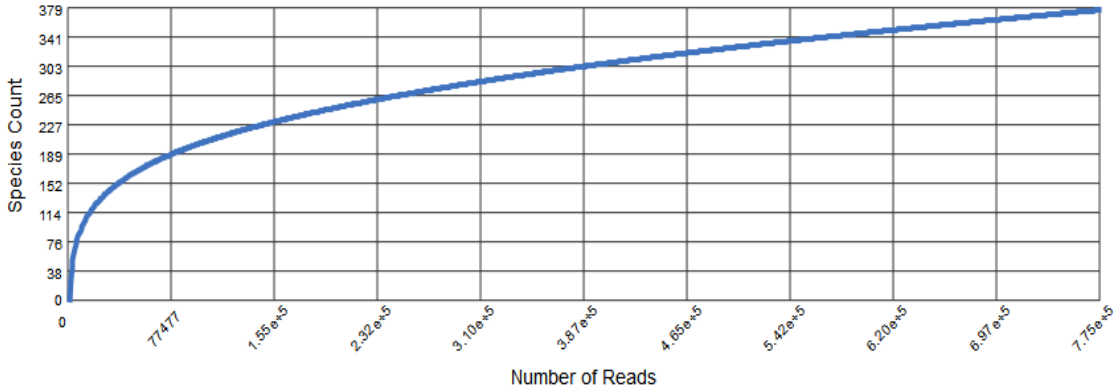

Figure 4. Rarefaction curve generated by MG-RAST of control diet cecum sample J pertaining to annotated richness of species. The distinct species annotation total number is a function of the original sampled sequence number. Curve data is calculated from generated species abundance, and curves are representative of the complete dataset's subsample annotations of the average number of differing species.

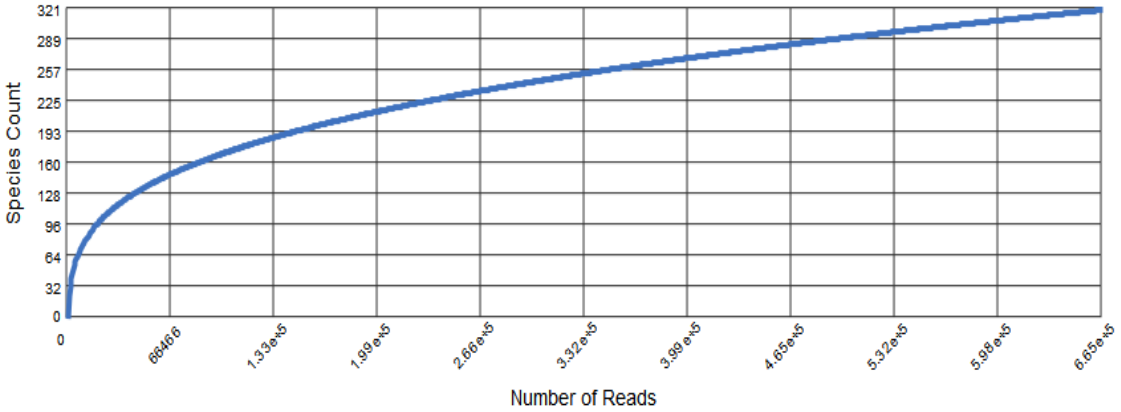

Figure 5. Rarefaction curve of control diet cecum sample C pertaining to annotated richness of species. The distinct species annotation total number is a function of the original sampled sequence number. Curve data is calculated from generated species abundance, and curves are representative of the complete dataset's subsample annotations of the average number of differing species.

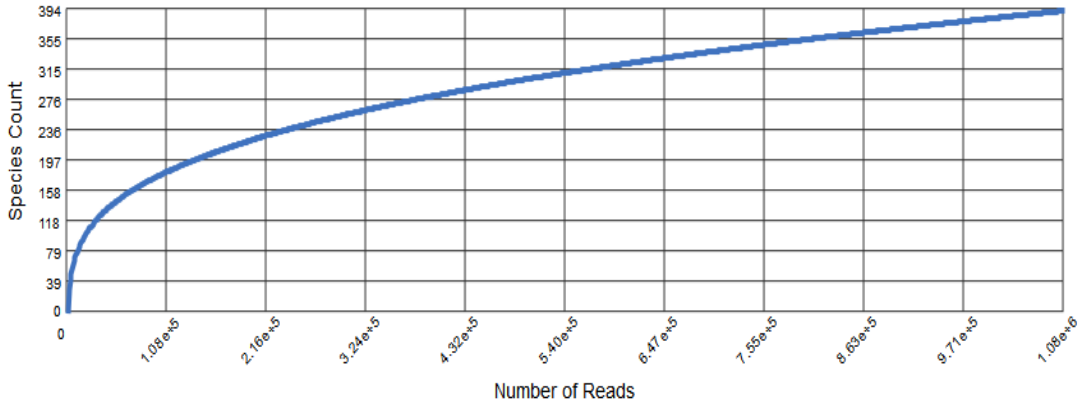

Figure 6. Rarefaction curve of control diet cecum sample F pertaining to annotated richness of species. The distinct species annotation total number is a function of the original sampled sequence number. Curve data is calculated from generated species abundance, and curves are representative of the complete dataset's subsample annotations of the average number of differing species.

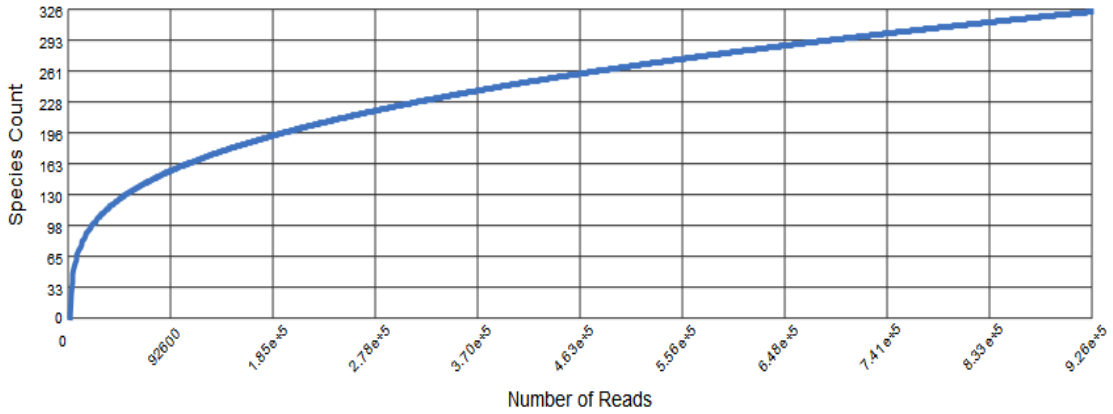

Figure 7. Rarefaction curve generated by MG-RAST of control diet cecum sample L pertaining to annotated richness of species. The distinct species annotation total number is a function of the original sampled sequence number. Curve data is calculated from generated species abundance, and curves are representative of the complete dataset's subsample annotations of the average number of differing species.

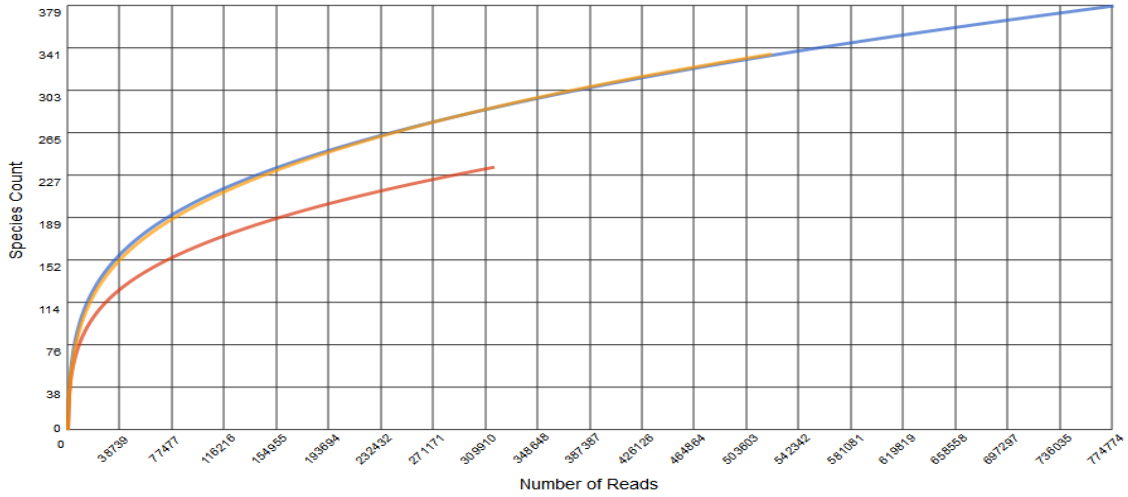

Figure 8. Rarefaction curve of 16S bacterial paired-end fragments of cecum microbiome of control diet mice, B, F, and J generated by MG-RAST. Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases).The top, blue line is representative of sample J, having an alpha diversity value of 55.72; middle, orange line is representative of sample B, having an alpha diversity value of 36.41; and bottom, red line mouse H, having an alpha diversity of 44.18 (data not shown).

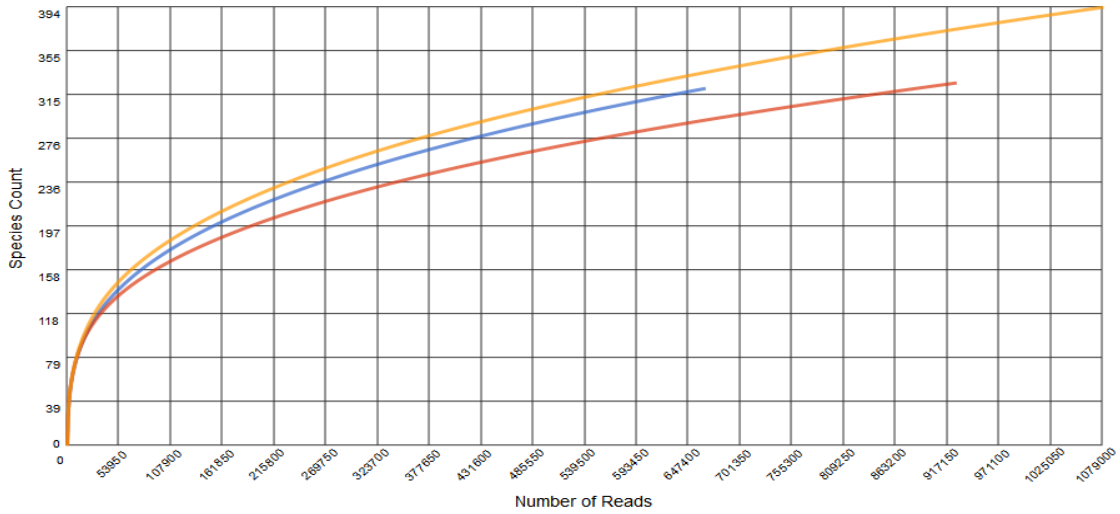

Figure 9. Rarefaction curve of 16S bacterial paired-end fragments of cecum microbiome of synbiotic diet mice, C, F, and L. Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases).The top, orange line is representative of sample F, having an alpha diversity value of 25.40; middle, blue line is representative of sample C, having an alpha diversity value of 43.36; and bottom, red line mouse L, having an alpha diversity of 38.17 (data not shown).

