

Investigating
the Gut
Microbiome in
Psychiatric
Illness

Investigating the gut microbiome of Generalized Anxiety Disorder, Major
Depressive Disorder and Bipolar patients

By: Ryan Potts, B.Sc.

A thesis submitted to the Department of Medical Sciences and the School of Graduate Studies of
McMaster University in partial fulfillment of the requirements for the degree of Master of
Science.

© Ryan Potts, September 2017
All Rights Reserved

Master of Science (2017)

McMaster University

(Medical Sciences)

Hamilton, Ontario, Canada

Title: Investigating the microbiome of Generalized Anxiety Disorder, Major Depressive Disorder and Bipolar Disorder

Author: Ryan Potts
B.Sc. Specialization, Microbiology
University of Waterloo (2015)

Supervisors: Dr. Rebecca Anglin & Dr. Michael Surette

Number of pages: xiii, 118

Lay Abstract

Mental health disorders including Generalized Anxiety Disorder (GAD), Major Depressive Disorder (MDD) and Bipolar Disorder (BD) affect somewhere between five and ten percent of Canadians, with the World Health Organization recently labeling MDD the leading cause of disability worldwide. Despite decades of research, we still have a poor understanding of what factors may contribute to causing these disorders etiology. Another hallmark of many mental health conditions are increased rates of gastrointestinal symptoms relative to healthy individuals. Recently, there has been a great deal of research about how the trillions of bacteria that live in the digestive tract play a role in brain development and behaviour. This study aimed to better understand the organisms in the digestive tract of mental health patients in hopes of better understanding how some of these organisms could be contributing to patients' poor mental health. Additionally, through better understanding the nature of these communities, recommendations could potentially be made about therapeutic interventions to restore a healthy community.

Abstract

The global burden of mental health disorders is rising with the world health organization recently having recognized major depressive disorder as the leading cause of disability worldwide. Nearly one in five Canadians are now estimated to struggle with a mental health disorder and Generalized Anxiety Disorder (GAD) Major Depressive Disorder (MDD) and Bipolar Disorder are three of the most prevalent. Despite significant research, the pathophysiology and underlying etiology of these diseases remains largely undiscovered. Recent research has highlighted the potential role of the gut microbiota in mental health, in particular in connection with anxiety. Our research aims to investigate this link in a cohort of GAD, MDD and bipolar patients. 71 GAD, 18 MDD, 17 euthymic MDD and 23 Bipolar patients provided fecal samples from which DNA was extracted, the V3 region of the 16S rRNA gene was amplified and sequenced using the Illumina Miseq platform. Sequencing data was analyzed through an in-house pipeline to construct community profiles of patients and age and sex matched healthy controls. My work involved analyzing the data to identify signature organisms that might identify putative disease associated microbial communities for further hypothesis generation about possible roles in disease. Furthermore, an extensive culturing effort was undertaken to identify and characterize some of the *Bacteroides* strains which were enriched in the GAD patient population. This study presents novel insights into some of the organisms that may be markers for a number of different diseases as well developing a better understanding of the *Bacteroides* that were correlated with anxiety.

Acknowledgements

I would like to begin by thanking my parents. From day one you've both supported and encouraged me to seek out my own opportunities and to apply myself to give back. I really feel I've grown as a scientist and as a person during these past two years and that opportunity was made possible through your support.

I would also like to offer an immense thank you to Merwa Amer, Jessica Gillard, Cindy James, Melanie Wolfe, Mathew Tersigni, the team in Toronto and everyone else who played a role in recruiting participants for the study. I have been incredibly fortunate to work on such a rich dataset and these data simply would not have existed without all your many hours of work.

I would like to thank Dr. Jen Stearns and Dr. Pat Schloss for their guidance in R. I had not imagined that I would become an amateur programmer when I began my graduate studies and yet I am so thankful for this skillset your guidance helped me to discover. Similarly, I would like to thank everyone I've worked, and likely complained alongside while trying to get code to work. In my experience programming is not a particularly pretty or easy road, but certainly one I've found to be very rewarding, and I would not have gotten nearly as far along the road without the support of programming and squash partner Jess Wallace.

To my friends back home who have supported me on weekends, helped with editing and humored my likely too lengthy discussion of feces, I continue to be humbled by how lucky I am to have friends like yourselves and I hope you will all be in my life for many years to come.

I would like to thank Dr. Deb Sloboda for her support on my committee. I have really enjoyed your mentorship and know I've improved immensely as a scientist through your guidance!

I of course am incredibly indebted to the members of the Surette lab for your guidance, edits, support and friendship. Pat, I'm incredibly thankful to have sat beside you in the office, on the way to Detroit and to have learned a great deal about science and life from you. Daph you have taught me about how to be a great scientist but also how to be a great friend, improve my organization and to always throw great parties. Steve, I feel I've learned a lot about lab and a little about life from you, and have had a lot of fun talking hockey in-between. Jen and Fiona, I have learned a lot about being scientists and grown-ups from you both and will certainly be a better professional for it. To our newer members Rachelle, Sharok, Brian and Ophelie, it has been a pleasure getting to know you although I've been busy, thank you for your kindness and best of luck going forward. Thank you to all of our undergraduate volunteers, in particular to Mr. Jeffery Ye and Ms. Kim Ng, simply put, without your help this project wouldn't have gone nearly as smoothly as it has.

I owe an incredible debt of gratitude to my supervisors Dr. Mike Surette and Dr. Rebecca Anglin. It has been a privileged to be your student. Thank you for all your time and support, I'm honoured to bring this chapter to a close, and while I'm elated to begin this next chapter I wouldn't have gotten to where I am today without you both. Thank you

I would also like to take a moment to recognize my wonderful roommate, occasional role model and constant friend for teaching me an immense deal about how to read, how to listen and how to live. I only hope I've offered you half as much as you've so generously given to me.

Finally, I would like to dedicate this work to my little sister Sydney. You are one of the reasons I became interested in this field of research and you are consistently my inspiration to keep at it. Your courage is rivaled only by your compassion and I'm truly blessed to call you my sister. Best of luck with everything that comes next kiddo, I hope I've made you happy in your heart 😊.

Abbreviations and Symbols

5-HT	5-Hydroxy Tryptophan (serotonin)
AMC	Amoxicillin-Calvulinic Acid
BBE	Bacteroides Bile Exculin Agar
BHI	Brain Heart Infusion
BP	Bipolar Disorder
BP1	Bipolar Disorder Type 1
BP2	Bipolar Disorder Type 2
CBA	Columbia Blood Agar
CBT	Cognitive Behavioural Therapy
CIP	Ciprofoxacin
CNA	Collistin Nalidixic Acid Agar
DMEM	Delbecco's Modified Eagle's medium
E	Erythromycin
eMDD	euthymic Major Depressive Disorder
FAA	Fastidious Anaerobe Agar
FCS	Fetal Calf Serum
FEP	Cefepine
FOX	Cefoxitin
GAD	Generalized Anxiety Disorder
H2O	Water
Igepal	octylphenoxypolyethoxyethanol (a detergent)
KCl	Potassium Chloride
LEV	Levofloxacin
LPS	Lipopolysaccharide
MDD	Major Depressive Disorder
MEM	Meropenem
MET	Metrinidazole
MgCl₂	Magnesium Chloride
PCR	Polymerase Chain Reaction
Penstrep	Penicillin Streptomycin
SXT	Trimethoprim-Sulfamethoxazole
TOB	Tobramycin
Tris-HCl	Tris-Hydrochloric Acid
VA	Vancomycin

Declaration of Academic Achievement

Ryan Potts contributed to the conception and design of experiments in this study, performed all the experiments, and analyzed and interpreted the data. Ms. Michelle Shah and Ms. Laura Rossi contributed to DNA extraction and sequencing of study participant samples.

Ms. Merwa Amer and Ms. Jessica Gillard recruited the participants for the study.

Mr. Saad Syed contributed to the analyses of whole genome data in chapter 3.

Drs. Michael Surette and Rebecca Anglin contributed to the conception and design of the experiments in this study and the interpretation of the results.

All experiments involving human samples were approved by the Hamilton Health Sciences research ethics board (HiREB) and written informed consent was obtained from all participants.

Table of Contents

Lay Abstract	iv
Abstract.....	v
Acknowledgements.....	vi
Abbreviations and Symbols	viii
Declaration of Academic Achievement	ix
Table of Contents	x
List of Tables.....	xii
List of Figures	xii
Chapter 1 Introduction:.....	1
1.1 Introduction:	1
1.2 Psychiatric Illnesses Addressed in this Thesis.....	2
1.2.1: Generalized Anxiety Disorder	2
1.2.2 Major Depressive Disorder	3
1.2.3 Bipolar Disorder	5
1.3 The Microbiota-Gut-Brain Axis.....	7
1.4 The Gut Microbiota affects the Brain and Behavior in Mice.....	10
1.5 The Gut Microbiome and Psychiatric Illness.....	10
1.6 Bacteroides	13
1.7 Experimental Questions and Hypothesis.....	15
Chapter 2 The GAD and MDD associated microbiomes:	16
2.1.1 Introduction.....	16
2.2 Hypothesis	16
2.3 Materials and Methods.....	17
2.3.1 Participant Recruitment	17
2.3.3 Sample Collection:	19
2.3.3 Sample Processing and DNA extraction.....	20
2.3.4 DNA Amplification and Sequencing.....	20
2.3.5 Data Processing for Analysis.....	22
2.3.6 Diet Data Processing.....	23
2.4 Results	24
2.4.1 Summary Statistics of Study Participants	24
2.4.2 Summary of Sample Data Processing.....	26
2.4.3 The Microbiome Distinguishes GAD patients from MDD patients and Healthy Controls.....	26
2.4.4 There is no significant difference in the variance of significantly different microbial community signatures.....	28
2.4.5 Metrics of alpha diversity do not distinguish participant samples	29
2.4.6 Abundance of specific genera distinguishes the microbiomes of patients and controls.....	31

2.4.7 Specific OTUs identify microbial communities associated with participant diagnoses.....	35
2.4.8 Marker OTU abundance appears to correlate with a measure of disease severity	41
2.5 Discussion	43
Chapter 2 Acknowledgments:.....	49
Chapter 3 The Bipolar and MDD microbiomes:.....	50
3.2 Hypothesis	51
3.3 Materials and Methods.....	51
3.3.1. Patient Recruitment	51
3.3.3 Sample Processing, DNA Extraction and Amplification and Sequencing.....	52
3.3.4 Data Processing for Analysis.....	53
3.3.5 Diet Data Processing.....	54
3.4 Results	54
3.4.1 Summary Statistics of Study Participants	54
3.4.2 Summary of Sample Data Processing.....	55
3.4.3 No evidence of significantly distinct microbial communities associated with bipolar disorder compared to healthy volunteers and depression patients.	56
3.4.4 Metrics of alpha diversity do not distinguish participant samples	57
3.4.5 Abundance of specific genera distinguishes the microbiomes of patients and controls.....	59
3.4.6 Investigating the identity of OTU 168.....	65
3.5 Discussion:.....	67
3.6 Chapter 3 Acknowledgements.....	73
Chapter 4: Analyzing <i>Bacteroides</i> strains isolated from a GAD patient and healthy volunteer	74
4.1 Introduction	74
4.2 Methods.....	75
4.2.1 Strain collection construction	75
4.2.2 Whole genome sequencing:	78
4.2.3 Growth Assays.....	78
4.2.4 Antibiotic resistance assay.....	79
4.2.5 Serotonin Production Assay	80
4.3 Results	82
4.3.1 Isolated strains represent a variety of species	83
4.3.2. Phenotypes distinguish strains, even within the same species	84
4.3.5 <i>Bacteroides</i> strains display variable antibiotic resistance within species	89
4.3.7 Analysis of whole genome sequence data suggests contamination	91
4.4 Discussion.....	92
Chapter 4 Acknowledgements	98
Chapter 5: Moving Forwards	99
5.1 Challenges in interpreting microbiome data.....	99
5.2 Experiments to continue the story.....	105
Works Cited	113

List of Tables

Table	Table Title	Page #
T1	Summary Statistics of eMDD, MDD, GAD and HC study participants	25-26
T2	Chapter 2 Data processing summary	27
T3	Summary statistics of Bipolar I, Bipolar II, MDD and HC study participants	54-55
T4	Chapter 3 Data processing summary	56
T5	Summary of efforts to confirm the identity of <i>Streptophyta</i> OTU 168	67
T6	Summary of the <i>Bacteroides</i> strain collection including 16s and whole genome sequencing IDs	84-85
T7	A summary of phenotypic variability observed across the <i>Bacteroides</i> strain collection	86-87

List of Figures

Figure	Figure Caption	Page #
1	Diagnosis explains significant separation of the GAD and HC -associated microbiomes	28
2	Within group variance is not significantly different between GAD, MDD and HC study participants	29
3	Measures of alpha diversity to not significantly distinguish the GAD, MDD, eMDD and HC microbiomes	30
4	The top 25 most abundant taxa at the genus level in all study participants	32
5	The top 25 most abundant taxa at the genus level averaged across the diagnostic groups of the GAD dataset	33
6	The genus <i>Bacteroides</i> is significantly enriched in GAD patient microbiomes relative to healthy controls	34
7	OTU 8 <i>Bacteroides</i> is significantly enriched in GAD patients in comparison with healthy controls	36
8	OTU 13 Ruminococcus is significantly enriched in patients relative to healthy controls	37
9	OTU 167 family <i>Rikenellaceaea</i> is significantly elevated in MDD patients relative to healthy controls	38
10	Enrichments of OTU 224 family <i>Erysipelotrichaceae</i> and OUT 237 genus <i>Clostridium</i> distinguish depression patients from healthy controls	39
11	OTU 109 family <i>Ruminococcaceae</i> and OTU 221 genus <i>Clostridia</i> are significantly enriched in MDD patients relative to healthy controls.	41
12	Abundance of GAD and HC distinguishing organisms correlate with a measure of disease severity	42
13	Disease does not explain a significant portion of variation observed within the bipolar microbiome dataset	57
14	Measures of alpha diversity to not significantly distinguish the GAD,	58

	MDD, eMDD and HC	
15	The top 25 most abundant taxa at the genus level in all bipolar study participants	60
16	The top 25 most abundant taxa at the genus level averaged across the diagnostic groups of the bipolar dataset	61
17	An enrichment of <i>Sarcina</i> distinguishes healthy controls from bipolar patients	63
18	Differing abundance of the genus <i>Collinsella</i> distinguish bipolar I & II patients	64
19	The family <i>Lachnospiraceae</i> and genus <i>Lachnospira</i> are more abundant in MDD patients when compared with bipolar patients	65
20	Bipolar patients are enriched in the order <i>Streptophyta</i> when compared with healthy controls	66
21	Photographic evidence of variable esculin hydrolysis within the <i>Bacteroides</i> collection	88
22	Photographic evidence of agarose activity from strains 2 and 28 occurring on Columbia Blood Agar	89
23	A side view of the plat in figure 22 to try to better demonstrate agarase activity	90
24	Antimicrobial resistance is variable across the <i>Bacteroides</i> strain collection	91
25	Antimicrobial resistance is variable within the WGS assigned species <i>B. thetaiotamicron</i>	92
26	Screen capture of progressive mauve analysis of putative contamination in strain 28 WGS data	93

Chapter 1 Introduction:

1.1 Introduction:

A subject of increasing popularity in healthcare is that of the community of microorganisms living on and inside the human body, commonly termed the human microbiome. Microorganisms are found throughout the digestive tract, in particular the colon, where estimates are that there are nearly 10^{13} bacterial cells, approximating the number of human cells found in the body ¹. In the GI tract alone, this diverse community is estimated to be comprised by over 1 000 different species of bacteria containing nearly 150 times more genes than their human hosts ². The gut microbiome also contains multitudes of viruses (predominantly bacteriophage, or bacterial viruses) along with fungi and archaea³⁻⁵. While it is probable that the non-bacterial microbes play a role in maintaining composition of the gut microbiome, investigation thus far has focused primarily on the bacterial residents of the digestive tract.

As technology has advanced to allow us to explore and understand the microbial communities that inhabit our bodies, we have begun to unravel the complex nature of their interactions with each other, and with their hosts. Microorganisms are now understood to play a variety of roles in human health including regulating fat storage⁶, producing vitamins⁷, offering protection against some pathogens⁸, bile acid metabolism⁹⁻¹¹, in the development of a functional immune system¹² and potentially much more. One area of particular recent interest is the study of putative connections between the bacteria that inhabit our digestive systems (the gut microbiome) and mental health¹³⁻¹⁶. This thesis has investigated putative microbial signatures in the gut community of patients with Generalized Anxiety Disorder (GAD), Major Depressive Disorder

(MDD), and Bipolar Disorder patients along with further investigation into some of the organisms that may be associated with GAD.

1.2 Psychiatric Illnesses Addressed in this Thesis

1.2.1: Generalized Anxiety Disorder

Generalized Anxiety Disorder (GAD) as defined by the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), is a chronic disorder in which the patient struggles with excessive and uncontrollable anxiety and worry, more days than not, over a six month period, about a number of events or activities in their lives¹⁷. Patients report impairment in occupational, social or other significant areas of function resulting from their worrying¹⁷. Furthermore, their anxiety and worry must be accompanied by at least three of the following symptoms; restlessness, becoming easily fatigued, concentration difficulties, irritability, muscle tensions and disturbed sleep. Finally, the anxiety and worry must not be confined to the features of another disorder, nor a prior medical intervention or substance dependency¹⁷.

GAD frequently co-occurs with Major Depressive Disorder (MDD), Dysthymic Disorder, other anxiety disorders (eg. Panic and Social Phobia) and with substance related disorders. As well, GAD patients frequently report other stress related conditions including headaches and irritable bowel syndrome (IBS)¹⁷. Between fifty-five and sixty percent of patients who are diagnosed with

GAD are female¹⁸. Both cognitive behavioral therapy (CBT) and medicinal intervention with selective serotonin reuptake inhibitors (SSRI) have been shown to ameliorate this condition¹⁷. Current guidelines for the treatment of GAD recommend that first-line treatment for patients consist of antidepressants, either a SSRI or a selective serotonin noradrenaline reuptake inhibitor (SNRI). In greater than fifty percent of diagnosed cases GAD onset is after 20 years of age¹⁷. The course of the disorder is chronic but fluctuating, with the condition of patients typically worsening during periods of elevated stress. Evidence for genetic factors influencing the occurrence of GAD in patients is thus far limited, however it has been suggested that factors influencing the risk of GAD may be associated with those for major depressive disorder (MDD)¹⁷.

1.2.2 Major Depressive Disorder

Major Depressive Disorder (MDD) as defined by the American Psychiatric Association's Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), is a chronic disorder in which the patient struggles with persistent low mood, loss of enjoyment and interest in routine activities, resulting in varying levels of occupational and social disruption¹⁷. Patients report at least five of the following nine symptoms, and at least one of the first two, having been present during the same two-week period and, that these symptoms must represent a deviation from previous function. The symptoms include: 1. Depressed mood; 2. Markedly diminished interest or pleasure in most or all activities; 3. Significant weight loss when not dieting, loss of appetite or weight gain; 4. Insomnia or hypersomnia; 5. Psychomotor agitation or retardation; 6. Diminished ability to think or concentrate or indecisiveness; 7. Feelings of worthlessness or

excessive or inappropriate guilt; 8. Fatigue or loss of energy, and 9. Recurrent thoughts of death (not simply a fear of dying), often accompanied by recurrent suicidal ideation with or without a specific suicide plan or a suicide attempt¹⁷. Symptoms one through eight must be observed by the patient or an observer nearly every day throughout the two-week span. The episode must not be attributable to the effects of substance misuse or another medical condition. Previously, bereavement had been an exclusionary clause for diagnosis of MDD, however, the DSM-5 now stipulates that a diagnosis of a depressive episode may be made following a significant loss at the discretion of the supervising clinician¹⁷. Further a manic or hypomanic episode would shift a patient's diagnosis to one of Bipolar Disorder including if the episode is induced or attributable to a medication or other mind-altering substance. A variety of treatment options are available to MDD patients with varying levels of success, however, as with GAD, combined or stand-alone pharmaceutical and psychotherapeutic interventions are currently the first-line treatment options¹⁹.

Patients who have had a prior depressive episode but currently remitted are referred to as euthymic (eMDD)²⁰. Although eMDD patients are not currently depressed, there is some evidence to suggest these patients maintain different personality traits and clinical characteristics than mentally well individuals and they may have future episodes of depression²¹. As in the case of GAD, the cause of MDD is unknown. Some early results suggested a possible link between mutations in the serotonin transporter gene (5-HTTLPR) and increased risk for depression, however, a recent meta-evaluation of studies on this putative link found no significant evidence for 5-HTTLPR mutation and increased risk of MDD²². Another proposed mechanism for the cause of MDD is the monoamine hypothesis. It suggests that depression may

result from decreased levels of neurotransmitters like serotonin²³. This hypothesis stems from the observation that some effective antidepressants enhance synaptic levels of these molecules whilst also improving depressive symptoms. The observation that through pharmaceutical intervention monoamine levels can be corrected in patients weeks before there is a significant improvement in symptomology, however, suggests that other systems may be involved in MDD outside of neurotransmitter imbalance²³. Other systems could include the immune system. This idea is based on the observation that LPS administration in rodents stimulates the release of pro-inflammatory cytokines and subsequent depression-like behaviour^{23,24}. Another potential system which could be involved in MDD is the hypothalamic-pituitary-adrenal axis of HPA-axis²⁵. Briefly, the HPA-axis consists of a series of stimulatory hormones and feedback loops and is involved in the body's response to stress²⁵. HPA-axis stimulation results in the production of glucocorticoids like cortisol which bind intracellular glucocorticoid receptors which then migrate to the nucleus and modify gene transcription and thus regulate cellular processes²⁵. MDD patients have been observed to have higher serum levels of glucocorticoids^{26,27}. Sudo et al. observed that germ-free mice had more HPA-axis activity in response to restraint stress when compared with conventionally colonized mice demonstrating critical role for the microbiome in HPA-axis development and thus potentially for its role in disorders in which the HPA-axis is perturbed²⁸. Accordingly, there is a currently significant interest in the potential role of the microbiome in depression potentially through mediating inflammation or through effects on the HPA-axis.

1.2.3 Bipolar Disorder

Bipolar disorder, more accurately described by its two sub-divisions, Bipolar Disorder I and Bipolar Disorder II are psychiatric disorders in which patients experience episodes of hypomania or mania and may also experience depression¹⁷. The two subtypes of bipolar disorder are distinguished on the basis of patients experiencing manic (type I) and hypomanic (type II) episodes¹⁷. During mania, patients experience a combination of symptoms including inflated self-esteem or grandiosity, decreased need for sleep, more frequent and sometimes pressured talkativeness, flight of ideas or racing thoughts, lack of focus or increased distractibility, an increase in goal-directed activity and excessive involvement in high risk behaviors or activities for example, spending sprees, gambling, and increased promiscuity or sexual indiscretion. These symptoms are observed during a period of persistently elevated (requiring at least three observed symptoms for diagnosis), expansive or irritable (must be accompanied by four observed symptoms for diagnosis) mood lasting at least a week and present most of the day¹⁷. Episodes in which these symptoms are so overwhelming that they significantly impair patients' capacity to function in either occupational or social settings are called mania (or manic episodes)¹⁷. Conversely, hypomanic episodes are ones in which patients experience similar symptoms but to a lesser degree without marked impairment in function¹⁷. Contrasting the periods of mania (and hypomania) patients generally also struggle with bouts of depression during which they experience many of the previously described symptoms of depression. . The main stay of treatment of Bipolar disorder is medication, including mood stabilizers, antipsychotics and antidepressants.²⁹ Furthermore, therapy has been found to be helpful in treating specific symptoms, however, the particular therapy prescribed is often case specific²⁹. One of the challenges with pharmaceutical intervention in bipolar disorder is poor adherence, often due to challenging side effects like weight gain and decreased sexual drive²⁹. Accordingly,

investigations which could yield a better understanding of the disorder's etiology could be helpful in better tailoring treatment regimens, potentially reducing polypharmacy and through this increasing adherence. Finally, emerging evidence supports a continuity or spectrum between bipolar disorder II and MDD³⁰. Accordingly, whilst this study is a pilot, it may offer an interesting opportunity to investigate co-occurrence of microbiome signatures across MDD and bipolar disorders³⁰.

1.3 The Microbiota-Gut-Brain Axis

Whilst the study of putative connections between the microbiome and mental health has gained particular traction in last decade, studies suggesting this link are now more than 100 years old³¹⁻³³. For example, in 1910, Dr. G. Phillips published his results outlining the successful treatment of depression (then called melancholia) with a cocktail of lactic acid bacteria³¹. More recently, there has been growing interest in the microbiota-gut-brain axis, which refers to the many signaling pathways connecting the brain to the gut and its resident bacteria²⁵. The microbiota-gut-brain-axis consists of bidirectional connections including signaling via the immune, humoral and neuronal systems²⁵. Host associated bacteria can communicate with the brain in a variety of ways including some species which can either influence secretion or metabolism of neurotransmitters including dopamine, serotonin and gamma-aminobutyric acid (GABA)³⁴⁻³⁶. Interestingly, recent work by Yano and colleagues showed a relationship between spore forming bacteria from both the human and murine gut microbiomes and their capacity to stimulate serotonin production³⁶. Indigenous spore forming bacteria were found to stimulate serotonin

production in enterochromaffin cells. Additionally, specific fecal metabolites from these bacteria could replicate this effect independent of bacterial presence. Furthermore, it has been shown that germ-free mice have significantly lower levels of circulating serotonin when compared with traditionally colonized specific pathogen free (SPF) mice³⁷. These findings represent just a few examples demonstrating the link between the gut microbiome and the host's production of important neurological signaling molecules.

A significant portion of the microbiome to brain communications appear to happen via the vagus nerve, and numerous studies have been conducted in rodents to demonstrate the importance of this pathway. Research has demonstrated the critical role of vagal signaling in mediating response to acute intestinal infection and inflammation^{38,39}. Similarly, vagotomized rodents lose the putative probiotic anti-depressant and anxiolytic effects observed following the administration of species like *Bifidobacterium longum* and *Lactobacillus rhamnosus* JB-1^{40,41}. Further, the probiotic *Lactobacillus reuteri* was shown to accelerate wound healing through triggering oxytocin release from the hypothalamus⁴². Interestingly, vagotomy also prevented development of anxious behaviour in DSS-colitis model mice⁴⁰. These results serve to demonstrate the importance of the vagus in both detrimental and beneficial signaling between the gut and brain. It is important to note, however, that the vagus nerve is not the exclusive conduit for gut to brain communication as other work involving the probiotic effects of *Lactobacillus reuteri* and *Bifidobacterium infantis* was shown to be independent of vagal integrity⁴³. Furthermore, vagotomized mice were observed to have similar increases in anxiety-like behavior when compared to sham controls following chronic parasite infection⁴⁴. Accordingly, further

investigation is required to fully understand the many channels through which the gut microbiota can communicate with the brain.

Evidence for the bi-directionality of communication must demonstrate the capacity of the microbiota to alter the brain, and reciprocally, the capacity of the brain to impact the microbiome. Accordingly, studies demonstrating that stress can alter the microbiome have been important in demonstrating this capacity^{45,46}. These changes are typically correlated with shifts in levels of inflammatory biomarkers⁴⁷. Furthermore, increases in intestinal permeability have been demonstrated in association with stress⁴⁵. During periods of increased intestinal permeability more bacteria are able to cross the epithelial barrier of the gut and interact with the host immune and enteric nervous systems⁴⁷. Additionally, studies have demonstrated that the gut microbiota contribute to programming the development of the stress response during a critical period in early life²⁸. Another route through which the brain can interact with the microbiome is through secretion of cortisol⁴⁸.

Additional evidence supporting the link between the microbiota and the brain come from shifts in the intestinal community observed in frequently occurring gastrointestinal disorders like irritable bowel syndrome (IBS) which often co-occur with psychiatric illness^{49,50}. Furthermore, through the use of animal models, research has now suggested a role for the commensal microbiota in advancing disease progression in the auto-immune neurological disorder, multiple sclerosis⁵¹.

1.4 The Gut Microbiota affects the Brain and Behavior in Mice

Use of mouse models has enabled extensive study of changes, both behavioral and biochemical, in the host following alteration to the gut microbiota. Studies comparing germ-free or gnotobiotic mice to specific pathogen free mice (SPF) show significant changes in behaviors associated with the anxiety-phenotype⁵². Tests of the anxiety phenotype include behavioral tests like the elevated plus maze test or the light-dark box test⁴⁸. Increases and decreases in anxiety phenotypes have been observed in mice with specifically altered gut microbiota⁵². These findings are corroborated by associated shifts in neuronal signaling molecules like brain derived neurotrophic factor (BDNF)⁵². Furthermore, interventions including the addition of antibiotics and probiotics to the diet of mice have also been shown to change behavior and signaling molecule levels^{40,41}. Thus, the evidence of changes, both behavioral and biochemical, in mice with initially different microbiota, and with microbiota altered through intervention, strongly support the link between the gut microbiota and the brain.

1.5 The Gut Microbiome and Psychiatric Illness

In spite of the mounting body of evidence linking mental health disorders and the gut microbiome in animals, there remains a dearth of projects focusing on links between the gut microbiota and psychiatric illness in humans. The majority of work to date has been conducted in healthy volunteer populations, looking at interventions like probiotics. Promisingly, a

combination of probiotic strains of *Bifidobacterium longum* and *Lactobacillus helveticus* was shown to reduce psychological distress and to change brain activity in response to emotional stimuli in healthy volunteers⁵³. This supported a study investigating the effects of consuming a probiotic fermented milk (containing strains of *Bifidobacterium*, *Streptococcus*, *Lactobacillus* and *Lactococcus*) by healthy women which demonstrated a capacity to positively alter their response to negative stimuli⁵⁴. Finally, probiotic pills and probiotic yogurts were independently found to have positive effects on the mental health of petrochemical workers¹⁵.

