THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF VETERINARY AND BIOMEDICAL SCIENCES

ACTIVATION OF THE ARYL HYDROCARBON RECEPTOR MODULATES THE GUT MICROBIOME

ROBERT NICHOLS
Spring 2014

A thesis
submitted in partial fulfillment
of the requirements
for baccalaureate degrees
in Immunology and Infectious Disease and Toxicology
with interdisciplinary honors in Immunology and Infectious Disease and Toxicology

Reviewed and approved* by the following:

Andrew Patterson
Assistant Professor of Molecular Toxicology
Thesis Supervisor

Gary Perdew
Director of the Center for Molecular Toxicology and Carcinogenesis
John T. and Paige S. Smith Professorship in Agricultural Sciences
Thesis Supervisor

Pamela Hankey
Professor of Immunology
Honors Adviser

* Signatures are on file in the Schreyer Honors College.
of the requirements

ABSTRACT

2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) is an environmental concern. It is extremely persistent so it can stay in the environment for a long time. TCDD binds the aryl hydrocarbon receptor and causes many different effects including cytochrome p450 activation, induction of phase two detoxifying enzymes and immune cell regulation. Many of these effects are associated with inflammation and inflammatory signaling. The gut microbiome is an integral part of normal physiology and regulates many different processes like complete food digestion, protection from harmful bacteria, insulin sensation, satiety, and even obesity prevention. The gut can also have a large role on inflammation, either attenuating it or inducing it. In this study, I explored the idea that TCDD affects the gut microbiome and this can lead to a state of inflammation. We used cecal DNA which was then sequenced at the Penn State Genomics Facility with 16S rRNA gene specific primers. The sequenced 16S rRNA was then analyzed on a terminal-based software called Mothur to obtain taxonomic data. The taxonomic data was then used to quantitatively and qualitatively determine if certain species increased or decreased after treatment with 2,3,7,8-tetrachlorodibenzofuran (TCDF). TCDF was used because it is a less toxic and more rapidly cleared chemical compared to TCDD. Overall 2 major phyla, a family and a species were examined: *Bacteriodetes* increased after treatment, *Firmicutes* decreased after treatment, *Clostridia* decreased after treatment and *Oscillibactor* decreased after treatment. Along with an analysis of the major results, an in-depth guide of 16S rRNA gene analysis is discussed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First off I would like to thank my honors advisor Dr. Andrew Patterson for taking me into his lab and letting me work on my thesis with him. Thank you for guiding me along the right path and giving me insight when I needed it. Thank you for trusting me with this project and giving me time to learn the new methods. Also I would like to thank him and my whole lab for putting up with my awful hand writing. I also would like to thank Dr. Limin Zhang who I worked with most of the time. Thank you for letting me share this project with you. Also I would like to thank my honors advisor Dr. Pamela Hankey. Thank you for putting up with my countless questions about what classes to take and if certain classes could count for other ones.

I would next like to thank Dr. Gary Perdew for being part of my thesis committee. Thank you for taking the time out of your busy schedule and research to go over my thesis. Thank you to Dr. Istvan Albert and Aswathy Sebastian over at the Bioinformatics Consulting Center. Thank you for helping me learn two new languages of code and walking me through the 16S rDNA gene analysis.

I would also like to thank the other members of my lab; Jared Correll, Kerry Belton, Jessica Montanez, Chris Chiaro and Phil Smith. Thank you for putting up with my countless number of questions of where things are, if I am doing this protocol right, and if Limin's standards are here yet. I would also like to thank them for being extremely helpful and very enjoyable, making my time in the lab unforgettable.

I would also like to thank all my friends for keeping me sane while I was writing this thesis. Also I would like to thank them for keeping me on track while I wrote this and bothering me about writing it. Along with my friends, I would like to thank my family who are always supportive and jokingly told me "The lab never closes" every time I said I was working on my thesis. Thank you for always putting a smile on my face.

Lastly I would like to thank all the professors here at Penn State for doing a tremendous job on imparting all the information you have given to me. Without teachers and professors there would be no learning, so I thank you for taking time out of your research to help us students, who strive for knowledge, to learn. Without you I have no idea where myself or anyone else at this university would be. Thank you, and thank you again to everyone I mentioned above.

Chapter 1 :LITERATURE REVIEW

1.1 The Microbiome and Its Functions

1.1.1 Overview of the Microbiome

Over the last 5 to 10 years interest in the human microbiome has grown. In fact according to Serkirov and his colleagues, between 1990 and 2009, there has been "a near fivefold increase in the yearly publication rate" on papers talking about the microbiome (2). Also between 2009 and 2010 there was a 28% increase in these publications (2). The human body needs the microbiome for myriad different activities, such as food digestion, immune system support and protection of the gut from harmful bacteria (1). Over 1 million species of bacteria are present in human microbiomes but to date little is known about how they interact with each other and how different changes in the body can change the microbiome (1). Also every person has a slightly different microbiome composition, and it is not known at what composition the microbiome becomes unhealthy for the human (1). However now that scientists have genomic, proteomic and metabolomic methods, the microbiome can be looked at much more in depth (2). One of the major ways this is done is via 16S rRNA gene analysis (2). With these methods researchers are able to see how diverse each individual gut microbiome is and all of the key functions the species do in the gut (2). This is discussed in the methods section of the thesis.

There are several sections that make up the human gut which include: the mouth, oesophagus, stomach, and intestines (2). The amount of diverse species increases as one moves from the oral cavity to the lower colon, giving the colon the most amount of bacteria (2). When analyzing newborn babies' development of the microbiome, it was discovered that F*irmicutes* dominated the early gut but was soon overtaken by *Bacteroidetes* and these two phyla made up the majority of the early gut microbiota (2). In fact in a recent study it was seen that these two species make up approximately 97% of the total gut bacteria in humans (2). However, there was a variability of distribution between *Firmicutes* and *Bacteroidetes.* At one extreme *Bacteroidetes* dominated with almost 90% concentration, but at the other extreme *Firmicutes* dominated with 95% (2). This massive diversity at the phylum level means that looking for any trend at the

species level will be very difficult (2). Certain trends were seen with different age groups and a significant difference between the elderly and the young was also measured (2). After doing 16s rRNA gene sequencing it was seen that there were two species that were found in almost all 210 samples and they were both from *Firmicutes* in the family *Lachnospiraceae (2)*. In only two years research has taken the idea of a core group of bacterial species and made these findings much more translatable to human health and disease.

This reality was achieved with the introduction of quantitative methods and it was seen that there is a dominating core of bacteria (3). Like with the other studies it was seen that *Firmicutes* and *Bacteroidetes* dominated the population (3). Even with the limitations in current day analysis this group found that there was several stable species of bacteria that was seen at similar levels in all three subjects (3). These species were *Bifidobacterium longum, Bacteroides vulgatus, Bacteroides putredinis* and *Odoribacter splanchnicus* just to name a few (3). Throughout the yearlong study it was seen that these and all other species fluctuated, which makes it extremely hard to nail down a stable core (3). That being said it has also been seen that major changes in the microbiota can lead to diseases such as obesity and inflammatory disorders (3).

Inflammatory bowel disease (IBD) is a major problem in the modern world and is one of the most uncomfortable and annoying diseases to date. Recent studies have shown that the microbiome could have a direct effect in causing IBD by interacting with the immune system(7). It has been noted that there was a lower amount of *Firmicutes* in the microbiome in patients that suffered from IBD (7). Most studies look at this decrease in *Firmicutes* as a direct cause of IBD but the methanogens could also have a role in causing these diseases(7). *Methanobrevibacter smithii* and *Methanosphaera stadmanae* have both been seen to activate the tumor necrosis factor alpha (TNF alpha) response in the body(7). *M. stadmanae* has been seen to induce a much higher TNF response(7). *M. stadmanae* was also seen to be more prevalent in patients that suffered from

IBD (7). This theory of methanogens being a major constituent of IBD is still being researched but evidence looks promising that this could be true.

Obesity is a major problem worldwide and it is hypothesized that the microbiota has a direct role in obesity. It was first seen that in mice that are genetically predisposed to develop obesity (ob/ob) when fed the same diet as a wild type mouse, that the ob/ob mice had an increase in *Firmicutes* but a decrease in *Bacteroidetes* populations (4). The wild type mice could be made obese by feeding them a diet with a much higher fat content, which lead to similar microbial changes that were seen in the ob/ob mice (4). In human studies more extreme differences between the two phyla have been observed but some studies show that there is no difference between the two (4). A possibility that has been seen when combining numerous studies is that obesity is less dependent on the phylogenetic changes rather than the metabolic changes that the microbiome causes(4). This means that it is less important which species of bacteria is changing and more important what the new species does. So even if gram negative (*Bacteroidetes)* bacteria is decreasing as a phyla, certain *Bacteroidetes* species that over produce lipopolysaccharide bacteria could be increasing leading to inflammation and obesity. The gut bacteria has been seen to cause an increase in the degradation of food in the ob/ob mice (4). The gut bacterial makeup can be transferred via a fecal transplant to a germ free mouse and similar increases in digestion of food will occur(4). Two major species of bacteria that were found to increase in obese mice were *Bacteroides Thetaiotaomicron* and *Methanobrevibacter smithii*, these species enhanced the digestion of the food but also increased the formation of adipose tissue(4). Another aspect that was tested was if a high fat diet could cause this increase in *Firmicutes* and a decrease in *Bacteroidetes* without the genetic disposition for obesity, and it was seen that the high fat diet did cause the same changes (4). In a study in which germ free mice were inoculated with human fecal matter to make a humanized mouse, a shift in the microbiome was seen after a day when the diet was changed from low fat to high fat (4). This suggests that the human microbiome is directly related to obesity, due to extreme changes in composition (4).

Along with diet, chemicals also could have a major effect on the microbiome and could change the composition, which can lead to numerous issues. These chemicals fall under the category of xenobiotics, which are any foreign chemicals the body is exposed to (5). Xenobiotics are seen in the food we eat, the water that we drink, and the air that we breathe; they are everywhere. Xenobiotics are part of the metabolome, which represents all biological compounds that are made in the millions of reactions that happen in the body on a daily basis (5). The gut microbiota adds unique metabolites, such as short chain fatty acids, to an individual's metabolome and these chemicals can be used to sense disruptions in the normal flora (5). A normal flora is very important to humans because they break down food that humans cannot and they also can help to biotransform xenobiotics (5). The main way the microbiome can detoxify is via sulfation of the chemicals that they encounter(5). Along with the detoxification, the normal gut microbiota can produce certain xenobiotics of its own (5). One of these xenobiotics could be lipopolysaccharide which could increase inflammation in the gut. These can be up regulated or down regulated due to environmental changes, diet changes and even exposure to chemicals.