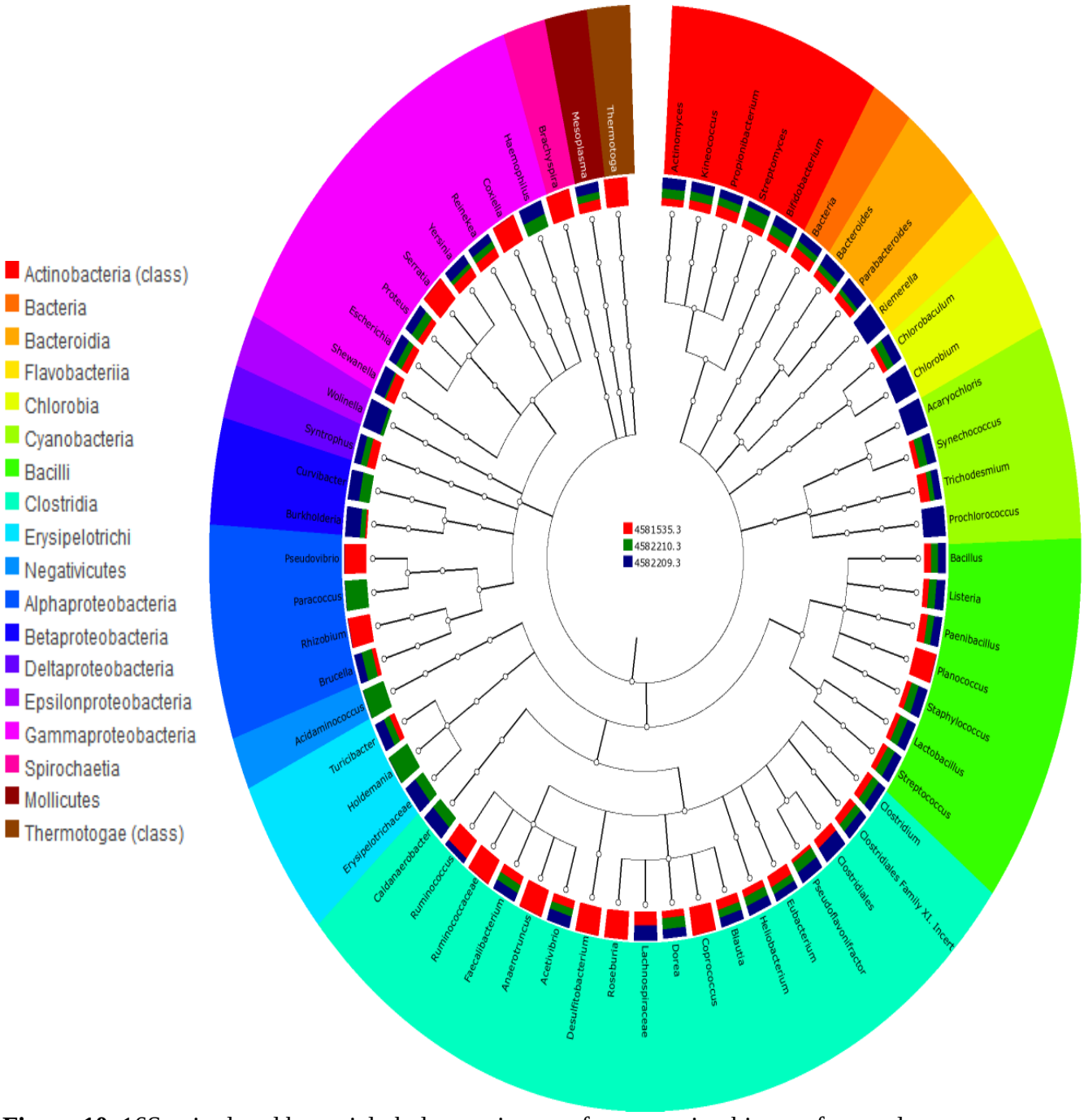

Figure 10. 16S paired-end bacterial phylogenetic tree of cecum microbiome of control diet mice, samples B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3) having stacked bar leaf weights, a genus maximum level, and coloring according to class**.** Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases). Order membership was indicated by the color of the species names.

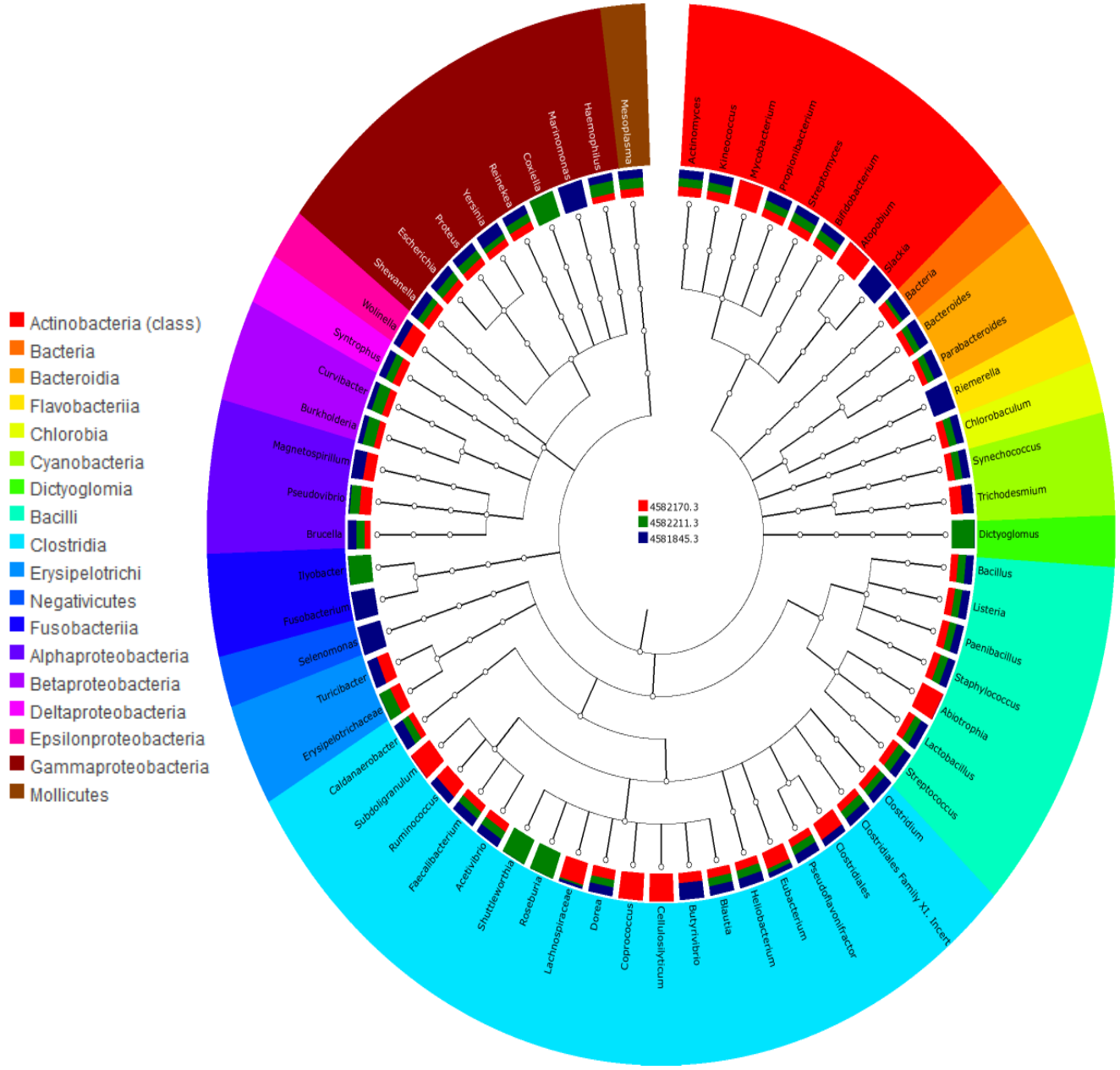

Figure 11. 16S paired-end bacterial phylogenetic tree of cecum microbiome of synbiotic diet mice, samples C (MG-RAST accession 4581845.3), F (4582170.3), and L (4582211.3) having stacked bar leaf weights, a genus maximum level, and coloring according to class**.** Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases). Order membership was indicated by the color of the species names.

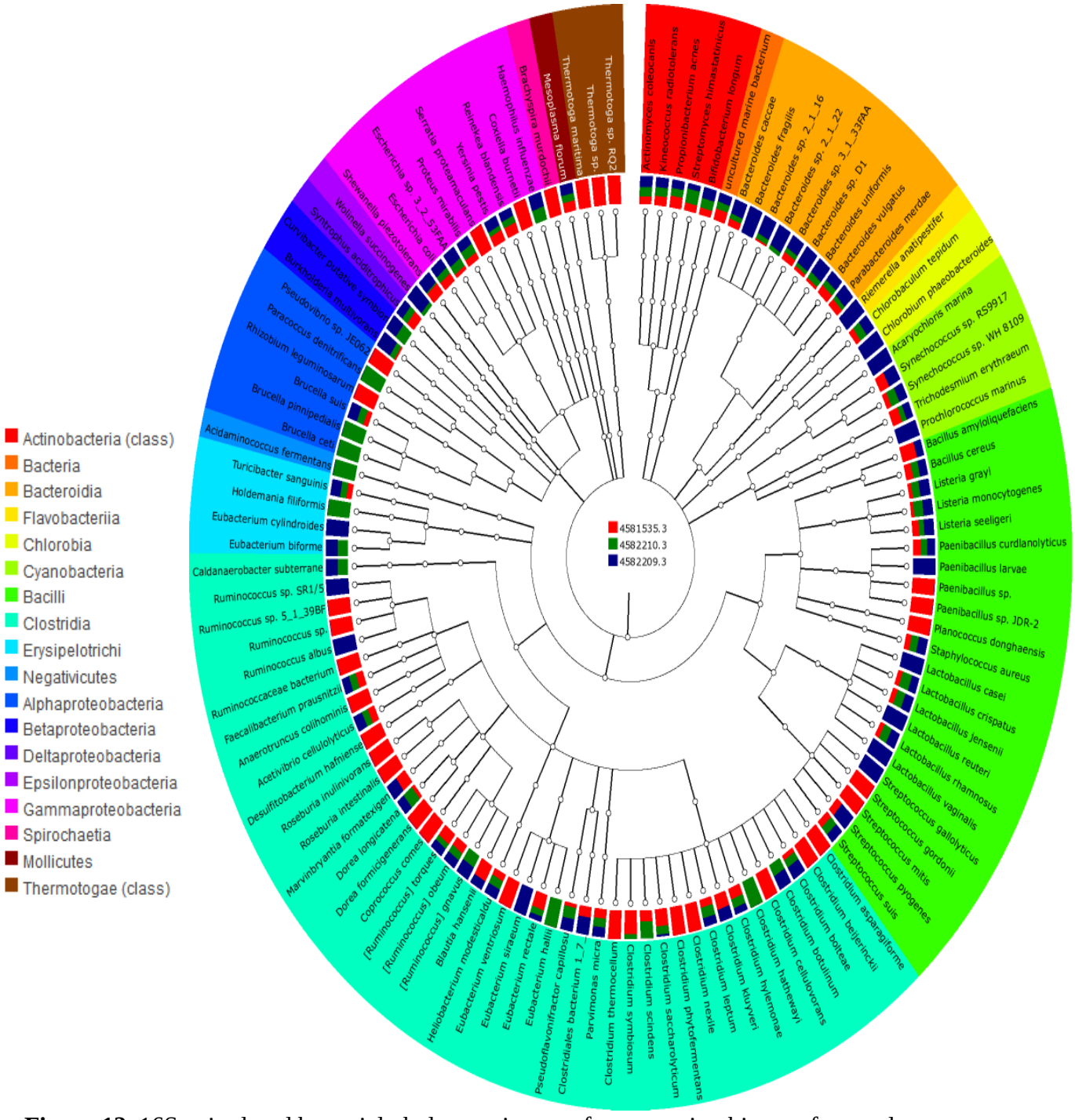

Figure 12. 16S paired-end bacterial phylogenetic tree of cecum microbiome of control diet mice, samples B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3) having stacked bar leaf weights, a species maximum level, and coloring according to class. Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases). Order membership was indicated by the color of the species names.