One disorder that has received particular attention with regards to the gut microbiome is autism spectrum disorder (ASD). Interest in the gut microbiome of children with ASD was inspired in part by the observation of high gastrointestinal affliction reported in these patients⁵⁵. Autistic children have been observed to have greater bacterial diversity in their gut communities when compared to neurotypical children⁵⁶. Of particular interest is the observation that levels of some *Clostridia* are higher in autistic children when compared with neurotypical children, and furthermore, that levels of these *Clostridia* appear to correlate with severity of autistic symptoms^{57,58}. Additional evidence supporting the putative role of *Clostridia* in ASD came from a study which treated autistic children with vancomycin, an antibiotic effective against the *Clostridia* amongst other bacteria⁵⁹. The children's symptoms did appear to ameliorate during the course of the antibiotic in an open study, however, these improvements quickly waned following stopping the drug. A new study published at the start of 2017 investigated the effect of fecal microbial transplant in an open label trial with autistic children, and while it was only a small trial, the results seem to reinforce the significant role of the microbiome in autism⁶⁰.

Preliminary efforts have been undertaken to characterize the gut microbiome of major depression disorder (MDD) patients^{61,62}. These studies fail to agree on a consistent signal or microbial community amongst MDD patients. In the first study by Naseribafrouei and colleagues it was reported that the *Bacteroidetes* were underrepresented in depression patients relative to the controls. Conversely, Jiang *et al.* found the *Bacteroidetes*, as well as the *Proteobacteria* and *Actinobacteria* to be over represented in their depressed cohort. This could be influenced by a number of factors including the relatively small sample size of these studies and the heterogeneity of the disorder. Further, the first study by Naseribafrouei and colleagues compared MDD patients to outpatients from a neurological unit, contrasting the study by Jiang who compared to healthy subjects^{61,62}. Additionally, Jiang's depressed cohort contained both active and remitted depression patients. A further potential confounder in comparing these studies is diet, as Naseribafrouei's participants were recruited in Norway whilst Jiang's were recruited in China. Unfortunately, neither study collected dietary data from patients and thus it is impossible to estimate and effect the difference between traditional Western and Eastern diets may have had on these participants. Interestingly, Maes *et al.* demonstrated increased circulating levels of immunoglobulins A and M (IgA & IgM) against the LPS of *Enterobacteriaceae* in depressed patients. This could suggest increased translocation of bacteria or bacterial products like lipopolysaccharide (LPS) across the gut of MDD patients into the circulatory system promoting an inflammatory response that could contribute to the depression⁶³. This study further explores the microbiome in MDD patients as well as investigating the microbiome in GAD and Bipolar disorder, both of which are believed to be novel investigations.

1.6 Bacteroides

Organisms from the genus *Bacteroides* were found to be of significance in some of these investigations and thus are largely the focus of chapter three. Accordingly, the following provides a brief background on this genus.

Bacteroides are Gram-negative bacteria and one of the many genera within the phylum *Bacteroidetes*, one of the two major bacterial phyla in the human gut⁶⁴. Initial interest in this phylum within the context of gut health is likely due to the *Bacteroidetes* to *Firmicutes* ratio, in particular, this ratio possibly influencing obesity⁶⁵. The phylum *Bacteroidetes* contains a considerable number of including the genera *Bacteroides*, *Parabacteroides*, *Alistipes* and *Prevotella*⁶⁴. The genus *Bacteroides* consists of Gram-negative bacilli that do not form spores⁶⁶. Motility is variable across species within the genus⁶⁶. Current literature suggests this group of organisms to be obligate anaerobes⁶⁶.

The *Bacteroides*, like many of the genera found in the human digestive tract are understudied, however, they have been recently proposed as an ideal model organism for studying dynamics in the gut community⁶⁷. One reason this group is of particular interest is their incredible capacity for carbohydrate metabolism which may be both beneficial and detrimental to the host⁶⁸. Recent publication showing different gut derived strains of *Bacteroides* to be capable of almost completely digesting the most structurally complex glycan, rhamnoglacturonan-II, and a variety of algae derived polysaccharides including carrageenan and agarose demonstrate the phenomenally diverse carbohydrate metabolism of this genus^{68,69}. Through degrading

carbohydrates that cannot be used by the host, these bacteria may flourish and help the host through remodeling of the host community with an extensive arsenal of bacteriocins and the Type VI secretion system^{70,71}. Their extensive capacity for carbohydrate production, however, may also be detrimental to the host as they can degrade mucins when they are starved of other carbohydrates⁷². This reduces the mucous barrier protecting host cells thus exposing them to potential pathogens like *Escherichia coli*. Interestingly, however, one member of this genus, *B. thetaiotomicron*, has also been found to sequester vitamin B₁₂ and in doing so limit the pathogenicity of some *E. coli* strains⁷³. Similarly, a number of studies have demonstrated interactions between *Bacteroides* and the immune system. The *Bacteroides* have been shown to be anti-inflammatory in a variety of conditions including inflammatory bowel disease^{74,75}. As well, the *Bacteroides* are believed to play an important role in immune system maturation; however, evidence suggests the timing of this interaction may be important. A study of the gut microbiome in eastern European children suggested that early exposure to the lipopolysaccharide of *Bacteroides*, could put children at a greater risk for developing diabetes⁷⁶. Later in life, however, Ivanov et al. demonstrated an important role for the *Bacteroides* in stimulating anti-inflammatory Th-17 immune cells in the host⁷⁷. Thus, though these data may seem contradictory, they serve to highlight the complex nature of microbe-microbe and microbe-host interactions within the gut community, as well as the potential diversity within this group of bacteria. Further, they suggest a number of hypotheses about how these organisms could influence the microbiota-gut-brain axis. Accordingly, my research was intended to better understand some of the species within this genus and their putative role in influencing microbiota-gut-brain communications.

1.7 Experimental Questions and Hypothesis

Given the relative dearth of information surrounding the etiology of GAD, MDD and Bipolar disorder, and of data regarding the specific microbial communities found in these conditions, my research aimed to preliminarily categorize these communities. Further, in response to some pilot data collected early in my work, I hoped to isolate and better understand some of the *Bacteroides* that were found to be overrepresented in the GAD patient population such that I might be able to generate hypotheses about how these organisms contribute to the onset, or perpetuation of the anxious state in patients. My experiments were designed to test the hypothesis that the gut communities of GAD, MDD and bipolar patients are significantly different from those of healthy controls. Further experiments were undertaken to investigate heterogeneity within and between closely related *Bacteroides* species to generate hypotheses about putative mechanisms of influencing anxiety in the host.

Chapter 2 The GAD and MDD associated

microbiomes:

2.1.1 Introduction

Given the increasing prevalence of mental health disorders (especially Major Depressive Disorder and Generalized Anxiety Disorder), the dearth of information regarding their etiology, the observation of increased rates of comorbid gastrointestinal distress in these patients, and increasing evidence supporting putative role(s) for the gut microbiota in modulating brain function and behaviour, this study was undertaken in an effort to investigate the bacterial communities of mental health patients to see if they are distinct from those of healthy volunteers. Further analyses were conducted to determine which bacteria might be distinguishing patient communities from those of healthy volunteers in an attempt to generate hypotheses about how organisms of interest could be impacting their host.

2.2 Hypothesis

I hypothesize that the composition of the microbiome will significantly differentiate healthy volunteers from participating patients with diagnoses of GAD, MDD and eMDD, furthermore that the composition of the microbiome will also differentiate patients of differing diagnoses.

2.3 Materials and Methods

2.3.1 Participant Recruitment

Between June 2014 and 2016, participants were recruited through several venues within the St. Joseph's Healthcare system in Hamilton, Ontario. Advertisements for participants were placed throughout the city of Hamilton, in particular at three of the city's hospitals and throughout the campus of McMaster University. All advertisements were approved by the HiREB. Additionally, participants were approached during the second week of a twelve-week cognitive behavioral therapy (CBT) course offered through the Anxiety Treatment and Research Center and Mood Disorders Clinic, both located 100 West 5th St. Campus of St. Joseph's Hospital. Patients actively dealing with GAD (Principal diagnosis of Generalized Anxiety Disorder diagnosed using the Mini International Neuropsychiatric Interview (MINI) 6.0⁷⁸), major depressive disorder (MDD) (confirmed via the Montgomery Asberg Depression Rating Scale – MADRS score >19) and patients with remitted or euthymic MDD (eMDD) (MADRS <8) were all recruited. The study aimed to recruit a total of 100 GAD patients, and 30 of both MDD and eMDD patients. The study was forced to conclude prior to successfully meeting recruitment targets and thus final populations of 71 GAD patients 17 MDD patients and 16 eMDD patients were recruited. Additionally, healthy volunteers were recruited with the goal to age and sex matching patient participants. A match of age was acceptable within five-year window of the matched patient's age. In total 86 eligible healthy controls (HC) completed the study. Fifty-one GAD patients were matched by healthy controls. For the analyses conducted in this chapter, these fifty-one matched controls, along with the remaining controls were included such that the final analyzed data set consists of 71 GAD patients, 86 healthy controls, 18 MDD patients and 17 eMDD patients.

Inclusion criteria for participants stipulated that patients needed to be between the ages of 18 and 65, that they were able to provide and sign informed consent, and finally that they have a principal diagnosis of GAD, MDD or eMDD using the Mini International Neuropsychiatric Interview (MINI) 6.0⁷⁸. Inclusion criteria for healthy controls were similar, however, healthy controls were required to have no psychiatric illness. Exclusion criteria for participants included:

1. Antibiotic or probiotic use 4 weeks prior to participation (this included the consumption of probiotic fortified dairy products)
2. A competing psychiatric diagnosis that was felt to be principal,
3. Having a diagnosis of, or fulfilling the diagnostic criteria for a lifetime history of bipolar disorder, schizophrenia or other psychotic disorders, delirium, dementia and amnesic or other major cognitive disorders
4. Active substance abuse
5. Diagnosis or any major organic medical illness that could reasonably be believed to impact the gut microbiome. Specifically, any organic gastrointestinal disorder (e.g. Crohn's disease, ulcerative colitis), autoimmune rheumatological disorder (e.g. Rheumatoid arthritis) or immunological disorder (e.g. Lupus).
6. Active eating disorder
7. Pregnancy or breastfeeding

Informed consent was obtained from all participants either in person, or over the phone.

Participants were then screened over the phone for study eligibility. Eligible participants were required to undergo the following:

1. Collection of background information including demographics, past medical and specifically psychiatric history, medication and allergies and Body Mass Index (BMI)
*BMI data was only collected starting in August of 2015 and thus BMI data is unavailable for 51 of 71 GAD participants and 41 of 86 healthy controls. BMI data was collected from all MDD and eMDD participants.
2. Completion of standardized rating scales including the depression anxiety and stress scale (DASS) ⁷⁹, the Childhood Trauma Questionnaire (CTQ) ⁸⁰, the Penn State Worry Questionnaire (PSWQ) ⁸¹, Hamilton Depression Rating Scale (HAMD) ⁸² and Montgomery Asberg Depression Rating Scale (MADRS) ⁸³
3. Assessment of GI symptoms using the gastrointestinal symptom rating scale (GSRS), short form Leeds dyspepsia questionnaires (SF-LDQ) ⁸⁴ and Rome III ⁸⁵ criteria for IBS
4. Completion of the clinical global impression severity scale (CGI-S) ⁸⁶ for severity of GAD (for GAD patients only)
5. Assessment of diet using the simplified food frequency questionnaire ⁸⁷
6. Collection of fecal and blood samples as later described.
*All analysis of inflammatory biomarkers in participants' blood was performed by another student and thus is not included in this analysis.

2.3.3 Sample Collection:

Patients were provided the requisite materials and instructions to safely collect fecal material in a sterile screw capped sample jar. Participants were provided with a collection kit that also contained an anaerobic sachet, ice pack and sterile plastic bag such that the sample could be transported anaerobically (in an effort to preserve viability of the multitude of anaerobic gut

bacteria) and at a relatively controlled temperature to the lab. Most samples were transported with the help of Hamilton's Blue Line taxi company; however, some participants delivered samples directly to the hospital. Ideally, samples were delivered and processed within four hours of collection. In some rare cases, however, accommodations were made for patients such that they could provide a frozen sample rather than a fresh sample. In these instances, samples were not transported anaerobically as this is not known to have any significant effect on DNA integrity.

2.3.3 Sample Processing and DNA extraction

Once samples arrived to the lab, they were processed anaerobically (under 85% N₂, 10% H₂ and 5% CO₂). Approximately 100 mg of sample was added to a genomic prep tube containing 100 uL GES buffer (0.1 M glycine-NaOH, pH 9.0, 50 mM NaCl, 1mM EDTA, pH 8.0, 0.5% (v/v) Triton X-100) and 800 uL NaPO₄ buffer as well as 0.1mm glass beads and 2.8 mm ceramic beads. Further, extra fecal sample was stored in up to eight cryo-vials (amount of sample dependent, approximately 2g fecal sample per tube) and stored at -80 °C. Initially, the protocol stated that two cryo-vials of extra sample should be stored; this was increased to eight in October of 2015 to accommodate a greater number of potential downstream experiments. DNA was then extracted from the sample prepared for genomic extraction using an in-house protocol on a MagMaxTM robot (Thermo Fischer Scientific) and final concentration was measured using a NanoDrop2000 (Thermo Fischer Scientific).

2.3.4 DNA Amplification and Sequencing

PCR was performed using an in-house method heavily adapted from Bartram et al.⁸⁸ In a sterile, PCR workstation, the following mix was prepared for each sample: 5µl 10x PCR buffer (Life Technologies, #10342020), 1.5µl 50mM MgCl₂ (Life Technologies, #10342020), 2µl 10mg/mL BSA (made in ddH₂O, aliquoted and irradiated on a UV transilluminator for 30 mins to eliminate and contaminating DNA), 1µl 10mM dNTPs (New England Biolabs, #N0447L), 5µl 1µM V3F barcoded primer (5 pmoles), 5µl 1µM V3R_mod2 primer (5 pmoles), 0.25µl Taq polymerase, xµl template DNA (30-50ng) (This varied between 0.5 µl and 2 µl dependent on sample concentration) and finally sufficient dH₂O to bring final reaction volume to 50ul. Once mixed, samples were split into triplicates (16.7µl each) for the reaction the PCR program was as follows: two minutes at 94°C for initial denaturation followed by twenty-five cycles of thirty seconds at 94°C, thirty seconds at 50°C and thirty seconds at 72°C, followed by a final elongation during which the reaction was held at 72°C for ten minutes. 5µl of each sample from the reaction was run on a 2% agarose gel to visualize the expected 150bp product.

All PCR products were then sequenced on the Illumina MiSeq (San Diego, CA). Illumina sequencing involves the attachment of target DNA molecules to the surface of a flow cell. DNA molecules are replicated through bridge amplification many times resulting in clonal amplification of all loaded DNA molecules. Reverse read molecules are washed away and then forward reads are sequenced. Sequencing occurs in a stepwise fashion with fluorescently labeled cytosine, thymine, guanine and adenine added to the sequence strand based on complementarity with the template. Nucleotides are added one at a time, and after each round of addition, a fluorescent image is taken. Given the unique fluorophores on each of the nucleotides, sequences can be composed based on the sequence of fluorescent images. Special modifications to the 3'

end of the nucleotides prevent the addition of more than one nucleotide at a time. Following each fluorescent image, termination chemistry is reversed so that another cycle of nucleotide addition can occur. Current Illumina MiSeq technology allows for up to twenty-five million reads with two times three hundred (forward and reverse read) base pair read lengths per run. The addition of barcodes during sample preparation allows for multiple samples to be sequenced on a single run.

The data was then processed by an in-house bioinformatics pipeline (SLIP, publication is forthcoming) that incorporates quality filter using Sickle (q-score > 30)⁸⁹, Cutadapt (which removed adaptor sequences from the Illumina sequencing)⁹⁰, PandaSeq (which aligned forward and reverse reads)⁹¹, AbundantOTU (which assigns OTU identities to grouped reads)⁹² and QIIME⁹³. Taxonomic assignments were made using the RDP classifier with the Greengenes 2011 training set⁹⁴.

In total 191 samples were sequenced to an average depth of approximately seventy-thousand reads. (min: 19 reads, max: 171 543 reads).

2.3.5 Data Processing for Analysis

Prior to analysis the data set was curated in the following ways, firstly, samples with less than eight thousand reads were eliminated from the data set. Seven samples were trimmed from the dataset as they had been sequenced on a run that was identified as having poor quality by the sequencing center (all samples had fewer than 10 000 total reads). The seven samples represented four healthy controls and one each of the GAD, MDD and eMDD patient samples.

Sequences from the remaining 184 samples were then filtered such that OTUs in the data set had to be present in at least 10% of the original data set (19 samples), had to have a minimum read count of at least thirty across the entirety of the data set and finally had to be present at a prevalence of at least one hundredth of a percent within the data set. These curations were made in an attempt to minimize the effect of sequencing noise on the data in order to achieve the most reliable results. Stringency cutoffs were established based on previous work completed within the Surette laboratory as well as discussion with the supervisors. Data set curation was performed in MACQIIME ⁹³ using the functions `filter_samples_from_otu_table.py` and `filter_otus_from_otu_table.py`. This curation resulted in a final data set used in analyses with the following characteristics: 184 samples (70 GAD, 17 MDD, 15 eMDD, 83 Healthy Controls) with 235 observed OTUs, 12796062 total reads, average read count per sample 69544 (min: 16941 reads, max: 151 646 reads). Analyses were conducted in R ⁹⁵ using a custom script incorporating the ANCOM ⁹⁶, phangorn ⁹⁷, phyloseq ⁹⁸, vegan ⁹⁹, ggsignif ¹⁰⁰, and stats ⁹⁵ packages. Figures were composed in R using the ggplot2 ¹⁰¹, Hmisc ¹⁰², gridExtra ¹⁰³, and cowplot ¹⁰⁴ packages unless otherwise stated.

2.3.6 Diet Data Processing

Diet data was collected using the DQES v2 Food Frequency Questionnaire provided by the Cancer Council Victoria in Australia ⁸⁷. Completed questionnaires were sent to Australia for processing. Food intake, measured in g/day, was assessed for 102 different food stuffs. Individual food stuffs were grouped into thirty food groups (eg. The food group red meat consisted of beef, veal, lamb, pork, ham, meat pies and hamburgers) upon which a principal component analysis (PCA) was completed using the psych package in R. The PCA allows

investigators to identify food whose consumption is correlated (loading score > 0.2), and in doing so identify putative dietary patterns. For example, one of the identified dietary patterns was characterized by having high correlations in consumption of high and low FODMAPS fruits and vegetables, leafy greens, eggs and dairy and thus likely represents a vegetarian diet. These dietary patterns are then mapped back against individual participant data to determine with which pattern their diet most closely conforms. This allows the investigator to assign a dietary pattern to each participant such that putative patterns of diet can be compared between groups of participants. In total four dietary patterns were identified, and these values were then used as a categorical variable in a PERMANOVA to investigate putative associations between dietary pattern and microbiome.

2.4 Results

2.4.1 Summary Statistics of Study Participants

Table one shows the demographic data of patient and healthy volunteer participants in the study. Given the effort to match GAD patients and healthy controls, both cohorts are nearly identical in terms of average age, as well as in percentage of female participants. The approximate 2:1 ratio of females to males reported in GAD and MDD populations is true also in our dataset¹⁷. Interestingly, the eMDD participants are composed of an equal number of male and female volunteers; however, this is likely the result of the relatively small participant sample size. Participants using birth control is reported as a percentage of potentially eligible participants. Participants were identified as eligible to be using birth control if they were a

female under 52 (the average age at which women begin menopause according to the Canadian Women’s Health Network) years of age¹⁰⁵.

Table 1: Summary statistics of eMDD, MDD, GAD and HC study participants

	Participant Diagnosis			
	Euthymic	MDD	GAD	HC
n	16	18	71	86
n (female)	8	13	59	57
% Female	50.00	72.22	83.10	66.28
n with Height and Weight Data	16	18	50	56
Average Age (IQR)	40.38 (35.00)	43.50 (17.25)	37.70 (17.50)	35.51 (25.00)
Average Height cm (IQR)	164.13 (13.75)	168.28 (11.88)	163.45 (12.50)	171.35 (48.13)
Average Weight lb (IQR)	161.94 (40.75)	199.42 (104.00)	172.20 (55.25)	158.36 (53.50)
Average BMI (IQR)	27 (7)	32 (17)	29 (9)	24 (6)
Average PSWQ	47	59	66	35
Average Depression Score (DASS)	9	29	16	1
Average PSWQ Group	Moderate	Moderate	High	Low
Average Depression Score Group (DASS)	Normal	Extreme	Moderate	Normal
Percentage of Participants Reporting Cigarette Smoking (n)	6 % (1)	11 % (2)	10 % (7)	6 % (5)
Percentage of Eligible Participants Reporting Birth Control Medications (n/n possible)	33 % (1/3)	17 % (1/6)	20 % (10/49)	28 % (13/46)

2.4.2 Summary of Sample Data Processing

Table two presents a brief summary of how many samples from each population were collected and of those how many were analyzed in the dataset presented in this thesis. Furthermore, it presents a summary of the distribution of read-depth within the samples that were analyzed.

Table 2: Chapter 2 Data Processing Summary

	Participant Breakdown				
	Total	GAD	MDD	eMDD	HC
# Samples Sequenced	191	71	18	16	87
# Samples < 10 000 reads	7	1	1	1	4
# Samples > 10 000 reads	184	70	17	15	83
Final Analyzed Data Set Statistics					
# Samples	# OTUs	Min. Read #	Max. Read #	Avg. Read #	Median Read #
184	235	16 941	151 646	69 544	73 198

2.4.3 The Microbiome Distinguishes GAD patients from MDD patients and Healthy Controls

With samples collected and processed, and data refined, the first step in the analysis was to compare the microbiomes of the diagnostic groups of the participants to one another to determine if any of the putatively associated microbiomes were distinct. This analysis was performed using a PERMANOVA test, specifically the adonis function within the vegan package⁹⁹ of R. The analysis was performed on a Bray-Curtis dissimilarity distance matrix calculated from the

relative abundance of OTU data for each of the samples. Diagnosis was found to distinguish the putatively GAD-associated microbiome from the control-associated and MDD-associated microbiomes, with diagnosis explaining 3.7% and 2.4% of observed variation respectively ($p=0.001$ and $p=0.019$). The eMDD patient-associated microbiome was not observed to be significantly distinct from any of the other groups, nor was the MDD-associated microbiome distinct from the microbiome of the healthy volunteers.

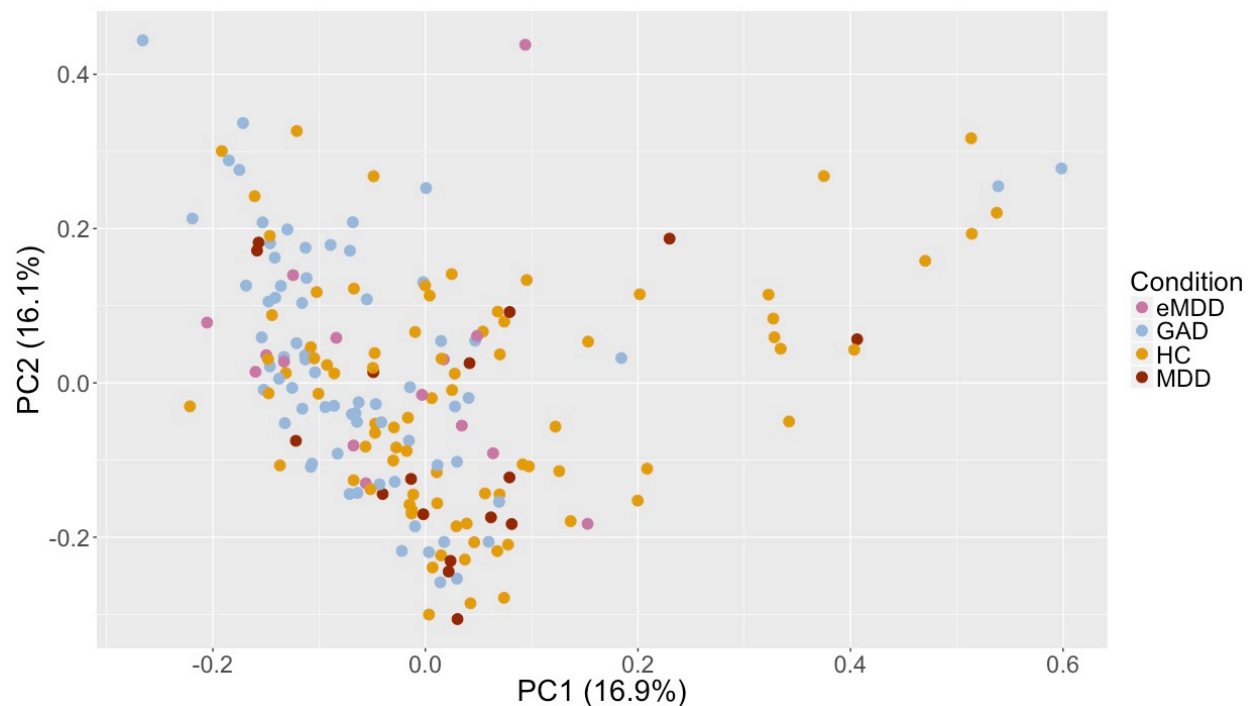
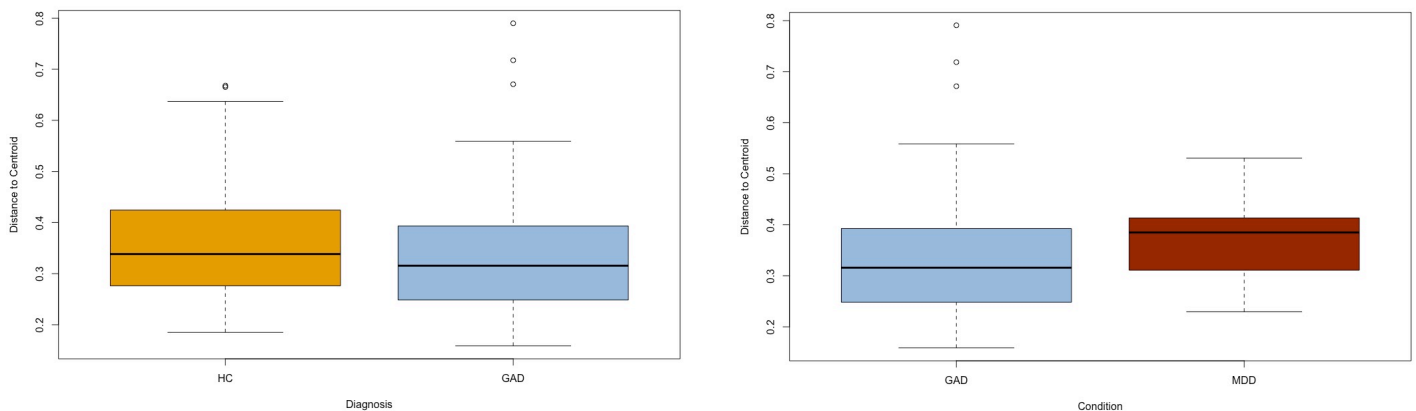


Figure 1: A PCOA plot constructed using the Bray-Curtis dissimilarity matrix comparing samples of GAD patients, healthy controls, eMDD patients and MDD patients. Each point on the plot represents a sample from a study participant. Points that are closely grouped together represent communities that are more similar, while point that are spread farther apart are more dissimilar. The sample colour scheme is maintained throughout the chapter.

2.4.4 There is no significant difference in the variance of significantly different microbial community signatures

One of the assumptions of a PERMANOVA test is that the groups being compared do not have significantly different variances. Accordingly, the within-group variances of the two previous identified results (GAD vs. HC and GAD vs MDD) were calculated using the betadisper function from the vegan package⁹⁹ of R and then compared with an ANOVA test. The finding that within group variance was not significantly different between the diagnostic groups validated the



findings of the previous analysis.

Figure 2: Boxplots summarizing the average length to centroid within the grouped samples.

Significance testing was performed using ANOVA. The betadisper function calculates an artificial center point within all of the samples in a given group; in this case, all of the samples from participants with the same diagnosis. This point is computed and then the distance from

each of the samples within a group to the grouping's center are calculated. These distance to centroid values are used to approximate within group variance.

2.4.5 Metrics of alpha diversity do not distinguish participant samples

Another common measure through which putatively different communities can be compared is alpha diversity. Alpha diversity metrics measure the amount of diversity within individual samples. Two metrics of alpha diversity were compared, richness which is the number of distinct groups, in this case OTUs observed within a sample, and evenness, which is a measure of how evenly distributed individuals are amongst observed groups. Evenness was calculated using the inverse Simpson method. Testing for significant difference was done with the ANOVA test using the stats package in R⁹⁵. As shown in figure 3, alpha diversity did not differentiate any of the diagnosis groups of participants.