Normal function of the microbiota starts with the essential degradation of certain polysaccharides bonds that hold the monosaccharides together, such as cellulose (6). The species that live in the *Bacteroidetes* phyla are responsible for digesting cellulose (6). It has also been noted that people that live in rural areas and have a main diet of plants have a much higher amount of *Bacteroidetes* in their guts (6). As mentioned, obesity is accompanied by an increase in the *Firmicutes* and a decrease in the *Bacteroidetes* populations(5). It could be a possibility that a high plant diet could change the microbiome to a predominately *Bacteroidetes* population, which would help reduce obesity progression and possibly reverse it (6). *Bacteroidetes* has also been seen to de-conjugate bile acids which could have an effect on physiology considering that the bile acids are a major subset of signaling molecules (6). Specific species can be used as markers for disease such as a decrease in *Faecalibacterium* is directly related to Crohn's disease (6). Also a high prevalence of sulfur bacteria can lead to an over production of hydrogen sulfide which is

toxic to humans and could be a precursor to colon cancer (6). Another species called *Oxalobacter formigenes* has been seen to be involved in kidney stone formation due to its oxalate degradation capability (6). Currently research is only scratching the surface of the human microbiome and there is still much to be discovered, considering we only know 20% of the microbiome's function (6).

One of the major discoveries in the past century has been the introduction of antibiotics. Antibiotics have saved countless number of lives and continue to be an important medicine today. The problems arise when doctors over prescribe antibiotics for viral infections. This can cause serious problems for the normal microbiome. One of the major problems is that antibiotic therapy can result in antibiotic resistant bacteria (15). These antibiotic resistant bacteria can thrive in the gut a very long time (15). When exposed to antibiotics populations of *Bacteroidetes* decreased while *Firmicutes* increased in the gut (15). The gram negative bacteria during a treatment regimen were significantly decreased after a couple of days (15). Protein expression from the microbiota also decreased when under antibiotic treatment, mainly important proteins used for glycolysis, GTP hydrolysis and iron uptake (15). A major gene category involved in the bacterial immune response was also down regulated in the presence of antibiotics (15). This category is known for extrachromosomal and mobile element functions (15). This category includes all plasmids, and some of these plasmids can give the bacteria antibiotic and immunity resistance. Many more studies are being done on this area of the microbiome and it is a very important topic. Also as more gut bacterial species are being discovered their functions can be determined and these functions could be vital to human health.

1.1.2 TCDD and the Aryl Hydrocarbon Receptor

The aryl hydrocarbon receptor (AHR) is a unique soluble receptor because it involves a DNA binding domain called a basic helix loop helix domain and a Per-Arnt-Ser (PAS) ligand binding domain (26). This receptor has also been seen to be highly conserved between invertebrates and vertebrates which shows that it has a very important biological function (14). AHR is found in the cytoplasm of many different types of cells and is activated when ligands bind and interact with the PAS domain (11). AHR then translocates into the nucleus where it finds its binding partner Arnt (26). The AHR-Arnt complex then finds the appropriate response element on the DNA and induces gene expression (11). The major consequence of AHR binding is the activation of cytochrome p450 1A1 (CYP1A1) phase one detoxifying enzyme (11). AHR transcription factor activates different genes in different tissues and also is highly dependent on the ligand (11). Some other responses that come about when AHR is activated are epidermal growth factor and vascular endothelial growth factor (26). This pathway also involves a negative feedback loop where the AHR/Arnt ligand complex is exported from the nucleus back to the cytoplasm to be recycled and degraded by ubiquitin ligases (14). One of the major pathways involved with AHR signaling is the regulation of the immune system (11). It has been seen that AHR can modulate levels of regulatory T cells which would suppress the immune system (27). It has also been seen to be largely expressed in the gut epithelial barriers and be largely involved in the regulation of the microbiome of the body (12). Various chemicals activate the AHR pathway and they mostly belong to the family of aromatic hydrocarbons (27).

One of these aromatic hydrocarbons, 2,3,7,8-Tetrachlorodibenzo-p-dioxin (TCDD), has been seen to be one of the most potent AHR agonists discovered (11). When exposed to TCDD numerous toxic events occur in the body (8). Some of these include immunosuppression, tumor promotion and serious reproductive defects (8). The main binding of TCDD occurs at the AHR and once activated AHR then affects numerous genes including CYP1A1 (8). The cytochrome

p450s are mainly used to detoxify xenobiotics in the body (8). In rats it has been seen that TCDD induced necrosis and swelling of the liver (8). This is likely due to the fact that sometimes when a cytocrome detoxifies its substrate, it actually makes metabolites that are more harmful to the body, a process call metabolic activation (8). Some of these harmful metabolites include reactive oxygen species or ROS, and these directly affect DNA and other macromolecules in the cell (8). Along with ROS, hydrogen peroxide is also generated from TCDD activation of AHR, which can disrupt proteins and cause harmful effects to the cell (8). TCDD has around a 10 year half- life in the environment which makes it an extreme problem for human and animal populations (13). TCDD can easily cross membranes due to its lipophilic nature (13). Once exposed the half -life of TCDD in humans is only around a year and is excreted via the urine and the feces(13). This long half-life, along with the very high affinity TCDD has for AHR, leads to the major toxic effects that are seen with TCDD exposure (13). In 2004 Victor Yushchenko, the former Ukrainian president, was allegedly poisoned with TCDD (13). It was seen that he actually had levels of 50,000 times the average exposure (13). One of the most notable effects from this exposure was lesions on his face called chloracne (13). Interestingly enough when exposed to TCDD, TCDD is stored in fat tissue, so the more fat tissue that is on a person the higher the resistance to TCDD (11). Also the fact whether or not TCDD is a carcinogen is under debate because it is classified as one by the World Health Organization but some researchers argue on the potential of this idea (11).

Researchers have discovered that when exposed to TCDD, AHR activates Fox3p+ regulatory T cells which are important for suppressing the immune system (10). In mice that have AHR knocked out, it is seen that they have a higher chance of a bacterial infection (12). In one study, researchers saw that when exposed to TCDD, AHR increased regulatory T cells(10). This increase helped suppress a form of encephalitis induced in mice(10). However, when the aryl hydrocarbon receptor was activated by another ligand, it repressed the regulatory T cell response which made the encephalitis worse (10). This means that AHR is a ligand specific receptor. With

these regulatory T cells, it was seen that they were needed in newborns and AHR is a large regulator of this (12). AHR has also been seen to be a necessity to bring massive amounts of T cells to the mucosa (12). Along with that it has been seen to be a regulator of Th17 T cells(10). It is known that mice that are deficient of AHR were deficient in IL22, an important inflammatory mediator(10). It was also seen that IL22 helps regulate microbiota in external sites like the skin, and has a role in wound healing (12). With this fact it was seen that an endogenous ligand for AHR could be cinnabarinic acid (10). Another group of ligands that have recently been discovered in cabbage include indole-3 carboinol and sulforaphane (11). These ligands activate the aryl hydrocarbon receptor and actually have beneficial effects (11). This ligand is used to activate the synthesis of IL 22 in the body(10). These are only some of the viable options for an endogenous ligand for AHR and there are still others that are under investigation.

Along with this T cell regulation AHR is a major player in metabolite detoxification (14). When activated, AHR activates phase one detoxifying enzymes like CYP1A1, CYP1A2 and CYP1B1(14). The phase one detoxification involves the addition of functional groups that make the xenobiotic substance more polar (14). AHR also activates phase two detoxifying enzymes like Glutathione S-Aryltransferase A1(14). Phase two detoxification involves the addition of glutathionyl groups to the functional group added in phase one (14). Overall both of these result in a more polar form of the xenobiotic substance which is more easily excreted (14).

Recently there have been experiments to prove that AHR signaling and circadian rhythm of the body are related (14). Mainly the fact that the circadian rhythm is altered when AHR is activated (14). When humans are exposed to PCB's and other AHR agonists, it disrupts the natural circadian rhythm of the body (14). The effect this activation has on the body is an overall inhibition of responses to day/night changes(14). However this effect has yet to be seen in a lab but it was noticed that when exposed to TCDD, mice induce more Cytocrome P450 1A1 at night (14). It has however been seen that there is a link between AHR and the circadian rhythm but the specifics have not been discovered (14). They have discovered that AHR can activate the

peroxisome proliferator-activated receptor alpha (PPAR alpha) in the liver which then regulates the circadian rhythm (14). One of the major facts is that some ligands, like cyclic adenosine 3'-5' monophosphate (cAMP), that are used with AHR are extremely regulated by circadian rhythm (14). This regulation of the circadian rhythm can lead to some serious disorders or diseases when the aryl hydrocarbon receptor is either over expressed or repressed.

One of the major diseases associated with this dysregulation is IBD which has been discussed briefly in the above chapter on the microbiome. Most patients that have IBD have a major disruption in their circadian rhythm (14). This disease may be associated with a high ratio of tryptophan metabolites to tryptophan (14). The metabolism of tryptophan to its metabolites is partially due to a high activity of AHR which promotes inflammatory factors (14). The major pathway that AHR recruits these inflammatory factors is by the activation of Th17 cells which are predominantly inflammatory (14). Interestingly enough, when exposed to TCDD, mice will have a semi-resistance to Crohn's disease because of the shift from Th17 cells to Treg cells (14). Most people with IBD are obese and this increase in inflammation increases their risk for leptin resistance(14). Leptin resistance comes about when certain over expressed microbiota species repress GLP-1 (25). GLP-1 is made in response to leptin to suppress hunger and obesity (25). So in some cases even with high leptin levels it cannot induce fullness, which could cause obesity (25). The bacteria that cause this inhibition of GLP-1 have yet to be discovered (25). This resistance in leptin causes increased cAMP levels that could cause an increased aryl hydrocarbon receptor activation, leading to IBD (14). Also this increased leptin can cause an increase in cell proliferation and a decrease in apoptosis which is a staple of IBD(14). This however has yet to be proven in the field but the theory is there to support it(14). Another major player in IBD is mast cells which get activated by AHR and increase inflammation(14). There are a few treatments for IBD and they likely alter the pathways involved, like AHR, but it its not know specifically how this occurs(14).