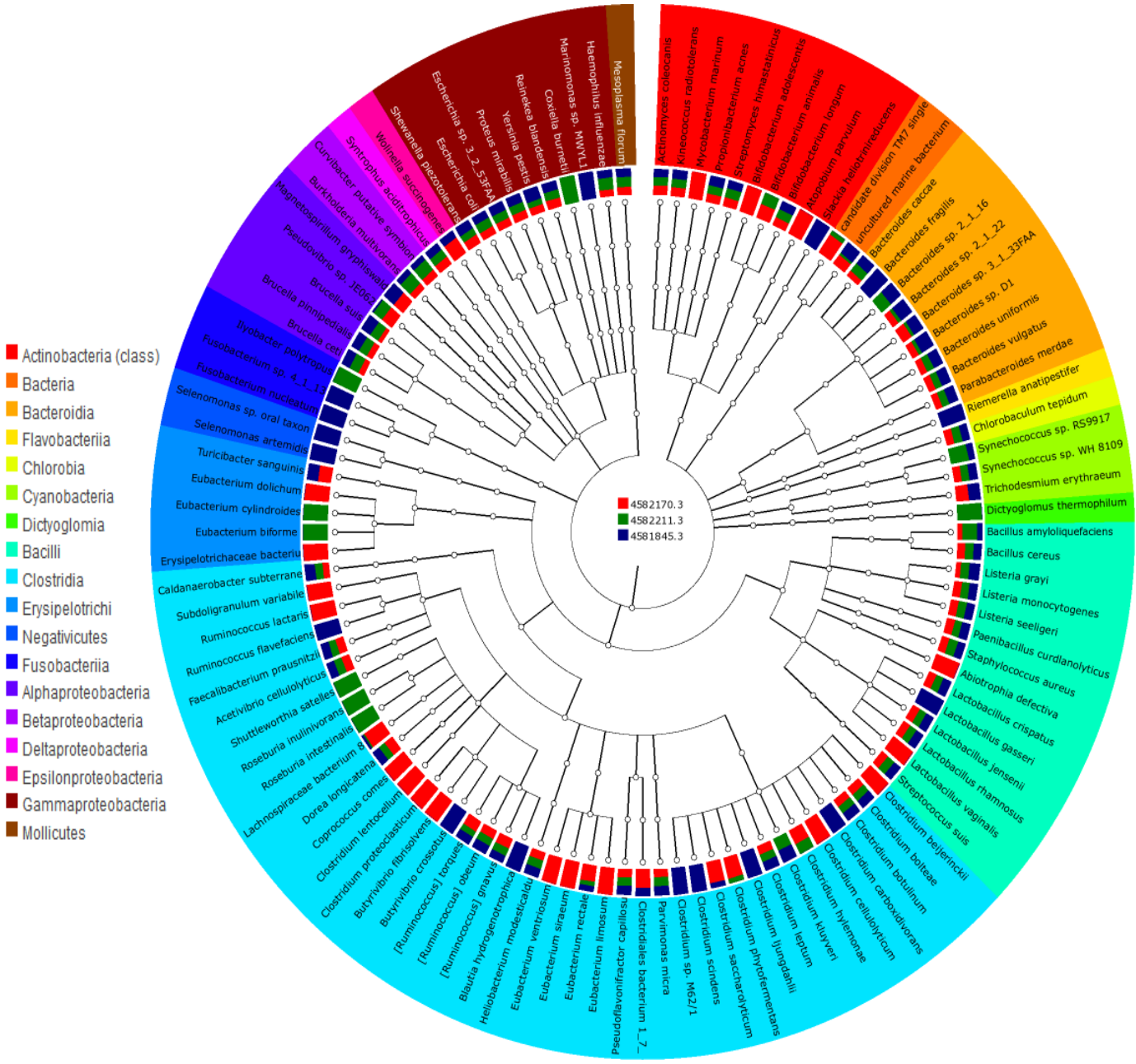

Figure 13. 16S paired-end bacterial phylogenetic tree of cecum microbiome of synbiotic diet mice, samples C (MG-RAST accession 4581845.3), F (4582170.3), and L (4582211.3) having stacked bar leaf weights, a species maximum level, and coloring according to class**.** Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases). Order membership was indicated by the color of the species names.

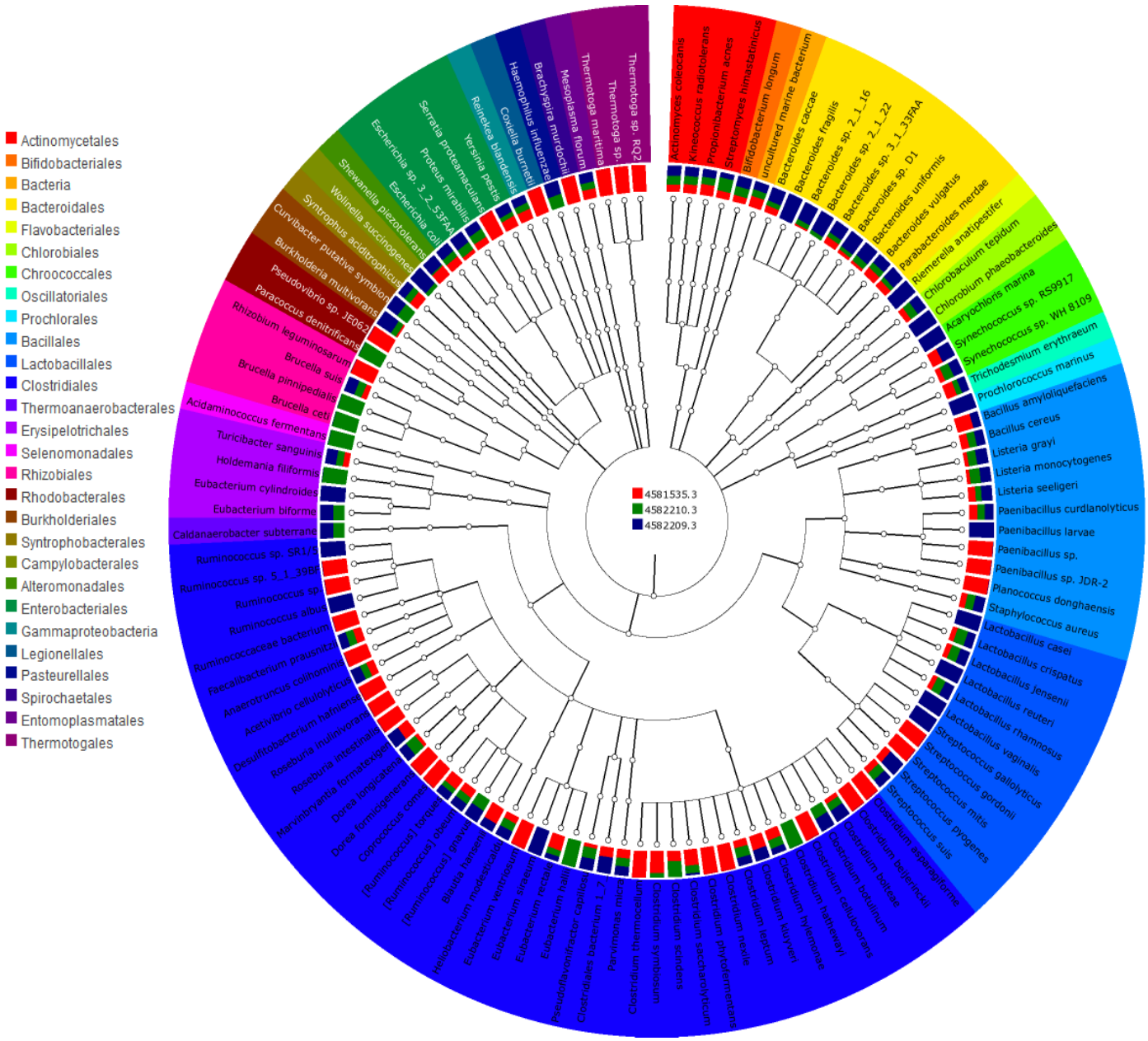

Figure 14. 16S paired-end bacterial phylogenetic tree of cecum microbiome of control diet mice, samples B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3) having stacked bar leaf weights, a species maximum level, and coloring according to order. Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases). Order membership was indicated by the color of the species names.