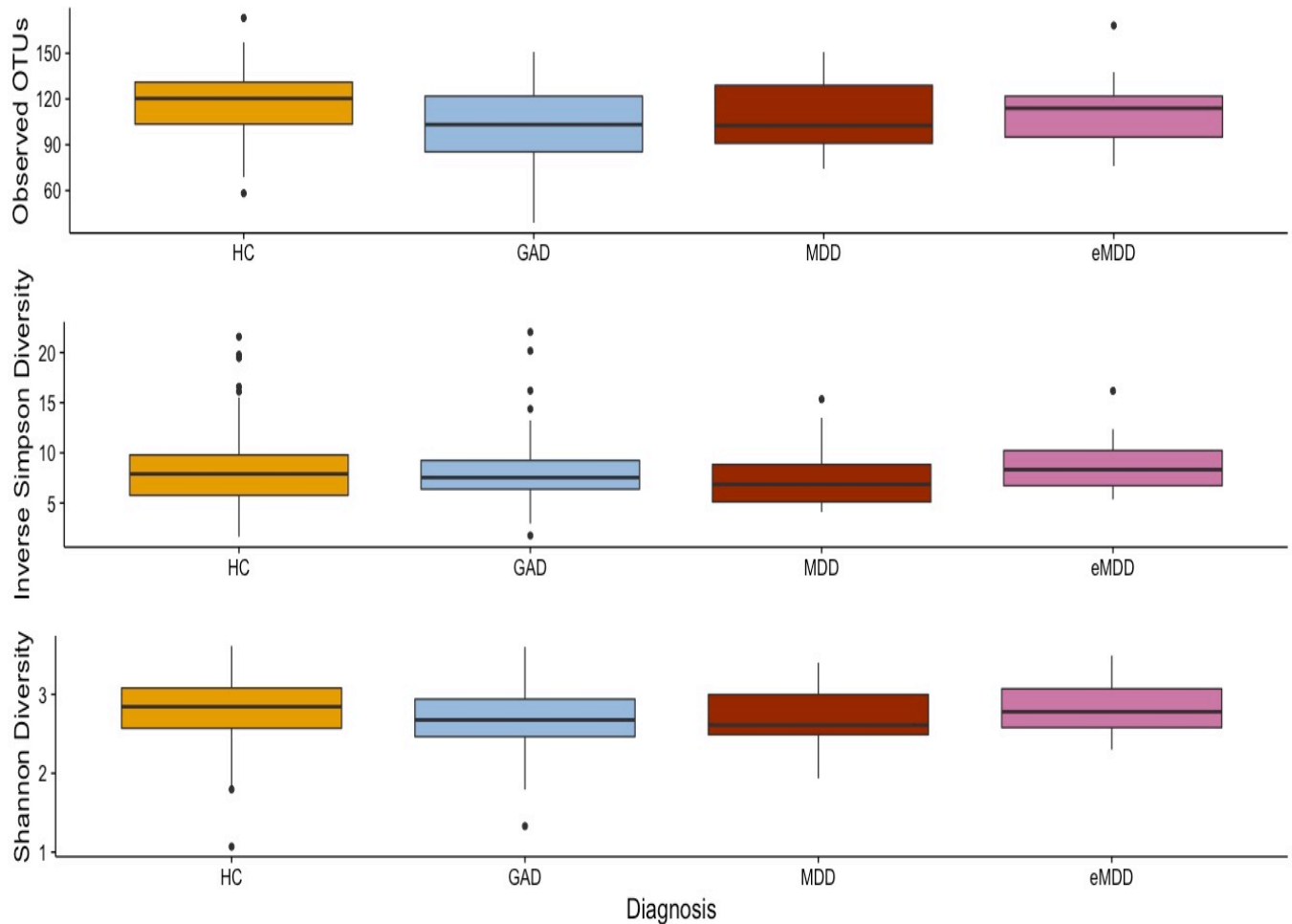


Figure 3: Comparison of OTU richness and evenness between the four diagnosis groups of study participants. The richness, Inverse Simpson and Shannon diversity measures were generated by first rarifying (by means of sampling without replacement) the number of reads in each of the samples to the number of reads in the smallest sample (16 941). This process is repeated 100 times and diversity measures for each sample are averaged across the 100 repeats. This method is adapted from Berry's phyloseq tutorial¹⁰⁶ which itself is inspired by Hughes et al¹⁰⁷. That the averaged measures of alpha diversity between diagnostic groups is statistically indistinguishable (Kruskal-Wallis test, $p = 0.45, 0.53, 0.26$) suggests that diagnosis does not significantly impact observed community richness or diversity.

2.4.6 Abundance of specific genera distinguishes the microbiomes of patients and controls

To further interrogate the differences in microbial communities between the diagnostic groups, the next step was to compare communities at the genus and OTU levels. These analyses were all conducted in R. The GAD and HC associated microbiomes were compared at both the genus and OTU levels, while the remaining groups were compared only at the OTU level. These comparisons were made despite not having identified significantly different communities (save for in the comparison of GAD to MDD) in an attempt to identify putative marker organisms. Given the small sample size of the eMDD and MDD participant groups it was deemed worthwhile to investigate possible markers that could drive hypothesis generation for future experiments. Additionally, the eMDD patients and MDD patients were pooled when compared to the GAD and HC groups, as well as being compared separately to identify or distinguish organisms putatively associated with active and remitted depression. Figures 4 and 5 which follow are summary charts of the relative abundance of the top 25 most prevalent genera identified in participant samples.

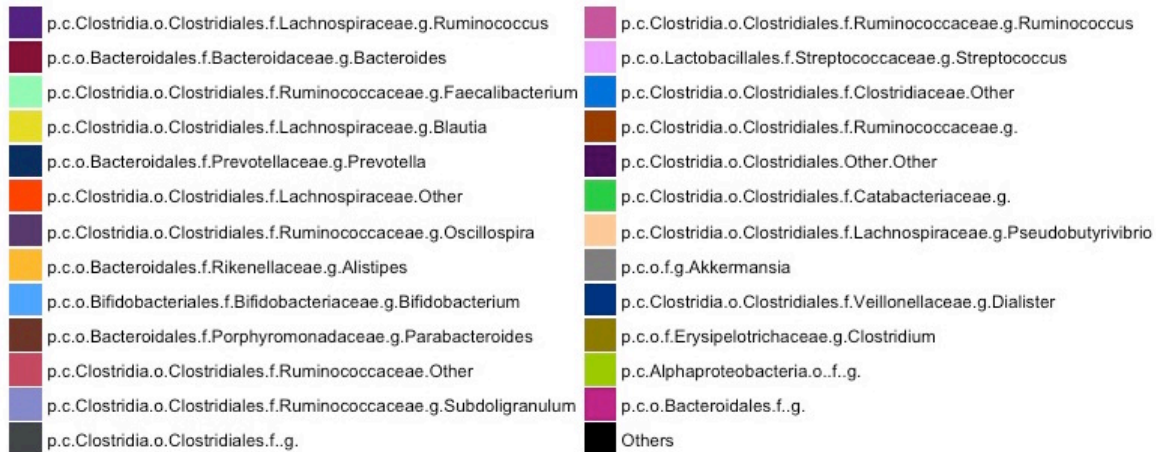
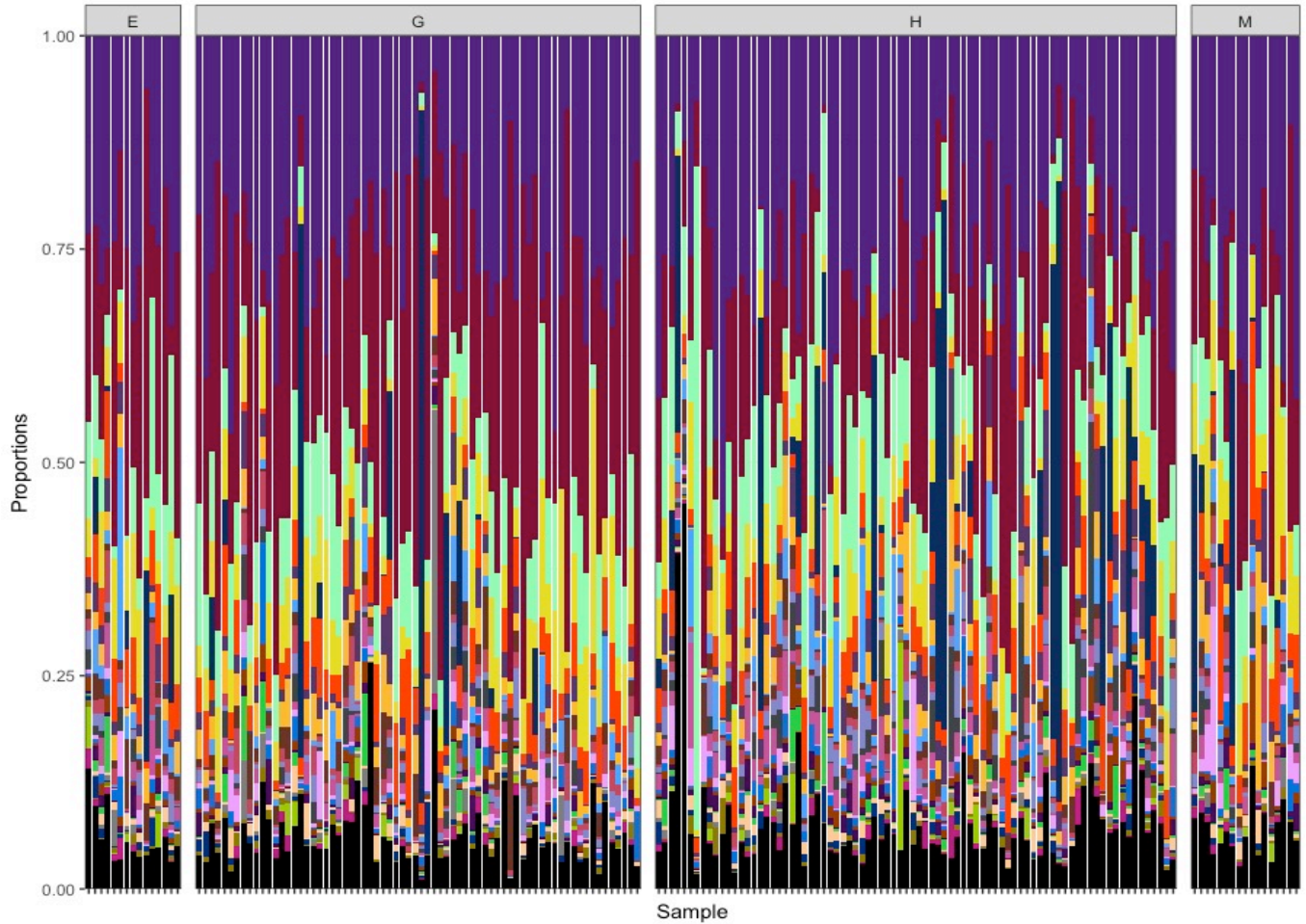


Figure 4: A comparison of the top twenty-five most abundant genera within individual samples, grouped by diagnosis.

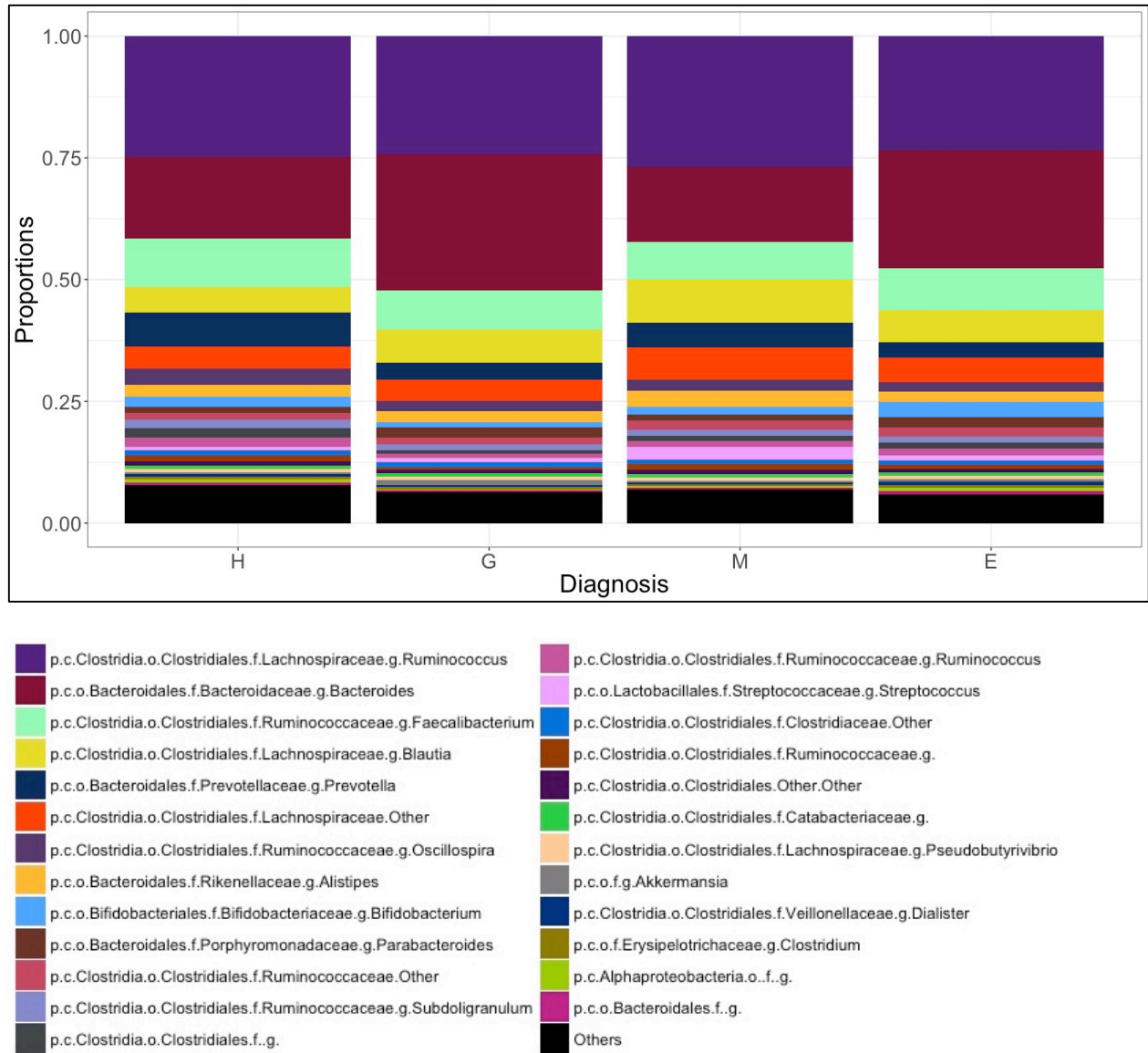


Figure 5: A comparison of the top twenty-five most abundant genera within each of the four participant diagnosis groups. Individual coloured portions represent the average relative abundance of individual genera while the bottommost section of the charts, indicated by the black portion, represents the combined abundance of the remaining genera. Abundance of each of the genera within the samples is calculated and then averaged within the groups. The genus *Bacteroides* (shown in maroon) was found to be significantly enriched in GAD patients relative to healthy controls (identified using ANCOM software⁹⁶). The MDD and eMDD groups are

shown in this plot but were not compared at the L6 level (another way of saying the genus level) to the HC or GAD participants as their small sample size affords limited statistical power which was instead allocated to investigate potential distinguishing OTUs.

The first comparison run in the ANCOM software⁹⁶ compared the GAD and HC associated microbiomes at the L6 level. This analysis identified the genus *Bacteroides* as being significantly elevated in GAD patients relative to controls. Data from MDD and eMDD participants was included in this graph as well, however, they were not tested for statistical significance.

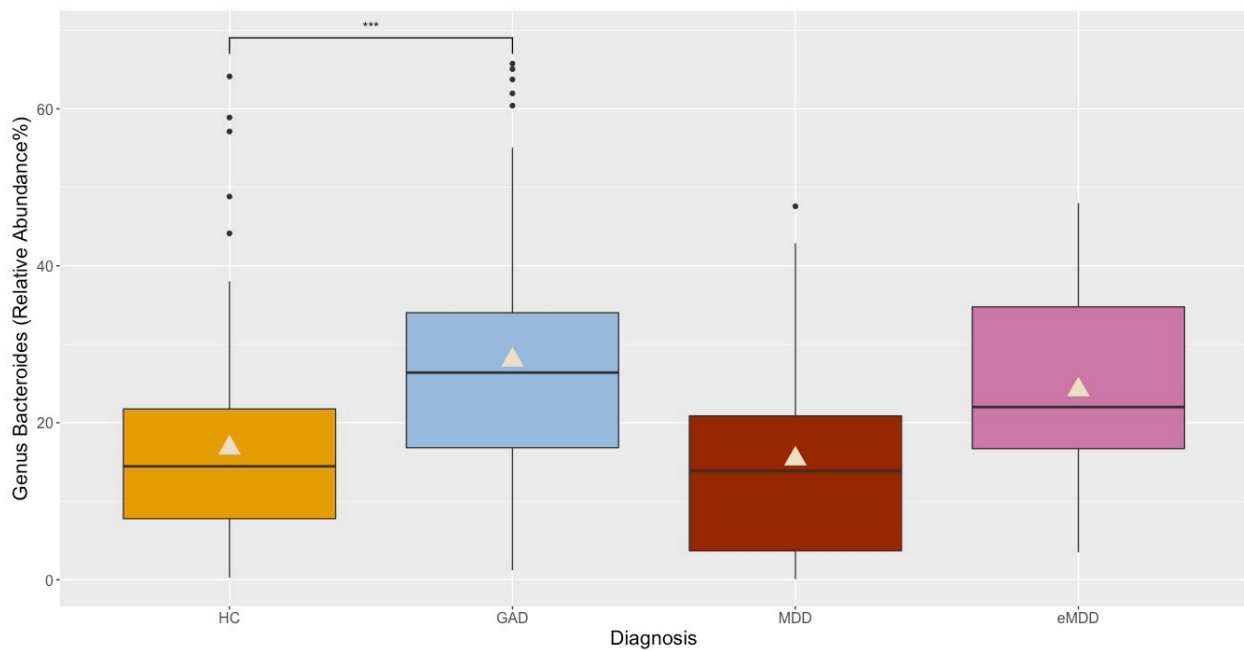


Figure 6: A comparison of the relative abundance of the genus *Bacteroides* across diagnostic groups. Boxplots show IQR while mean abundance is indicated by the triangle. This genus was found to be significantly elevated in GAD patients relative to controls using ANCOM software at a significance threshold of 95% following FDR correction for multiple testing. Although the plot suggests a possible trend of an increased abundance of *Bacteroides* in eMDD patients relative to MDD patients or healthy controls, this trend was not investigated for statistical significance.

2.4.7 Specific OTUs identify microbial communities associated with participant diagnoses

Following analysis at the genus level, communities were compared at the OTU level. A total of seven OTUs were found to have significantly different abundances between groups: GAD vs. HC, 2, GAD vs. MDD, 1, HC vs. Pooled Depression, 2 and MDD vs. HC, 4 (which included the two OTUs that distinguished HC from pooled depression). Figures 7 through 11 highlight these significant findings. It should be noted that given the small sample size of the MDD and eMDD groups, it is difficult to comment with a high degree of certainty on the validity of these purportedly significant OTUs, however, given the significance of these participants to the study and this thesis, the findings have been included. Each of the following figures have boxplots showing the IQR of the relative abundance of the given OTU within the respective groups as well as points which demonstrate the spread of the participants within groups and the severity of their symptoms. Each point is scaled proportionally to the measure of symptomology in the participants (anxiety via the PSWQ in GAD patients, and depression via the DASS in MDD patients).

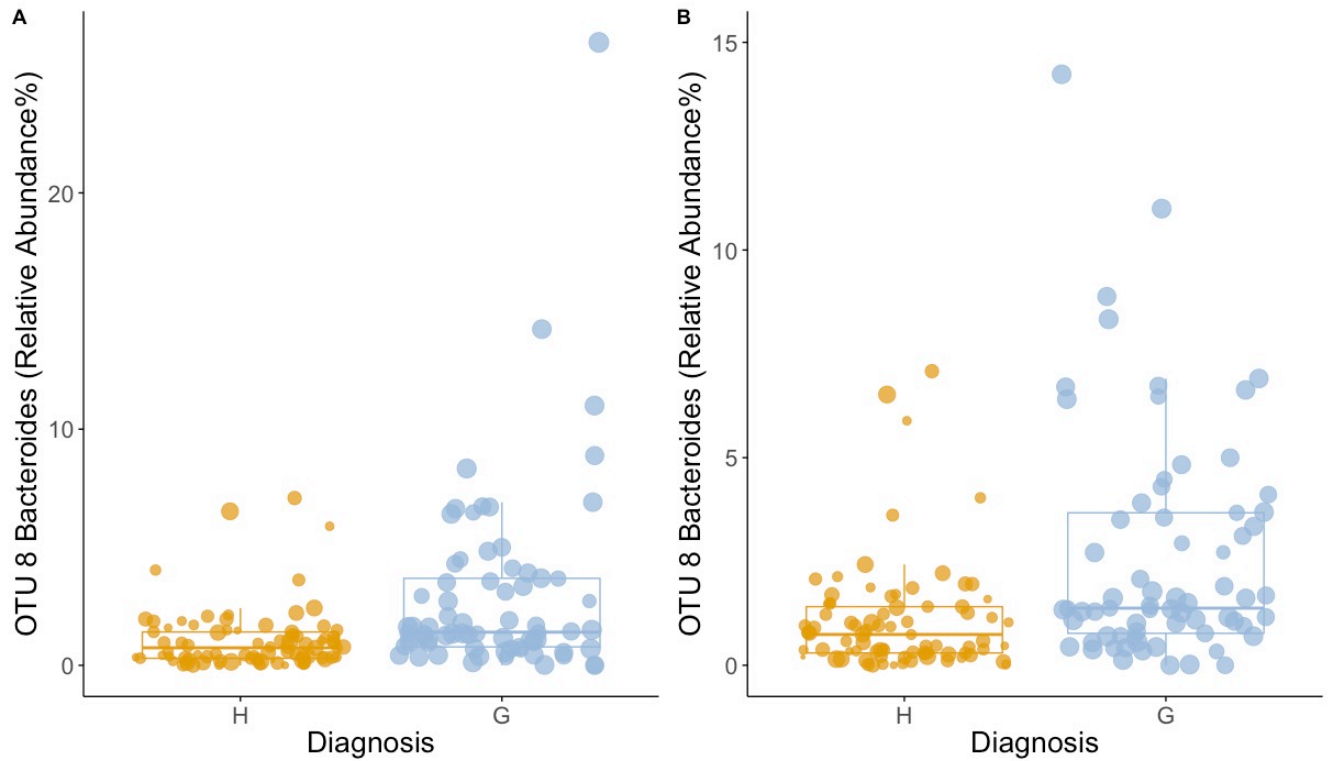


Figure 7: Boxplots comparing the relative abundance of the genus *Bacteroides* OTU 8 in GAD patient and healthy controls samples. This OTU was identified as being significant at a significance threshold of 95% following FDR multiple testing correction using ANCOM software. An increased abundance of the OTU distinguishes GAD patient samples from HCs within the dataset. Panel **A** shows data from all the GAD patient and control participants while panel **B** plots the same participants with the uppermost GAD patient removed allowing for better resolution of the difference in abundance between patients and controls. Points show abundance of the OTU within individual samples and are scaled to represent anxiety score as assessed by the PSWQ with larger points representing higher and thus more severe anxiety scores.

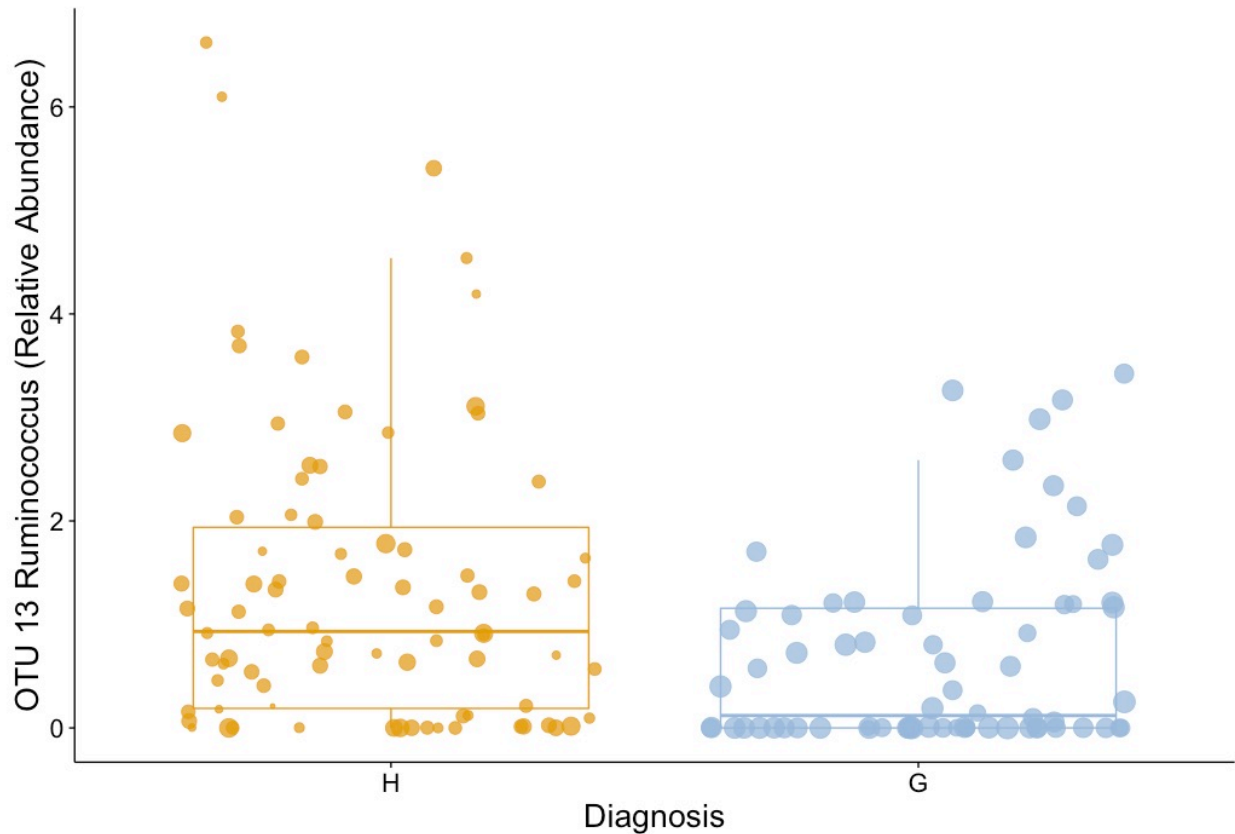


Figure 8: A comparison of the relative abundance of the genus *Ruminococcus* OTU 13 in GAD patient and healthy controls samples. This OTU was identified as being significant at a significance threshold of 95% following FDR multiple testing correction using ANCOM software. A decreased abundance of the OTU distinguishes GAD patient samples from HCs within the dataset. Points show abundance of the OTU within individual samples and are scaled to represent anxiety score as assessed by the PSWQ with larger points representing higher and thus more severe anxiety scores.

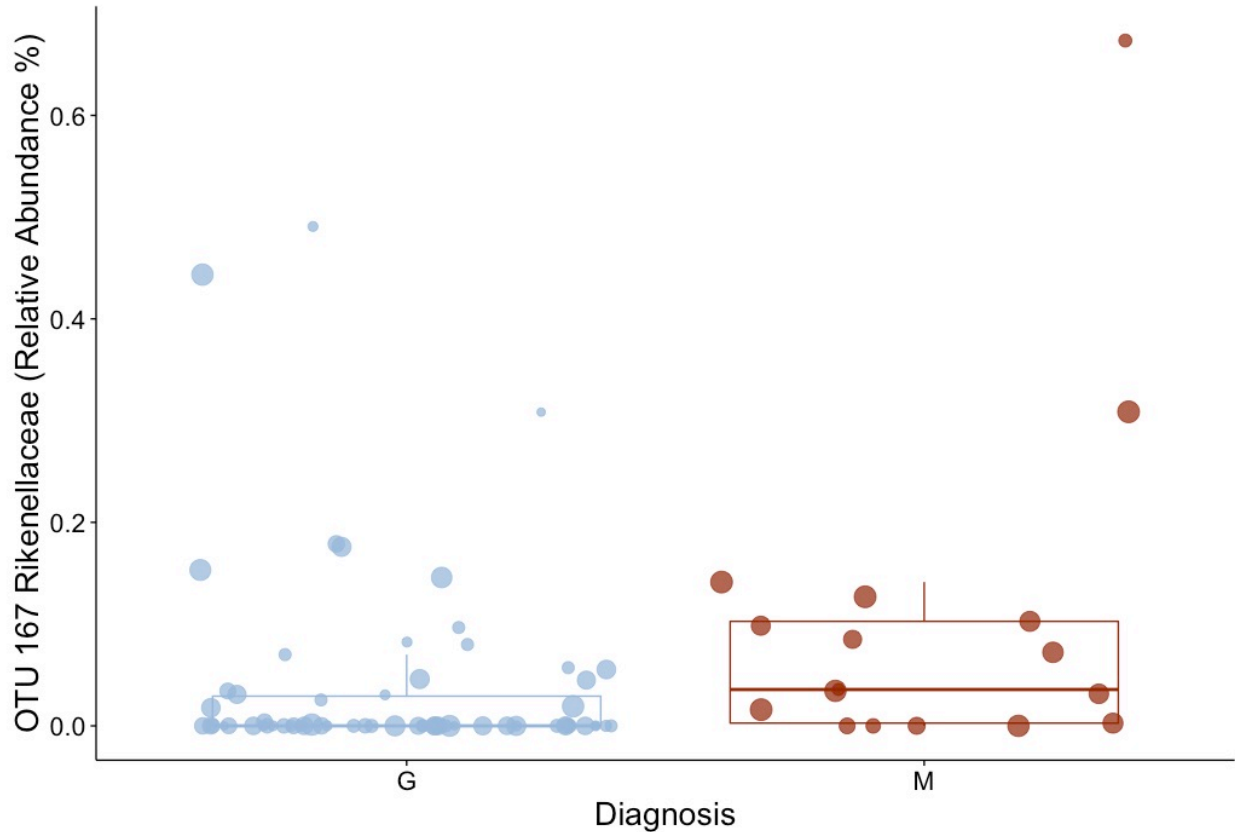


Figure 9: A comparison of the relative abundance of the family *Rikenellaceae* OTU 167 in GAD and MDD patient samples. This OTU was identified as being significant at a significance threshold of 95% following FDR multiple testing correction using ANCOM⁹⁶ software. A decreased abundance of the OTU distinguishes GAD patient samples from MDD patients within the dataset. Points show abundance of the OTU within individual samples and are scaled to represent depression score as assessed by the DASS with larger points representing higher and thus more severe depression scores.