Interestingly enough one of the treatments for these inflammatory diseases is the activation of AHR with certain ligands. There has been research into the activation of AHR by TCDD and it can actually prevent graft verses host disease and decrease the amount of pro inflammatory cytokines(16). TCDD has also been seen to help combat Crohn's disease(16). On that note it was seen that when mice were given 2,4,6-trinitrobenzenesulfonic acid which induces Crohn's disease, TCDD suppressed the outbreak of the disease(16). This all occurs because TCDD induces the expression of regulatory T cells which ultimately decrease inflammation (16). The problems arise when AHR is stimulated to make Th 17 cells which are discussed above and pro inflammatory cells (14). Another, safer group of anti inflammatory antagonist of the aryl hydrocarbon receptor are flavonoids like Quercetin (16). Quercetin is seen in many fruits and vegetables (16). Quercetin inhibited the activation of cytochrome p450 CYP1A1(16). This also reduces the amount of NF-kB which is a potent inflammatory disease(16). It is interesting that TCDD can be such a problem chemical but then again it is seen as being a useful treatment to some serious inflammatory diseases.

Chapter 2: THE EFFECT OF TCDF ON THE ARYL HYDROCARBON RECEPTOR MODULATES THE GUT MICROBIOTA

2.1 Materials and Methods

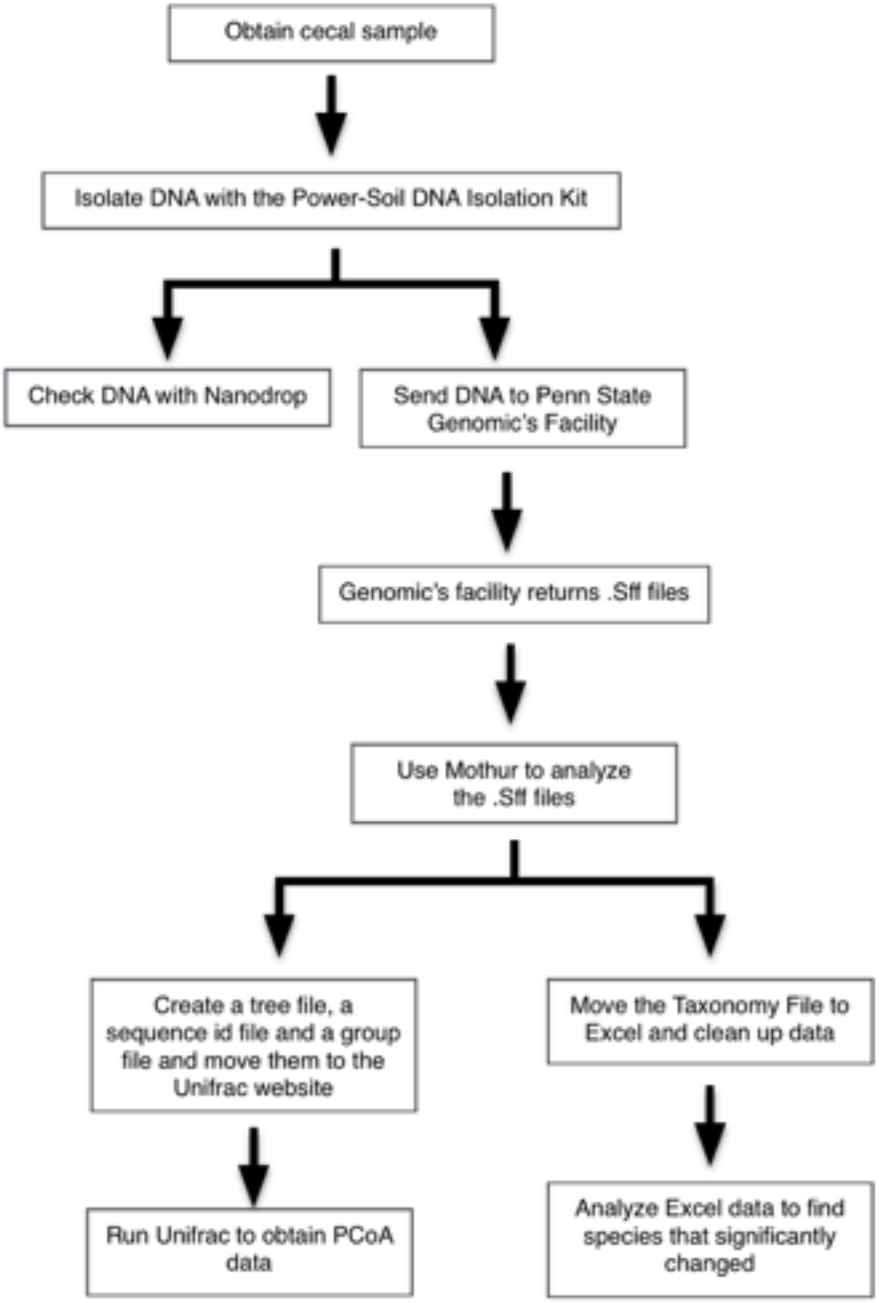


Figure 1: Overview of the Methods Used in This Experiment

2.1.1 DNA isolation

DNA isolation was performed using the MO BIO Laboratories, Inc. Power-Soil DNA Isolation Kit. About 50 mg of cecal matter was used for each sample and placed into a screw top tube with 400 uL of solution C1 and 6 or 7 1mm diameter Zircoma/Silica beads. Samples were homogenized at 6500 rpm for 30 seconds twice using a Bertin Precellys 24 Homogenizor. They were then centrifuged at 10,000xg for 30 seconds at room temperature in an Eppendorf Centrifuge 5430 R. The supernatant was transferred to a new tube with 250 uL of C2. Samples were vortexed and incubated at 4C for 5 minutes with a Denville Scientific Inc. Vortexer 59A and a Benchmark Multi-therm, respectively. They then were centrifuged for one minute at 10,000xg with the Eppendorf Centrifuge 5430 R. Supernatant was then transferred to a new tube with 200 uL of C3. Samples were vortexed, incubated and centrifuged again with a Denville Scientific Inc. Vortexer 59A, a Benchmark Multi-therm, and an Eppendorf Centrifuge 5430 R, respectivly. The supernatant was transferred to a new tube with 1.2 mL of C4 solution. This was vortexed with a Denville Scientific Inc. Vortexer 59A and 675 uL were transferred to a filtered tube. This was spun at 10,000 xg for one minute with an an Eppendorf Centrifuge 5430 R. The flow through was discarded and the next 675 uL of sample was added to the filter tube. It was spun and the flow through was discarded again with an Eppendorf Centrifuge 5430 R. The remaining sample was added, spun and the flow through was discarded a third time with an Eppendorf Centrifuge 5430 R. 500 uL of C5 was added to the filter tubes, centrifuged for 30 seconds at 10,000 xg and the the flow through was discarded with an Eppendorf Centrifuge 5430 R. Filter was transferred to a final tube and 100 uL of C6 was added. This was spun at 10,000xg for 30 seconds with an Eppendorf Centrifuge 5430 R and the spin filter was discarded and the DNA was measured with Nanodrop.

2.1.4 16S rRNA Gene Analysis

The Isolated DNA was sent to the Genomics Facility at Penn State University for sequencing and they returned this file, IDJ8VR304.sff. This is just a unique barcode system that can easily be renamed in terminal. I renamed the file data.sff for clarity. To open and work with this file a special terminal based program is needed and this file is called Mothur (17). This program should be run on an external server so multiple processors can be used, to increase the speed of the commands. The code to open the data is *sffinfo(sff=data.sff, flow=T)*, the flow command brings up flow files to be used. This command will give three files; data.fasta, data.qual and data.flow. I will be using the data.qual and data.fasta for this experiment. The flow files are used to make shhh tagged files that usually lead to a more accurate analysis but occasionally they do not work, it all depends on the data set. The flow files did not add to the accuracy of this data set so the code will be mentioned but was not used. To check on the progress there is a command called summary.seqs and the input is a .fasta file and a .name file. The .fasta file is where all the data is and as this program runs, the name on the .fasta file will get larger and larger. The first summary that should be run has the code *summary.seqs(fasta=data.fasta, processors=8)*, the resulting output should look like **Figure 2**.

Figure 2: Summary After the Sffinfo Command

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	39	39	0	2	1
2.5%-tile:	1	71	71	0	3	7832
25%-tile:	1	425	425	0	5	78318
Median:	1	552	552	0	5	156636
75%-tile:	1	624	624	0	5	234953
97.5%-tile:	1	765	765	1	6	305439
Maximum:	1	1352	1352	33	28	313270
Mean: 1	503.062	503.062	0.0539311		5.08859	
# of Seqs:	313270					

The next step is to run either a flow command or a quality command. The code for the flow command goes as follows: *trim.flows(flow=data.flow, oligos=newbarcodes.txt pdiffs=2, bdiffs=1, processors=2*). The oligos file should be provided by the company that sequences the rRNA, however you may have to make it into a .txt file. The oligos file should include a forward and a reverse primer and bar codes for each sample. As an example my barcodes file looks like this.

forward AGAGTTTGATCMTGGCTCAG

#reverse CCCCGTCAATTCMTTTGAGTTT

barcode TGATACGTCT 202.rt

barcode CATAGTAGTG 203.rt

barcode CGAGAGATAC 301.rt

barcode ATACGACGTA 302.rt

barcode TCACGTACTA 303.rt

barcode CGTCTAGTAC 304.rt

barcode TCTACGTAGC 402.rt

barcode TGTACTACTC 404.rt

barcode TACTCTCGTG 601.rt

barcode TAGAGACGAG 602.rt

barcode TCGTCGCTCG 603.rt

barcode ACATACGCGT 604.rt

The hashtag on the reverse primer makes it invisible to the program because only one primer is needed for the data analysis, if the forward and reverse are used problems may occur. The bdiffs and the pdiffs commands are an input to control the amount of differences between the barcodes and the primers. In this case I put in two differences between the primer sequence in the oligos file and the primer sequence in the data and only one difference between the barcode in the oligos file and the barcode in the data. The whole purpose of this command is to get rid of the barcode

and the primer for each of the sequences in the data. The command to create shhh files is *shhh.flows(file=data.flow.files, processors=8)* and this will usually take about 6 hours to run and was not used in this experiment. The code that I used in this experiment went as followed *trim.seqs(fasta=data.fasta, oligos=newbarcodes.txt, qfile=data.qual, maxambig=0, maxhomop=8, flip=T, bdiffs=1, pdiffs=2, qwindowaverage=35, qwindowsize=50,processors=8)*. Most of the commands are the same from the flow command but the qfile command is the input for the quality file that was obtained from the ssfinfo command. The maxambig is the maximum number of ambiguous sequences there can be and in this case it is zero. The maxhomop is the maximum number of homopolymers and in this case it was 8. The flip command looks at the reverse sequence. The qwindowsize command lets the user determine a window of bases in the sequence that is used for quality control. The qwindowaverage is the minimum average value that is used over the window for quality control. In this case the window is 50 bases and the average is 35, so 35 out of 50 bases have to match up from the sequence to the quality file. At this time it is much easier to run the summary command in this fashion, *summery.seqs(name=current)*. This uses the current .fasta file and the current name file, and this makes finding summaries much easier. The current summary can be seen in **Figure 3**.