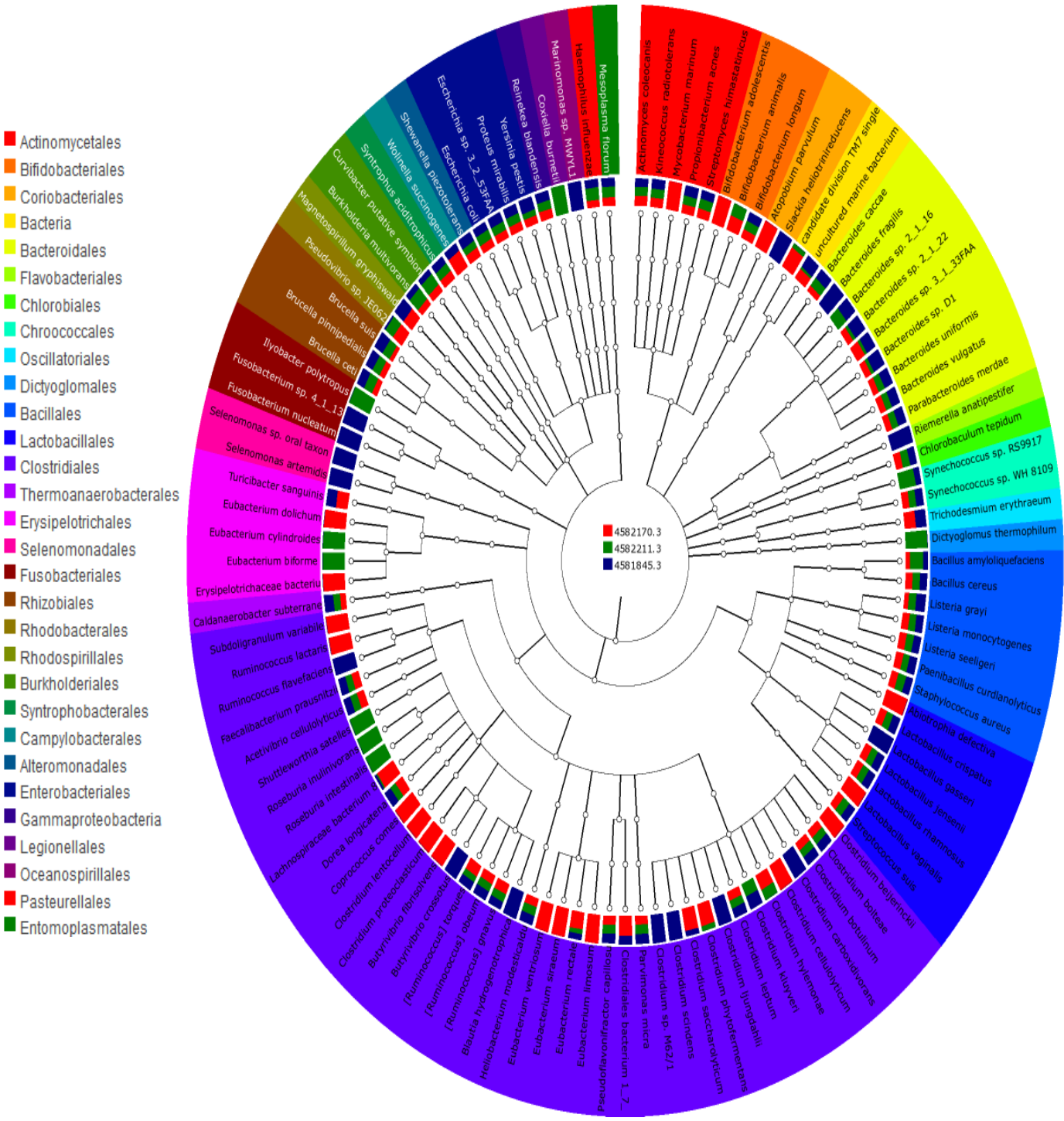

Figure 15. 16S paired-end bacterial phylogenetic tree of cecum microbiome of synbiotic diet mice, samples C (MG-RAST accession 4581845.3), F (4582170.3), and L (4582211.3) having stacked bar leaf weights, a species maximum level, and coloring according to order**.** Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases). Order membership was indicated by the color of the species names.

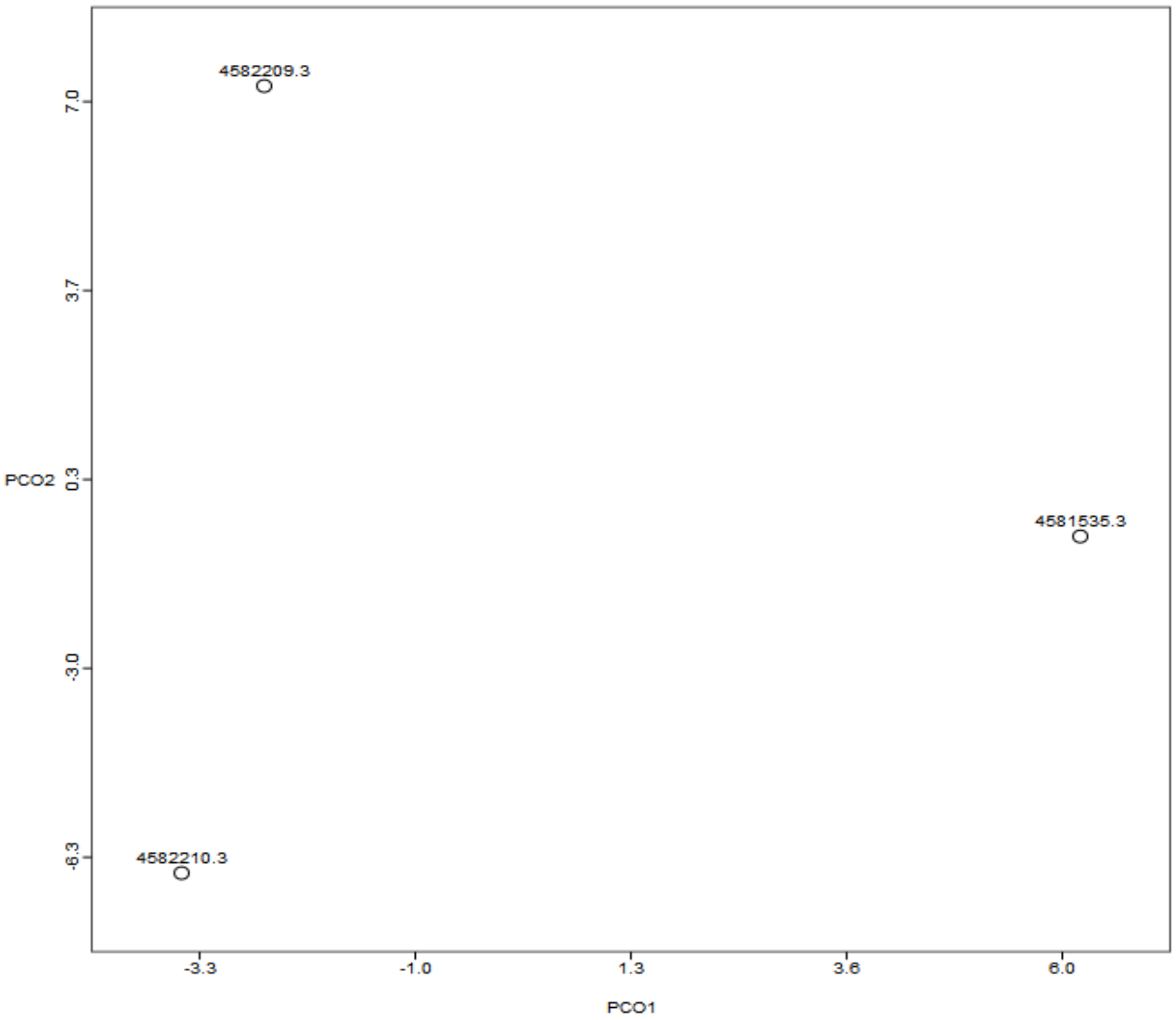

Figure 16. A 16S paired-end Bray-Curtis PCOA analysis of control diet mouse cecum microbiome content of samples B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3). Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases), and having normalized data values between 0 and 1. PCO1 is listed on the x-axis and has an r^2 value of 0.62852 while PCO2 is represented on the y-axis and has an r^2 value of .37148.

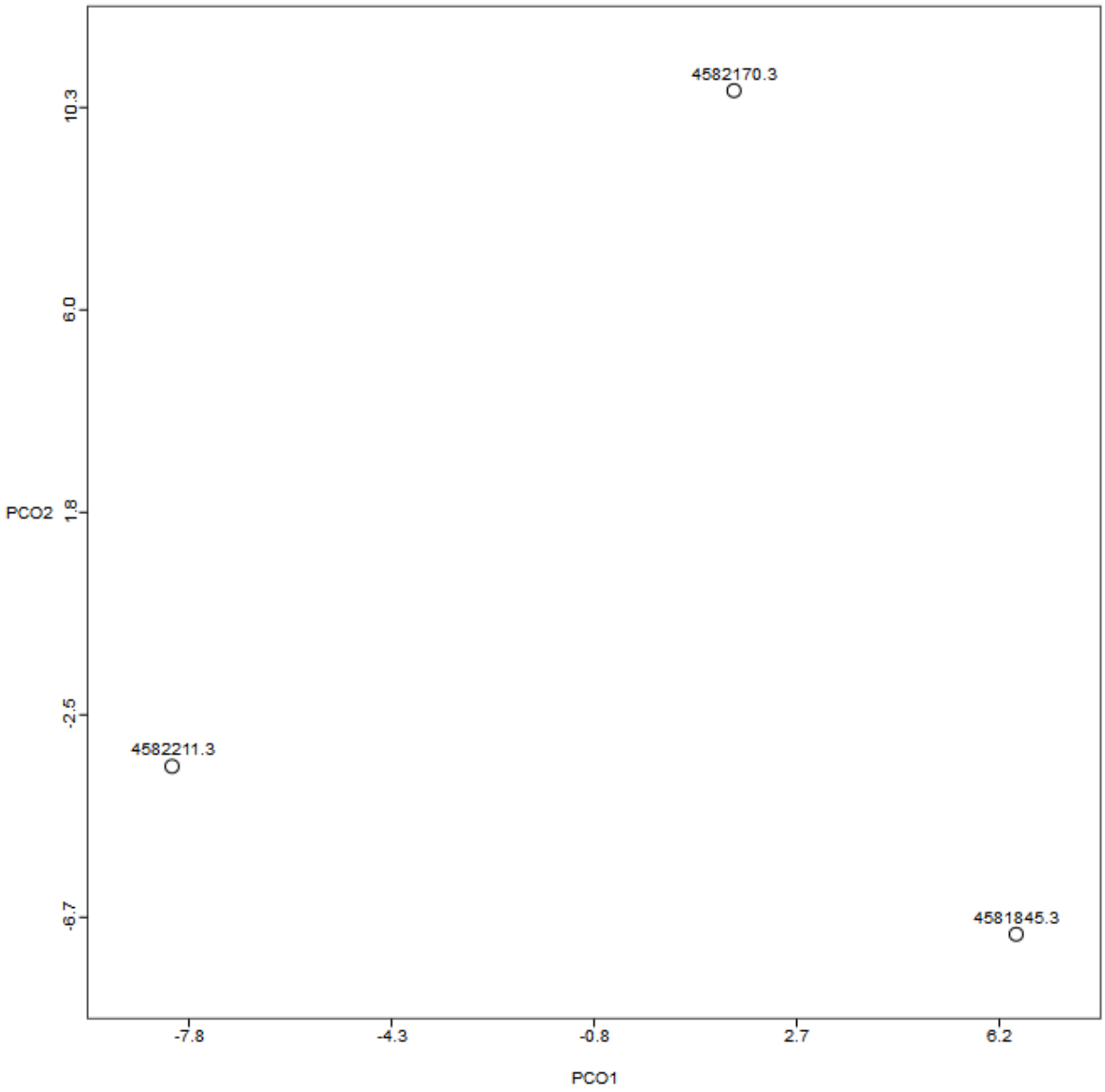

Figure 17. A 16S paired-end Bray-Curtis PCOA analysis of synbiotic diet mouse cecum microbiome content of samples C (MG-RAST accession 4581845.3), F (4582170.3), and L (4582211.3). Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases), and having normalized data values between 0 and 1. PCO1 is listed on the x-axis and has an r^2 value of 0.61386 while PCO2 is represented on the y-axis and has an r^2 value of .38614.