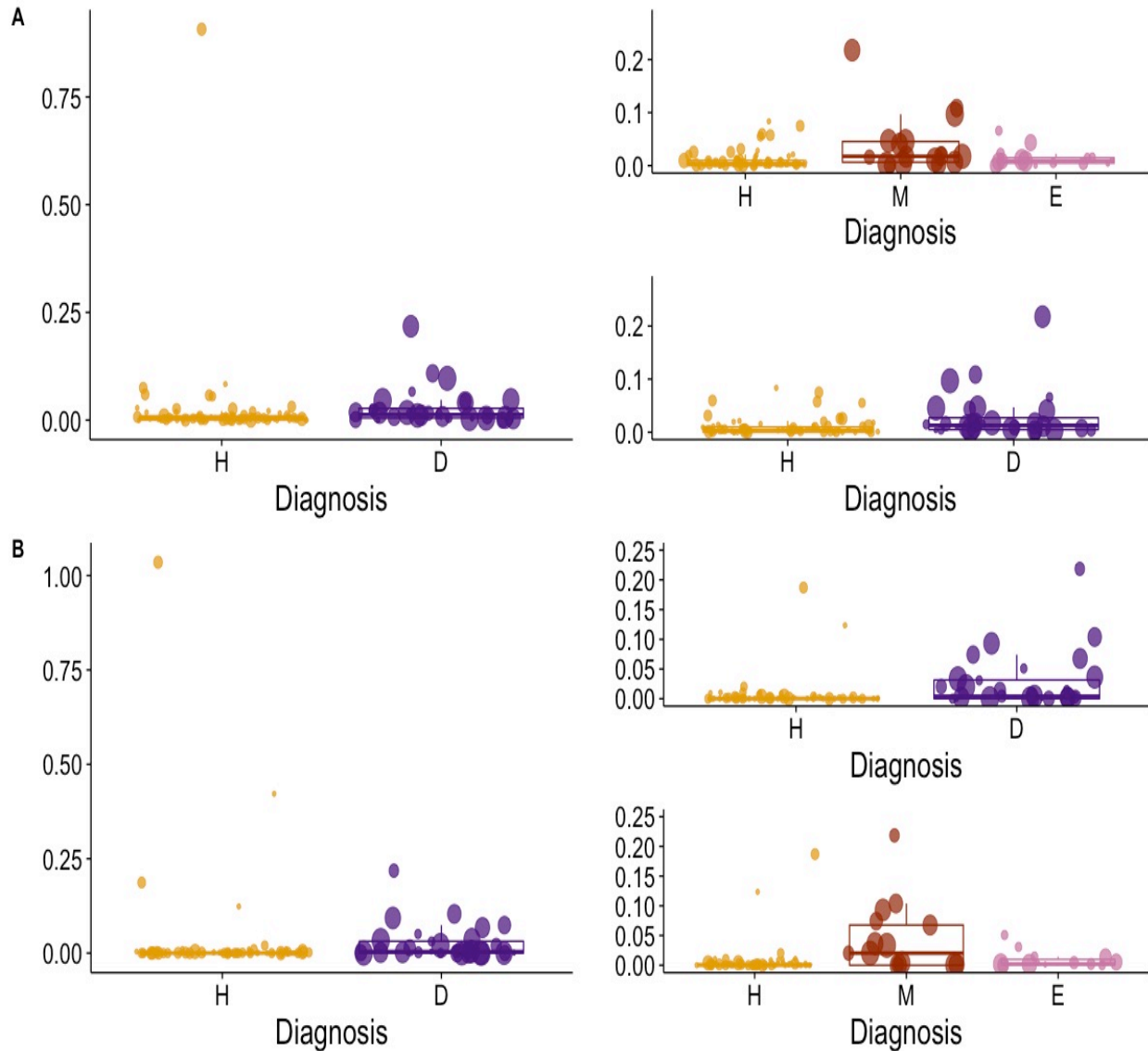


Figure 10: Boxplots representing the relative abundance of OTUs 224 family

Erysipelotrichaceae (in panel A) and 237 genus *Clostridium* (in panel B) in healthy control and pooled depression (plotted in purple) and separate MDD and eMDD patients (shown in the usual red and pink respectively). In both cases, an enrichment of the OTU was observed in the microbiome of depression patients when compared to that of healthy volunteers. Percent relative abundance of the organisms within the samples is plotted on the y-axis. Each of the horizontal

panel A and B consist of three plots, one large plot on the left side of the panel and two stacked on top each other on the right side of the panel. The large plot on the left side of the panel represents all of the patient and control data and thus the spread of the data is more difficult to observe due to the effects of a couple healthy volunteers with relatively high abundances. Accordingly, the top plot on the right-hand side of the panels shows the same graph again but with the uppermost outliers removed. The third graph plotted in the bottom right of the panels shows the same data but with MDD and eMDD patients separated demonstrating that the observation of a greater abundance of these OTUs amongst depressed participants is being driven by their relatively higher abundance in MDD patients. Points show abundance of the OTU within individual samples and are scaled to represent depression score as assessed by the DASS with larger points representing higher and thus more severe depression scores.

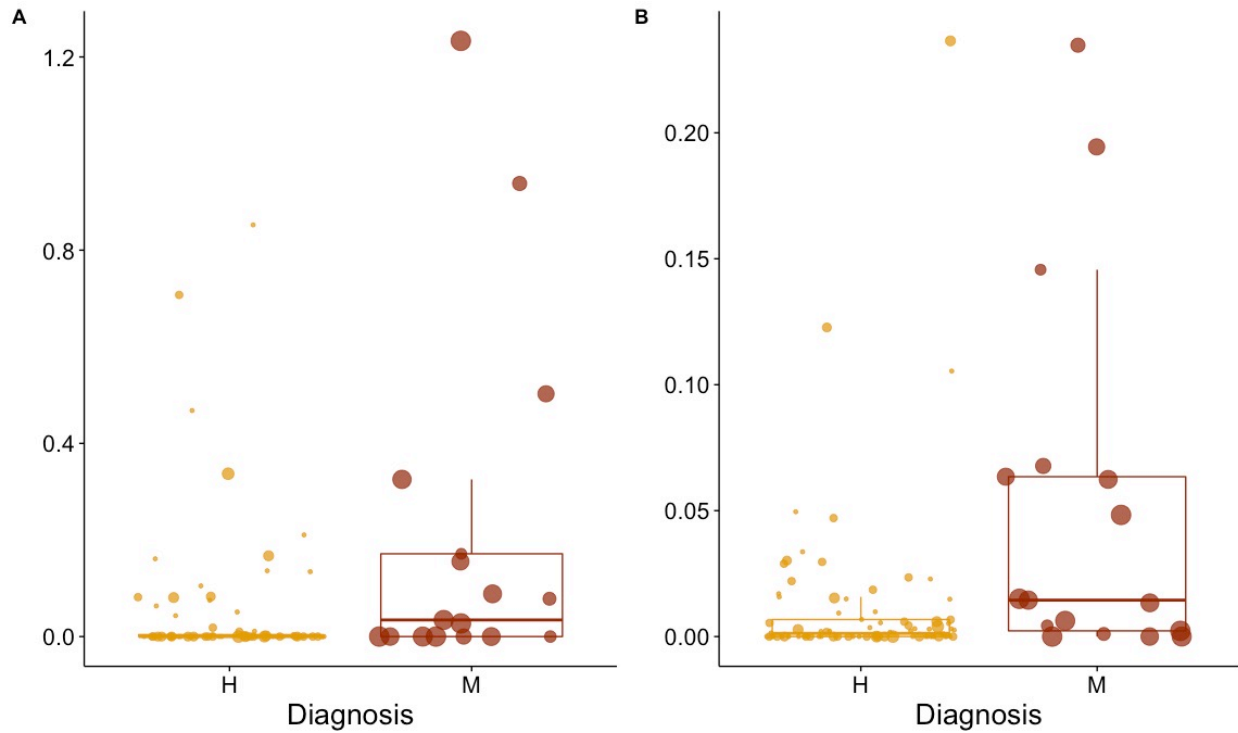


Figure 11: Boxplots representing the relative abundance of the two additional (along with OTUs 224 and 237) OTUs found to distinguish MDD patients from healthy controls. They represent OTU 109 family *Ruminococcacea* (in panel A) and OTU 221 order *Clostridia* (in panel B).

2.4.8 Marker OTU abundance appears to correlate with a measure of disease severity

The final analysis conducted on this dataset was the correlation of the two OTUs which were found to distinguish GAD and health volunteer populations with the PSWQ scores of participants. This analysis was performed to investigate if the OTUs that were putatively correlated with the presence or absence of the disease would correlate with a validated measure

of disease severity and thus lend further credence to the putative significant and possible biological relevance of the correlation.

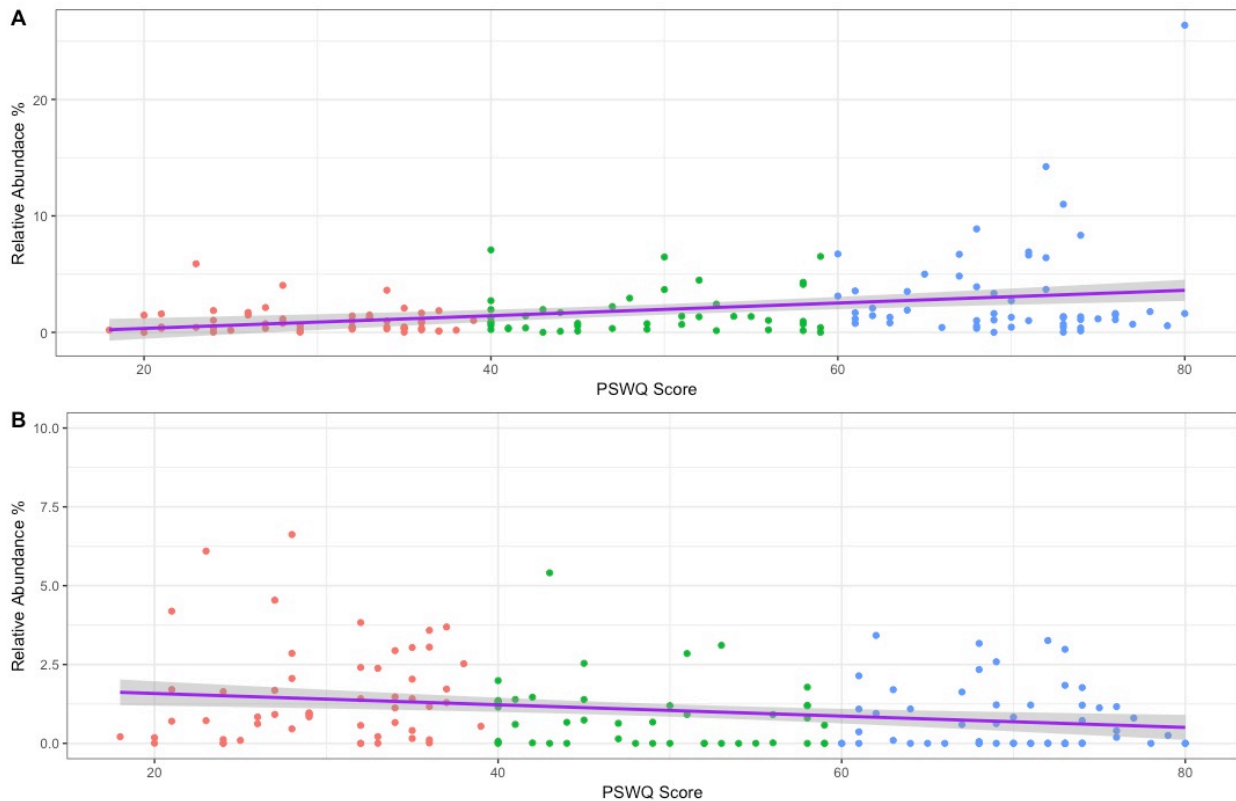


Figure 12: Significant correlation of abundance genera and OTUs of interest (OTU 8 - *Bacteroides* in panel A and OTU 13 - *Ruminococcus* in panel B) with a measure of anxiety support the hypothesis of a correlation of biological relevance in GAD. The individual plots represent the relationship between anxiety score (as measured by the PSWQ) correlated with the relative abundance of OTUs 8 and 13 within GAD patient and control samples. Participants with low, medium and high PSWQ scores are coloured red blue and green respectively. Trend lines are shown in purple and were both found to be significant relationships using the `lm` function from the base stats package in R. The correlations are as follows: OTU 8 adjusted $R^2 = 0.099$, $p = 4.81 \times 10^{-5}$, OTU 13 adjusted $R^2 = 0.056$, $p = 1.98 \times 10^{-3}$.

2.5 Discussion

The goal of this chapter was to investigate the microbiomes of study participants in hopes of differentiating a putatively healthy gut community from that of patients with GAD, MDD and eMDD. No evidence for a significant differentiation of the eMDD-associated microbiome from any other of the investigated groups was identified. Similarly, the putatively MDD-associated microbiome could not be distinguished from that of healthy volunteers. Evidence presented herein successfully demonstrates statistically significant differences in the putatively associated microbiome of GAD patients when compared with healthy controls and when compared with that of MDD patients. Furthermore, several possible marker species differentiating depression patients from healthy controls were identified.

The GAD and healthy volunteer populations were the primary focus of this study and thus the most heavily recruited. Accordingly, these participants represent a majority of the dataset and the comparison from which the most significant findings could be derived. Microbiome analyses, as has been conducted in this experiment, can only provide results implying correlation; nonetheless, they are valuable for their capacity for future hypothesis generation and more targeted investigation. To demonstrate the correlation and reinforce its validity three questions were asked: firstly, is the putatively GAD associated microbiome statistically distinct from that of the healthy volunteers? Secondly, can an organism(s) be identified as driving this distinction? Finally, with putatively associated microbes identified, can they be correlated with a measure of disease severity to support the validity of a meaningful correlation over a spurious one?

To answer the first question, the GAD and healthy volunteer microbiomes were compared using PERMANOVA and found to differ significantly when compared on the basis of diagnosis. This comparison variable was found to explain approximately four percent of the variation between samples. This distinction can be seen in figure 1 wherein the GAD samples, represented in blue are clustered more towards the left-hand side of the plot relative to the healthy volunteer samples that are distributed more throughout the middle third of the plot. A critical step in confirming these results was to ensure the two groups of data met the assumption of equal variance. The PERMANOVA test is unable to distinguish the difference observed between two distinct tight clusters of samples from that observed between one tightly clustered data set compared to one that is very diffuse. Accordingly, figure 2 compares the heterogeneity of variance between the GAD patients and healthy volunteers, as well as between GAD and MDD patients. In both cases, as confirmed via ANOVA, there is not a statistically significant difference in variance between the groups, therefore, the findings of the PERMANOVA can be considered valid. A second validation step was to determine, again via PERMANOVA, that neither BMI or diet significantly interacted with diagnosis in explaining variation in the samples. These two variables, while not an exhaustive list, are two of the more commonly studied variables proposed to impact the gut microbiome^{65,108,109}. Accordingly, that I was able to demonstrate a significant effect of diagnosis whilst correcting for these two variables further validates the existence of real differences between the putatively associated microbiomes investigated in this study.

Following the identification of distinct microbial communities, it was then desirable to try to identify an organism(s) that may be the driver for the observed differences. The genus *Bacteroides* was found to be enriched in GAD patients as can be seen in figures 6 and 7. In an

initial pilot dataset, the first *Bacteroides* OTU (OTU 2) was found to be significantly elevated in patients. Accordingly, when beginning analysis of the final dataset, this OTU was tested to corroborate the findings from the pilot data. This OTU was not identified as having a significantly different abundance in GAD patients when compared to healthy controls at a significance threshold of 95% in ANCOM, however, a different *Bacteroides* OTU, OTU 8 was found to be significantly elevated in patients. Furthermore, OTU 13 *Ruminococcus* was elevated in healthy controls relative to GAD participants, a relationship not observed in the pilot dataset.

In comparing the GAD and MDD microbiomes, an OTU assigned to the family *Rickenellaceae* was observed to be elevated in MDD patients. Similarly, a more abundant OTU of the family *Rickenellaceae* was found to distinguish MDD patients from healthy controls along with one belonging to the family *Erysipelotrichaceae* and two *Clostridia* (one assigned to the order *Clostridia* and one assigned the genus *Clostridium*).

In an effort to further demonstrate the putative biological relevance of the observed correlations, figures 7 through 11 were plotted with points representing the individual participants layered on-top of the boxplots and the size of the points was scaled relative to the severity of symptomology assessed by the Penn State Worry Questionnaire (PSWQ)⁸¹ for anxiety and the Depression Anxiety and Stress Scale (DASS)⁷⁹ for depression. These correlations are more clearly demonstrated in figure 12 which illustrates the statistically significant positive correlation of abundance of OTU 8 and PSWQ score as well as the significant negative correlation of OTU 13 and PSWQ score.

Given that this is the first analysis of the putative GAD microbiome, the findings of elevated *Bacteroides* and diminished *Ruminococci* cannot be compared to other GAD associated microbiome studies in the current literature. One comparison of interest can be made to the depression associated microbiome studies of Naseribafrouei *et al.*⁶¹ and Jiang *et al.*⁶². They reported a lower abundance of the phylum *Bacteroidetes* (the phylum that contains the *Bacteroides* genus) in depression patients relative to healthy controls. If both their observation and this observation of higher *Bacteroides* (one of the many genera within the phylum *Bacteroidetes*) in anxious patients are true, it could be helpful in distinguishing GAD and MDD patients (a feat that can be challenging given overlapping symptomology) and potentially therefore, in helping better prescribe treatment. While these data provide reasonable evidence for correlation, they cannot demonstrate biological relevance nor causation. Additionally, given the field's relatively limited understanding of the incredibly complex dynamics of the digestive ecosystem, it is difficult to predict what role if any these organisms might play in a healthy gut, let alone in the advancement or prevention of disease. Interestingly, however, in 2012 work by Ze *et al.*¹¹⁰ demonstrated a capacity for a *Ruminococcus* species, specifically, *R. bromii* to stimulate other gut bacteria to metabolize some resistant starch molecules that would otherwise be ignored. This phenomenon was observed even in conditions in which *R. bromii* itself could not grow. Despite having tested only a few additional bacterial species, one of the organisms *R. bromii* stimulated to grow was *Bacteroides thetaiotomicron*. Although the group of bacteria in which this stimulatory phenotype was explored was small, it could suggest a possible mechanism through which disruption of the balance of *Bacteroides* and ruminococci in the gut could disrupt homeostasis. Perhaps without the aid of the ruminococci, the *Bacteroides* must switch from metabolizing complex carbohydrates that they alone can metabolize and instead compete for

simpler sugars or metabolize mucin in the mucous barrier leading to inflammation in the host's epithelium. Again, the data reported in this thesis are purely observational and cannot be taken to show anything beyond a correlation of these organisms and patient phenotype; however, they do offer a basis for future hypothesis generation and testing surrounding potential mechanisms of disease.

Unlike the GAD microbiome, two previous studies have examined the MDD microbiome, the previously mentioned study by Naseribafrouei⁶¹, and one by Jiang⁶². Jiang reported increased Shannon diversity amongst depression patients relative to psychiatrically healthy controls, a finding these data do not corroborate. Interestingly, in their supplementary materials Jiang *et al.*⁶² include a figure showing no significant difference in diversity when assessed using the Inverse Simpson method, a finding which does align with these data. Furthermore, Jiang's control population was recruited from a hospital. While they do state that patients were deemed mentally healthy, they do not comment on what diagnoses had resulted in these patients' hospitalization. Additionally, they do not comment on patients' medications and while one would assume controls were not on antibiotics, they do not explicitly state that control patients had not undergone a recent course(s) of antibiotics or other medication (like a PPI) which could contribute to the decrease in diversity observed. Jiang⁶² also reported a number of OTUs which they found to be enriched in their patients, one of which was *Clostridia*. As was previously reported, two OTUs belonging to the *Clostridia* family were enriched in the MDD patient population relative to healthy controls potentially emphasizing the significance of the finding. Jiang also reported an enrichment of ruminococci in their control population, whereas these data find a *Ruminococcacea* (the family to which *Rumicoccocus* belongs) to be enriched in patients.

This discrepancy could stem from the differences in control populations used by the two studies. One common finding between the Jiang⁶² and Naseribafrouei⁶¹ studies was an observation of elevated levels of *Alistipes* being associated with depression. While this study did not find evidence in support of this relationship it is possible that this results from the small size of this study population and perhaps, with a larger study that trend would become more apparent. Regardless, the discrepancy between studies, and the relatively small sample size of all the studies emphasize the importance of further investigation of the microbiome in these populations. One unique aspect of this study relative to the two previous ones, is the capacity to compare the MDD population with that of the eMDD patients. The observation that only two of the four OTUs that distinguish the MDD and HC groups distinguish the pooled depression patients from healthy controls could suggest that the two OTUs unique in distinguishing the MDD patients are particularly deleterious and that as levels of these organisms, in particular the *Clostridia* start to return to healthy levels, depression improves. Furthermore, it could be that an enrichment of organisms within the *Clostridia* (the family containing three of the four OTUs that appear to distinguish MDD patients from healthy controls) play a role in depression and, therefore, the observation of only one OTU distinguishing pooled depression patients compared to three in comparing MDD patients to healthy controls is representative of a progression towards a healthier community. Equally, however, as can be seen in figure 10, the relative abundance of OTUs 224 and 227 in eMDD patients is nearly indistinguishable from that of healthy control. Thus, this could simply be an effect of these OTUs having the largest difference in abundance between depression patients and controls, and thus this trend is still statistically significant despite the low abundance of these OTUs in eMDD patients. Once again it is important to mention that these are simply hypotheses formulated from observational data, but

they could be helpful in generating future studies. For example, perhaps a longitudinal study of depression patients that following them from initial psychiatric consult through remission could monitor *Clostridia* levels, to see if they diminish with remission. This could suggest that lowering levels of *Clostridia* may help with healing in depression and while that certainly wouldn't prove causation could at least inform potential antibiotic or probiotic therapies to help hasten remission in depression patients.

In conclusion, these data putatively demonstrate that the GAD and healthy volunteer-associated microbiomes are distinct, as well of those of GAD and MDD patients. They identify some of the organisms that appear to drive the distinction of these communities, as well as some that appear to be markers of depression when compared to healthy controls. These results are by no means definitive, but can hopefully serve as the basis for hypothesis generation for experiments like the ones described above, or others moving forwards.

Chapter 2 Acknowledgments:

This chapter details the analysis of the combined GAD, MDD and eMDD patient data sets recruited through Hamilton Health Sciences between 2014 and 2016. I would specifically like to acknowledge the efforts of Ms. Merwa Amer, Ms. Jessica Gillard, Ms. Cindy James, Ms. Melanie Wolfe, Mr. Matthew Tersigni and Dr. Christian Avilla for their incredible efforts in recruiting the participants for the study. I would also like to acknowledge Drs. Russell de Souza and Michael Zulyniak for their guidance on analyzing the diet data. Finally, a sincere thank you to Michelle Shah and Laura Rossi for all of their time and guidance in DNA extraction and sequencing of the samples.

Chapter 3 The Bipolar and MDD microbiomes:

3.1.1 Introduction

Bipolar disorder, though less prevalent in the population than GAD or MDD (~1% for bipolar disorder compared with between 5 and 10% prevalence for GAD and MDD), has considerable impact on the Canadian population. Similar to GAD and MDD, bipolar disorder patients are more likely to report comorbid GI distress. Further, a 2016 study showed that symptoms of mania may be more common in patients following a course of antibiotics¹¹¹. Additionally, given the hypothesis that Bipolar disorder and MDD may exist on a spectrum, samples from MDD patients obtained in the chapter two study were also included to investigate possible correlations and potentially dose effects³⁰. Accordingly, these investigations were undertaken in similar fashion to those reported in chapter two in an attempt to uncover evidence of putative correlations between the disorder and the gut microbial community.

Patients (14 Bipolar I, 9 Bipolar II) were recruited between 2016 and 2017 through both Hamilton Health Sciences and the University Health Network in Toronto. As was the case with the study in chapter two, patients and healthy volunteers completed a number of questionnaires to confirm their diagnosis, ensure that they met inclusion and exclusion criteria of the study, and to collect significant metadata; in particular, to get an estimate of their diet. Study participants also provided blood and fecal samples intended for analysis of inflammatory biomarkers and analysis of microbial communities respectively. These investigations focused exclusively on analysis of the bacterial communities. Microbial community data was generated through extraction of DNA from the fecal samples followed by PCR amplification of the V3 region of the 16s rRNA gene which were then sequenced using Illumina technology (San Diego, CA).

Sequence data was compiled and analyzed *in silico* and was used to answer questions about the putatively distinct communities associated with Bipolar I, Bipolar II, and MDD as well as to identify specific organisms differentiating these communities.

3.2 Hypothesis

This chapter represents preliminary analysis of a recently recruited cohort of bipolar patients and thus these investigations aim to test the hypothesis that the gut microbiome of bipolar patients differs significantly from that of healthy volunteers. Furthermore, these analyses will hopefully yield target microorganisms that are enriched in either group and can lead to hypothesis generation about putative mechanisms of interactions in patients.

3.3 Materials and Methods

3.3.1. Patient Recruitment

Through 2016, participants were recruited from the St. Joseph's Healthcare system in Hamilton Ontario and the Mood Disorder Psychopharmacology Unit at University Health Network (UHN) in Toronto, Ontario. A total of twenty-three bipolar patients (ages 18-65) were recruited along with twenty-three age-and-sex-matched (+/- 5 years) healthy volunteers. Study participants provided informed voluntary consent. Patients were required to have a principle diagnosis of Bipolar I or II disorder which was confirmed using the M.I.N.I. International Neuropsychiatric Interview²⁵. Potential participants were excluded if they had any of the following: consumption antibiotics or probiotics in 4 weeks prior to participation, a primary diagnosis of any competing

psychiatric illness, a lifetime diagnosis of schizophrenia or schizoaffective disorder, an active eating disorder, and active substance misuse disorder, presence of a major neurocognitive disorder or delirium, a diagnosis of any major organic illness that would affect the gut microbiome and finally pregnancy or breast-feeding. Potential participants were screened for birth control medications; however, they were not excluded on this basis.

3.3.2 Sample Collection

Patients were provided the requisite materials and instructions to safely collect fecal material in a sterile screw capped sample jar. Participants were instructed to freeze their samples before delivering it either directly to the lab at McMaster University, or to one of either St. Joseph's West 5th campus in Hamilton or to the Mood Disorders Psychopharmacology Unit at UHN in Toronto. Samples deposited at the hospitals were collected by research assistants and delivered, still frozen, in bulk to the Surette laboratory at McMaster University.

3.3.3 Sample Processing, DNA Extraction and Amplification and Sequencing

Samples were processed identically to those in chapter one. Similarly, the same DNA extraction protocol and robot was used. PCR was performed with the same primers and protocol and the sequencing was all done at the same sequencing facility at McMaster University.

The data were processed by the same in-house bioinformatics pipeline that incorporates quality filter using Sickle (q-score > 30)⁸⁹, Cutadapt (which removed adaptor sequences from the Illumina sequencing)⁹⁰, PandaSeq (which aligned forward and reverse reads)⁹¹, AbundantOTU

(which assigns OTU identities to grouped reads)⁹² and QIIME⁹³. Taxonomic assignments were made using the RDP classifier with the Greengenes 2011 training set⁹⁴.

In total 61 samples (14 BP1, 9 BP2, 20 HC and 17 MDD) were sequenced to an average depth of approximately forty-six-thousand reads. (min: 45 reads, max: 107630 reads).

3.3.4 Data Processing for Analysis

Prior to analysis the data set was curated in the following ways First, the one sample with less than ten thousand reads was eliminated from the data set. This sample was the one MDD sample that was similarly trimmed from the chapter one dataset. Sequences from the remaining 60 samples were then filtered such that OTUs in the data set had to be present in at least 10% of the original data set (6 samples), had to have a minimum read count of at least thirty across the entirety of the data set and finally had to be present at a prevalence of at least one thousandth of a percent within the data set. All of these curations were made in an attempt to minimize the effect of sequencing noise on the data so as to be able to achieve the most reliable results. Stringency cutoffs were established based on previous work completed in the lab as well as discussion with the supervisors. Data set curation was performed in MACQIIME⁹³ using the functions `filter_samples_from_otu_table.py` and `filter_otus_from_otu_table.py`. This curation resulted in a final data set used in analyses with the following characteristics: 60 samples (23 Bipolar, 17 MDD, and 20 Healthy Controls) with 472 observed OTUs, 2720350 total reads, average read count per sample 45339 (min: 14869 reads, max: 99472 reads)

Analyses were conducted in R⁹⁵ using a custom script incorporating the ANCOM⁹⁶, `permute`, `phangorn`⁹⁷, `phyloseq`⁹⁸, `vegan`⁹⁹, `ggsignif`¹⁰⁰, and `stats`⁹⁵ packages. Figures were composed in

R using the `ggplot2`¹⁰¹, `Hmisc`¹⁰², `gridExtra`¹⁰³, and `cowplot`¹⁰⁴ packages unless otherwise stated.

3.3.5 Diet Data Processing

Diet data were collected and processed identically to those in chapter one.

3.4 Results

3.4.1 Summary Statistics of Study Participants

Table three shows the demographic data of patient and healthy volunteer participants in the study. Participants using birth control is reported as a percentage of potentially eligible participants. Participants were identified as eligible to be using birth control if they were a female under 52 (the average age at which women begin menopause according to the Canadian Women’s Health Network) years of age¹⁰⁵.