Figure 3: Summary After the Trim Command

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	50	50	0	2	1
2.5%-tile:	1	51	51	0	3	6526
25%-tile:	1	92	92	0	4	65253
Median:	1	141	141	0	5	130506
75%-tile:	1	181	181	0	5	195759
97.5%-tile:	1	248	248	0	5	254486
Maximum:	1	443	443	0	7	261011
Mean: 1	140.231	140.231	0	4.42833		
# of Seqs:	261011					

The next group of commands starts with a command called unique.seqs and looks as follows, *unique.seqs(fasta=data.trim.fasta)*. This is used to extract the unique sequences from the

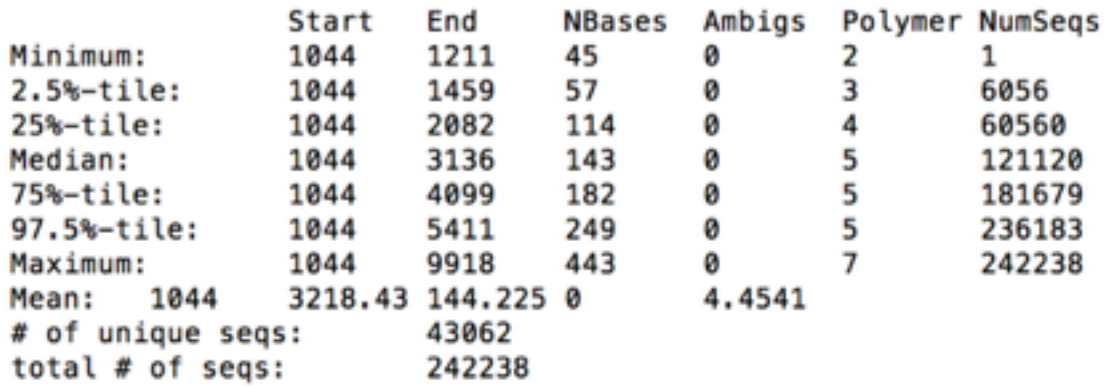
data set. The most important part of this analysis is to use the DNA from your data set to figure out what bacteria are present in your samples. This is done with the align.seqs command. The code goes as follows, *align.seqs(fasta=data.trim.unique.fasta, reference=silva.bacteria.fasta, flip=T, processors=8)*. The reference file can be downloaded on the internet and the silva reference file is the most common and most accurate reference file to date. The flip function is very important here because most of the time the DNA is in the reverse form and needs to be flipped to get an accurate reading on them. A summary of this command should look like **Figure 4**.

Figure 4: Summary After the Align Command

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1044	1048	3	0	1	1
2.5%-tile:	1044	1207	51	0	3	6526
25%-tile:	1044	2051	92	0	4	65253
Median:	1044	3135	141	0	5	130506
75%-tile:	1044	4104	181	0	5	195759
97.5%-tile:	1085	5441	248	0	5	254486
Maximum:	43100	43116	443	0	7	261011
Mean: 1242.26	3333.36	139.948	0	4.4274		
# of unique seqs:		46186				
total # of seqs:		261011				

This is then run through a screen command. The code that is used is *screen.seqs(fasta=data.trim.unique.align, name=data.trim.names, group=data.groups, start=1044, optimize=end, criteria=95, processors=8).* The group file is obtained from the trim output. The start command tells the sequence where to start and in this case since the majority of the sequences start at around 1044, that is why I used it as my criteria. The optimize command attenuates the start or the end, this case the end was used, to a specific criteria. This criteria is set by the criteria command and in this case its 95. That means that the program makes an end sequence location where 95% of all sequences end before that particular location. After this command your summary should look like **Figure five** on the next page.

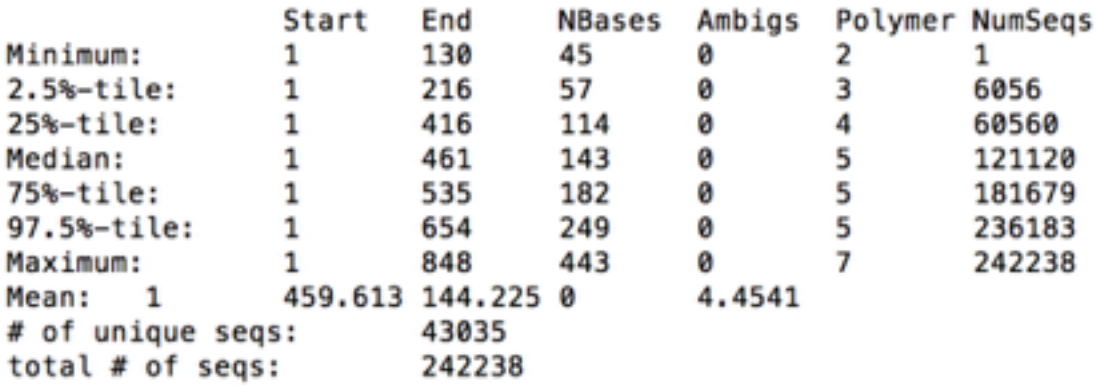
Figure 5: Summary After the Screen Command



	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1044	1211	45	0	2	1
2.5%-tile:	1044	1459	57	0	3	6056
25%-tile:	1044	2082	114	0	4	60560
Median:	1044	3136	143	0	5	121120
75%-tile:	1044	4099	182	0	5	181679
97.5%-tile:	1044	5411	249	0	5	236183
Maximum:	1044	9918	443	0	7	242238
Mean:	1044	3218.43	144.225	0	4.4541	
# of unique seqs:		43062				
total # of seqs:		242238				

The next command is a filter command and the code goes as follows, *filter.seqs(fasta=data.trim.unique.good.align, vertical=T, processors=8).* This command calls for a trump input but I found that using the trump causes more problems than the good it does. The next command is another unique command, to extract the unique sequences from the data. The code goes as follows, *unique.seqs(fasta=data.trim.unique.good.filter.fasta, name=data.trim.good.names).* If a summary.seq command is done on these .fasta and name files, it should look like **Figure 6**.

Figure 6: Summary After the Filter and Unique Command



	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	130	45	0	2	1
2.5%-tile:	1	216	57	0	3	6056
25%-tile:	1	416	114	0	4	60560
Median:	1	461	143	0	5	121120
75%-tile:	1	535	182	0	5	181679
97.5%-tile:	1	654	249	0	5	236183
Maximum:	1	848	443	0	7	242238
Mean:	1	459.613	144.225	0	4.4541	
# of unique seqs:		43035				
total # of seqs:		242238				

The last part of the analysis involves pre clustering and chimera removal. The most important aspect of this command is using at least eight processors because if not, both of these commands can take up to 2 days to run. The first command is a pre cluster command and the code goes as follows, *pre.cluster(fasta=data.trim.unique.good.filter.unique.fasta, name=data.trim.unique.good.filter.names, group=data.good.groups, diffs=2, processors=8).* This command merges sequences that are very close to each other, to make analysis much easier. The next command is a chimera removal step. The code is as follows *chimera.uchime(fasta=data.trim.unique.good.filter.unique.precluster.fasta, name=data.trim.unique.good.filter.unique.precluster.names, group=data.good.groups, processors=8).* This gives a chimera file and an accnos file. The next step is a removal command that removes all the chimeras that were found with the chimera.uchime command. The code goes as follows, *remove.seqs(accnos=data.trim.unique.good.filter.unique.precluster.uchime.accnos, fasta=data.trim.unique.good.filter.unique.precluster.fasta, name=data.trim.unique.good.filter.unique.precluster.names, group=data.good.groups, dups=T)*If a summary command was ran at this point it should look like **Figure 7**.

Figure 7: Summary After the Pre-Cluster and Chimera Removal Commands

	Start	End	NBases	Ambigs	Polymer	NumSeqs
Minimum:	1	130	45	0	2	1
2.5%-tile:	1	216	57	0	3	5989
25%-tile:	1	414	113	0	4	59881
Median:	1	461	143	0	5	119761
75%-tile:	1	535	181	0	5	179641
97.5%-tile:	1	653	249	0	5	233533
Maximum:	1	848	443	0	7	239521
Mean: 1	458.812	143.839	0	4.4493		
# of unique seqs:		30004				
total # of seqs:		239521				

This leads to the final step, classification, which gives a taxonomy table. Within this taxonomy table is the final results of the analysis. The code for this goes as follows *classify.seqs(fasta=data.trim.unique.good.filter.unique.precluster.pick.fasta,*

name=data.trim.unique.good.filter.unique.precluster.pick.names, group=data.good.pick.groups, template=trainset9_032012.pds.fasta, taxonomy=trainset9_032012.pds.tax, cutoff=85, processors=8). The template input and the taxonomy input can be found on the Mothur website and can be downloaded from there. The cutoff input is very important because it gives the percent of the sequence that matches with the template sequence. If this number was zero all sequences would be classified but the data could not be used because some of the data would have very low similarity. Using a cutoff of 85, that means 85% of the data is similar to the sequence, will lead to some undefined sequences but the data is very accurate. After this command is ran a file named data.trim.unique.good.filter.unique.precluster.pick.pds.wang.tax.summary is obtained. This is the table that is exported to your desktop to do further analysis on it with Excel or Numbers.

2.1.5 Unifrac Data Analysis

Unifrac is a website that takes several files that can be made in Mothur and creates PCoA data and runs other statistical programs on the data. PCoA stands for principal coordinates analysis and is made with distance files that show how close or how far away the data is from each other. Data should cluster with its specific group, in this case control or treatment. Usually the control will cluster well and the treatment will be more variable but will be relatively close to the other treatment data points. Unifrac uses a phylogenic tree, an ID mapping file, and a category mapping file for the PCoA function (18,19,20). The phylogenic tree is generated from Mothur from the final .fasta file that is used in the analysis (18,19,20). The ID mapping file is made from a .group file and a .count-table file. Both are made using Mothur. The count table gives every read found by the analysis but they are not attached to the specific sample. A small program can be made to combine each read to its appropriate sample. The output from this program is the ID mapping file. The category map is made by the user in .txt form and includes the sample names, the niche, and the type of sample. An example line from the category map would look like, 202.rt 16S Control. 202.rt is the sample name, 16S is the niche, and control is the type of sample. These three data files were uploaded onto the Galaxy website and run with the PCoA command which gave several graphs.