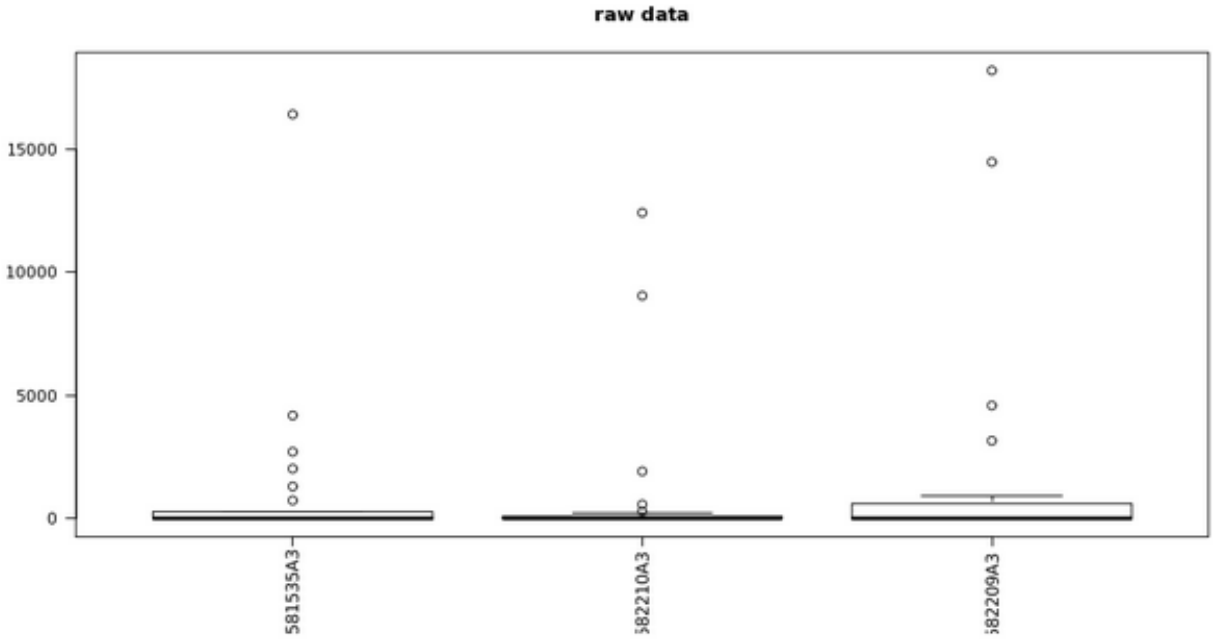

Figure 18. Raw abundance counts of control diet mouse cecum samples B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3) based on the PCOA analysis of figure 18.

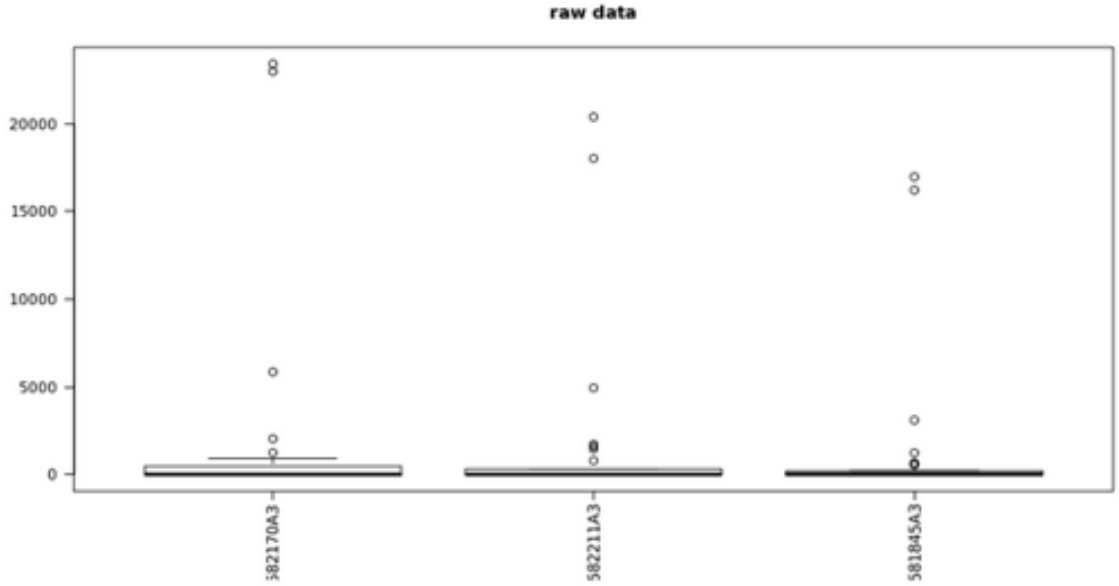

Figure 19. Raw abundance counts for synbiotic diet mouse cecum microbiome content of samples C (MG-RAST accession 4581845.3), F (4582170.3), and L (4582211.3).

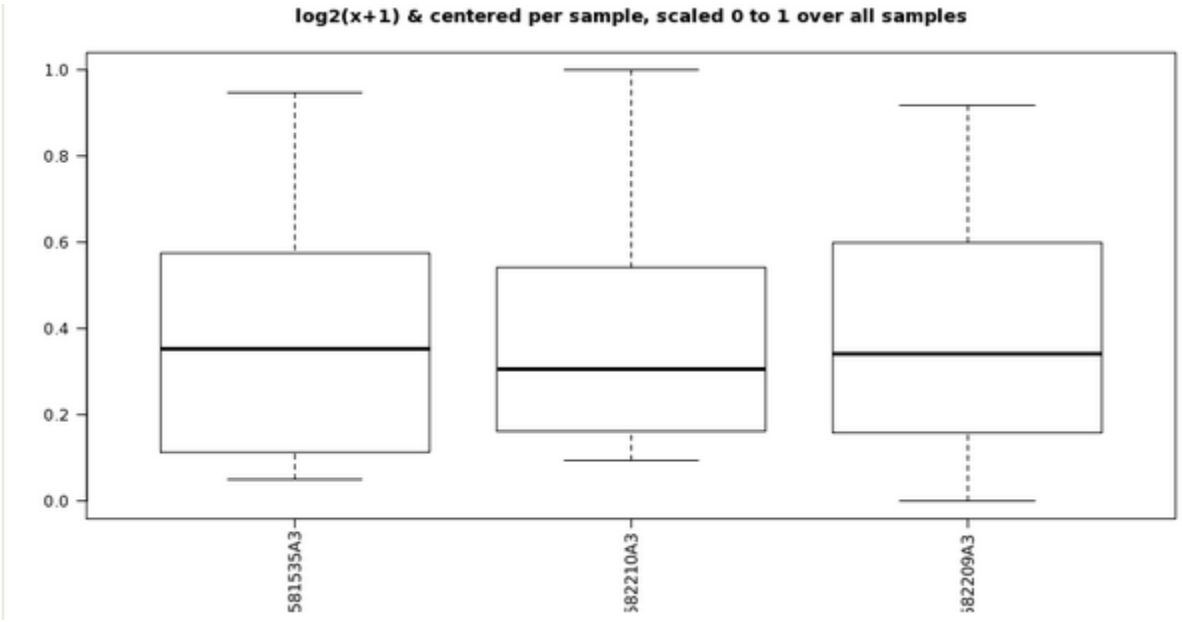

Figure 20. MG-RAST normalized abundance counts of control diet mouse cecum samples B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3).

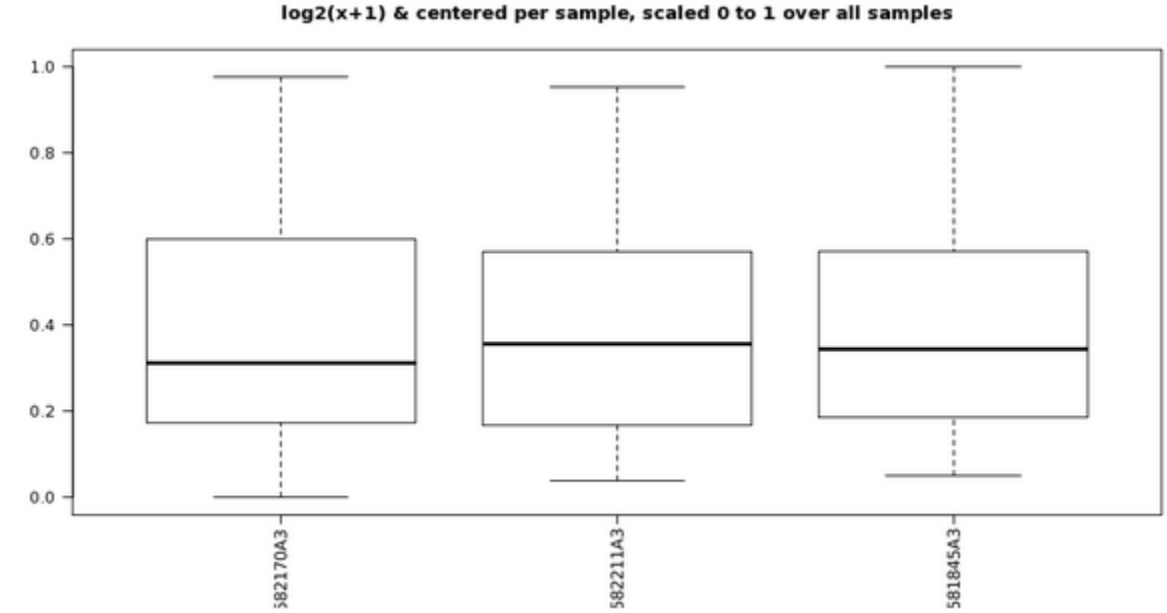

Figure 21. Normalized abundance counts for synbiotic diet mouse cecum microbiome content of samples C (MG-RAST accession 4581845.3), F (4582170.3), and L (4582211.3).

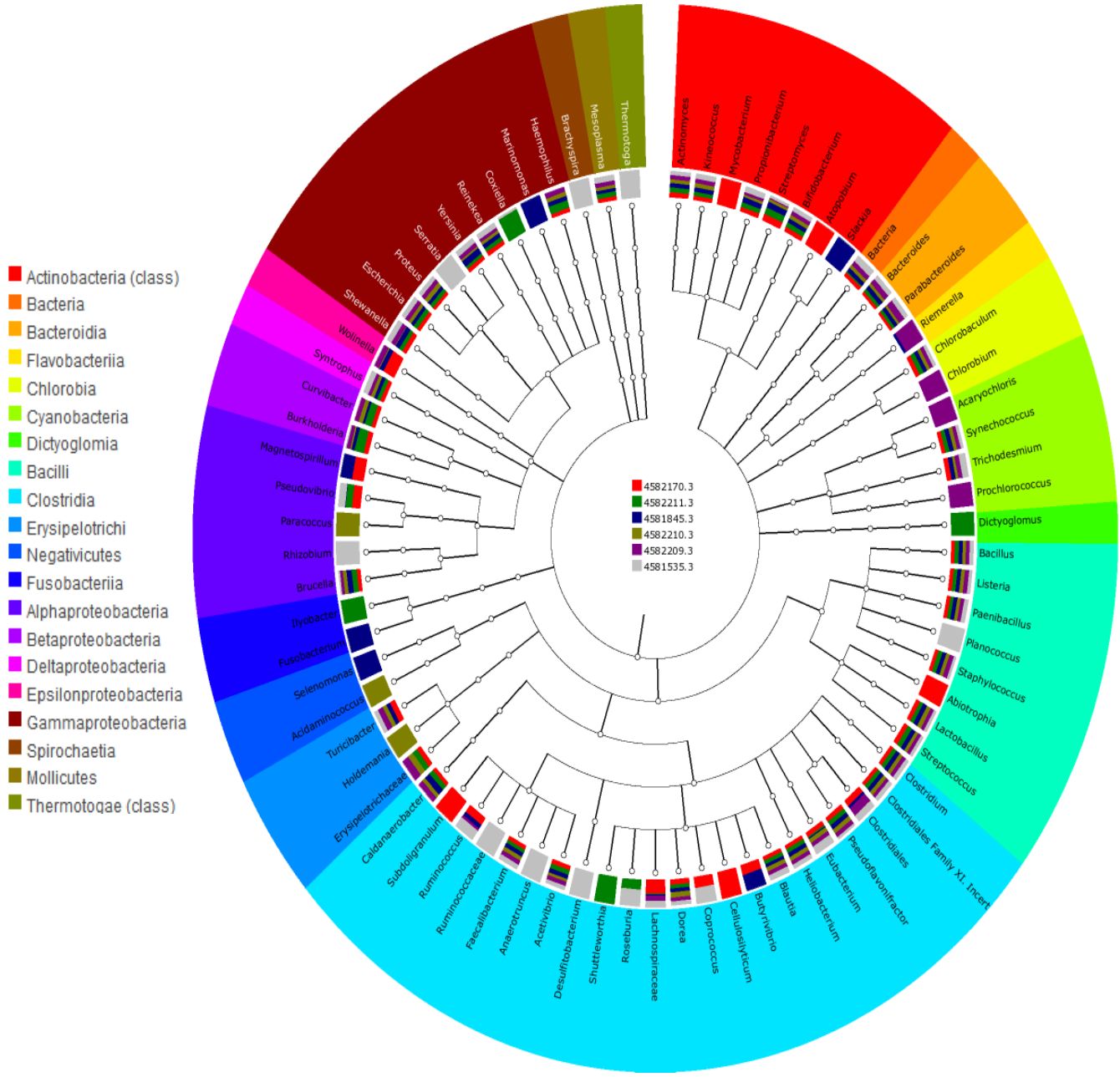

Figure 22. 16S paired-end bacterial phylogenetic tree of cecum microbiome of control diet mice B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3) and synbiotic C (MG-RAST accession 4581845.3), F (4582170.3), and L (4582211.3) having stacked bar leaf weights, a genus maximum level, and coloring according to class**.** Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases). Order membership was indicated by the color of the species names.