Table 3: Summary statistics Bipolar I, Bipolar II, MDD and HC study participants

	Participant Diagnosis			
	Bipolar I	Bipolar II	MDD	HC
n	14	9	18	20
n (female)	9	7	13	13
% Female	71.00	77.78	72.22	65.00
n with Height and Weight Data	10	7	18	20
Average Age (IQR)	43.5 (17.25)	46.6 (12.5)	43.50 (17.25)	43.75 (18.25)
Average Height cm (IQR)	164.67 (8.21)	165.3 (2.52)	168.28 (11.88)	169.89 (11.01)
Average Weight lb (IQR)	178.2 (41.5)	169.9 (29.5)	199.42 (104.00)	164.95 (74.25)
Average BMI	31 (10)	28 (5)	32 (17)	26 (7)

(IQR)				
Percentage of Participants Reporting Cigarette Smoking (n)	33 % (3)	0 % (0)	11 % (2)	10 % (2)
Percentage of Eligible Participants Reporting Birth Control Medications (n/n possible)	17 % (1/6)	0 % (0/4)	17 % (1/6)	23 % (3/13)

3.4.2 Summary of Sample Data Processing

Table four presents a brief summary of how many samples from each population were collected and of those how many were analyzed in the dataset presented in this thesis. Furthermore, it presents a summary of the distribution of read-depth within the samples that were analyzed.

Table 4: Chapter 4 Data Processing Summary

	Participant Breakdown					
	Total	Bipolar I	Bipolar II	MDD	HC	
# Samples Sequenced	61	14	9	18	20	
# Samples < 10 000 reads	1	0	0	1	0	
# Samples > 10 000 reads	60	14	9	17	20	
Final Analyzed Data Set Statistics						
# Samples	# OTUs	Min. Read #	Max. Read #	Avg. Read #	Median Read #	
60	472	14 869	99 472	45 339	42 532	

Figure 13: A PCOA plot mapped from the Bray-Curtis dissimilarity matrix comparing samples of Bipolar type 1 patients (shown in light green), Bipolar type 2 patients (shown in dark green), controls (shown in orange) and MDD patients (shown in red). Each point on the plot represents a sample from a study participant. The colour scheme of Bipolar type one patients represented in light green, Bipolar type two patients shown in dark green, healthy control data shown in orange and MDD patient data coloured red is maintained throughout the chapter.

3.4.4 Metrics of alpha diversity do not distinguish participant samples

To evaluate possible differences in alpha diversity, the four diagnostic groups were compared in richness (Observed OTUs) and evenness (Inverse Simpson Index and Shannon Index) differences were tested for statistical significance via an Kruskal-Wallis test using the stats package in R⁹⁵. None of the four tested groups were found to be statistically significantly different any of the three measures of alpha diversity.

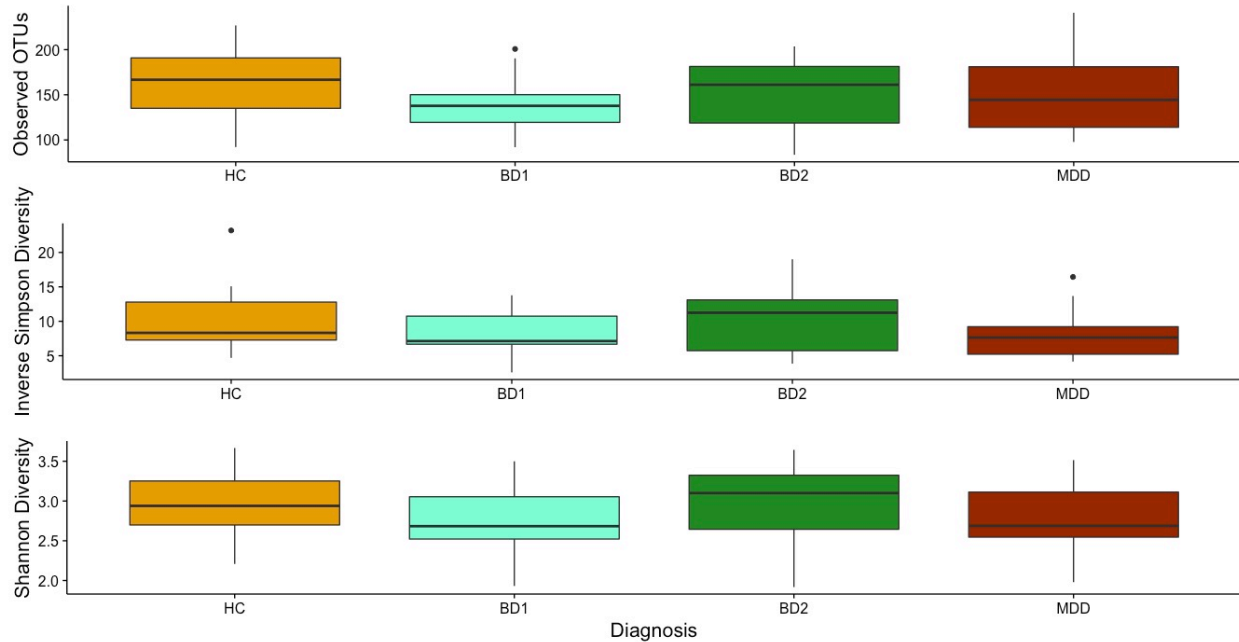
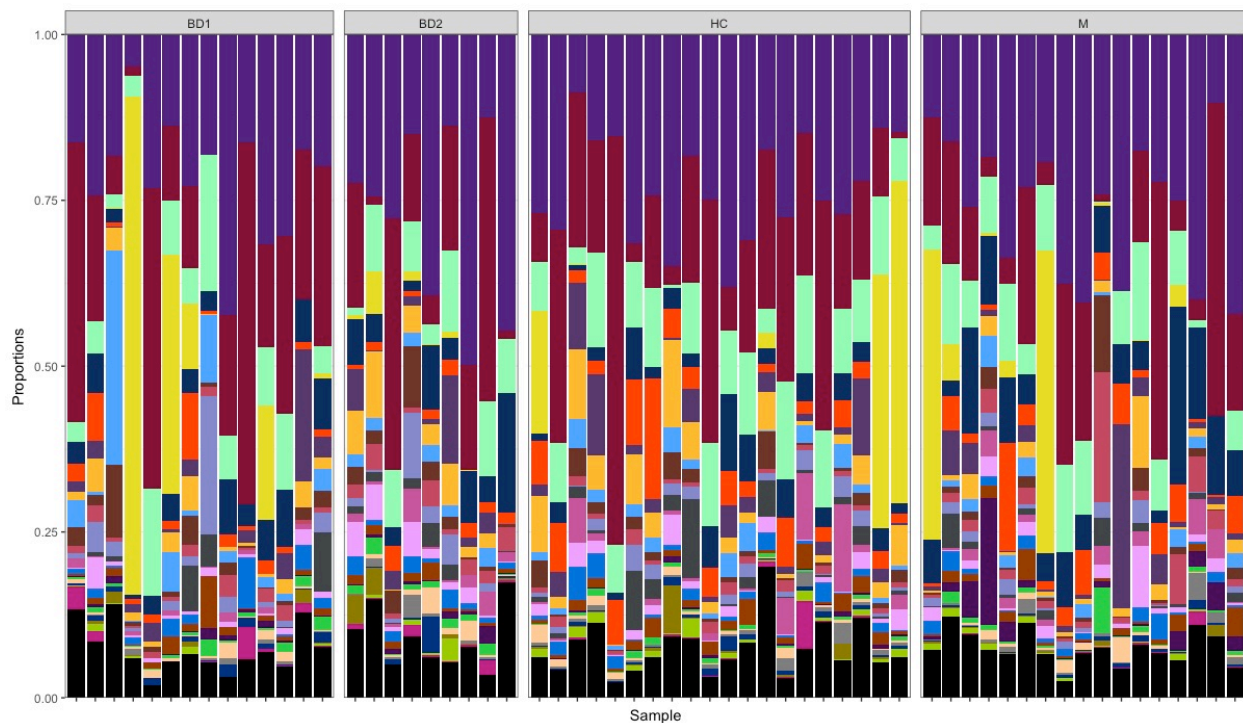


Figure 14: Comparison of OTU richness and evenness between the four diagnosis groups of study participants. The richness and evenness measures were generated by first rarifying (by means of sampling without replacement) the number of reads in each sample to the number of reads in the smallest sample (14 869). This process is then repeated 100 times and richness and evenness (measured with the Inverse Simpson Index and Shannon Index) for each sample is averaged across the 100 repeats. This method is adapted from Berry’s phyloseq tutorial¹⁰⁶ which itself is inspired by Hughes et al¹⁰⁷. That the mean and variance of observed species richness is statistically indistinguishable (Kruskall-Wallis Test, $p=0.45, 0.53, 0.26$) across the four groups suggests that diagnosis does not significantly impact observed community richness or evenness.

3.4.5 Abundance of specific genera distinguishes the microbiomes of patients and controls

The four diagnostic groups were then interrogated at the genus and OTU levels for defining organisms. Granted, a statistically significant clustering of microbiomes was not observed, but as this could be a problem stemming from the relatively small number of participants, and given the pilot nature of these investigations, putative differences were investigated. Bipolar I and II patients were pooled for comparison against the MDD and healthy volunteer populations, and then compared against each other. Figures four and five which follow are summary charts of the relative abundance of the top 25 most prevalent genera identified in participant samples.



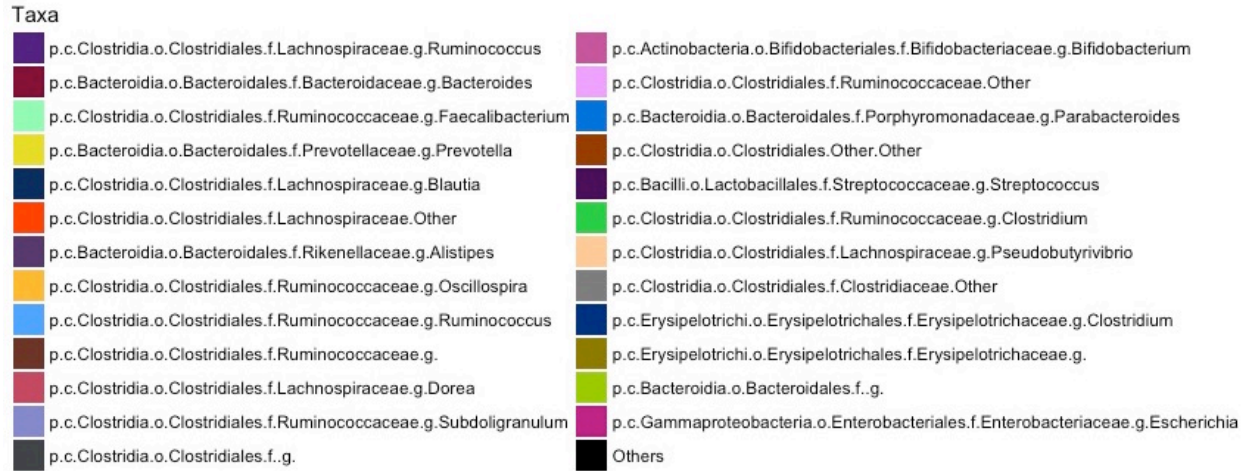
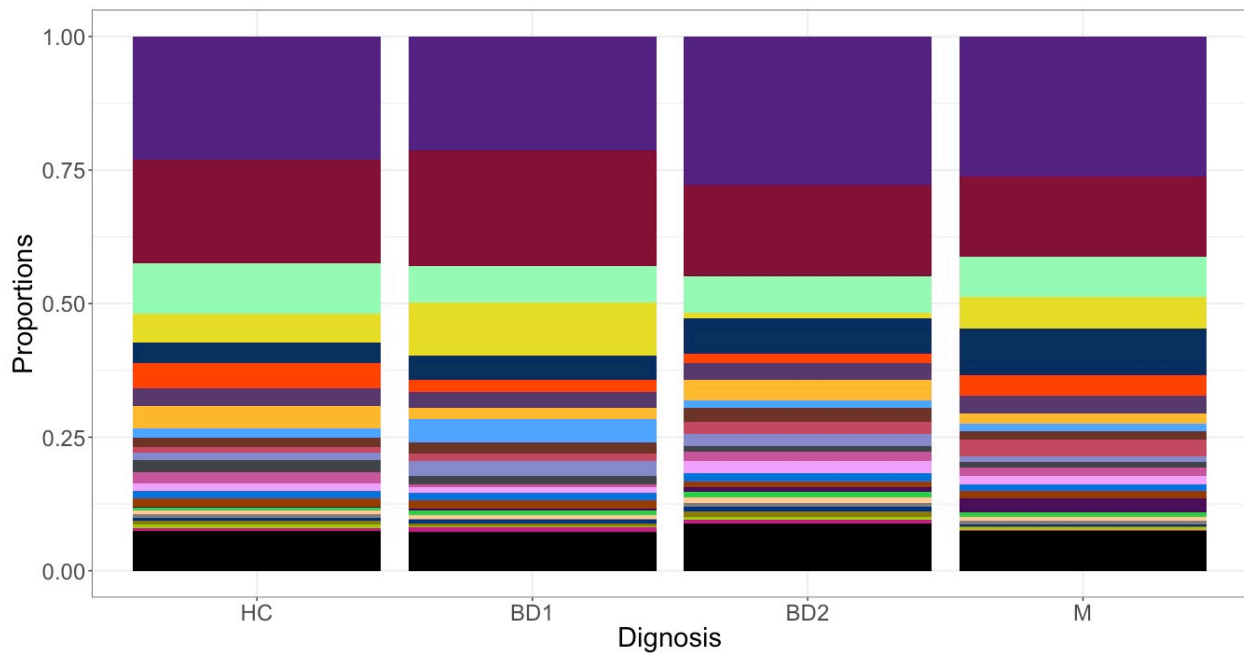


Figure 15: A comparison of the top twenty-five most abundant genera within individual samples, grouped by diagnosis.



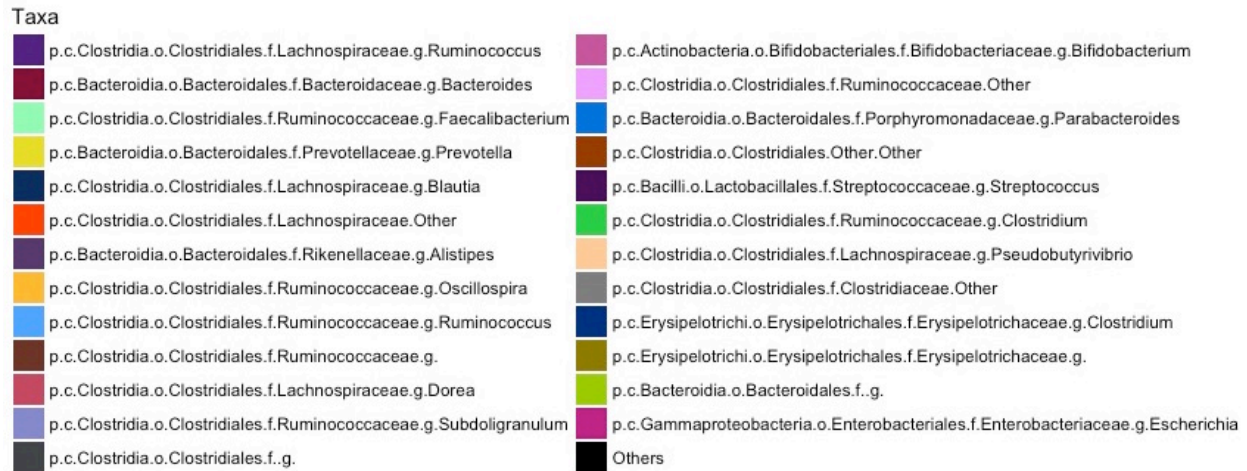


Figure 16: A comparison of the top twenty-five most abundant genera within each of the four participant diagnosis groups. Individual coloured portions represent the average relative abundance of individual genera while the bottommost section of the charts, indicated by the black portion, represents the combined abundance of the remaining genera. Abundance of each of the genera within the samples is calculated and then averaged within the groups. The family *Lachnospiraceae* were found to distinguish MDD patient samples from those of the bipolar patients and are shown above in orange. The remaining distinguishing OTUs that were identified were not among the 25 most abundant and thus are contained in the black portion of the plot.

In comparing the participant groups an enrichment of the genus *Sarcina*, as well as an OTU assigned to *Sarcina* were found to distinguish healthy volunteers from bipolar patients. An OTU from, as well as the genus *Collinsella* were found to be elevated in type two bipolar patients when compared to type one bipolar patients. A genus level group classified as family *Lachnospiraceae* genus other, and an OTU assigned to the genus *Lachnospira* (a genus found within the family *Lachnospiraceae*) were found to be elevated in depression patients when compared to bipolar patients. Finally, the bipolar patients were found to have an enrichment of a

bacteria belonging to the order *Streptophyta* when compared to the depression patients. All observed enrichments were found to be significant following FDR correction at a significance cutoff of 95% using ANCOM⁹⁶ software.

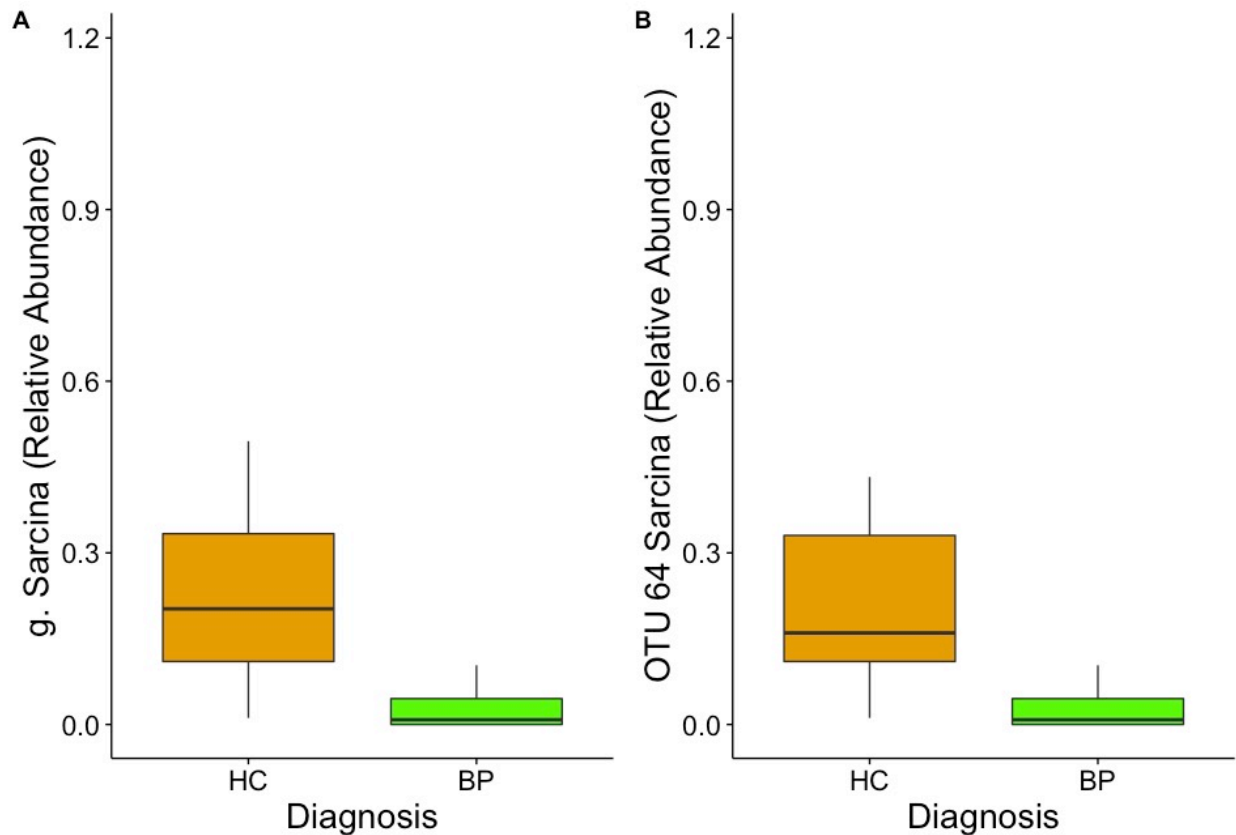


Figure 17: Boxplots comparing the relative abundance of the genus *Sarcina* (panel A) and a corresponding OTU 64 (panel B) in pooled Bipolar patients (shown in bright green). This OTU was identified as being significant at a significance threshold of 95% following FDR multiple testing correction using ANCOM software. An increased abundance of the OUT appears to distinguish type two bipolar patients from type 1 patients.

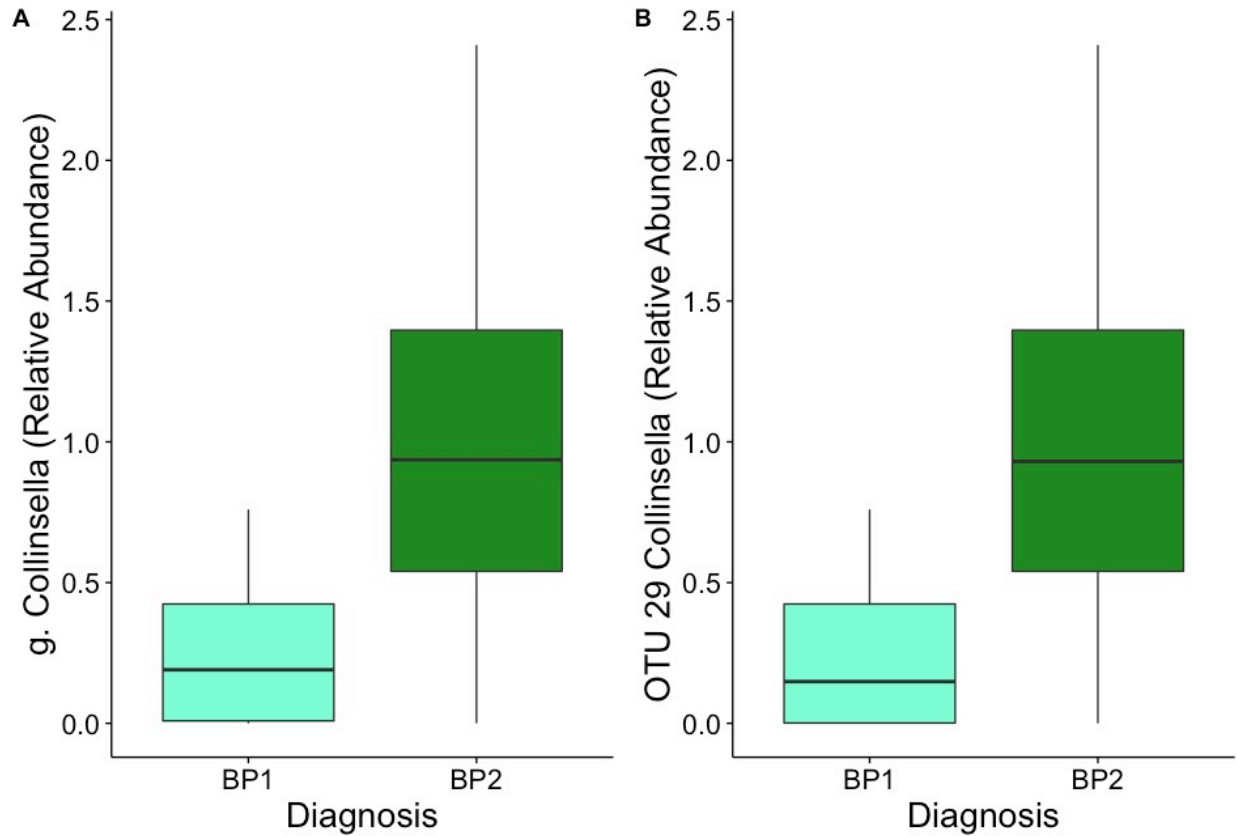


Figure 18: Boxplots comparing the relative abundance of the genus *Collinsella* (panel A) and a corresponding OTU 29 (panel B) in bipolar I and II patients. This OTU was identified as being significant at a significance threshold of 95% following FDR multiple testing correction using ANCOM software. An increased abundance of the OTU appears to distinguish bipolar II from bipolar I patients.

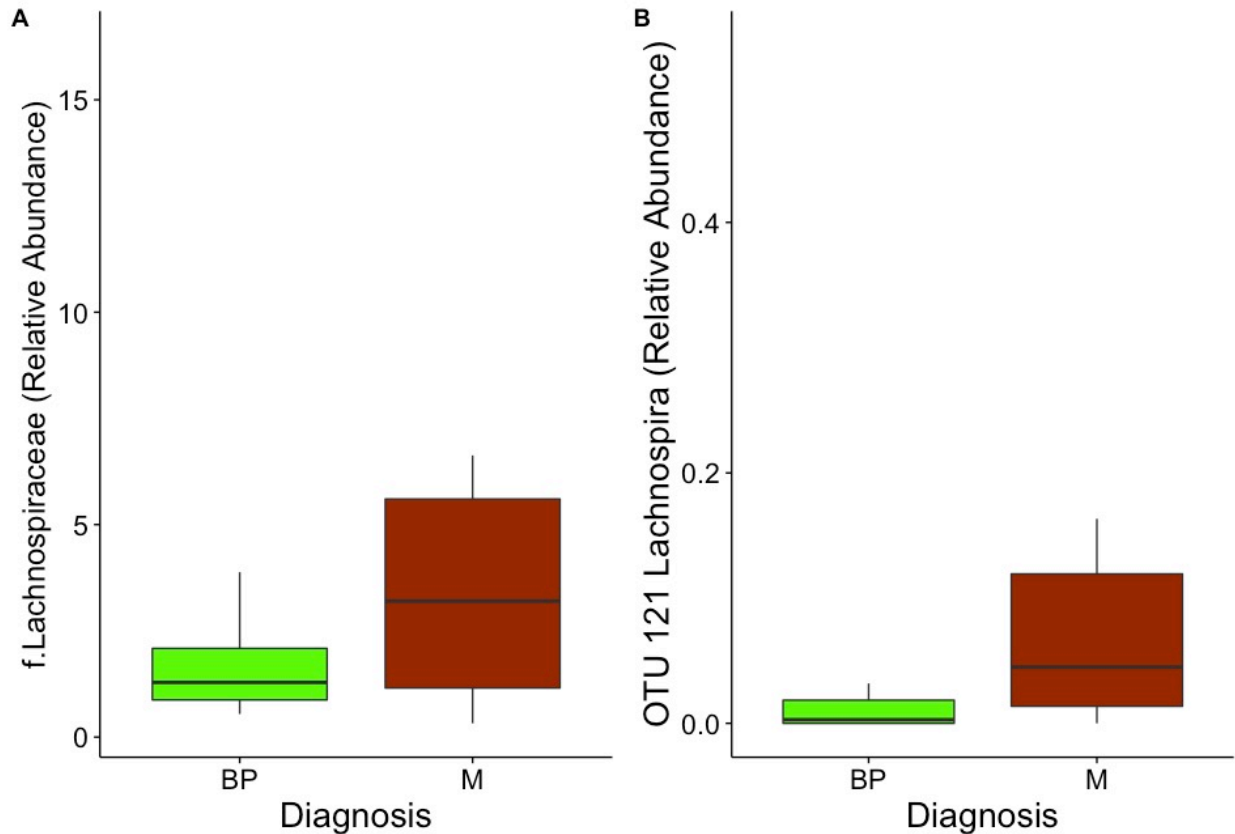


Figure 19: A comparison of the relative abundance of the family *Lachnospiraceae* (in panel A) and an OTU of the genus *Lachnospira* (a genus within the family *Lachnospiraceae*, and shown in panel B) in pooled bipolar (in bright green) and depression patients. This OTU was identified as being significant at a significance threshold of 95% following FDR multiple testing correction using ANCOM software. An increased abundance of the OTU distinguishes depression patients from bipolar patients within the dataset.

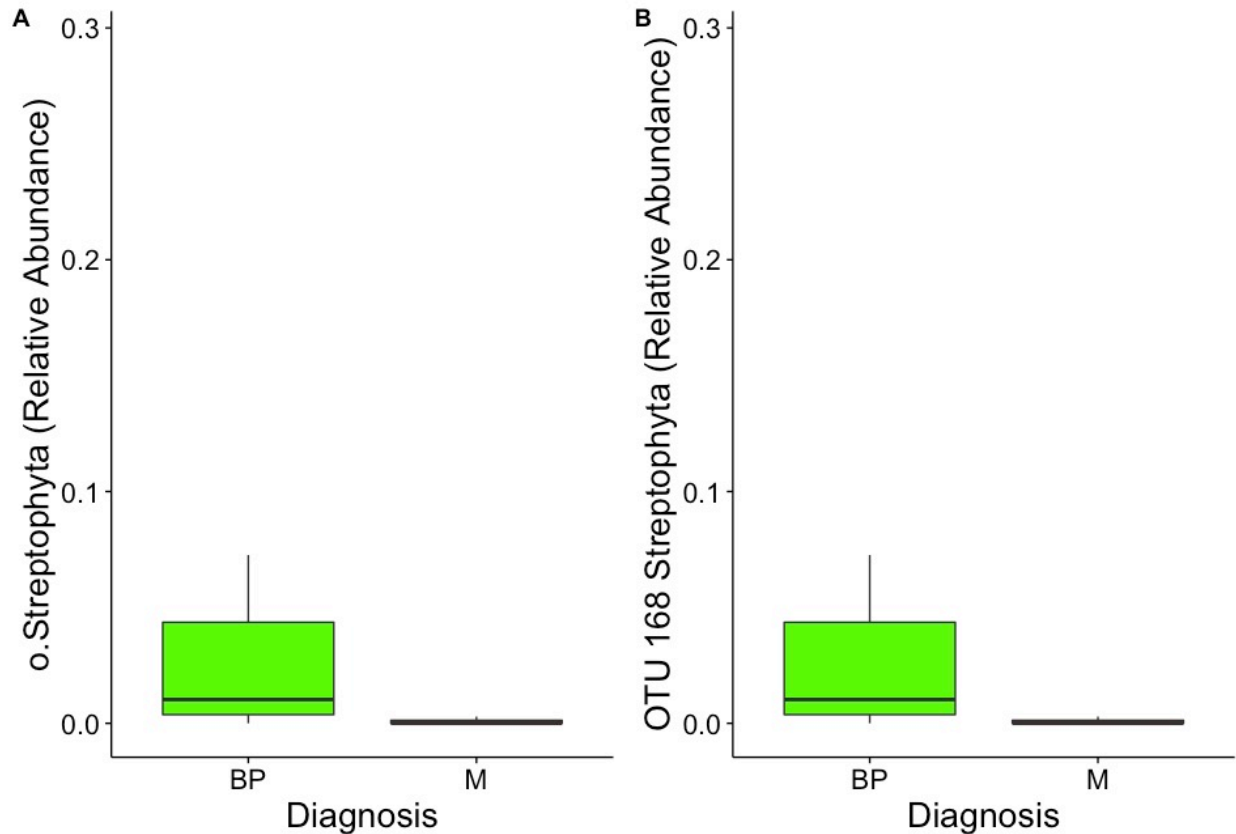


Figure 20: A comparison of the relative abundance of the order *Streptophyta* (in panel A) and its corresponding OTU 168 (in panel B) in pooled bipolar and depression patients. This OTU was identified as being significant at a significance threshold of 95% following FDR multiple testing correction using ANCOM software. An increased abundance of the OTU distinguishes bipolar patients from depression patients within the dataset.