2.2 Results

2.2.1 16S rRNA Gene Analysis

The main goal of this experiment was to look at the changes that TCDF caused in the gut of mice. This was done by two groups, control and treatment. The control group included six mice, 202-304 and the treatment group was 402-604. All mice were trained to eat dough pills before the treatment started. The treatment group was fed dough pills that had 0.6 ug of TCDF in them which gave a final dose of 24 ug/kg of TCDF per mouse. The control mice were just given plain dough pills. The group 601 was removed from this experiment because of the extreme outliers that it had in its data set. The best way to look at this data was with the analysis that was done on Mothur and the final file that was obtained. This was looked at mainly in Excel. **Figure 8** shows the different phyla levels in a percent form. This was one of the first things that was done with the data to show relative amounts of bacteria in each phyla. The changes may not look like much but when dealing with around 18000 reads a 5% change is fairly large. A number of different groups of bacteria were seen to have significant changes between the control group and the treatment group. There were four main groups that were looked at. The first group was the phylum *Bacteroidetes* and it was seen that the treatment group had significantly more than the control group as seen in **Figure 9**. The *Firmicutes* phylum was also looked at and the treatment group was seen to have less than the control group as seen in **Figure 9**. **Figure 10** shows the changes in the *Clostridia* family and the treatment group was seen to have less than the control group as well. **Figure 11** shows the species *Oscillibactor* and this showed that the treatment group had significantly less than the control group as well. The majority of the changes were a lower level of bacteria between the treatment and control groups. This lower level leads to a lower diversity of bacteria in the gut.

Mouse	Total Count
202	22117
203	19454
301	19175
302	20347
303	25216
304	19015
402	23235
404	22364
602	16723
603	15771
604	14484

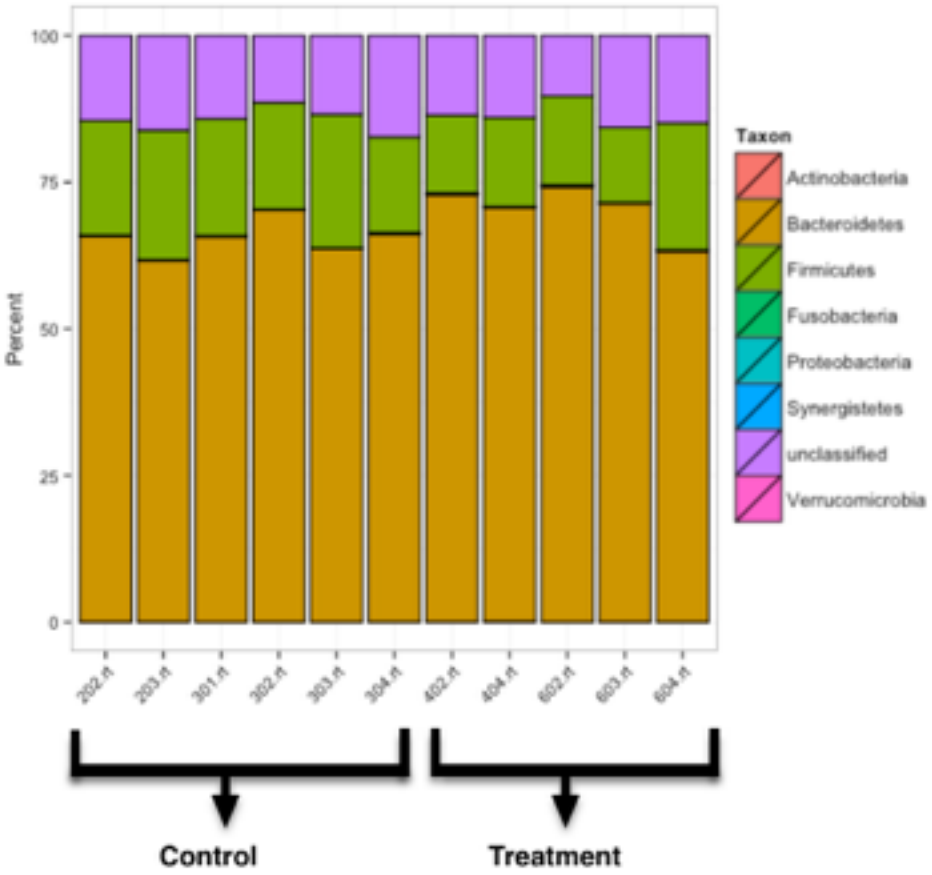


Figure 8: Levels of Different Phyla in all Samples

This is a graph of all the different phyla concentrations seen in the analysis. The purple group on top is the unclassified bacteria. The green in the middle is the *Firmicutes* population. The orange on the bottom is the *Bacteroidetes* population. The other phyla are existent at very low levels. This graph uses percentages taken by dividing the counts of a specific phyla by the total counts. The total counts are shown on the left side.

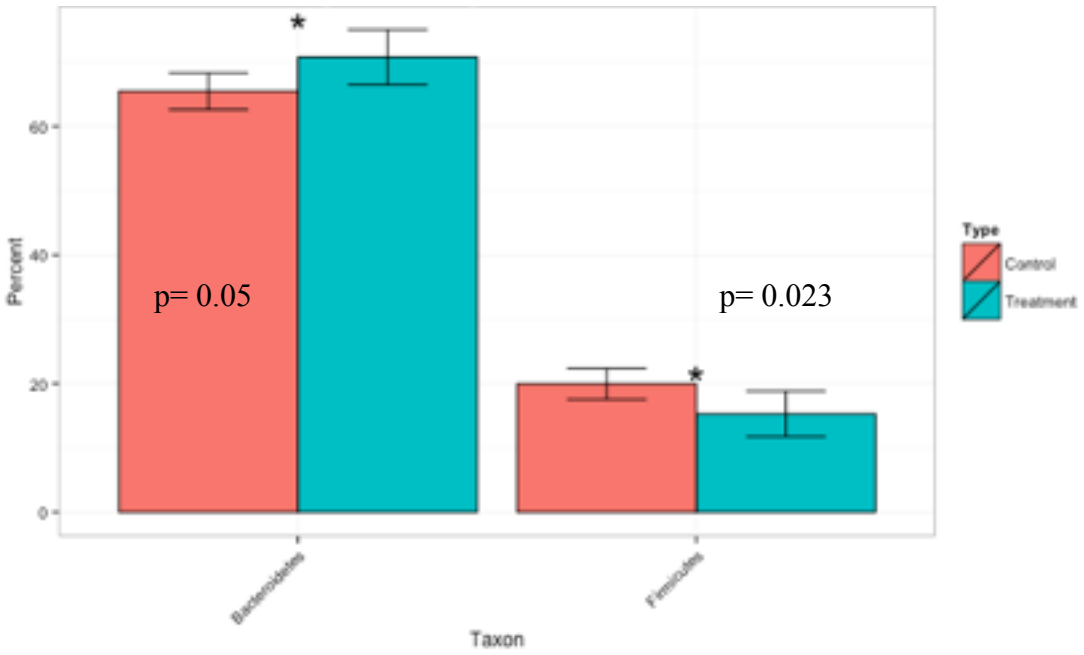


Figure 9: The Major Changes in Bacteria Phyla From Control to Treatment.

This graph shows the two major phyla in *Bacteroidetes* and *Firmicutes.* 601 was removed because it was an outlier. All analysis was done with Mothur and the graph design was done in R studio. Error bars are + and - one standard deviation.

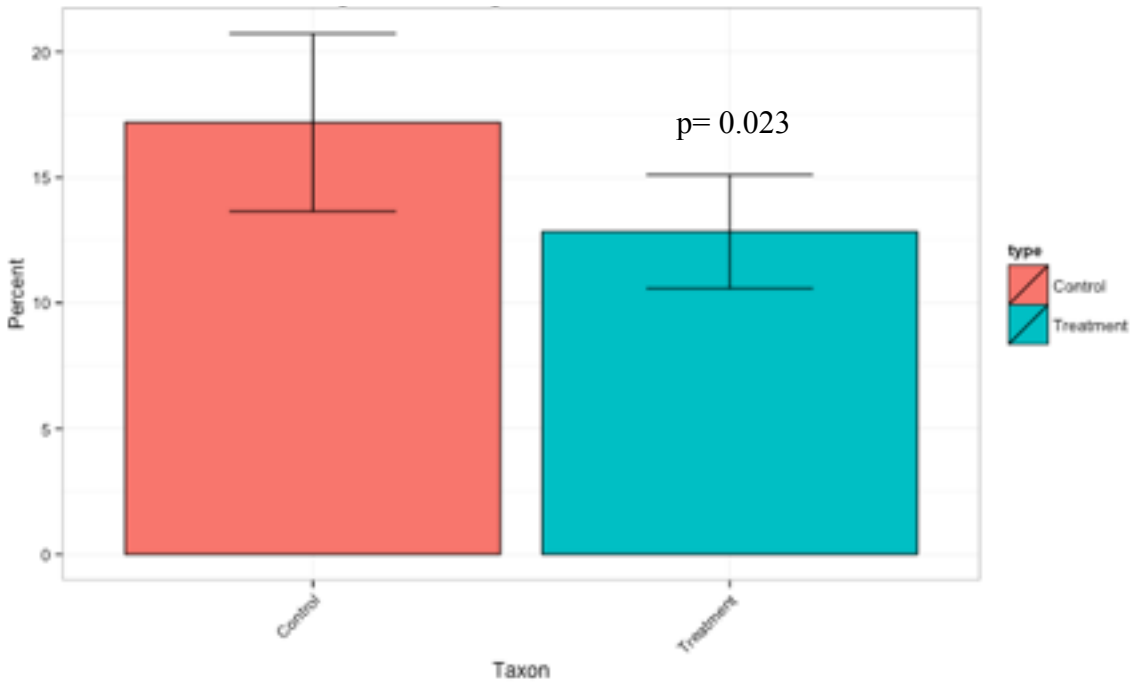


Figure 10: Significant change in the *Clostridia* Family.

This graph shows a major family, *Clostridia*. Again 601 was removed because it was an outlier. All analysis was done with Mothur and the graph design was done in R studio. Error bars are + and - one standard deviation.