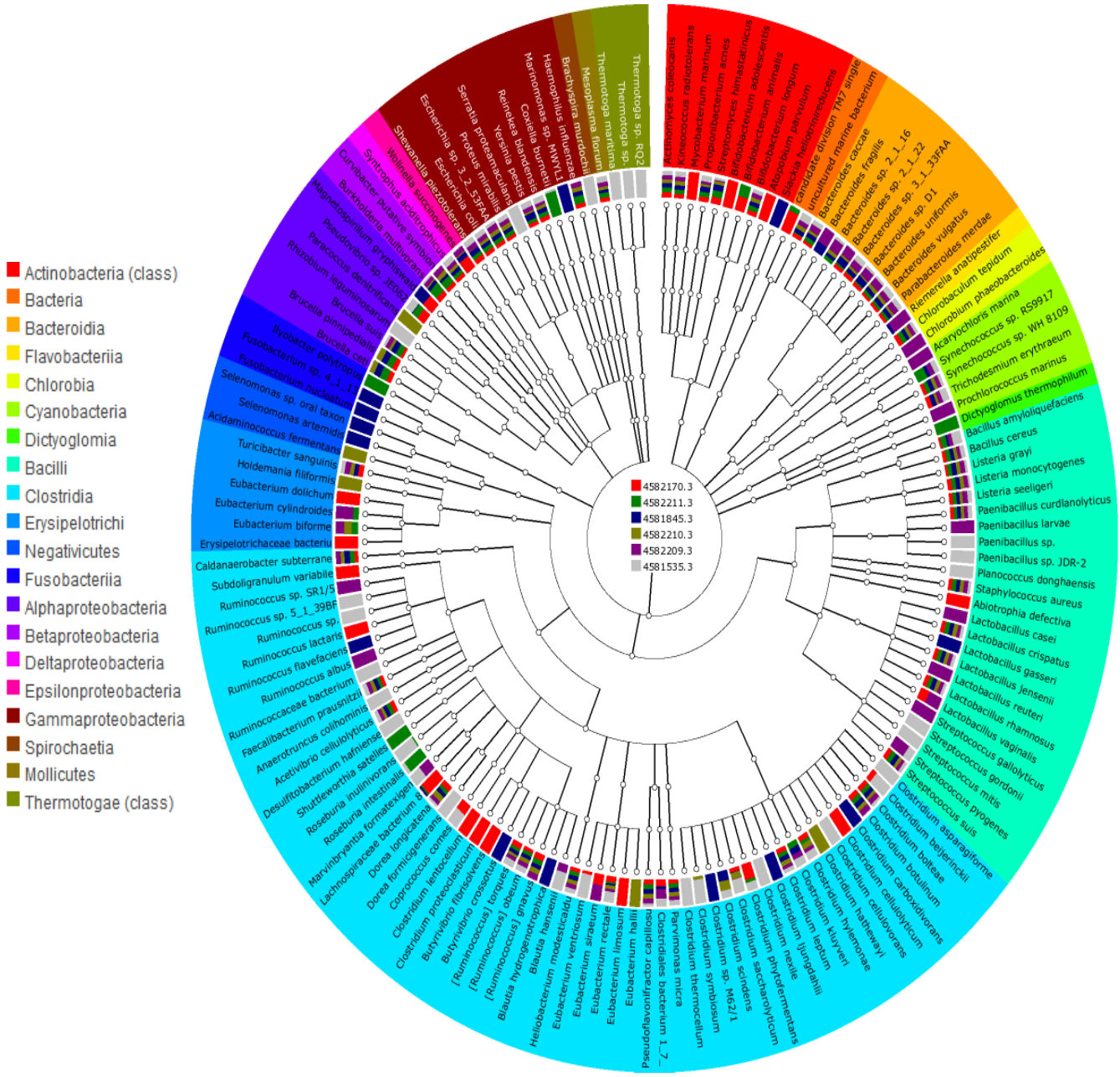

Figure 23. 16S paired-end bacterial phylogenetic tree of cecum microbiome of control diet samples B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3) and synbiotic C (MG-RAST accession 4581845.3), F (4582170.3), and L (4582211.3) having stacked bar leaf weights, a species maximum level, and coloring according to class**.** Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases). Order membership was indicated by the color of the species names.

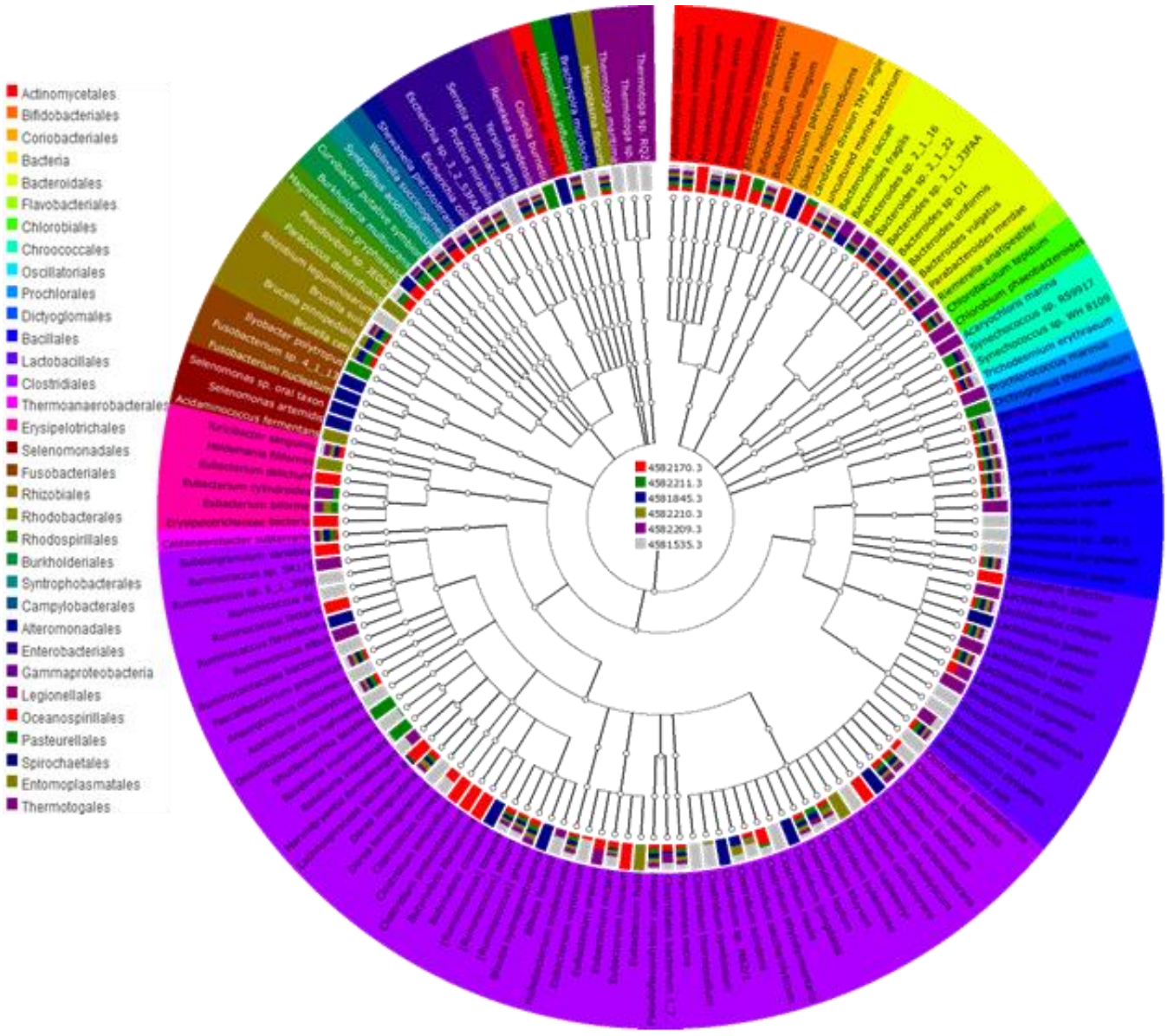

Figure 24. 16S paired-end bacterial phylogenetic tree of cecum microbiome of control diet mice B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3) and synbiotic samples C (4581845.3), F (4582170.3), and L (4582211.3) having stacked bar leaf weights, a species maximum level, and coloring according to order**.** Data was compared to M5NR according to a maximum e-value of 1e-5, 60% minimum identity, and an alignment minimum length of 15 (amino acid measurement according to protein and bp in RNA databases). Order membership was indicated by the color of the species names.