3.4.6 Investigating the identity of OTU 168

One of the challenges of working with 16s reference databases is that many are contaminated with mitochondrial and chloroplast DNA sequences due in part to the shared evolutionary history of these organelles and the *Rickencella* and *Cyanobacteria*⁹⁴. Accordingly, in an effort to better interpret the observation of *Streptophyta* enrichment in bipolar patients relative to MDD patients,

the reference sequence from OTU 168 was investigated in a number of publically available databases to more accurately classify its identity. Table five below details the tools utilized to investigate OTU 168 along with summarizing the findings of these efforts.

Table 5: Summary of efforts to confirm the identity of *Streptophyta* OTU 168

Tool Used	Settings	Result	Strength of Finding
NCBI BLAST	Neucleotide Collection exclude uncultured/environmental samples	100% coverage of chloroplasts from greater than 50 different plants	
NCBI BLAST	16S Ribosomal RNA database	<i>Chroococidiopsis thermalis</i>	
RDP	16S rRNA training set 16	Root [100%] Bacteria [100%] Cyanobacteria/Chloroplast [100%] Chloroplast [100%] Chloroplast [100%] Streptophyta [100%]	
SILVA	Default	No Match	N/A

3.5 Discussion:

In comparing the putatively bipolar-associated microbiome to that of healthy controls and MDD patients, a number of novel insights regarding potential marker species were made. Although these data do not support the notion of statistically distinct clustering of bipolar patient samples, these putative marker species have the potential to drive hypothesis generation about possible roles of the microbiome in bipolar disorder. The following discussion will elaborate on each of these organisms, as well as using them to further elaborate on some of the challenging aspects of microbiome data as discussed in chapter one.

Firstly, in comparing the samples of the pooled bipolar patients to those of the healthy controls, the genus, and an OTU belonging to the genus *Sarcina* were identified as elevated in patients. In this case, as is evident in the comparisons of OTUs identified in figure 17, in this data set there was only one OTU assigned to the *Sarcina* and thus the genus level and OTU level plot are identical. This could be a helpful observation in that isolating the constituent strains within the genus *Sarcina* patients is one taxonomic level more focused than for example the family *Lachnospiraceae*, which was observed to be elevated in Bipolar patients relative to patients with MDD. *Sarcina* (from the order *Clostridiales*) are Gram-positive, round or cocci-shaped bacteria found both on human skin and in the gut¹¹². Species within the genus are both facultative and obligate anaerobes. One particularly remarkable distinguishing trait is their unique, cuboidal style of cell division from which they get their name (*Sarcina* is the Latin word for pack or bundle). Unlike other bacteria that divide along either along one or two planes, or, irregularly along a seemingly random set of planes; *Sarcina* consistently divide along three equidistant planes resulting in the formation of cuboid aggregations of cells¹¹². This unique division strategy

resulted in the initial categorization of a polyphyletic group of *Sarcina*^{112,113}. It has relatively recently been recognized that bacteria with the *Sarcina* style of cell division exists within the orders *Bacillales* (*Sporosarcina*), *Actinomycetales* (*Micrococcus*), *Kineosporiales* (*Kineococcus*) and even within the domain *Archaea* (*Methanosarcina*). Despite the novelty and relative scarcity of this division strategy with prokaryotes, however, no clinical significance has been associated with this trait. In fact, there is almost no mention of *Sarcina* in the context of human health within the literature. A review published in 2016 highlighted that since its initial identification in humans in 1843, only 19 cases of the type strain of genus, *S. ventriculi* have been reported in the English literature^{114,115}. This particular bacterium is often identified in the stomach of patients and is thought to be the cause of some ulcers, however, there is some debate as to whether *Sarcina* may be a benign commensal in the human stomach¹¹⁵⁻¹¹⁷. Additionally, in 1971, a culturing effort by Crowther found *Sarcina* to be recoverable from the stool of vegetarians at a much greater rate than from stool of omnivorous volunteers (75 of 106 vs. 2 of 123)¹¹⁸. It should be noted that it appears technology for anaerobic transportation of samples was not available in 1971, and further, that many of the vegetarian participants froze their stool samples before sending them to the lab. Both the stress of oxygen exposure and a freeze/thaw cycle would dramatically decrease the viability of the community and accordingly, the results must be viewed with a healthy degree of skepticism. Additionally, samples were cultured on two only two media, thus while samples were cultured at a variety of pHs and temperatures, the growth conditions provided can hardly be considered an exhaustive list¹¹⁸. A final shortcoming of this study was that *Sarcina* colonies were identified via microscopy only if morphology was identical to a previously described morphology of *S. ventriculi* colonies. Accordingly, it is possible that strains with different morphologies could have been missed. Furthermore, while the investigation of a

putatively vegetarian microbiome is its relative infancy¹¹⁹, there does not yet appear to be evidence corroborating the association of *Sarcina* and vegetarianism. One relatively more consistent finding concerning the *Sarcina* is their association with delayed gastric emptying^{115,116}. All of the observations supporting this association are case reports, however, and are made in severely unwell patients who present with conditions like stomach cancer, abdominal perforation and ulceration¹¹⁵. Accordingly, without further evidence¹¹⁵, it is difficult to determine whether *Sarcina* is a pathogen in the stomach, or if this portion of the literature is biased in that the only people whose stomach microbiomes are assessed, are patients who present as severely unwell. In the case of these data, while ten of the twenty healthy volunteers diet aligned most closely with the putatively vegetarian diet, diet was nonetheless, not found to significantly cluster microbiome data. Additionally, with the relative dearth of knowledge surrounding the genus and its role in gut health it is difficult to speculate as to how an absence of this organism could impact bipolar disorder. These organisms have, however, been observed to ferment carbohydrates and produce short chain fatty acids *in vitro*¹¹², accordingly given their tolerance of low pH perhaps they have a role in colonizing and SCFA signaling in the small intestine an area that is believed to be too acid for many of the organisms found through the rest of the digestive tract.

Collinsella was the organism observed to be more abundant in bipolar II patients when compared bipolar I patients. These organisms are also carbohydrate fermenters and are found in the human gut and on the surface of the tongue¹²⁰. *Collinsella*, in particular its type strain, *C. aerofaciens*, is believed to be the most abundant Gram-positive, non-spore-forming bacillus in the human gut¹²⁰. Despite its purported abundance, little is known about this organism in the context of human

health. Of note are the observations that abundance of this organism appears to be severely decreased in preterm infants given macrolide antibiotics¹²¹, it has been correlated with insulin levels in pregnant women¹²², is absent in the gut some cystic fibrosis patients¹²³ and is significantly reduced in the guts of IBS-C, IBS-D and smoking Crohn's disease patients relative to healthy volunteers^{124,125}. None of these correlations, however, are particularly helpful in explaining or inferring the significance of the discrepancy in abundance bipolar I and II patients

Two organisms were found to distinguish the microbiome of the bipolar cohort from that of the MDD patients; the bipolar patients showed an enrichment of *Streptophyta* while the microbiome of the MDD patients was enriched in the *Lachnospiriceae* and specifically, in an OTU belonging to the genus *Lachnospira*. Addressing first the *Streptophyta* OTU (a member of the photosynthetic *Cyanobacteria*), upon further reviewing the assigned taxonomy, this seems to be either a misidentification of DNA from chloroplasts, or alternatively, the misclassification of an organism(s) with a shared ancestor of the prokaryotes that became chloroplasts. This is an acknowledged shortcoming of the Greengenes database from which the OTU assignments are drawn⁹⁴, and speaks to the challenge of assigning identity to such a massively diverse community that potentially still contains organisms that have yet to be properly identified. Often when examining samples taken from eukaryotes, for example biopsies, a significant challenge is removing data sequenced from host DNA. It is possible that this OTU is truly representative of a cyanobacteria living in the gut, however, it is also possible that this DNA is sequenced from some incompletely digested plant material in participants' diets. Recently, however, a new phylum of bacteria which are believed to be closely related to the *Cyanobacteria*, the *Melanibacteria*, was identified in groundwater and human gut samples¹²⁶. Despite their

proposed shared ancestry, the *Melanibacteria*, are not believed to be photosynthetic, but are proposed to play a role in helping to supply their host with B and K vitamins. As was detailed in table 5, efforts were undertaken to align the *Streptophyta* 16s sequence with all available *Melanibacteria* sequences. These investigations were unable to yield a match, further suggesting that the sequence may not have originated in a bacterium, but rather from fruits and vegetables in participants' diets. This example highlights a challenge in working with these data, however, in that despite the best efforts of the scientists curating these databases they are not representative of every organism as many remain to be sequenced. Accordingly, at this point I suspect that this OTU represents DNA from participants' diets, especially since it aligned perfectly with the chloroplast sequences of a number of tree species in NCBI. There is some speculative literature surrounding the possible immunostimulatory capacity of cyanobacterial LPS, however, there is agreement that more research on LPS specifically belonging to these bacteria needs to be performed before even this hypothesis can be considered valid ¹²⁷.

The *Lachnospiraceae* and *Lachnospira* organisms are straight or slightly curved rod shaped cells that are strict anaerobes. Some species within the family and genus are motile, and most if not all type strains have been isolated from the digestive tracts of humans and domesticated mammals. Members of the *Lachnospiraceae*, including the *Lachnospira* are producers of butyrate. Butyrate production may be helpful in preventing colon cancer ¹²⁸, and is involved in signaling within the gut and along the gut brain-axis ^{129,130}. In the case of both groups of organisms whose abundance was found to differ between bipolar and MDD patients, abundance was also investigated between type bipolar one and two patients. No significant results or trends could be observed; however, this could be a result of the relatively small participant sample size.

A limitation of our study is the small sample size, increasing the chance of type 2 errors. In addition, there are many confounding factors including medication for which we are insufficiently statistically powered to control. Smoking, psychiatric medications and birth control medications represent a few examples of factors that could conceivably be expected to affect the gut microbiome through their effects on immunosuppression, GI motility and hormone levels respectively. Participant data regarding these factors, among others, were collected, however, recent research has shown that properly controlling for these variable in microbiome datasets can require hundreds of study participants⁴⁹. Accordingly, while table one shows rates of smoking and birth control use to be relatively similar between participant groups, with the current participant numbers it is impossible to accurately assess the possible statistical effects of these confounders on the microbiome. Additionally, data pertaining to patients' mental status was collected through questionnaires like the HAM-D, and a larger data set would allow the statistical power required to investigate possible correlations between abundance of the identified genera and quality of patients' mood.

While observational data like these cannot imply causation, the *Lachnospiraceae*, *Collinsella* and *Sarcina* are all known ferment carbohydrates and thus could be producing SCFAs that in turn are affecting gut brain signaling. Perhaps a metabolic by-product of *Sarcina* has a calming effect that contrasts a particularly stimulatory by-product produced by the *Collinsella*. Metabolomics of the SCFAs or untargeted metabolomics could offer another route for analyzing the samples that have been collected to better support hypotheses generated here-in. The

observation of the *Lachnospira* OTU specifically being enriched in the depression patients when compared to the bipolar patients could suggest that this strain is a marker, or potentially a contributor to depressed state. Accordingly, targeted culturing could be performed to isolate this organism so that it could be more thoroughly investigated phenotypically, or used in *in vitro* or *in vivo* experiments with mammalian cells or model organisms respectively to investigate putatively depressive effects. These data also highlight that while, as was detailed extensively in the introduction, there are promising data to suggest a possible role of the microbiota in maintaining or possibly disturbing mental health, there is no evidence, at least within this dataset, to suggest a single putative organism is the cause. More data would not only allow the opportunity to corroborate results, but could also allow statistical power to investigate hypothetical phenomena like an enrichment of certain groups of bacteria during or after a patient experiences a manic episode. While it would require significant time and effort on the part of the patient, data from samples collected over a course of a number of months, or even years might inform a better understanding of how a bipolar individual's gut microbiota fluctuate with their mood. These patterns could offer insights into how to accelerate the completion of a manic episode once its begun, or perhaps eventually, allow for a prediction of when a manic episode is about to occur such that a patient could seek support if not eventually preemptive treatment.

3.6 Chapter 3 Acknowledgements

With regards to this chapter I would like to specifically acknowledgment goes to Ms. Merwa Amer, Ms. Jessica Gillard and Ms. Cindy James for their efforts in recruiting in Hamilton and to Mehala Subramanieapillai and for their help recruiting in Toronto. Again, I am also indebted to Ms. Laura Rossi and Ms. Michelle Shah for their help processing and sequencing the fecal

samples. Given the great deal of overlap, particularly in methods, identical methodology will not be reprinted, rather, the materials and methods section will serve to highlight the differences between the methods used here and the ones used in chapter one.

Chapter 4: Analyzing *Bacteroides* strains isolated from a GAD patient and healthy volunteer

4.1 Introduction

The following experiments were undertaken to better understand the bacteria within the genus *Bacteroides* that appeared to be overrepresented in GAD patients following analysis of the pilot thirty and thirty (patients and matched healthy controls) data set. Phenotypic analyses were undertaken to better characterize isolated strains with the purpose of generating hypotheses about how these strains might contribute to both GI symptoms and gut-brain-microbiome cross-talk and in doing so perpetuate the disorder. Additionally, whole genomes of the isolates were amplified and sequenced such that they could be investigated for differences that might drive further hypothesis testing of putative host-impairing traits.

4.2 Methods

4.2.1 Strain collection construction

Samples used for culturing were identified based on review of 16s rRNA gene profiles generated in chapter one. Ultimately, patient and control samples were selected on the basis of having a large abundance of the overrepresented *Bacteroides* genus, as well as on the basis of having a large abundance of the genus *Alistipes*, as these bacteria were being investigated by Mr. Ye.

100 mg of fecal sample was collected from one of the stock containers of frozen sample using a Disposable Biopsy Punch (Integra™, Miltex ®. Integra York PA INC.) under anaerobic conditions and subsequently suspended in 900 uL of BHI broth with 0.05% L-cysteine. The samples were serially diluted in the BHI broth with L-cysteine and 100 uL aliquots of the 10^{-2} , 10^{-4} and 10^{-5} dilutions were plated in duplicate on Columbia Blood Agar (CBA), Colistin Nalidixic acid Blood Agar (CNA), Bacteroides Bile Esculin Agar (BBE), and Fastidious Anaerobe Agar (FAA). Plates were then incubated anaerobically at 37° Celsius for seventy-two hours. Following incubation, nearly 1500 colonies were picked both by hand and with the help of the RapidPick SP from Hudson Robotics (Hudson Robotics, Springfield NJ) all under anaerobic conditions. Colonies were picked into the 75 uL aliquots of the same BHI broth with 0.05% L-cysteine solution in 96 well plates and allowed to grow over night. Following overnight growth, 20 uL of culture was added to 50 uL of Worm Lysis Buffer (5% 1M KCl, 1% 1M Tris-HCl (pH 8.3), 0.25% 1M MgCl₂, 0.45% Igepal, 0.45% Tween-20, 93% H₂O all mixed and then added to a 10mM solution of Proteinase-K at a ratio of 99:1(recipe courtesy the lab of Dr. L. MacNeil)) and boiled at 99° Celsius for fifteen minutes. After removing 20 uL of culture, 50 uL of 20% skim

milk solution was added to each of the wells as a cryoprotectant and plates were stored at -80° Celsius. 5uL of the resulting boiled solution was then used as template for PCR following the previously described protocol to prepare samples for Illumina sequencing. The only modification made from the original protocol was that the V3-V4 region of the 16s gene was amplified instead of just the V3 region. Given that individual wells were inoculated by single colonies and individually sequenced, sequencing results were used to identify wells containing bacteria from the genus *Bacteroides*. In total 18 OTUs assigned to the genus *Bacteroides* were identified, from which an initial screen of 32 wells or isolates was completed. Under anaerobic conditions, wooden sticks were used to scrape some of the still frozen milk-culture mix from each of the target wells and the scraping was streak for isolation on CBA agar plates. Once strains were successfully grown, they were inoculated into 5 mL of BHI broth with 0.05% L-cysteine and grown over night. Following the overnight growth, 1 mL of culture was mixed with 1 mL of 20% skim milk to serve as a stock, 1 mL of culture was added to 333 uL of 60% glycerol to serve as a glycerol stock, 1 mL of culture was added to a 1.5 mL Eppendorf tube and 20 uL of culture were added to worm lysis buffer for confirmatory PCR. The milk and glycerol stocks were frozen immediately following preparation. Milk and glycerol were both initially used as cryoprotectants as there was a dearth of literature regarding successful storage of the genus *Bacteroides* at -80° Celsius. Milk was found to be a suitable cryoprotectant for all strains isolated and thus to minimize use of freezer space subsequent stocks were made only in skim milk. Cultures in the 1.5 mL Eppendorf tubes were spun at 16 000 RPM for 5 minutes using a benchtop centrifuge to pellet cells. Following centrifugation, supernatants were carefully pipetted off of the pelleted cells and stored separately alongside stocks and pelleted cells for potential future analysis. Pelleted cells were used as template for DNA extraction for whole genome sequencing.

5 uL of the boiled culture was used as template for full 16s PCR using the 8f and 926R primers. PCR products were visualized on a 2% agarose gel with ethidium bromide and then sent to GENEWIZ (South Plainfield, NJ) for sequencing.

From the original set of 32 isolates, 17 were determined to be *Bacteroides*. Wanting to expand the collection beyond 17 isolates, a further 45 wells (from the initial collection of 1500 strains) were selected. The same protocol was undertaken to isolate, preserve and identify the strains, save for the glycerol stock which was deemed redundant. Full 16s PCR was performed on these strains as previously described and the PCR products were sequenced by GENEWIZ (South Plainfield, NJ). Ultimately these efforts yielded a collection of thirty-three strains which are stored as individual pure culture stocks in the -80° Celsius freezers of the Surette lab.

V3-V4 data from the initial sequencing efforts were also used in conjunction with sequence data from patient and control samples (detailed extensively in chapter 2) and run through the Oligotyping pipeline from A. Murat Eren¹³¹. This tool is used to dissect OTUs to see if there is evidence for more than one sequence featuring prominently within the OTU. Given that OTUs are constructed in our pipeline based on 97% sequence similarity, it is almost always the case that each OTU contains two or more putatively real sequences. Thus, by breaking down the OTUs one is better able to understand which sequences constitute the OTUs. Relative to the computer generated representative sequence, the sequences belonging to the oligotype subgroups are more accurate thus allowing for better comparison of sequences between sequencing runs. Given that the V3-V4 data was from putatively isolated colonies, there was not significant information to be gained through oligotyping this data. In the case of the patient and control data,

however, the top four *Bacteroides* OTUs were successfully separated into 23 distinct sequences. A full outline of the oligotyping results can be seen in supplemental TABLE X, however, OTUs were broken following the method outlined in the paper¹³¹. Oligotyping continued until purity scores were greater than 0.9 (out of a possible 1) and sequences were only retained if they contained greater than one percent of the reads found in the original OTU. Most importantly, this analysis confirmed the successful isolation and culture of strains that shared 100% sequence identity with the *Bacteroides* OTU (OTU 2) found to be overrepresented in the pilot data set, as well as a number of other *Bacteroides* strains.

4.2.2 Whole genome sequencing:

Pelleted cells were used as the template for whole genome sequencing. As was previously mentioned, these cells were pelleted from 1 mL of a pure culture of cells grown anaerobically in BHI with 0.05% L-cysteine for forty-eight hours. Genomic DNA was extracted from the cells with the help of the Wizard DNA extraction kit (Promega, Madison, WI) following the protocol included with the kit. Purified DNA was then measured using the NanoDrop 2000 (Thermo Fischer Scientific, Waltham, MA), standardized to a concentration of 1ng/uL and sent to the McMaster sequencing center for sequencing on the Illumina HiSeq machine. Genomes were assembled using A5-miseq pipeline and annotated using Prokka^{132,133}. The software KmerFinder was used to assign taxonomy to the whole genome sequence data set.

4.2.3 Growth Assays

All phenotypic assays were carried out anaerobically at 37° Celsius unless otherwise stated. The broth media used for growth was always BHI supplemented with 0.05% L-cysteine. Pure culture

stocks were streaked from frozen onto CBA plates and let to incubate for 48hrs to grow. After 48 hours, well isolated colonies from these streaks were inoculated into five milliliters of BHI broth with 0.05% L-cysteine and grown over night. Following growth, 500uL of the overnight cultures was added to 1 mL of fresh media in individual wells within a 96-well plate such that they could then be stamped onto agar plates for phenotypic analyses. Media used for phenotyping included CBA¹³⁴ and FAA¹³⁵ both supplemented with 5% of either defibrinated sheep's blood, or a mix of human blood collected from volunteers in the lab to test hemolysis, 20% skim milk supplemented BHI agar to test proteolysis, unsupplemented M9 salts minimal media, M9 salts media supplemented individually with, and with the combination of 20% glucose and case amino acids to investigate growth on minimal media and agarose activity¹³⁶, BBE agar to investigate esculin hydrolysis¹³⁷, and finally, Prussian blue agar to investigate hydrogen peroxide production¹³⁸.

4.2.4 Antibiotic resistance assay

The antibiotic resistance assay was adapted from the standard Kirby-Bauer disc diffusion method, following the methods of Bodner *et al.* such that it could be performed on the anaerobic *Bacteroides* strains^{139,140}. To conduct the antibiotic resistance assay 60 mL large diameter plates were made using CBA supplemented with 5% defibrinated sheep's blood. 100uL of overnight cultures of each of the strains were individually diluted in 900uL of sterile 5% NaCl solution and optical density at 600 nm was measured. These readings were used to calculate the necessary dilution such that cultures of each strain could be diluted to an OD 600nm of 0.1. Standardized cultures were then streaked in duplicate using a sterile cotton swab applicator onto the plates to create a lawn of growth. Immediately following application of the strains, antibiotic disks were

applied to the plate. Eleven of the twelve antibiotics used were in the form of purchased standardized discs. These eleven antibiotics were Ciprofloxacin (CIP, 5ug), Amoxicillin-Clavulnic Acid (AMC 30 ug), Vancomycin (VA, 30 ug), Tobramycin (TOB, 10 ug), Cefoxitin (FOX, 50ug), Cefapime (FEP, 30ug), Meropenem (MEM, 10ug), Tetracycline (TET, 30ug), Trimethoprim-Sulphamethoxazole (SXT, 25 ug), Erythromycin (E, 15ug), Levofloxacin (LEV, 5ug). Our lab did not own discs of the twelfth antibiotic used, Metronidazole, accordingly, a stock solution of 10 ug/uL was prepared and applied at a volume of 10uL to a sterile paper disc such that it was at the same amount 50 ug as in Metronidazole discs sold for this assay by Thermo Scientific. Following disc application, plates were incubated for forty-eight hours and then photographed. Break-point data was not available when assessing antibiotic resistance for the *Bacteroides*, accordingly organisms were identified as resistant to a given antibiotic if they grew up to the edge of the disc. Approximate estimations of relative inhibition were made through visual comparison of antibiotic inhibition zones within strains. Antibiotics were ranked per strain based on the size of their zone of inhibition relative to other antibiotics.

4.2.5 Serotonin Production Assay

In collaboration with the lab of Dr. Kahn, an exposure assay was conducted in which cultures of Bon cells were exposed to the *Bacteroides* strain collection to investigate if the strains would differentially promote serotonin expression. Given the importance of serotonin as a signaling molecule in the gut, this assay was undertaken in hopes of exploring strain differences in capacity to stimulate serotonin production.

Bon cells were derived from a carcinoid tumor in 1994 by the group of Evers et al. and are used as a model system of enterochromaffin cells¹⁴¹. Given that enterochromaffin cells are known to be one of the primary producers of serotonin in the gut, were some of the strains able to significantly increase or decrease serotonin production *in vitro* this could have significant effects on the gut environment.

For this experiment, Bon cells were cultured by Dr. Dewan from the Kahn lab overnight at 37° Celsius in Dulbecco's modified Eagle's medium (DMEM) supplemented with fetal calf serum (FCS, 10% v/v), and penicillin and streptomycin (Penstrep, 1% v/v). Cells were collected and concentration was measured using a hemocytometer. Cells were then seeded into 24 well plates at an approximate concentration of five-hundred-thousand cells per well. In total one-hundred and nineteen such wells were seeded across five plates. One milliliter of fresh DMEM medium supplemented with FBS and Penstrep and cells were then incubated overnight at 37° Celsius. Concurrently, *Bacteroides* cultures were grown overnight in BHI broth from frozen stocks as previously described. For this experiment cultures of *Bacteroides* were not standardized. In future experiments, standardization could be achieved by measuring optical densities of the cultures and then diluting them all to a standard concentration as was done in the antibiotic resistance assay. Given, however, that this initial assay was a pilot to see if any interactions of interest would be observed it was decided that overnight cultures added to the BON cells at a controlled volume would suffice.

In preparation for the addition of the bacterial cells, the BON cells were first examined under a microscope to ensure morphology was consistent with that of healthy growing cells. Following

this examination, the spent culture medium was carefully pipetted off of the cells. Cells were then gently washed with two-hundred microliters of PBS in an effort to remove any remaining antibiotic. One milliliter of fresh DMEM without FBS or Penstrep was then added to each of the wells and finally ten microliters of *Bacteroides* cell culture was added to each of the wells in triplicate. Accordingly, the one-hundred and nineteen wells were used as follows: ninety-nine were used for triplicate assays investigating response to the addition of ten microliters of culture of each of the thirty-three strains. Three wells across two plates were used as growth/viability controls and only contained the Bon cells and fresh DMEM. Four wells were supplemented with two and four microliters methyl alpha-D-glucopyranoside (75 mmol/L) as a positive control for serotonin production. Twelve wells were used to assay potential dose effects of adding five or twenty microliters of culture in duplicate using strains ten, fifteen and twenty-eight. These strains were selected as they represented three different species from within the strain collection. Finally, one well contained sterile DMEM without Bon cells as a media control.

Cell cultures were then again incubated overnight at 37° Celsius. Following incubation, supernatants were carefully pipetted off of cell cultures and given to Dr. Dewan to use as the substrate for the 5-HT ELISA sourced from Beckman Coulter. Cells were then vigorously re-suspended in TRIZOL and stored at -80° C for potential future mRNA work. Unfortunately, at the time of writing, the ELISA kit had not arrived and thus the completed results of this assay cannot be reported.

4.3 Results

4.3.1 Isolated strains represent a variety of species

Upon completion of the lengthy colony picking procedure, the 16s gene of the isolates were amplified to determine their species. The taxonomy obtained from 16s sequencing as well as that obtained via KmerFinder analysis¹⁴² of the whole sequenced genomes revealed that a number of different species were included in the collection of thirty-three strains. Species identities, as well as the sample of origin for each of the strains in the collection is summarized in table 3.