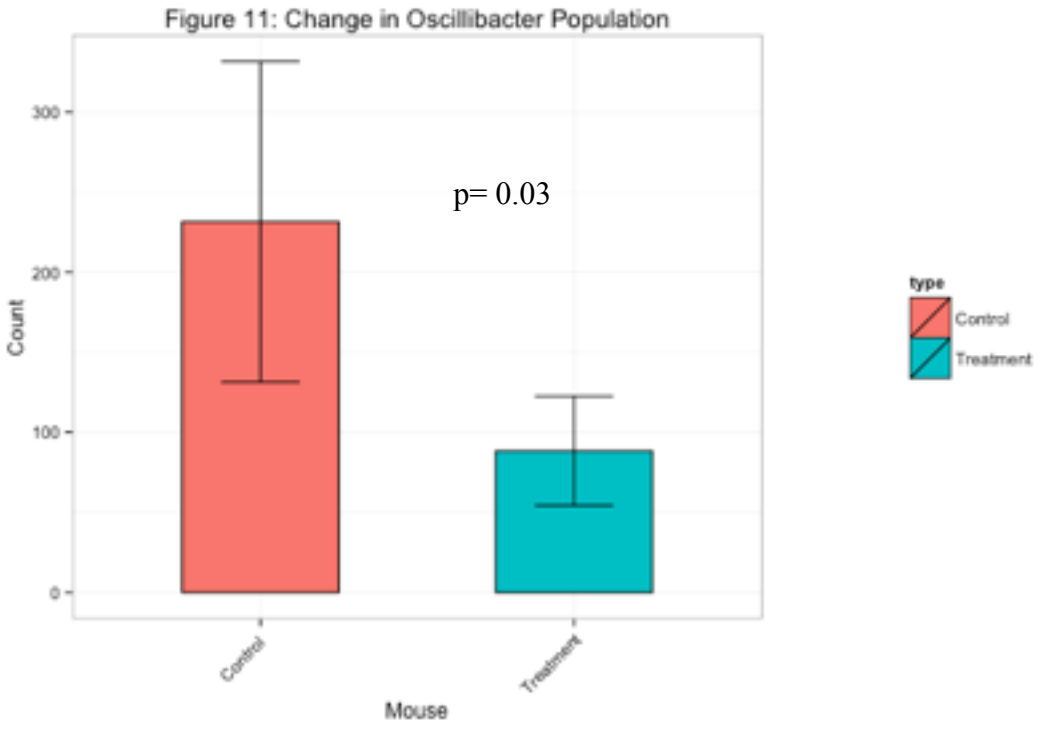


Figure 11: Change in *Oscillibactor* Population

This was another group that had significant changes, unfortunately converting counts into percentages would not do this change justice. That is the reason it is left in counts rather than converting it to percentages. Again 601 was left out because it was an outlier. All analysis was done with Mothur and graph design was done in R studio. Error Bars are + and - one standard deviation The error bar is so high on the control because there was a very large drop in *Oscillibactor* in one of the control groups. This is due to the fact that all guts are different and some have more of a certain species than others.

2.2.2 PCoA data

The next step that was done was a PCoA analysis to determine the similarity of the samples and to see grouping. As seen from **Figure 12** and **Figure 13** the clustering of the data is what is expected. The control group is very clustered together while the treatment group is highly variable. When using the Unifrac software the weighted and unweighted measurements are taken into consideration. When using weighted measurements it takes into account when certain sequences has a multitude of reads and normalizes the data in that fashion. When our lab uses Unifrac it is done with unweighted and no normalization step is taken.

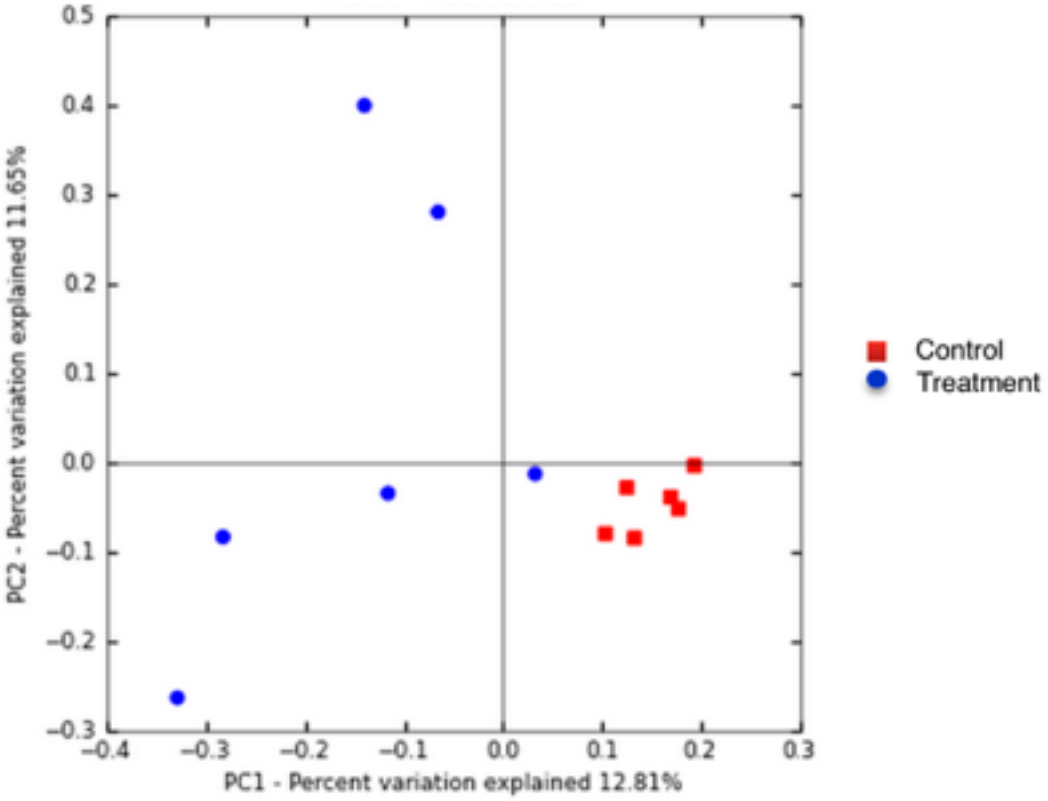


Figure 12: PC1 vs PC2 PCOA data.

This shows the variation of the samples between each other. It can be seen that the control group clusters up very nicely and the treatment group seems to be quite sporadic. 601 is the group closest to the control data and this is why it was removed. We decided that crossing the 0 line makes it too much of an outlier. This graph was made from the Unifrac website.

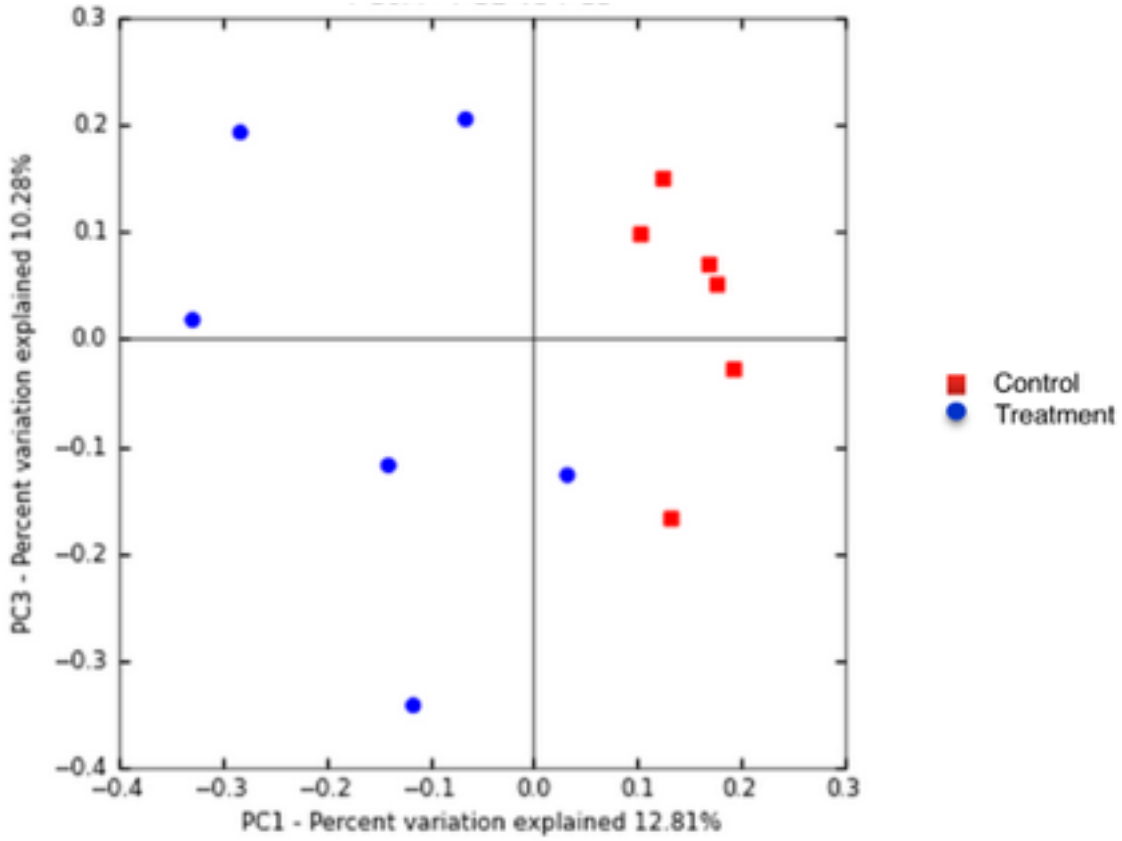


Figure 13: PC1 vs PC3 PCoA data

This is another PCoA plot. It shows the data more sporadic but besides 601 both groups are confined to either the positive side or negative side. Again this plot along with figure 10 was the reason that 601 was removed. This graph was made from the Unifrac website.