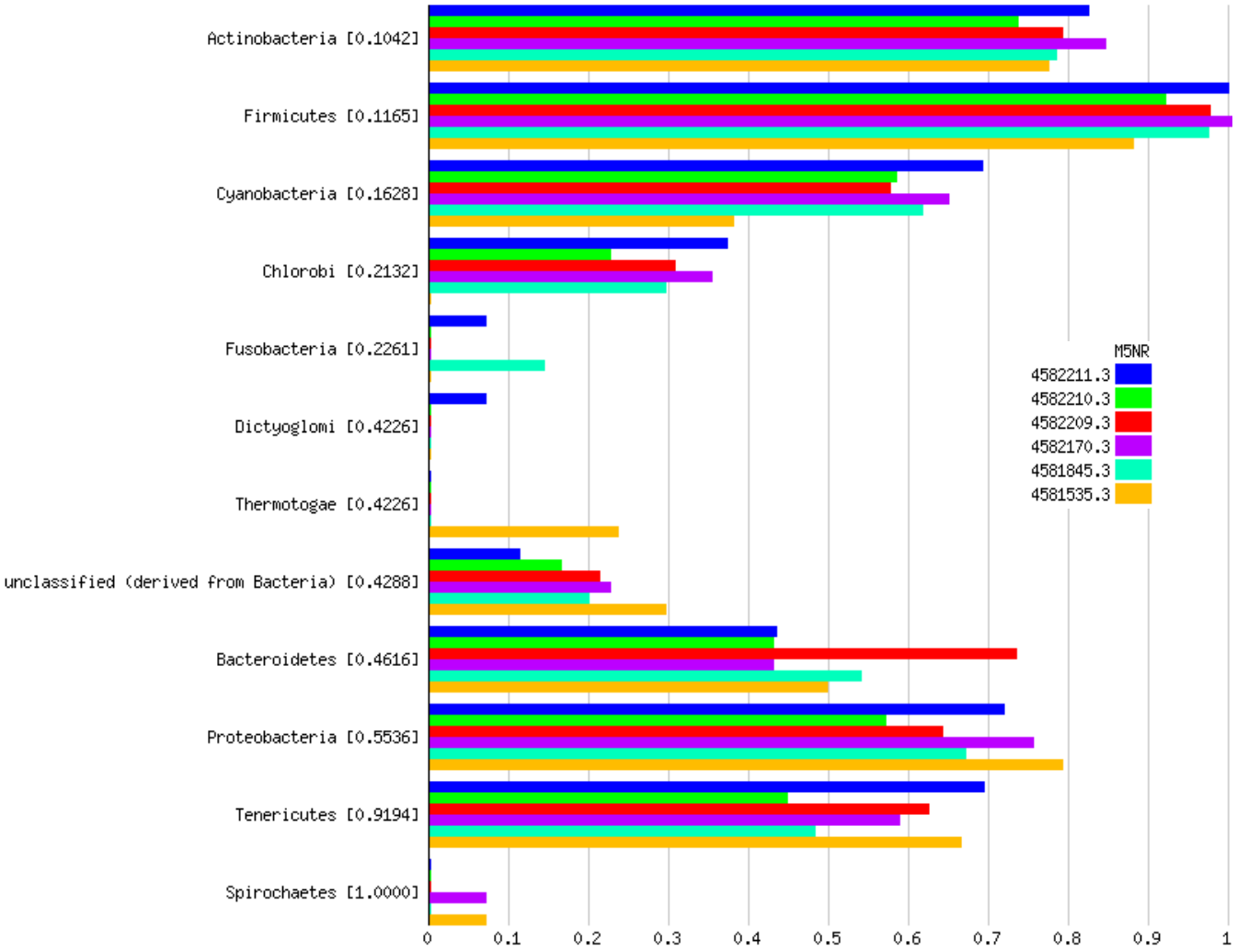

Figure 25. A normalized bar chart of control diet samples B (MG-RAST accession 4581535.3), J (4582209.3), and H (4582210.3) and synbiotic diet C (MG-RAST accession 4581845.3), F (4582170.3), and L (4582211.3) for detected bacteria in addition to the corresponding p-values according to diet groupings listed in brackets. M5NR was compared to the data with e-value maximum of 1e-5, 60% minimum identity, minimum alignment length of 15bp in RNA databases.

CHAPTER 4

Future Directions

The purpose of this project was to analyze and determine any phylogenetic or diversity differences in mouse cecum 16S rRNA genomic DNA samples that were fed a synbiotic diet comprised of probiotics and prebiotics when compared to a control diet. This research showed preliminary diversity data based on culture-independent AMMI models and diversity indices obtained by terminal restriction length fragment polymorphism profile data that were later more acutely verified by next generation sequencing technologies utilizing the Illumina MiSeq platform for analysis of the 16S V3/V4 region of the same cecum samples. By said methodologies, it was concluded that the synbiotic composition utilized did indeed colonize, and affect the microbial community structure of the cecum while maintaining homeostatic balance and diversity.

The verification of preliminary T-RFLP results by differing analysis methodology was required due to the inherent biases apparent within sample preparation, lack of universality of available primers, and errors resulting from T-RFLP analysis algorithms for the resulting data. Despite the error resulting from innate biases within next-gen sequencing, its high-throughput, high resolution, and robust nature in addition to an integration of a quality control pipeline make the process more reliable and representative of measured diversity and taxonomic assignment for the exponentially expanding prokaryotic sequence databases when compared to that of T-RFLP analyses. However, this analysis is still not exhaustive of the potential diversity present within the samples analyzed, as results are dependent on quality and completeness of available data bases

and computing capacity of software packages and algorithms for sequence annotation. As technologies improve, the variables apparent will become less influential.

Although direct analysis of human host cecum is the most ideal to measure the effectiveness of synbiotic treatments as a therapeutic option to maintain a state of host health and combat dysbiosis, the numerous variables present in the human model confound results, making the mouse model applicable for further research. Additional samples could be integrated to verify the statistical significance of the results obtained in the present study.

Although the synbiotic utilized seemed to colonize and affect the relative number of bacteria within the cecum, further research in additional synbiotic combinations of prebiotics and probiotic could be beneficial in deriving the most effective treatment. It would also be useful to test various combinations on differing degrees of dysbiosis and disease states or other host variables in order to test the rang of efficiency and optimal activity of the combination utilized on endogenous host gut flora. Integration of metabolomic studies to elucidate the metabolic profiles of the bacteria can provide insights in the potential metabolic activities of the gut microbial flora. The data resulting from these studies could be integrated to current treatment options in addition to the development of new, more effective means.

References:

1. Abdo, Z., Schuette, U.M.E., Bent, S. J., Williams, C. J., Forney, L. J., and P. Joyce. (2006) Statistical methods for characterizing diversity of microbial communities by analysis of terminal restriction fragment length polymorphisms of 16S rRNA genes. *Environmental Microbiology*. 8:929-938.
2. Applied Biosystems. (2005) Application Note: TRFLP on the 3130/3730. Terminal Fragment Length Polymorphism (T-RFLP) Analysis on Applied Biosystems Capillary Electrophoresis systems.
3. Bartram, A. K., Lynch, M. D., Stearns, J. C., Moreno-Hagelsieb, G., and J. D. Neufeld. (2011) Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Applied and Environmental Microbiology*. 77(11)3846-3852. DOI: 10.1128/AEM.02772-10.
4. Bäckhed, F., Ding, H., Wang, T., Hooper, L. V., Koh, G. Y., et al. (2004) The gut microbiota as an environmental factor that regulates fat storage. *Proc Natl Acad Sci.* 101:15718-15723.
5. Blackwood, C.B., Marsh, T., Kim, S., and E. A. Paul. (2003) Terminal Restriction Fragment Length Polymorphism Data Analysis for Quantitative Comparison of Microbial Communities. *Applied and Environmental Microbiology*. 69:926-932.
6. Borody, T. J., and A. Khoruts. (2012) Fecal microbiota transplantation and emerging applications. *Nat Rev Gastroenterol Hepatol*. 9:88-96.
7. Brook, I. (2005) The role of bacterial interference in otitis, sinusitis and tonsillitis. *Otolaryngol Head Neck Surg*. 133:139-146.

8. Cammarota, G., Ianiro, G., Bibbò, S., and A. Gasbarrini. (2014) Gut microbiota modulation: probiotics, antibiotics or fecal microbiota transplantation? Internal and Emergency Medicine. 9:365-373.
9. Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N., and R. Knight. (2010) Global Patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *PNAS*. 108:4516-4522.
10. Chang, J. Y., Antonopoulos, D. A., Karla, A., et al. (2008) Decreased diversity of fecal microbiome in recurrent Clostridium difficile-associated diarrhea. *J Infect Dis.* 197:435-438.
11. Chen, J., He, X., and J. Huang. (2014) Diet effects in gut microbiome and obesity. *Journal of Food and Science*. Doi: 10.1111/1750-3841.12397
12. Clarke, K. R., and R. N. Gorley. (2006) Chapter 8: Diversity Measures, Dominance Curves and other Graphical Analyses. *Primer v6: User Manual/ Tutorial*. Primer-E LTD. Plymouth.
13. Clement, B. G., Kehl, L. E., Debord, K. L., and C. L. Kitts. (1998) Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities. *Journal of Microbiological Methods*. 31:135-142.
14. Costello, E.K., et al. (2009) Bacterial community variation in human body habitats across space and time. *Science*. 326:1694-1697.
15. Cotter, P. D. (2011) Small intestine microbiota. *Curr Opin Gastroenterol*. 27:99-105.

16. Cowan, D.A., Makhalanyane, T. P., Dennis, P. G., and D. W. Hopkins. (2014) Microbial Ecology and Biogeochemistry of Continental Antarctic Soils. *Frontiers in Microbiology*. 5:154. Doi:10.3389/fmicb.201400154
17. Culman, S. W., Bukowski, R., Gauch, H. G., Cadillo-Quiroz, H., and D. H. Buckley. (2009). T-REX: Software for the Processing and Analysis of T-RFLP data. *BMC Bioinformatics*. 10:171.
18. Culman, S. W., Gauch, H. G., Blackwood, C. B., and J. E. Thies. (2008) Analysis of T-RFLP data using analysis of variance and ordination methods: A comparative study. *Journal of Microbiological Methods*. 75 55-63.
19. Curtis, T. P., and W. T. Sloan. (2004) Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. *Ecology and Industrial Microbiology*. 7:221-226.
20. de Wit, N., Derrien, M., Bosch-Vermeulen, H., Oosterink, E., Keshtkar, S., et al. (2012) Saturated fat stimulates obesity and hepatic steatosis and affects gut microbiota composition by an enhanced overflow of dietary fat to the distal intestine. *Am J Physiol Gastrointest Liver Physiol*. 303:589-599.
21. Dethlefsen, L., Huse, S., Sogin, M. L., and D. A. Relman. (2008) The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biol*. 6(**11**): 2383-2400.
22. Dethlefsen, L., McFall-Ngai, M., and D. A. Relman. (2007) an ecological and evolutionary perspective on human-microbe mutalism and disease. *Nature*. 449:811-818.

23. Di Bella. (2013) Fecal microbiota transplantation: the state of the art. Infectious Disease Reports. 13(**5**):43-45.

24. Dos Santos Dias, C. T. and W. J. Krzanowski. (2006) Choosing components in the additive main effect and multiplicative interaction (AMMI) models. *Scientia Agricola*. 63(**2**): 169-175.

25. Druart, C., Neyrinck, A. M, Dewulf, E. M., De Backer, F. C., Possemiers, S., et al. (2013) Implication of fermentable carbohydrates targeting the gut microbiota on conjugated linoleic acid production in high-fat-fed mice. *Br J Nutr*: 1-14.

26. Druart, C., Neyrinck, A. M., Vlaeminck, B., Fievez, V., Cani, P. D., and N. M. Delzenne. (2014) Role of the Lower and Upper Intestine in the Production and Absorption of Gut Microbiota-Derived PUFA Metabolites. *PLoS ONE*. 9(**1**): e87560. Doi: 10.1371/journal.pone.0087560

27. Dunbar, J., Ticknor, L. O., and C. R. Kuske. (2001) Phylogenetic Specificity and Reproducibility and New Method for Analysis of Terminal Restriction Fragment Profiles of 16S rRNA. *Applied and Environmental Microbiology*. 67(**1**):190-197.

28. Edgar, R. C. (2004) Local homology recognition and distance measures in linear time using compressed amino acid alphabets. *Nucleic Acids Res*. 32:380-385.

29. Evans, M., Brown, J., and M. McIntosh. (2002) Isomer-specific effects of conjugated linoleic acid (CLA) on adiposity and lipid metabolism. *J Nutr Biochem*. 13:508.

30. Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Grotman, R. M., and J. Ravel. (2014) An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*. (2)6.