Table 6: Species identities of the organisms within the *Bacteroides* strain collection

Strain Number	Isolated From	16s ID	WGS ID
1	GAD pt.	<i>B. vulgatus</i>	<i>B. vulgatus</i>
2	HC	<i>B. uniformis</i>	<i>Bifidobacterium adolescentis</i>
3	HC	<i>B. xylanosolvens</i>	<i>B. thetaiotaomicron</i>
4	HC	<i>B. xylanosolvens</i>	<i>B. thetaiotaomicron</i>
5	HC	<i>B. ovatus</i>	<i>B. thetaiotaomicron</i>
6	HC	<i>B. ovatus</i>	<i>B. thetaiotaomicron</i>
7	HC	<i>B. ovatus</i>	<i>B. thetaiotaomicron</i>
8	HC	<i>B. vulgatus</i>	<i>B. vulgatus</i>
9	GAD pt.	<i>B. massiliensis</i>	<i>B. vulgatus</i>
10	HC	<i>B. xylanosolvens</i>	<i>B. thetaiotaomicron</i>
11	HC	<i>B. uniformis</i>	<i>B. vulgatus</i>
12	HC	<i>B. vulgatus</i>	<i>B. vulgatus</i>
13	HC	<i>B. vulgatus</i>	<i>B. vulgatus</i>
14	HC	<i>B. vulgatus</i>	<i>B. vulgatus</i>
15	GAD pt.	<i>B. vulgatus</i>	<i>B. vulgatus</i>
16	HC	<i>B. vulgatus</i>	<i>B. vulgatus</i>
17	GAD pt.	<i>B. dorei</i>	<i>B. vulgatus</i>
18	GAD pt.	<i>B. vulgatus</i>	<i>B. vulgatus</i>
19	GAD pt.	<i>B. caccae</i>	<i>B. thetaiotaomicron</i>
20	GAD pt.	<i>B. vulgatus</i>	<i>B. vulgatus</i>
21	GAD pt.	<i>B. caccae</i>	<i>B. thetaiotaomicron</i>
22	GAD pt.	<i>B. vulgatus</i>	<i>B. vulgatus</i>
23	GAD pt.	<i>B. vulgatus</i>	<i>B. vulgatus</i>
24	GAD pt.	<i>B. caccae</i>	<i>B. thetaiotaomicron</i>
25	GAD pt.	<i>B. caccae</i>	<i>B. thetaiotaomicron</i>
26	GAD pt.	<i>B. faecis</i>	<i>B. thetaiotaomicron</i>
27	GAD pt.	<i>B. caccae</i>	<i>B. thetaiotaomicron</i>
28	HC	<i>B. uniformis</i>	<i>Enterococcus faecium</i>

29	GAD pt.	<i>B. vulgatus</i>	<i>B. vulgatus</i>
30	GAD pt.	<i>B. uniformis</i>	<i>B. helcogenes</i>
31	GAD pt.	<i>B. vulgatus</i>	<i>B. vulgatus</i>
32	HC	<i>B. vulgatus</i>	<i>B. vulgatus</i>
33	HC	<i>B. vulgatus</i>	<i>B. vulgatus</i>
Totals:	GAD pt. –	<i>B. vulgatus</i> - 15	<i>B. vulgatus</i> – 18
	17	<i>B. uniformis</i> - 4	<i>B. thetaiotaomicron</i> – 12
	HC - 16	<i>B. caccae</i> - 5	<i>Enterococcus faecium</i> - 1
		<i>B. xylanosolvens</i> – 3	<i>Bifidobacterium adollescentis</i> - 1
		<i>B. ovatus</i> – 3	
		<i>B. faecis</i> – 1	
		<i>B. massiliensis</i> – 1	
		<i>B. dorei</i> – 1	

Due to some delays at the sequencing center, whole genome sequencing results were not obtained until early May. Accordingly, all the other experiments were completed on the collection assuming all the strains were *Bacteroides*. The validity of the KmerFinder assignment of strains 2 and 28 will be further elaborated upon in the discussion.

4.3.2. Phenotypes distinguish strains, even within the same species

The first of the four sets of experiments performed on the *Bacteroides* collection was phenotyping. The collection was stamped on a number of different media so that different phenotypes could be observed. Each assay was performed in duplicate and each of the phenotyping growth experiments were performed at least twice meaning that phenotypes were consistently observed a minimum of four times. Serendipitously while conducting the assays, an unexpected phenotype was observed, agarase activity, which describes a phenotype in which the bacteria are able to metabolize the agar in the media as a carbon source and thus form a dimple in the plate. Hemolysis was tested on both 5% defibrinated sheep's blood plates and pooled human blood plates. Given that the strains did not grow at equal rates, for these experiments

growth was semi-standardized as all strains were allowed 24 hours to grow in broth prior to the assay. Given, however, that this does not allow for a true standardization or control of number of cells inoculated onto the phenotyping plates and magnitude of phenotypes was not compared

Table 7: A summary of phenotypic variability observed across the *Bacteroides* strain collection

Strain #	Species 16s ID	Species WGS ID	Hemolysis	Proteolysis	Esculin Hydrolysis	H ₂ O ₂ Production	Agarase Activity
19	<i>caccae</i>	<i>B. thetaiotaomicron</i>	-	-	+	-	-
21	<i>caccae</i>	<i>B. thetaiotaomicron</i>	-	-	+	-	-
24	<i>caccae</i>	<i>B. thetaiotaomicron</i>	-	-	+	-	-
25	<i>caccae</i>	<i>B. thetaiotaomicron</i>	-	-	+	-	-
27	<i>caccae</i>	<i>B. thetaiotaomicron</i>	-	-	+	-	-
17	<i>dorei</i>	<i>B. vulgatus</i>	-	-	+	-	-
26	<i>faecis</i>	<i>B. thetaiotaomicron</i>	-	-	+	-	-
9	<i>massiliensis</i>	<i>B. vulgatus</i>	-	-	-	-	-
5	<i>ovatus</i>	<i>B. thetaiotaomicron</i>	-	-	+	-	-
6	<i>ovatus</i>	<i>B. thetaiotaomicron</i>	-	-	-	-	-
7	<i>ovatus</i>	<i>B. thetaiotaomicron</i>	-	-	-	-	-
2	<i>uniformis</i>	<i>Bifidobacterium adolescentis</i>	-	-	-	-	+
11	<i>uniformis</i>	<i>B. vulgatus</i>	-	+	-	-	-
28	<i>uniformis</i>	<i>Enterococcus faecium</i>	-	+	+	+	+
30	<i>uniformis</i>	<i>B. helcogenes</i>	-	-	+	-	-
1	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-
8	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-
12	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-
13	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-
14	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-
15	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-

16	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-
18	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	+	-	-
20	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	+	-	-
22	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-
23	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-
29	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	+	-	-
31	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-
32	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	-	-	-
33	<i>vulgatis</i>	<i>B. vulgatus</i>	-	-	+	-	-
3	<i>xylanosolvans</i>	<i>B. thetaiotaomicron</i>	-	-	+	-	-
4	<i>xylanosolvans</i>	<i>B. thetaiotaomicron</i>	-	-	+	-	-
10	<i>xylanosolvans</i>	<i>B. thetaiotaomicron</i>	-	-	+	-	-

Figures 21, 22 and 23 are photographs taken of a BBE and CBA plate included to illustrate the variability in esculin hydrolysis and evidence of agarose activity respectively. The strain collection is arrayed vertically in columns of eight.

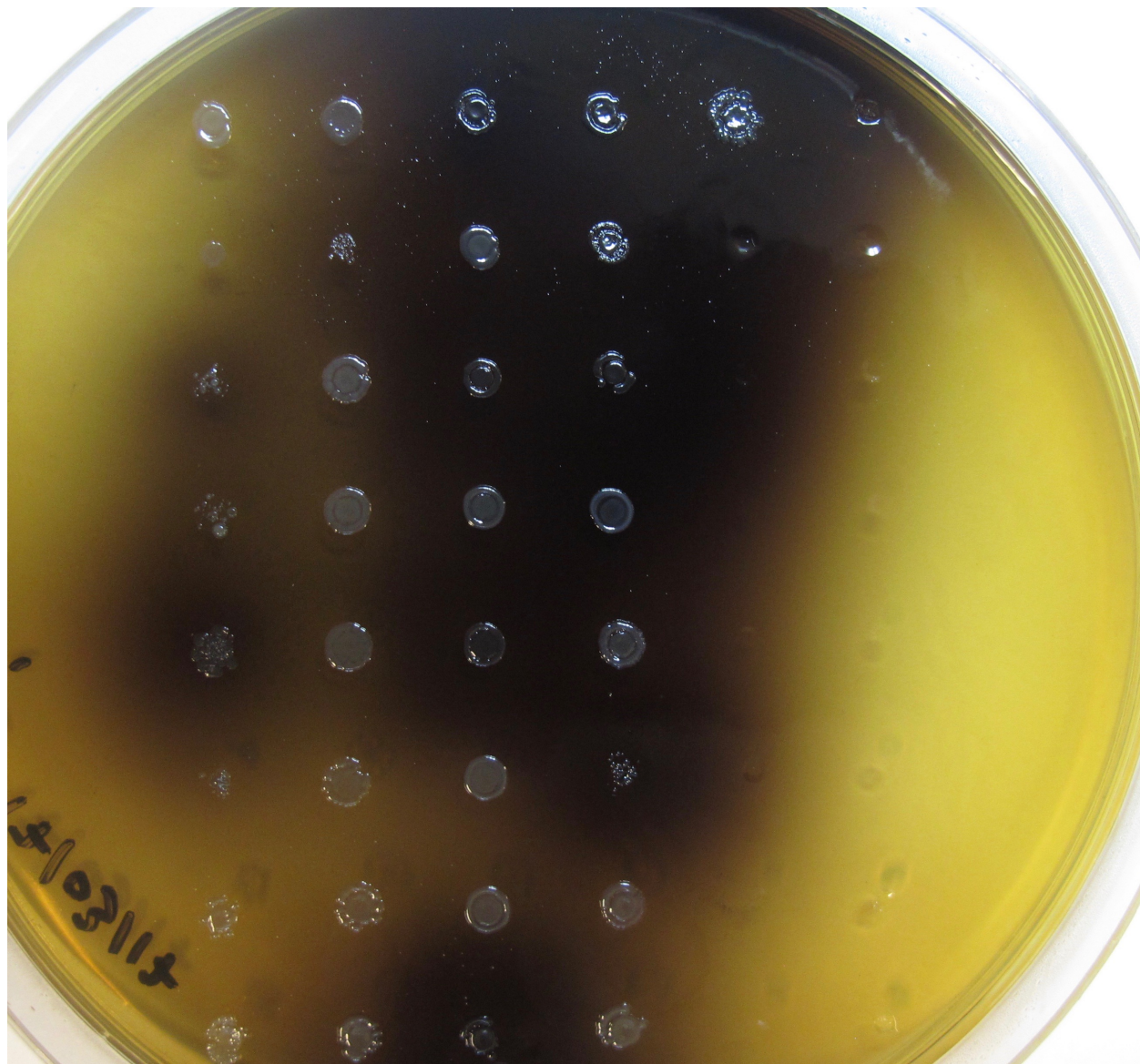


Figure 21: A sample BBE plate used to assay esculin hydrolysis in the *Bacteroides* strain collection. Esculin is hydrolyzed to dextrose and esculetin which reacts with ferric iron ions in the medium turning the media around the colonies a dark brown or black.

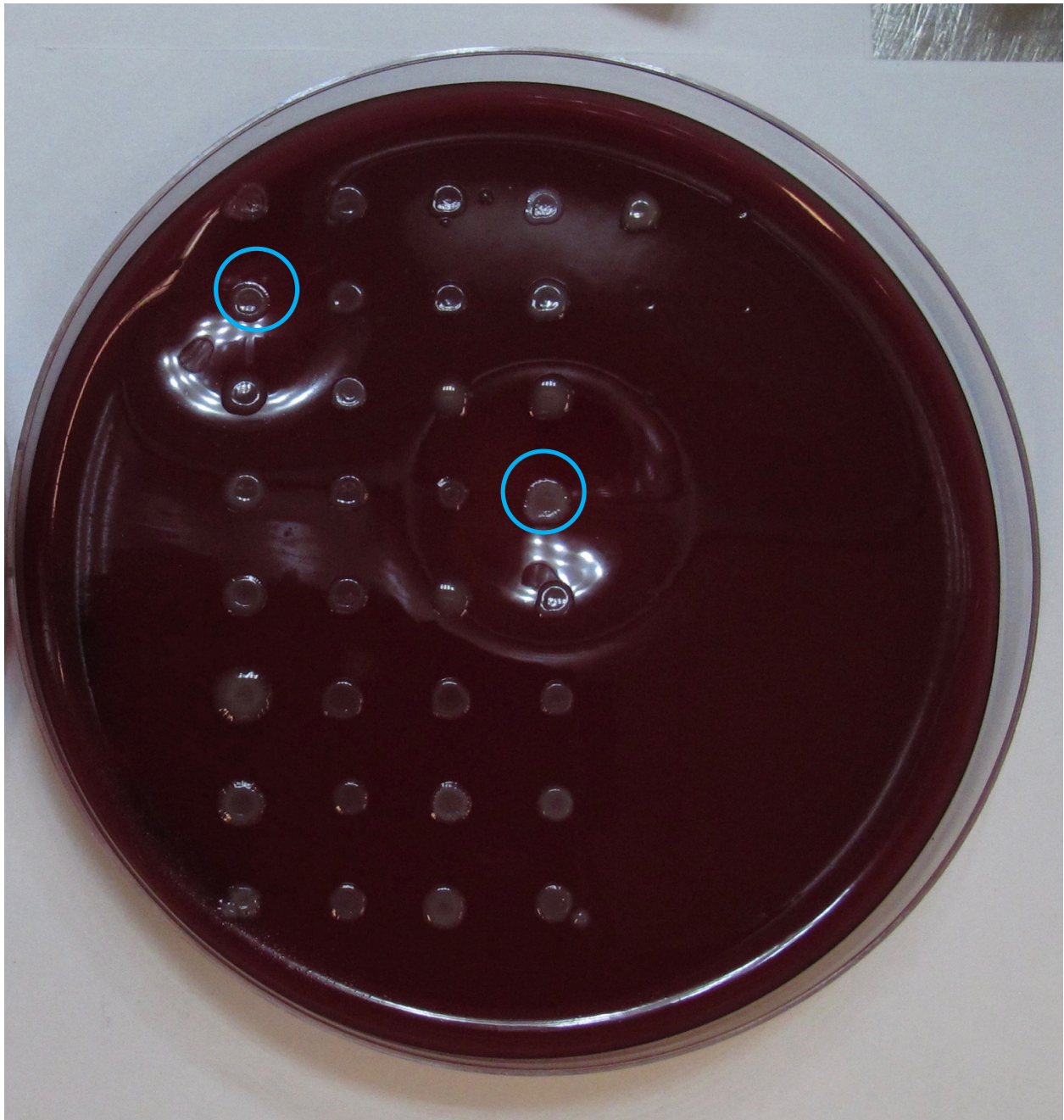


Figure 22: An overhead photograph of a CBA plate showing agarose activity in strains two and 28. Although agarose activity is a bit challenging to photograph it is most visible in this photo as the dimple or halo in the media surrounding the two strain that are highlighted with blue circles.

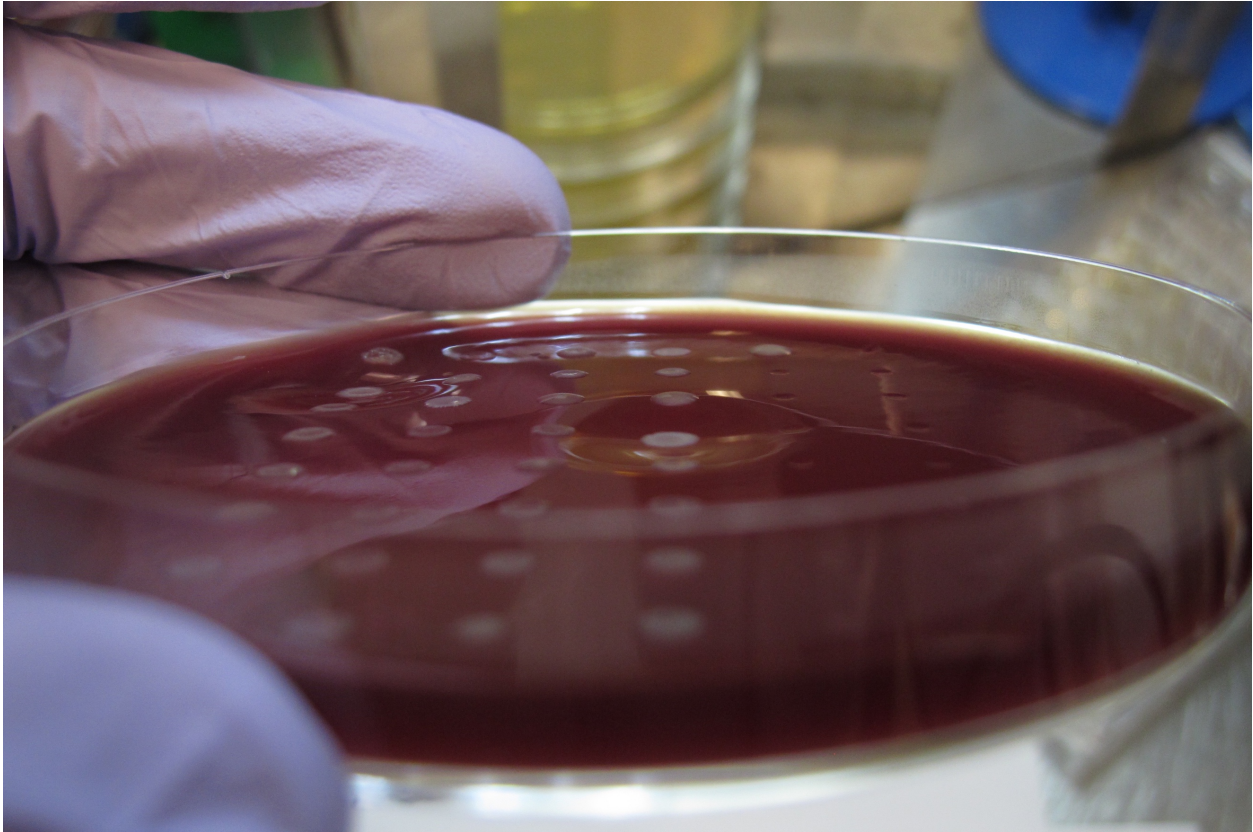


Figure 23: A side profile shot of the same plate shown in figure 2, intended to further demonstrate the dimple created by the agarase positive colonies.

4.3.5 *Bacteroides* strains display variable antibiotic resistance within species

The following figures summarize the findings of the antibiotics assay, they are grouped by the species assignments from both the 16s data and the KmerFinder results. Regardless of grouping there is evidence for differences in antibiotic resistance. Furthermore, these data suggest that strains of putatively the same species isolated from both the GAD patient and healthy volunteer vary in their antibiotic resistance.

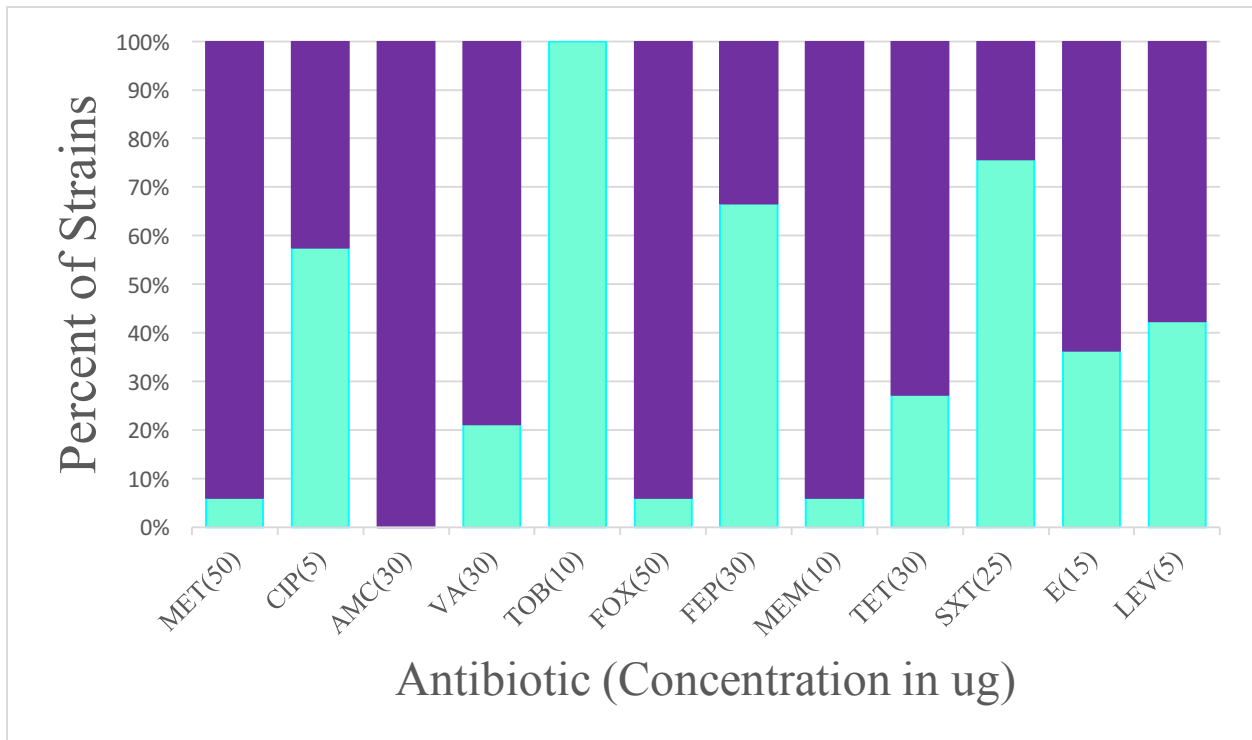


Figure 24: Stacked bar charts illustrating the percentage of strains within the *Bacteroides* strain collection found to be resistant to the twelve assayed antibiotics. Resistant strains are shown in teal while not resistant strains are shown in purple.

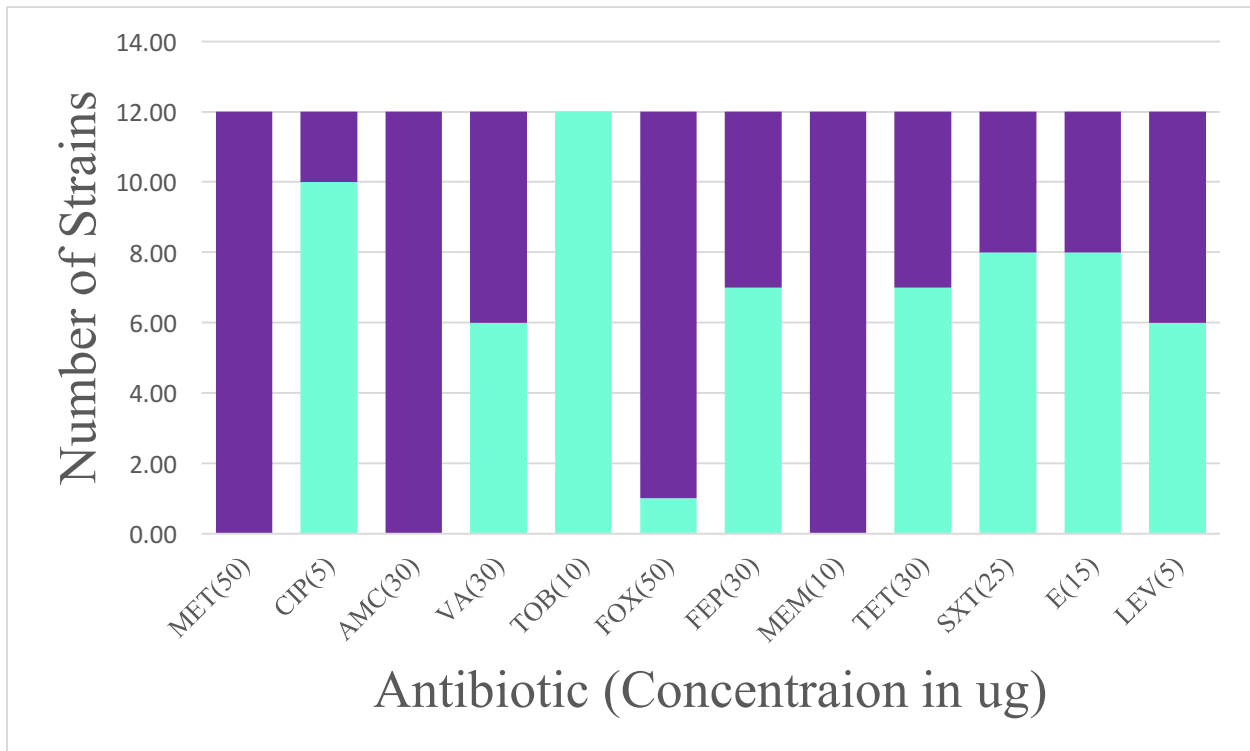


Figure 25: Stacked bar charts illustrating the resistance patterns of strains within the collection that were identified as *B. thetaiotaomicron* through analysis of the whole genome sequence data. These results demonstrate that even within strain of putatively the same species, antibiotic resistance is variable in three quarters of the antibiotics assayed.

4.3.7 Analysis of whole genome sequence data suggests contamination

The seemingly contaminated strain 28 was run through a metagenomics analysis software called *progressivemauve*¹⁴³ with the hope that this could offer better insight into what DNA had been sequenced.

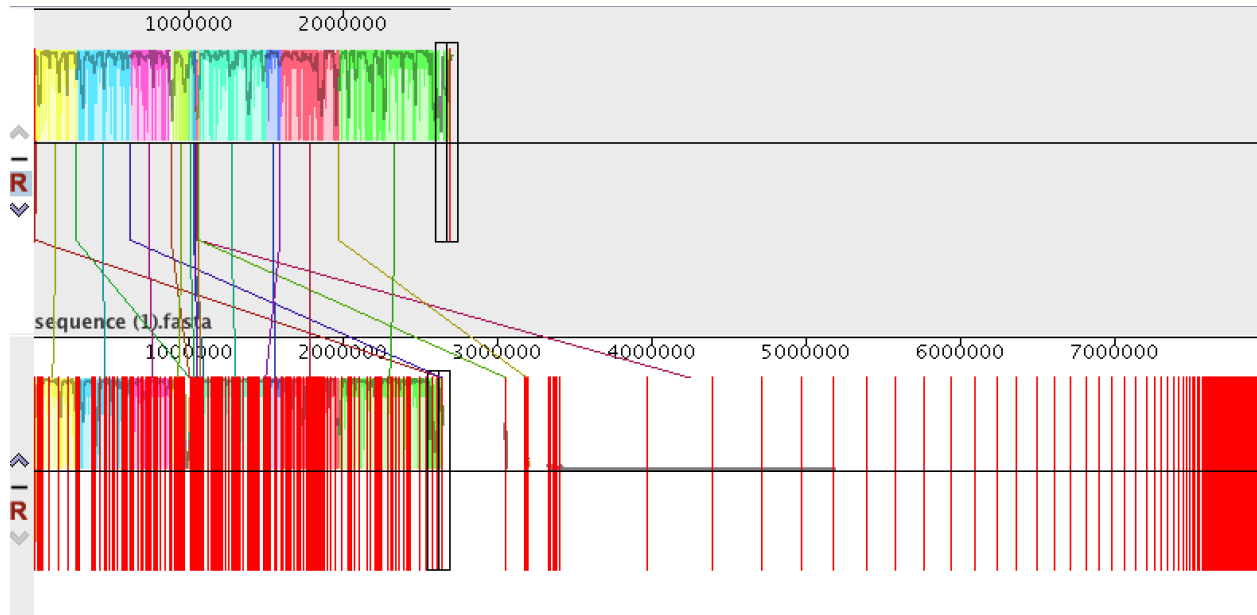


Figure 26: Screen capture of the metagenomics comparison figure produced using *progressivemauve*¹⁴³ software. A reference genome for *Enterococcus faecium* is shown in a variety of colours in the top portion of the figure. The whole genome sequence data from isolate 28 is shown in the bottom half of the figure. In the lower half of the figure the red bars represent sequenced portions of the genome often referred to as contigs. Lines connecting the two plots identify overlapping areas between the sequence data and the *E. faecium* reference genome. Additionally, the unmatched contigs, particularly towards the right-hand side of the image are likely suggestive of contamination and could represent a co-cultured *Bacteroides* isolate within this sample.

4.4 Discussion

The results in this chapter highlight some of the exciting and challenging aspects of lab based microbiology. Utilizing the *Bacteroides* collection I was able to demonstrate heterogeneity across a host of phenotypes, and have likely also had some results compromised by contamination issues.

This experiment was predicated on the successful identification and isolation of *Bacteroides* strains from patient and control samples. Months of work were put into this project and yet, as can be seen in table 3, two of the strains may not have been *Bacteroides*, but rather strains of other common gut bacteria *Bifidobacterium* and *Enterococcus*. This finding was somewhat discouraging and could have occurred in a number of ways. Firstly, it is possible that during the initial colony picking, these two organisms were accidentally included. This, however, seems exceedingly unlikely given that culture purity was repeatedly examined through streaking cultures to isolate single colonies, and further through amplification of the whole 16s gene. Certainly, there is some ambiguity in 16s gene sequencing, however, while this data cannot always be used to resolve organisms to the species level, it should not be mistaking bacteria between different phyla. Additionally, *Enterococcus faecium* is a Gram-positive organism, and thus its growth should have been prevented on the BBE plates this media is selective against Gram-positive bacteria¹³⁷. That a colony was repeatedly observed at the 28th position in the array suggests that at least some of the culture contained the correct organism. To further investigate the possibility of co-culture contamination, with the help of lab-mate Saad Syed, the assembled genome data for strain 28 was run through the comparative genomics software *progressivemauve*¹⁴³. This program enables the separation of complex and putatively contaminated metagenomics datasets. While the results shown in figure 26 corroborate KmerFinder's identification of an *Enterococcus faecium* within the sequence data for strain 28, they also highlight a significant portion of unmatched DNA sequence, nearly 5 MB. This is near the size of some *Bacteroides* genomes and could suggest that at some point, prior to whole genome sequencing the strain 28 stock became contaminated^{144,145}. This potential

contamination, as well as that observed in strain 2 will be further investigated with another metagenomics program called LMAT¹⁴⁶ which should allow for definitive resolution of any possible contaminants in both samples. Unfortunately, given that DNA was extracted for whole genome sequencing prior to performing any of the other assays, none of the results obtained for strains 2 and 28 can be reliably interpreted. With that caveat in mind, these data do still offer some valuable takeaways about this strain collection.

Firstly, the phenotyping data, in particular the esculin hydrolysis assay, demonstrate that whether strains are grouped into species using 16s data or KmerFinder analysis of whole genome data, there is heterogeneity within species. The finding of phenotypic heterogeneity within microbial species is not novel in the field of microbiology, but serves as an important reminder that marker gene sequencing data, like those involving 16s sequencing cannot fully appreciate the diversity of microorganisms in a community. No conclusions can be drawn about the relative amounts of esculin hydrolysis, however, a follow-up experiment in which cultures are standardized via dilution in salt water to a common OD could allow for the generation of quantitative results in future. Furthermore, because these bacteria simply represent a small sampling of the *Bacteroides* found in patient and control samples, these data cannot be used to say more of a given phenotype is observed in bacteria from patients or controls. In order to try to establish those sorts of findings, a collection of likely hundreds of isolates across multiple patient and controls samples would need to be built and even then, the generalizability of the data would be difficult to predict.