2.2.3 Different options that can be used in Data analysis

In the methods section it was mentioned that the SILVA reference file is used when aligning the sequences that are obtained from the genomics facility to a template. SILVA is the most commonly used file but there are still some others that are used occasionally. Greengenes was mentioned in the Mothur website as an alternative. We wondered how different the two files are. **Table 1** is a very brief overview of the most important factors that are taken into consideration when looking at data. The first criterion is total species and that is how many taxons were given after cleaning up the data. The data is cleaned up by removing any taxon that has a population count less than 20. The next criteria that is looked at is total average counts. This is an average of the counts for each mouse. The most important criteria is the number of significantly different species that are seen in each group. The average number of unclassified bacteria that are seen are also included in **Table 1**. Also along with different alignment templates, different cutoffs are investigated. 85% is a normal cutoff that gives very good results. 0% can lead to very attractive results but they cannot be used in making conclusions. As described in the methods section the cutoff is used when comparing how the sequence matches up with a template. If an 85% cutoff is used that means that 85% of the sequence has to match the template to be classified. This leads to some unclassified sequences. Using 0% cutoff means that everything is classified and no unclassified sequences will be gained. This is unethical because some of the data that got sequenced could be mutated or incomplete which means it falls under the unclassified category. If it is said that everything is classified then it is assumed that everything is sequenced which is not the case. Thus classifying data with a 0% cutoff is false representation of data.

When looking at the data it can be seen that significantly less species are classified using 85% but that is to be expected. We are getting less total species when using the 85% cutoff because as mentioned above sequencing is not perfect and not every piece of data that was given

to be sequenced can not be sequenced. There could be some artificial elements that are in the data that will not be sequenced or the genomic sequence could be incomplete, or mutated from the extraction process. This is why there is less total species from using a higher cutoff rate. Also there are many more specific changes in bacteria when using a 0% cutoff. When comparing the two reference files it can be seen that they lead to similar number of taxons and number of specific species. However the Greengene is almost three times more unclassified per mouse than the SILVA file.

Table 1: Differences Between reference files and classification criteria.

This table uses the SILVA reference file and the Greengenes reference file. Both of these files can be obtained from the Mothur website. Also this table shows the differences between different cutoffs that are used when making the taxonomy file. Normally 85% is used but sometimes people try to get away with using 0%.

	Total species	Total average counts	number of species that are significantly different between control and treatment	number of unclassified bacteria
Silva 85% classification	91	19809	9	2793
Silva 0% classification	325	19809	59	0
GreenGenes 85%	64	20319	8	6665
GreenGenes 0%	320	20337	60	0

2.3 Discussion

This was a relatively simple experiment with very complex analytical methods. Our group wanted to see the effects of TCDF, which clears from the mice faster and has a much shorter half life in mice than TCDD. Cecal contents were retrieved and sequenced which lead to the analysis of the gut microbiome based on a computational analysis of high throughput sequencing data. The other aspects that were worked on were bile acid regulation, cytochrome P450 activation and other tissue specific tests. The analysis of the 16S rRNA gene data was done on a computer, which required the learning of two whole new computer languages, Terminal and R. This involved months of trial an error to figure out the correct parameters to run the data at which would give the best, and most reliable results possible. These final parameters were displayed in the 16S rRNA gene analysis subsection of the methods section. **Figures 2-7** provide a summery at the end of each major step. These perimeters were based off a template that was given on the Mothur website (17). This also took the longest time because again it was a whole new language and it took time to figure out how to properly run the code. As mentioned in the methods section if this code is not run with 8 processors some commands can take up to two full days to run. Currently some of the more advanced analysis techniques that can be done with the Mothur program are still being optimized. These techniques include, amova (analysis of molecular variance) analysis, nmds (non-metric multidimensional scaling) and PCoA (Principal Coordinates)(17). All of these methods are used to determine the significance of the data that was analyzed.

The most important result from this experiment so far was the taxonomy file that was generated in the final step of 16S rRNA gene sequencing. This was first view in a graph of the total phyla concentrations among the 11 groups as seen in **Figure 8**. **Figure 9, Figure 10,** and **Figure 11,** show all of the significant changes that were seen between the treatment and control groups. From this it can be seen that most significant changes are in a decrease of population. The

only group that increased was the phylum of *Bacteroidetes*, as seen in **Figure 9.** Since *Bacteroidetes* is a gram negative bacteria it can be hypothesized that there is more lipopolysaccharide being released in the mouse gut. This increased level of lipopolysaccharide could cause a higher level of inflammation in the gut and lead to diseases like obesity or Crohn's disease (21). Also it was seen that there was a decrease in the *Clostridia* family as seen in **Figure 10**. Recently it has been seen that *Clostridia* plays an important role in inducing regulatory T cells (21). This also leads to an attenuation of inflammatory factors and a reduction in inflammation (21). This means that an increase in gram negative *Bacteroidetes* and a decrease in *Clostridia* could greatly lead to an inflamed state in the mice. Also the *Firmicutes* as a phylum was seen to decrease after the treatment, seen in **Figure 9**. *Firmicutes* has been seen to be a very helpful bacteria in relation to keeping the gut healthy. They have been seen to produce short chain fatty acids that have been observed preventing inflammation (23). Short chain fatty acids have also been seen to help prevent obesity, mainly by butyrate producing bacteria (23). It has been reported that a negative correlation exists between the short chain fatty acid producing bacteria and inflammation (23). That means as the short chain fatty acid producing bacteria in *Firmicutes* decreases, the amount of inflammation increases. The last group that changed was in the family of *Oscillibacter* and it was seen that it decreased during treatment as seen in **Figure 11**. It has been noted that as *Oscilibacter* decreases, body weight and body mass index increases (24). Since obesity has been seen to occur from an inflamed state, it can be assumed that a decrease in *Oscilibacter* will lead to an inflamed state. Over all the four major changes that were seen in the bacterial populations all lead to a higher level of inflammation in the gut which could lead to various diseases like diabetes, obesity and Crohn's disease.

After the in depth analysis of the taxonomic data, we wanted to see how the data clustered in a PCoA plot. When looking at PCoA data the optimal result is to have control data tightly clustered far away from a treatment group that is also tightly clustered. Realistically PCoA data looks like **Figure 12** and **Figure 13**. **Figure 12** is the main plot and shows a tight cluster of

the control groups. The treatment groups are all over the place but that is to be expected because the gut is a very fickle entity. Also most guts will act differently when perturbed so this variability was to be expected. We did decide that group 601 was too variable to keep in the analysis mainly because it was over the midline in both **Figure 12** and **Figure 13**. This data showed us that overall it was clustering correctly and the taxonomic data that was seen should be very representative of what occurred. Another way to show that this data is representative is by looking at the error bars of **Figure 9, Figure 10,** and **Figure 11. Figure 9** and **Figure 10** have very good error bars that show very little deviation from the mean. This means that all samples in these groups have similar values to the average which are used in the graph. **Figure 11** however has a very large error bar on the control group. This is due to the fact that there are differences in every gut microbiome so one large deviation like this is common and the decrease in *Oscillibacter* is still significant.

The last thing that was looked at was a comparison between reference files and cutoffs. This was done because we wanted to show that the best reference file available was being used and also the allure of using a 0% cutoff. When looking at **Table 1** it can be seen that the SILVA and the Greengene files have a similar amount of taxons and specific species changes. However Greengene gave almost three times the amount of unclassified species. This means that SILVA is 3 times better at classifying the same data at the same cutoff. This is also why the SILVA file is the best reference file currently in use. When dealing with cutoffs, you want a cutoff that has some wiggle room but not so much that it makes your data unusable. Some people want to use a 0% cutoff to have no unclassified species. This leads to amazing results, and this can be seen in **Table 1** because using the same data at 85% cutoff only has 9 significantly different species but using a 0% cutoff leads to 59 significantly different species. However those 9 species are 85% certain to match and be different. Those 59 species are 0% certain to match and be different. This is why 9 may seem like a very small amount of species but they are almost 100% sure to match and actually be significantly different. Even with today's sequencing capabilities

we can only look at a fraction of the gut microbiome. Also not every species of bacteria will be affected by the treatment. So when using an 85% cutoff we can see that about 10% of the species changed with treatment. Even when classifying every sequence by using a 0% cutoff, less than 20% of the bacteria was significantly changed with treatment. So since we are getting about half that and using a much higher cutoff, that means this data is good and can be used to represent the whole microbiome.

Chapter 3: CONCLUSIONS

Overall this experiment was about learning two different computer languages and helped to establish a new methodology for the lab. This past summer I was handed several text books that went into the basis of Terminal based coding and I had to learn this language on my own. I did have some help from the Bioinformatics facility and they guided me on the right path. To actually learn this language it took about two or three months. Once I got the language down I could actually start to code with Mothur which then took another 2 to 3 months. It took so long because this was my first example that was not specifically made for teaching so the data will not be perfect an I have to optimize the method to suite my own needs. Once I figured out what each and every command meant I could start looking at my analysis and get the data that I used today. By using Mothur I made sure that the SILVA classification file was the best one to use to date. Also I looked at the importance of using a higher cutoff of 85%. If one used a classification of 0% there would be no undefined sequences but certain artificial elements that are in the sequence, would be classified leading to a false representation of the data.

When looking at the actual data four species were looked at. First it was seen that *Bacteroidetes* increased after treatment. This could increase the amount of lipopolysaccharide leading to inflammation. Next *Firmicutes, Clostridia,* and *Oscillibacter* were all seen to decrease after treatment. *Firmicutes* has been seen to produce short chain fatty acids that can help reduce inflammation. *Clostridia* has been seen to produce regulatory T cells that help reduce inflammation. *Oscillibacter* has been seen to be inversely related with inflammatory factors. All of these factors could lead to a greater level of inflammation seen in the gut. This lead to the next group of tests that were done on this subject. These involved looking at cytochrome P450 activation and cytokines in digestive tissue, liver tissue, and blood levels.

For me the next steps would be to stick inside the microbiome. After writing this thesis and another research paper for another class on the microbiome, I am fascinated by it. I really want to look more into the different parts of the microbiome that can be affected by chemicals that we are exposed to every day. I want to look at how these changes can affect the overall health

of the mouse. Since the microbiome has just recently been discovered to be huge proponent in human health most of these questions remain unanswered. With my new skills in 16S rRNA gene analysis I feel like I can look into these questions and try my best to answer them. To date I have looked into how the gut is a key regulator in obesity and inflammatory diseases. I think looking more into IBD and Crohn's disease would be very interesting because not much is known about the onset of both of these diseases. I would really like to see how bacteria takes part in these diseases and if there is anything that can be done to help attenuate them. Recently there has been a push in guiding children's microbiomes to prevent obesity later in life and I think something like this could also be done to help prevent IBD and Crohn's disease.