31. Gibson, G. R., and M. B. Roberfroid. (1995) Dietary modulation of the human colonic microbiota: introducing the concept of Prebiotics. *The Journal of Nutrition*. 125: 1401-1412.
32. Guarner, F. (2006) Enteric flora in health and disease. *Digestion*. 73:5-12.
33. Hayashi, H., Takahashi, R., Nishi, T., Sakamoto, M., and Y. Benno. (2005) Molecular analysis of jejunal, ileal, caecal and recto-sigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *J Med Microbiol*. 54:1093-1101.
34. Hilton, A., and R. A. Armstrong. (2006) Statnote 6: one-way analysis of variance (ANOVA). *Microbiologist*. (pp33-37).
35. Hilton, A., & Armstrong, R. A. (2006) Statnote 6: post-hoc ANOVA tests. *Microbiologist*. (pp. 39-44).
36. Hooper, L. V. (2004) Bacterial contributions to mammalian gut development. *Trends Microbiol*. 12:129-134.
37. Imaizumi, K., Akutsu, T., Miyasaka, S., and M. Yoshino. (2007) Development of species identification tests targeting the 16S ribosomal RNA coding region in mitochondrial DNA. International Journal of Legal Medicine. 121:184-191.
38. Jernberg, C., Lofmark, S., Edlund, C., and J. K. Jansson. (2007) Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *ISME J.* 1:56-66.
39. Khoruts, A., Dicksved, J., Jansson, J. K., and M. J. Sadowsky. (2010) Changes in the composition of the human fecal microbiome after bacteriotherapy for

recurrent Clostridium difficile-associated diarrhea. *J Clin Gastroenterol*. 44:354-360.

40. Kitts, C. L. (2001) Terminal Restriction Fragment Patters: A Tool for Comparing Microbial Communities and Assessing Community Dynamics. *Biological Sciences*:69.
41. Klemashevich, C., Wu, C., Howsmon, D., Howsmon, D., Alaniz, R. C., Lee, K., and A. Jayaraman. (2014). Rational identification of diet-derived postbiotics for improving intestinal microbiota function. *Current Opinion in Biotechnology*. 26:85-90.
42. Konstantinidis, K. T., and J. M. Tiedje, (2005) Genomic insights that advance the species definition for prokaryotes. *Proc Natl Acad Sci USA.* 102:2567-2572.
43. Kozich, J. L., Westcott, S. L., Baxter, N. T., Highlander, S. K, and P. D. Schloss. (2013) Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*. 17(**79**):5112-5120.
44. Langille, G. I., M., Meehan, C. J., Koenig, J. E., Dhanani, A. S., Rose, R. A., Howlett, S. E., and R. G. Beiko. (2014) Microbial shifts in the aging mouse gut. *Microbiome*. 2:50.
45. Lazarevic, V, et al. (2009) Metagenomic study of the oral microbiota by illumine high-throughput sequencing. *J Microbiol Methods.* 79:266-271.
46. Lee, J. C., and R. R. Gutell. (2012) A comparison of the crystal structure of eukaryotic and bacterial SSU ribosomal RNAs reveals common structural features in the hypervariable regions. *PLoS ONE*. 5(**7**):1-6.

47. Leser and Molbak. (2009) Better living through microbial action: the benefits of the mammalian gastrointestinal microbioata on the host. *Environ. Microbial*. 11:2194-2206.

48. Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D., et al. (2005) Obesity alters gut microbial ecology. *Proc Natl Acad Sci USA.* 102: 11070-11075.

49. Ley, R. E., Peterson, D. A., and J. I. GordonI. (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell*. 124:837-848.

50. Li, M., Wang, B., Zhang, M., Rantalainen, M., Wang, S., et al. (2008) Symbiotic gut microbes modulate human metabolic phenotypes. *Proc Natl Acad Sci USA*. 105:2117-2122.

51. Liu, W. T., Marsh, T. L., Cheng, H., and L. J. Forney. (1997) Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding the 16S rRNA. *Applied and Environmental Microbiology*. 63(**11**):4516-4522.

52. Liu, W. T., Marsh, T. L., L. J. Forney. (1998). Determination of the microbial diversity of anaerobic-aerobic activated sludge by a novel molecular biological technique. *Water Science and Technology*, *37*(**4**), 417-422.

53. Magrone, T., and E. Jirillo. (2013) The interaction between gut microbiota and age-related changes in immune function and inflammation. *Immunity & Ageing*. 10:31.

54. Mazmanian, S. K., Liu, C. H., Tzianabos, A. O., and D. L. Kasper. (2005) An immunomodulatory molecule of symbiotic bacteria directs maturation of the host immune system. *Cell*. 122:107-118.

55. McGarr, S. E., Ridlon, J. M., and P. B. Hylemon. (2005) Diet, anaerobic bacterial metabolism, and colon cancer: a review of the literature. *J Clin Gastroenterol*. 39: 98-109.

56. Meyer, F., Paarmann, D., D'Souza, M., Olson, R., Glass, E. M., Kubal, M., Paczian, T., Rodriguez, A., Stevens, R., Wilke, A., Wilkening, J., and R. A. Edwards. (2008) The metagenomics RAST server- a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics*. 9:386.

57. Moreno-Indias, I., Cardona, F., Tinahones, and F. J. Queipo-Ortuño. (2014) Impact of the gut microbiota on the development of obesity and type 2 diabetes mellitus. *Frontiers in Microibology*. 5:1-10. doi:10.3389/fmicb.2014.00190

58. Mujico, J. R., Baccan, G. C., Gheorghe, A., Diaz, L. E., and A. Marcos. (2013) Changes in gut microbiota due to supplemented fatty acids in diet-induced obese mice. *Br J Nutr*. 1-10.

59. Nicholson, J. K., Holmes, E., and I. D. Wilson. (2005) Gut microorganisms, mammalian metabolism and personalized health care. *Nat Rev Microbiol*. 3:431-438.

60. Nicholson, JK., Holmes, E., Kinross, J., Burcelin, R., Gibson, G., et al. (2012) Host-gut microbiota metabolic interactions. *Science*. 336:1262-1267.

61. Niot, I., Poirier, H., Tran, T. T., and P. Besnard. (2009) Intestinal absorption of long-chain fatty acids: evidence and uncertainties. *Prog Lipid Res*. 48:101-115.

62. Noverr, M. C., and G. B. Huffnagle. (2005) The microflora hypothesis' of allergic responses. *Clin Exp Allergy*. 35:1511-1520.

63. O'Hara, A. M., and F. Shanahan. (2006) The gut flora as a forgotten organ. *EMBO Rep*. 7:688-693.

64. O'Shea, E. F., Cotter, P. D., Stanton, C., Ross, R. P., and C. Hill. (2012) Production of bioactive substances by intestinal bacteria as a basis for explaining probiotic mechanisms: bacteriocins and conjugated linoleic acid. *Int J Food Microbiol*. 152: 189-205.

65. Raju, B. M. K. (2002) A study on AMMI model and its biplots. *Journal of the Indian Society of Agricultural Statistics*. 55(**3**):297-322.

66. Rankoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and R. Medzhitov. (2004) Recognition of commensal microflora by toll-like receptors is required for intestinal homeostasis. *Cell*. 118:229-241.

67. Reddy, B. V. B., Kallifidas, D., Kim, J. H., Harlop-Powers, Z., Feng, Z., and S. F. Brady. (2012) Natural Product Biosynthetic Gene Diversity in Geographically Distinct Soil Microbiomes. *Applied and Environmental Microbiology*. 78(**10**):3744-3752.

68. Rees, G. N., Baldwin, D. S., Watson, G. O., Perryman, S., and D. L. Nielsen. (2004) Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: application in multivariate statistics. *Antonie van Leeuwenhoek*. 86:339-347.

69. Reid, G., and A.W. Bruce. (2006) Probiotics to prevent urinary tract infections: the rational and evidence. *World J Urol*. 24:28-32.

70. Sadeghi, S. M., Samizadeh, H. Amiri, E., and M. Ashouri. (2011) Additive main effects and multiplicative interactions (AMMI) analysis of dry leaf yield in tobacco hybrids across environments. *African Journal of Biotechnology*. 10(**21**):4358-4364.

71. Salnier, D. M., Ringel, Y., Heyman, M. B., Foster, J. A., Bercik, P., Shulman, R. J., Versalovic, J., Verdu, E. F., Dinan, T. G., Hecht, G., and F. Guarner. (2013) The intestinal microbiome, probiotics and prebiotics in neurogastroenterology. *Gut Microbes*. 1(**4**):17-27.

72. Sartor, R. B. (2008) Microbial influences in inflammatory bowel diseases. *Gastroenterlogy*. 134:577-594.

73. Schötte, U. M. E., Abdo, Z., Bent, S. J., Shyu, C., Williams, C. J., Pierson, J. D., and L. J. Forney. (2008) Advances in the use of terminal restriction fragment length polymorphism (T-RFLP) analysis of 16S rRNA genes to characterize microbial communities. *Applied Microbiology and Biotechnology*. 80:365-380.

74. Schloss, P. D. (2008) Evaluating different approaches that test whether microbial communities have the same structure. *The International Society for Microbial Ecology Journal*. 2:265-275.

75. Schloss, P. D. (2009) A high-throughput DNA sequence aligner for microbial ecological studies. *PLoS ONE*. 12(**4**):e8230.

76. Schloss, P. D. (2010) The effects of alignment quality, distance calculation method, sequence filtering, and region on the analysis of 16S rRNA gene-based studies. *PLoS Computational Biology*. 7(**6**):e1000844.

77. Schloss, P. D., and S. L. Westcott. (2011) Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. *Applied and Environmental Microbiology*. 10(**77**):3219-3226.

78. Schloss, P. D., Gevers, D., and S. L. Westcott. (2011) Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE*. 6(**12**):e27310.

79. Servin, A. L. (2004) Antagonistic activities of lactobacilli and bifidobacteria against microbial pathogens. *FEMS Microbiol Rev*. 28:405-440.

80. Smith, C. J., Danilowicz, B. S., Clear, A. K., Costello, F. J., Wilson, B., and W. G. Meijer. (2005) T-Align, a web-based tool for comparison of multiple terminal restriction fragment length polymorphism profiles. *FEMS Microbiology Ecology*. 54:375-380.

81. Stres, B. (2006) The first decade of terminal restriction fragment length polymorphism (T-RFLP) in microbial ecology. *Acta agriculturae Slovenica*. 88:65-73.

82. Szentkuti, L., and K. Lorenz. (1995) The thickness of the mucus layer in different segments of the rat intestine. *Histochem J*. 27:466-472.

83. Tringe, S. G., and P. Hugenholtz. (2008) A renaissance for the pioneering 16S rRNA gene. *Curr Opin Microbiol*. 11:442-446.

84. Walsh, C. J., Guinane, C. M., O'Toole, P. W., and P. D. Cotter. (2004) Beneficial modulation of the gut microbiota. *FEBS lett*. 588(**22**):4120-4130.

85. Werner, J. J., Zhou, D., Caporaso, J. G., Knight, R., and L. T. Angenent. (2012) Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gne amplicon surveys. *The International Society for Microbial Ecology Journal*. 6:1273-1276.

86. Zhang, X., Zeng, B., Liu, Z., Liao, Z., Li, W., Wei, H., and X. Fang. (2014) Comparative diversity of gut microbiota in two different human flora-associated Mouse strains. *Current Microbiology*.69:365-373.

87. Zoetendal, E. G., Raes, J., van den Bogert, B., Arumagam, M., Booijink, C. C., et al. (2012) The human small intestinal microbiota is driven by rapid intake and conversion of simple carbohydrates. *ISME J*. 6:1415-1426.