BIBLIOGRAPHY

"2020 Visions." Nature 463, no. 7277 (2010): 26-32. (1)

Aly, H. A., and O. Domènech. "Cytotoxicity and Mitochondrial Dysfunction of 2,3,7,8-Tetrachlorodibenzo-P-Dioxin (Tcdd) in Isolated Rat Hepatocytes." Toxicol Lett 191, no. 1 (2009): 79-87. (8)

Anderson, G., T. V. Beischlag, M. Vinciguerra, and G. Mazzoccoli. "The Circadian Clock Circuitry and the Ahr Signaling Pathway in Physiology and Pathology." Biochem Pharmacol 85, no. 10 (2013): 1405-16. (14)

Atarashi, K., T. Tanoue, K. Oshima, W. Suda, Y. Nagano, H. Nishikawa, S. Fukuda, T. Saito, S. Narushima, K. Hase, S. Kim, J. V. Fritz, P. Wilmes, S. Ueha, K. Matsushima, H. Ohno, B. Olle, S. Sakaguchi, T. Taniguchi, H. Morita, M. Hattori, and K. Honda. "Treg Induction by a Rationally Selected Mixture of Clostridia Strains from the Human Microbiota." Nature 500, no. 7461 (2013): 232-6. (22)

Blais Lecours, P., D. Marsolais, Y. Cormier, M. Berberi, C. Haché, R. Bourdages, and C. Duchaine. "Increased Prevalence of Methanosphaera Stadtmanae in Inflammatory Bowel Diseases." PLoS One 9, no. 2 (2014): e87734. (7)

Blankenberg, D., G. Von Kuster, N. Coraor, G. Ananda, R. Lazarus, M. Mangan, A. Nekrutenko, and J. Taylor. "Galaxy: A Web-Based Genome Analysis Tool for Experimentalists." Curr Protoc Mol Biol Chapter 19 (2010): Unit 19.10.1-21.(19)

Busbee, P. B., M. Rouse, M. Nagarkatti, and P. S. Nagarkatti. "Use of Natural Ahr Ligands as Potential Therapeutic Modalities against Inflammatory Disorders." Nutr Rev 71, no. 6 (2013): 353-69. (16)

F.S.Teixeira, T., Łukasz M. Grześkowiak, Seppo Salminen, Kirsi Laitinen, Josefina Bressan, and Maria do Carmo Gouveia Peluzio. "Faecal Levels of Bifidobacterium and Clostridium Coccoides but Not Plasma Lipopolysaccharide Are Inversely Related to Insulin and Homa Index in Women." Clinical Nutrition 32, no. 6 (2013): 1017-22. (21)

Giardine, B., C. Riemer, R. C. Hardison, R. Burhans, L. Elnitski, P. Shah, Y. Zhang, D. Blankenberg, I. Albert, J. Taylor, W. Miller, W. J. Kent, and A. Nekrutenko. "Galaxy: A Platform for Interactive Large-Scale Genome Analysis." Genome Res 15, no. 10 (2005): 1451-5. (20)

Goecks, J., A. Nekrutenko, J. Taylor, and Galaxy Team. "Galaxy: A Comprehensive Approach for Supporting Accessible, Reproducible, and Transparent Computational Research in the Life Sciences." Genome Biol 11, no. 8 (2010): R86. (18)

John, Kaarthik, Tejas S. Lahoti, Kelly Wagner, Jarod M. Hughes, and Gary H. Perdew. "The Ah Receptor Regulates Growth Factor Expression in Head and Neck Squamous Cell Carcinoma Cell Lines." Molecular Carcinogenesis (2013): n/a-n/a. (27)

Johnson, C. H., A. D. Patterson, J. R. Idle, and F. J. Gonzalez. "Xenobiotic Metabolomics: Major Impact on the Metabolome." Annu Rev Pharmacol Toxicol 52 (2012): 37-56. (5)

Kim, B. S., M. Y. Song, and H. Kim. "The Anti-Obesity Effect of Ephedra Sinica through Modulation of Gut Microbiota in Obese Korean Women." J Ethnopharmacol 152, no. 3 (2014): 532-9. (24)

Lahoti, T. S., J. M. Hughes, A. Kusnadi, K. John, B. Zhu, I. A. Murray, K. Gowda, J. M. Peters, S. G. Amin, and G. H. Perdew. "Aryl Hydrocarbon Receptor Antagonism Attenuates Growth Factor Expression, Proliferation, and Migration in Fibroblast-Like Synoviocytes from Patients with Rheumatoid Arthritis." J Pharmacol Exp Ther 348, no. 2 (2014): 236-45. (26)

Lowe, M. M., J. E. Mold, B. Kanwar, Y. Huang, A. Louie, M. P. Pollastri, C. Wang, G. Patel, D. G. Franks, J. Schlezinger, D. H. Sherr, A. E. Silverstone, M. E. Hahn, and J. M. McCune. "Identification of Cinnabarinic Acid as a Novel Endogenous Aryl Hydrocarbon Receptor Ligand That Drives Il-22 Production." PLoS One 9, no. 2 (2014): e87877. (9)

Marchesi, J. R. "Human Distal Gut Microbiome." Environ Microbiol 13, no. 12 (2011): 3088-102.(2)

Martínez, I., C. E. Muller, and J. Walter. "Long-Term Temporal Analysis of the Human Fecal Microbiota Revealed a Stable Core of Dominant Bacterial Species." PLoS One 8, no. 7 (2013): e69621. (3)

Pérez-Cobas, A. E., M. J. Gosalbes, A. Friedrichs, H. Knecht, A. Artacho, K. Eismann, W. Otto, D. Rojo, R. Bargiela, M. von Bergen, S. C. Neulinger, C. Däumer, F. A. Heinsen, A. Latorre, C. Barbas, J. Seifert, V. M. dos Santos, S. J. Ott, M. Ferrer, and A. Moya. "Gut Microbiota Disturbance During Antibiotic Therapy: A Multi-Omic Approach." Gut 62, no. 11 (2013): 1591-601. (15)

Quintana, Francisco J., Alexandre S. Basso, Antonio H. Iglesias, Thomas Korn, Mauricio F. Farez, Estelle Bettelli, Mario Caccamo, Mohamed Oukka, and Howard L. Weiner. "Control of Treg and Th17 Cell Differentiation by the Aryl Hydrocarbon Receptor." Nature 453, no. 7191 (2008): 65-71. (10)

Rajilić-Stojanović, M. "Function of the Microbiota." Best Pract Res Clin Gastroenterol 27, no. 1 (2013): 5-16. (6)

Remely, M., E. Aumueller, C. Merold, S. Dworzak, B. Hippe, J. Zanner, A. Pointner, H. Brath, and A. G. Haslberger. "Effects of Short Chain Fatty Acid Producing Bacteria on Epigenetic Regulation of Ffar3 in Type 2 Diabetes and Obesity." Gene 537, no. 1 (2014): 85-92. (23)

Schloss, P. D., D. Gevers, and S. L. Westcott. "Reducing the Effects of Pcr Amplification and Sequencing Artifacts on 16s Rrna-Based Studies." PLoS One 6, no. 12 (2011): e27310. (17)

Schéle, E., L. Grahnemo, F. Anesten, A. Hallén, F. Bäckhed, and J. O. Jansson. "The Gut Microbiota Reduces Leptin Sensitivity and the Expression of the Obesity-Suppressing Neuropeptides Proglucagon (Gcg) and Brain-Derived Neurotrophic Factor (Bdnf) in the Central Nervous System." Endocrinology 154, no. 10 (2013): 3643-51. (25)

Sorg, O. "Ahr Signalling and Dioxin Toxicity." Toxicol Lett (2013). (11)

Sorg, O., M. Zennegg, P. Schmid, R. Fedosyuk, R. Valikhnovskyi, O. Gaide, V. Kniazevych, and J. H. Saurat. "2,3,7,8-Tetrachlorodibenzo-P-Dioxin (Tcdd) Poisoning in Victor

Yushchenko: Identification and Measurement of Tcdd Metabolites." Lancet 374, no. 9696 (2009): 1179-85. (13)

Stange, J., and M. Veldhoen. "The Aryl Hydrocarbon Receptor in Innate T Cell Immunity." Semin Immunopathol 35, no. 6 (2013): 645-55. (12)

Tilg, H., and A. Kaser. "Gut Microbiome, Obesity, and Metabolic Dysfunction." J Clin Invest 121, no. 6 (2011): 2126-32. (4)

ACADEMIC VITA

Robert Nichols
rgn5011@psu.edu 602 South Pugh, State College, 16802

Experience/Education:

Metabolomics facility - Worked in Dr. Patterson's lab in the Metabolomics facility since the summer of 2013 until present. This has provided valuable experience in the field of toxicology. I have used HPLC, MS/MS, and Triple quadrupole machines daily in various experiments. Have used assays involved with real time PCR to investigate gene expression. Used 16s RNA sequence analysis programs to understand shifts in gut bacterial populations.

Attended Pennsylvania State University 2010-2014
In the Schreyer's Honor College 2012-2014
Dean's List all semesters at Pennsylvania State University
Accepted an offer to attend Graduate School at the Pennsylvania Stat University in the Molecular Toxicology program.

Work experience:

Lab assistant in Dr. Patterson's lab at Pennsylvania State University 2013-Present

Work on a project about the effects of TCDD through AHR activation. A focus was put on the gut microbiota and the shifts that occur when exposed to TCDD. An exploration will be done through a thesis on 16s RNA analysis thorough a variety of methods including a terminal based program called Mothur and R for data presentation.

Intern for Boston Scientific Summer of 2012

Helped lead an investigation into correction the major problem of pelvic organ prolapse. Worked in a small research and development group and had a major impact on decision making and gave a 40 minute presentation on my findings at the end of the summer.

Leadership:

Study Abroad in Ireland: Spring Semester - 2013
Studied in Ireland through The Pennsylvania State University and had to adapt to the major differences in cultures. Was largely on my own and even organized trips to Switzerland and Scotland.

Biomedical Science Club, University Park, PA 2010 - 2012
Helped plan and run the Biomedical Science Club Undergraduate Conference in 2011 (keynote was Dr. D. A. Henderson)
Head of the web design and development for the club and conference
Helping plan the Biomedical Science Club Undergraduate Conference for 2012

Mars United Presbyterian Church, Mars, PA: 2005-2010
Participated in several mission trips, including World Vision sponsored trip to Pickins, West Virginia and locally organized trip to Texas City, TX to assist in hurricane Katrina clean-up efforts as well as other local mission projects

Skills/Interests:

Terminal based programming and R based data presentation
Have used MS/MS, QTOF, Triple Quadrapole and HPLC machines in my research
Microsoft and Mac software including Word, Excel, Pages, Numbers
Watching Formula1 racing and all types of sports, avid reader