Unfortunately, three of the other phenotypes assayed, the proteolysis, hydrogen peroxide production and agarase activity were only observed in the two contaminated strains, and therefore these experiments would need to be repeated with isolates from a culture certain to be pure. One strategy for purifying strains 2 and 28, which were contaminated with Gram positive-*Bifidobacterium* and *Enterococcus* respectively would be to streak the culture for isolation on a gram negative selective media like BBE or eosin methylene blue (EMB) agar. There is a possibility that this contaminating strains previously transferred across a BBE plate, however, thus utilizing a different selective media like EMB would probably be more prudent. Agarase activity and proteolysis have both been reported in some cases in the *Bacteroides*, accordingly it is possible that a *Bacteroides* in strains 2 and 28 are responsible for these phenotypes in spite of the contamination^{68,69}.

Results from the antibiotic resistance assay corroborate the finding of heterogeneous phenotypes within theoretically homogenous groups. Figure 24 shows that only one antibiotic was found to be fully lethal (Amoxicillin-Clavulanic Acid), similarly, the 33 strains were only collectively resistant to Tobramycin. Interestingly, this trend held for a number of observed groups including the strains labeled *Bacteroides vulgatus* (by both the 16s and WGS data) as well as the strain identified as *B. tetaiotaomicron* in the whole genome sequencing data as can be seen in figure 25. This finding is particularly interesting given that both antibiotic resistant and susceptible strains were isolated from the same human gut communities. These data, especially without historical data regarding patient antibiotic use, cannot suggest how antibiotic resistant and susceptible strains have come to reside in the same ecosystem simultaneously. This observation, is compelling nonetheless and could basis for an investigation about how

thoroughly courses of antibiotics eradicate susceptible commensals in the host. Given the current excitement surrounding probiotics, especially during or after a course of antibiotics, one might assume that healthy commensals are completely wiped out, a paradigm that may not be entirely accurate. Conversely, it is possible that study participants acquired a combination of antibiotic susceptible and resistant strains through their diet and the co-existence of these strains within species is temporary.

Despite having identified two of the strains as contaminated, one of the major achievements of this chapter is nonetheless, the assembly and whole genome sequencing of the *Bacteroides* strain collection. Given the timing within my degree, I did not have sufficient time to learn to truly explore or manipulate this data, however, it is my sincere hope that a future student in the Surette lab will be better able to explore it to perhaps identify genes that appear to distinguish patient strains from those found in the controls.

To continue investigating these data, I would first propose to re-amplify the 16s gene from the frozen stocks of strains 2 and 28. The DNA extracted and amplified for whole genome sequencing was collected when the strains were first isolated. Accordingly, it is possible that while these initial stocks may have had contaminants, it is likely that during the course of replicating the phenotypic assays over the following months this contaminant could have been eliminated. Thus, by quickly re-amplifying the 16s gene the identity, *Bacteroides* or otherwise of the strains could be confirmed. Assuming they are in-fact *Bacteroides* the phenotypic assays

could be repeated to confirm the validity of the results, otherwise, if contamination persists the strains could be re-isolated as previously described.

To continue analysis of the WGS data, a number of tools could be used, for example, PHAST, and the recently released PHASTER could be used to search the genomes for evidence of bacteriophage sequences^{147,148}. Recent results suggest that phage could play an important role in maintaining and remodeling the microbiome, accordingly, investigating the distribution of phage sequences amongst patient and control samples could offer insight into a putative role of phage in tailoring this discrepancy. Perhaps *Bacteroides* phage are maintained at a higher level in healthy controls and as phage numbers are depleted, the expansion of the *Bacteroides* observed in the patients occurs. Further, a more sophisticated WGS analysis tool like LOOOKK, could be used to better assign identity to the strains such that their differences could be more thoroughly investigated. Perhaps, the strains are truly representatives of only the *Bacteroides vulgatus* and *B. thetaiotaomicron* species, or perhaps the collection contains more diversity at the genus level, a question that could be pursued with a second WGS analysis tool.

In conclusion, although I dealt with some contamination issues, and unfortunately had to discard some of my more unique observations, I feel these data represent the culmination of significant time and effort and do succeed in demonstrating phenotypic heterogeneity within species. Furthermore, I hope to complete the 5-HT assay, as well as replicate findings with certified pure stocks of strains 2 and 28 and prepare the collection and data for another excited young scientist to continue experimentation.

Chapter 4 Acknowledgements

I would like to acknowledge Mr. Jeffery Ye played a for his significant role in helping to grow and pick the nearly 1 500 colonies during the isolation of the *Bacteroides* strains. Additionally, I would like to thank Mr. Saad Syed for his help interpreting and analyzing the whole genome sequence data. Finally, thank you to Dr. Varun Dewan of the Kahn lab for supplying and helping to culture the Bon cells used in the 5-HT assay, and for performing the follow-up ELISA assay.

Chapter 5: Moving Forwards

5.1 Challenges in interpreting microbiome data

These findings in chapters two and three highlight some of the current big questions within (and limitations of) the field of microbiome research, particularly, how to curate the data, and how to interpret the significance of results. In considering data curation taking the chapter one two data as an example, of the nearly two and a half million sequence reads generated, approximately sixty-thousand, or just under three percent of them were discarded. The cutoffs used to trim the data were based off past results from the Surette lab, as well as seemingly reasonable assumptions about the nature of the data. When assessing the rationale behind the assumptions made surrounding the data, a requirement that an OTU be present in at least ten percent of samples to be kept seems reasonable when one considers the question being asked; under the hypothesis that microbes could be influencing GAD, and with over seventy GAD patients, an organism present in only fifteen individuals couldn't be contributing to illness in seventy. At the same time, however, there is active debate in the field regarding whether presence vs. absence comparisons are most relevant, or, if instead, we should consider functional niches within the host⁶⁴. Take for example the *Erysipelotrichaceae* these data show to be elevated in MDD patients relative to healthy controls; whilst this finding is novel in comparison to the two previously published studies, perhaps its novelty lies in the nature of the question being asked. The *Erysipelotrichaceae*, have been identified as being associated with inflammation in HIV patients, enriched in colorectal cancer patients and enriched in obese patients^{65,149,150}. All of these examples would seem to suggest that these organisms aren't ideal commensals, and also, that they are likely linked with, if not capable of exacerbating inflammation. Accordingly, while *Erysipelotrichaceae* may be contributing to inflammation and GI discomfort amongst this

study's MDD patients, perhaps that "disease niche" is filled by a different organism in the Jiang and Naseribafrouei patient populations^{61,62}.

Another challenge with data curation is that of rare or low abundance organisms. One of the curating steps for the data was that all OTUs with less than thirty total reads across the dataset were discarded. Again, given that within the smallest of the study's groups, the eMDD patients, there are sixteen participants, it seems unlikely that an organism present at roughly 2.8×10^{-5} percent abundance (30 reads / 16 patients / 70 000 reads (average read depth)) would be a main driver of disease. It is generally accepted that the colon houses somewhere near a trillion bacterial cells¹, accordingly, those reads could represent a community of twenty-eight million cells. This is merely a rough estimate and assumes that either each cell contains only one copy of the 16s gene (which is known not to be the case) or that the hypothetical organism has the mean number of 16s gene copies within the community. Further, it assumes that sequencing is sampling each member of the gut community, which it almost certainly isn't; nonetheless, this serves to illustrate that seemingly rare or low abundance data could be important in the context of the gut community. A conflict is reached again in that if no data is discarded, there is potential that spurious sequencing artifacts can be mistaken for OTUs. Furthermore, identifying significant taxa by performing multiple tests and then correcting for the thousands of OTUs tested makes reliably identifying significant results in small populations nearly impossible¹⁵¹.

A third challenge with data curation stems from the nature of the data; because sequencing depth is uneven, some sort of normalization, in this case, calculating the relative abundance of sequencing results, is necessary such that comparisons aren't biased by sequencing read counts.

Relative abundance data, however, is a simplex, meaning that possible abundance measures can only exist between 0 and 1. Accordingly, an increase the amount of one group in a sample inherently has to be accompanied by the proportional decrease of relative abundance of one or more other groups. Therefore, it is possible that a relative increase in the abundance of *Bacteroides* in patients is the result of, or causative of, the loss of a number of organisms in patients which could have functional redundancy but collectively contribute to a healthy gut-brain axis in healthy volunteers. It may be that the *Ruminococcus* is only one, of a group of organisms whose abundance is decreased in patients but the study is insufficiently powered to find these contributors. Other strategies exist to address uneven sample depth, for example rarifying as employed in calculating the alpha diversity of the samples. This method can also cause problems, including the potential to lose or ignore rare data ¹⁵². Accordingly, it is important to acknowledge that at this point the field is still limited in its handling of the large data sets generated in microbiome sequencing, not to mention, the active debate surrounding which tools and methods to use ^{96,153}. Consequently, it behooves scientists to be transparent with their methods, to make data curations strategies ahead of data analysis and to acknowledge that findings must be interpreted with a certain degree of skepticism.

The next area of focus was the interpretation of result significance. It is important to remember that given the observational nature of this data, at best, conclusions regarding putative correlation can be drawn. This is to say that while the data suggest that an enrichment of *Bacteroides* may be a hallmark of GAD, the data offer no insight into whether this organism is a driver of disease, or rather a benign opportunist that thrives in the previously existing conditions of an anxious gut. Equally, the *Ruminococcus* enrichment in healthy controls could help to combat anxiety, or

could be enriched in a non-anxious gut. Furthermore, it is possible that the organisms are anxiogenic (anxiety causing) and anxiolytic (anxiety preventing), respectively, and competition between them could contribute to the perpetuation or prevention of anxiety even if the putatively impactful strain has lower fitness. Finally, these observations cannot directly inform conclusions with respect to any potential interactions between the two organisms. For example, a possible explanation for the observed results could be that the *Bacteroides* are benign community members, especially given that figure six shows at least four healthy controls have a minimum of fifty percent relative abundance of this genus. In this example, the *Ruminococcus* could be anxiolytic and it is only when the *Bacteroides* out compete them for some nutrient(s) (for example, the *Bacteroides* could be capturing and metabolizing complex carbohydrates before they are digested decreasing the supply of the simpler carbohydrates like lactose and ethanol that ruminococci favour), that they are displaced from their niche, causing the host to become anxious. Further complicating the issue are the complexities of community dynamics. The gut microbiome is vastly more diverse than simply containing ruminococci and *Bacteroides*. Ruminococci are known to be capable of inter-species hydrogen transfer during fermentation. Accordingly, it could be possible that the *Bacteroides* are not in direct competition with the ruminococci, but rather with another organism participating in hydrogen transfer with the ruminococci¹¹⁰ and that when the *Bacteroides* displace this organism the ruminococci die-off without their fermentation partner. Another reasonably plausible explanation for these observations is that the bacteria may be affected by a behavioral change exhibited by patients which impacts the gut environment and thus one or both of the observed shifts could represent responses to these changes. In this case, either could serve as markers of a healthy or disordered state with neither playing a causative role. An example of how this phenomenon could occur is

through vitamin D availability¹⁵⁴. Vitamin D levels have been shown to impact gut microbiota and metabolic disorders, and one of the main sources of vitamin D is sunlight¹⁵⁴. Accordingly, if during disease progression GAD patients became increasingly anxious about leaving their house or about being in the sun for fear of getting skin cancer, they could dramatically decrease time in the sun and thus inadvertently decrease their vitamin D intake. Su *et al.* observed in a high fat diet-vitamin D deficient mouse model that the abundance of the putatively beneficial *Akkermansia muciniphilla* was decreased and mucous barrier integrity was lost¹⁵⁴. *Akkermansia* is thought to play a role in maintenance of a healthy mucous layer on the intestinal epithelium given its capacity to metabolize mucous. Given the phenomenal carbohydrate metabolizing capacity of *Bacteroides*^{68,72}, they perhaps take advantage of the more readily available mucous as a carbon source following a disturbance in *Akkermansia* and in doing so contribute to breakdown of the gut barrier. This could contribute to some of the GI symptoms reported in GAD patients. Certainly, the data presented are insufficient to reject or fail to reject any of the aforementioned hypotheses, nonetheless they are raised to highlight the potential levels of complexity still remaining in the ecosystem. Despite these considerations, however, correlations of these observations with measures of disease impact, especially after controlling for potential confounders like weight and diet, still support the idea that the correlation may be relevant for GAD.

A third limitation of the data is that they offer no course for interpreting or comparing the importance of the observed changes. The enrichment of both the genus *Bacteroides* and an OTU within this genus in patients could initially be (potentially mistakenly) interpreted as more valid or significant when contrasted with the *Ruminococcus* OTU observation, given the abundance of

the genus *Ruminococcus* is approximately equal between patients and controls. Further, the *Ruminococcus* observation could be the result of, and thus highlight an issue with data curation, in particular, the picking or assignment of OTUs. OTU picking algorithms are intended to best identify species-like groups within microbial sequence datasets (typically using the 97% similarity cut-off). As such, the reliability of sequence accuracy is vital, however, there is an observed trade-off between sequence length and quality. Shorter reads are more accurate but are not always able to uniquely identify organisms, whereas longer reads are typically more error prone but should allow for better taxonomic resolution^{155,156}. With these data, it is possible that the OTU picking algorithm has artificially split what should be sequences belonging to one *Ruminococcus* species into three or more OTUs. Accordingly, this observation could have resulted from a disproportionate number of the control-associated sequences being grouped in one OTU, in this case OTU 13, while the majority of patient associated *Ruminococcus* sequences were split between two or more OTUs and thus, not reaching an enrichment of statistical significance in any of them. Conversely, it is possible that the identified strain(s) of *Ruminococcus* grouped into OTU 13 could be anxiolytic. Thus, while the abundance of the genus *Ruminococcus* is approximately equal between groups, this particular strain could possess unique genes that are vital in preventing anxiety. With these data, however, the only conclusion that can be drawn is that there appears to be a larger proportion of enrichment of the *Bacteroides* genus in patients than there is enrichment of the *Ruminococcus* OTU in healthy controls. Similarly, when attempting to interpret the MDD data, it is possible that both OTUs 224 and 227, given their relative scarcity are random observations stemming from the relatively small size of the MDD patient population. Alternatively, they could represent incredibly insidious organisms who have systemic effects on the host, even in trace amounts. Either way, these limitations point to the

importance of using these data for driving hypothesis generation for future experiments of potential mechanisms.

These limitations are important to acknowledge when interpreting the findings in chapters two and three, but are common across all current microbiome studies, and they should not be misinterpreted as removing any significance from the findings. Rather, they serve to highlight the relevance of these findings as the basis for hypothesis generation about how best to investigate their putative role (or lack thereof) in disease. Accordingly, I will now outline a series of experiments I would propose to in an attempt to and address some of these limitations.

5.2 Experiments to continue the story

Firstly, I would propose to continue recruitment of patients in two ways. I would continue to recruit participants for the single participation study model as has been conducted to this point. This increase in sample size would serve to validate the findings of this study, and could also offer sufficient data for additional follow-up analyses such as random forest analysis using machine learning¹⁵⁷. Briefly, this method involves training a computer using an initial “training” data-set in which the user gives the computer a dataset (in this case microbiome data) and a variable(s) that classifies the data (in this case diagnosis). Using this training set, the computer tries to find factors that distinguish the groups and builds an equation to assign variable status based on data composition. This is followed by providing the computer with a follow-up dataset without the diagnostic variable, and then assessing the rate of success of data at predicting the variable. The better the success of prediction, the more confident one can feel about their data. An analogy would be to imagine that you are tasked with distinguishing bicycles from

automobiles. If you have data about the number of wheels you should always be able to distinguish a car from a bike. Conversely, being told the diameter of the wheels might be helpful at the extremes, but would likely prove less efficacious when separating small cars from large bicycles. In this instance wheel number would be a better variable for predicting whether something is a car or a bike than would wheel radius. Accordingly, if the data presented here were to serve as a training set, a follow-up analysis of an equal or bigger participant population could help inform whether the *Bacteroides* and *Ruminococcus* observations are replicable or simply random occurrences within the dataset. Further, they would allow for more thorough statistical analysis to control putative confounding by diet, medications and smoking amongst other variables. Along with continuing recruitment in Hamilton, I might try to expand recruitment to similar (and relatively nearby) programs in Toronto, London and Kitchener-Waterloo. While the distance from these cities to the lab could prohibit samples arriving at the lab anaerobically within four hours, and thus not be suitable for culturing experiments, this would hopefully lead to faster recruiting and might also help recruit a more diverse patient population. Stearns et al. recently demonstrated a link ethnicity to the microbiome in a comparison of white and South Asian Canadian infants¹⁵⁸. These data the potential importance of including participants from different cultural backgrounds as they may consume a different diet than that of our almost entirely white patient group recruited in this study. While a diversity of diets could complicate data analysis, it would be my hope that the results could lead to pan-cultural/pan-diet markers of a putative GAD microbiome and thus make findings applicable to as many patients as possible.

A second way I would propose to recruit patients would be through the course of a therapeutic intervention like Cognitive Behavioural Therapy (CBT). In an ideal study, I would propose recruiting participants who were willing to give multiple samples through the course of their therapy, but at a minimum, to provide a before and after sample. For example, patients could be recruited from the same twelve week CBT program offered by Hamilton Health Sciences as the patients in this study, and at minimum they would provide a sample at the outset and completion of their course. Ideally a sample would be collected through the course of the program, potentially around the six to eight-week mark, when improvement is first observed in participants. A fourth and final sample would be collected one to two months after completion of the program, concurrent with a self-reported assessment of symptomology. In theory, these data would provide a couple of advantages over the current data set. Firstly, having multiple samples from individuals can be used to eliminate some of the noise of individual heterogeneity and better focus on meaningful differences. ANCOM⁹⁶ is just one example of a software with the capacity to analyze repeated measures data and distinguish changes in individuals from differences between groups. Secondly, and admittedly ambitiously, it is my hope that longitudinal changes could be tracked in individuals and correlated with improvement in anxiety. It seems reasonable that if there is a top-down effect of the brain on the microbiome that as cognitive function improves with CBT, concurrent changes would happen in the microbiome. This would likely require a participant sample size at least equal to the one currently included in the study and thus could reasonably be predicted to take an additional two years to recruit. Furthermore, with follow-up assessment of mood, meaningful changes in patients that experience improvement could hopefully be distinguished from stochastic changes observed over the course of the study. Not only could this study provide important evidence for modification of

the microbiome by altered cognitive function, it could also potentially contribute to the field's understanding of how or why some subset of patients are treatment resistant. Perhaps these patients possess a community that is more resistant to change, or is deficient in some of the organisms that are enriched during therapy in responders and this could inform potentially microbiome based adjuvants to increase the efficacy of therapy in GAD. One of the benefits of this study is that while we do not completely understand why CBT works, we have significant evidence for its efficacy in a portion of the population, and accordingly, I think this population offers one of the best windows to study putative changes in the GAD microbiome that result in patient relevant improvements.

Theoretically, a conceptually similar study to the anxiety and CBT study could be conducted in patients who start psychiatric medications for their GAD or MDD. Similar to therapeutic interventions, pharmaceutical interventions typically have a delay of a month before patients experience improvements in symptomology. While it is possible this perceived delay could be microbially mediated, I feel this system has too many variables that cannot be controlled and therefore, is not the most promising place to begin investigation. For example, were a decrease or increase of a given bacteria observed in this group it would be impossible in an observational study to distinguish the direct effect of the drug on the microbial community from effects of the drug on the host that then impact the gut ecosystem. Seemingly, one of the only strategies to separate these effects would be to find a population of healthy volunteers willing to go on the anti-anxiety medication and see how it affects their microbiome. Even this solution, though, is problematic if for no other reason that given some of the less desirable side effects of some anti-anxiety medications (eg. weight gain, loss of sex drive) willing volunteers may be few and far

between, let alone the question of whether it would be ethical to expose volunteers to those side effects. Furthermore, even if the study had REB approval and volunteers were recruited, working under the hypothesis that a putatively distinct GAD associated microbiome exists, changes that might be observed in the microbiomes of the healthy volunteers could not be expected to be identical to those observed in GAD patients as the initial communities may be different. One possible solution to this issue could be to colonize germ-free mice with fecal samples from anxiety patients, and assuming the anxious phenotype is transferred, then treat one group of mice with an anti-depressant and see if changes are observed. Given the possibility that many very small changes may be responsible for these outcomes, and the relatively high cost and limited availability of germ-free mice, acquiring enough mice for this experiment to have reliable data, cost could be prohibitive for this experiment.

In continuing the work from chapter four I would propose a number of experiments. Firstly, I would propose an oxygen exposure assay as some of my observations while working with the *Bacteroides* suggest that all of the strains might not be strictly anaerobic, contrary to the literature dogma surrounding these organisms⁶⁷. Briefly, the organisms could be grown in broth in 96 well plates in duplicate inside the anaerobe chamber, while also exposing a series of replicate plates to increasing time outside of the chamber at atmospheric oxygen levels. Plates could then be incubated for growth both aerobically and anaerobically and then growth could be measured by measuring optical density of the cultures. Oxygen sensitivity is likely an important factor in modeling the gut community, thus a strain being a facultative aerobe could confer a significant advantage were it competing with obligate anaerobes.

I would next suggest a more thorough analysis of the whole genome sequence data. Although I was incredibly fortunate to encounter a host of new technologies and develop a number of proficiencies over the course of my project I did not have time to learn to analyze these data. Accordingly, my suggestions are likely more theoretical than instructional, nonetheless, these data could be investigated for a number of different questions. Investigating factors like single nucleotide polymorphisms, gene insertions and deletions, CRISPR systems and operon structure could offer insights into relatedness, functional capability, defense capacity and resource utilization capacity of the strain collection which could better inform hypotheses about community dynamics. Finally, taken in conjunction with the results of the serotonin production stimulation assay, these data could be used to identify strain of interest for a putative animal experiment. Ideally, if one of the strains proves to be particularly unique in its capabilities and possess putative capacity to influence the host, at least hypothetically, it could be used to monoculture an animal model, like a mouse, to assay the strain's anxiogenic potential.

Admittedly, any mouse experiment would be challenging to design, if for no other reason than, as has been extensively detailed, it is possible that there is no single "anxiety organism".

Additionally, while mice are an incredibly useful animal model, they are far from human, and differences in their digestive tract and brains could nullify a human meaningful impact.

Furthermore, monoculturing a mouse is significantly easier than trying to introduce a foreign organism into a mouse with a well-established microbiome (after all, one of the important roles the gut microbiome is believed to play in the host is preventing infection of unwanted and foreign strains), but the bacteria are unlikely to operate in monoculture as they would as a constituent of much more complex gut community. Assuming, however, that an strain of

particular interest were identified, I would propose a first experiment in which a monoculture of this organism was given to a litter of germ-free mice, in comparison with a litter inoculated with a mono-culture of another of the strains which the data suggest may be more benign as a control. Ideally, these experiments could be accompanied by ones in which groups of mice were inoculated with samples from the participants from which these bacteria were cultured. This would allow for comparison of an anxiety phenotype between the putatively influential strain and a seemingly more benign one, while hopefully also allowing for observation of any differences in effect when the organism exists as a member of a more complex community in comparison with a system in which it is the only bacteria. Samples from the mice inoculated with the whole human sample could also be used to establish that the strain of interest did in fact establish itself in the mice and thus could be reasonably believed to be contributing an anxiety phenotype putatively shared between the mice given the target strain and the whole sample but not the benign strain. Potentially, this experiment could be a first step in narrowing in on one (admittedly it seems possible there could be more than one bacteria and/or one pathway influencing anxiety) organism that is possibly contributing to anxiety and in doing so expand our understanding of both the causes of anxiety and hopefully, new strategies to combat it.

In conclusion, I would like to again thank everyone I've mentioned for helping me to bring together this project. I am incredibly thankful for my time and hope that my work will contribute to helping a group of people whom I feel are incredibly deserving of help. While my results do not come close to finding the cause of the various mental illnesses addressed, I hope that they can serve as a stepping stone on the path to a better understanding of these disorders.

Works Cited

1. Sender, R., Fuchs, S. & Milo, R. Are We Really Vastly Outnumbered? Revisiting the Ratio of Bacterial to Host Cells in Humans. *Cell* **164**, 337–340 (2016).
2. Balzola, F., Bernstein, C., Ho, G. T. & Lees, C. A human gut microbial gene catalogue established by metagenomic sequencing: Commentary. *Inflamm. Bowel Dis. Monit.* **11**, 28 (2010).
3. Reyes, A. *et al.* Viruses in the fecal microbiota of monozygotic twins and their mothers. *Nature* **466**, 334–338 (2010).
4. Breitbart, M. *et al.* Metagenomic Analyses of an Uncultured Viral Community from Human Feces. *J. Bacteriol.* **185**, 6220–6223 (2003).
5. Hoffmann, C. *et al.* Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *PLoS One* **8**, (2013).
6. Bäckhed, F. *et al.* The gut microbiota as an environmental factor that regulates fat storage. *PNAS* **101**, 15718–15723 (2004).
7. Bentley, R. & Meganathan, R. Biosynthesis of vitamin K (menaquinone) in bacteria. *Microbiol. Rev.* **46**, 241–280 (1982).
8. Fukuda, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543–547 (2011).
9. Tanaka, H., Hashiba, H., Kok, J. & Mierau, I. Bile salt hydrolase of *Bifidobacterium longum*-biochemical and genetic characterization. *Appl. Environ. Microbiol.* **66**, 2502–12 (2000).
10. Gopal-Srivastava, R. & Hylemon, P. B. Purification and characterization of bile salt hydrolase from *Clostridium perfringens*. *J. Lipid Res.* **29**, 1079–1085 (1988).
11. Elkins, C. A., Moser, S. A. & Savage, D. C. Genes encoding bile salt hydrolases and conjugated bile salt transporters in *Lactobacillus johnsonii* 100-100 and other *Lactobacillus* species. *Microbiology* **147**, 3403–3412 (2001).
12. Gensollen, T., Iyer, S. S., Kasper, D. L. & Blumberg, R. S. How colonization by microbiota in early life shapes the immune system. *Science (80-.)*. **352**, (2016).
13. Neufeld, K. M., Kang, N., Bienenstock, J. & Foster, J. A. Reduced anxiety-like behavior and central neurochemical change in germ-free mice. *Neurogastroenterol. Motil.* **23**, 255–265 (2011).
14. Neufeld, K.-A. M., Luczynski, P., Oriach, C. S., Dinan, T. G. & Cryan, J. F. What’s bugging your teen?—The microbiota and adolescent mental health. *Neurosci. Biobehav. Rev.* (2016). doi:10.1016/j.neubiorev.2016.06.005
15. Mohammadi, A. A. *et al.* The effects of probiotics on mental health and hypothalamic–pituitary–adrenal axis: A randomized, double-blind, placebo-controlled trial in petrochemical workers. *Nutr. Neurosci.* **19**, 387–395 (2016).
16. Zheng, P. *et al.* Gut microbiome remodeling induces depressive-like behaviors through a pathway mediated by the host’s metabolism. *Mol. Psychiatry* 1–11 (2016). doi:10.1038/mp.2016.44
17. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (5th ed.)*. (2013).
18. Hoge, E. a, Ivkovic, A. & Fricchione, G. L. Generalized anxiety disorder: diagnosis and treatment. *Br. Med. J.* **7500**, 1–6 (2012).
19. National Institute for Health and Care Excellence. Depression in adults: Recognition and

- management. *Clin. Guidel. [CG90]* (2016). at
<<https://www.nice.org.uk/guidance/cg90/chapter/1-Guidance>>
20. Blumberg, H. P. Euthymia, Depression, and Mania: What Do We Know About the Switch? *Biol Psychiatry* **71**, 250–257 (2012).
 21. Ekinçi, O., Albayrak, Y. & Ekinçi, A. E. Temperament and character in euthymic major depressive disorder patients: The effect of previous suicide attempts and psychotic mood episodes. *Psychiatry Investig.* **9**, 119–126 (2012).
 22. Risch, N. *et al.* Interaction Between the Serotonin Transporter Gene (5-HTTLPR), Stressful Life Events, and Risk of Depression. *J. Am. Med. Assoc.* **301**, 2462–2471 (2009).
 23. Berton, O. & Nestler, E. J. New approaches to antidepressant drug discovery: Beyond monoamines. *Nat. Rev. Neurosci.* **7**, 137–151 (2006).
 24. Yirmiya, R. Endotoxin produces a depressive-like episode in rats. *Brain Res.* **711**, 163–174 (1996).
 25. Foster, J. A. & McVey Neufeld, K.-A. Gut-brain axis: how the microbiome influences anxiety and depression. *Trends Neurosci.* **36**, 305–12 (2013).
 26. Gibbons, J. L. & McHugh, P. R. Plasma cortisol in depressive illness. *J. Psychiatr. Res.* **1**, 162–171 (1962).
 27. Parker, K. J., Schatzberg, A. F. & Lyons, D. M. Neuroendocrine aspects of hypercortisolism in major depression. *Horm. Behav.* **43**, 60–66 (2003).
 28. Sudo, N. *et al.* Postnatal microbial colonization programs the hypothalamic-pituitary-adrenal system for stress response in mice. *J. Physiol.* **558**, 263–75 (2004).
 29. Gross, R., Das, A. K. & Weissman, M. M. Bipolar disorder. *N Engl J Med* **351**, 2454–2455 (2004).
 30. Benazzi, F. Is there a continuity between bipolar and depressive disorders? *Psychother. Psychosom.* **76**, 70–76 (2007).
 31. Phillips, J. G. P. The Treatment of Melancholia by the Lactic Acid Bacillus. *BJPsych* **56**, 422–430 (1910).
 32. Hunter, W. & Moynihan, B. Chronic Sepsis as a Cause of Mental Disorder. *BMJ* **2**, 811–818 (1928).
 33. Stewart, F. H. Bacterial Change in Mental Disorder: Preliminary Note on Morgan's Bacillus. *BMJ* **74**, 269–275 (1928).
 34. Heijtz, R. D. *et al.* Normal gut microbiota modulates brain development and behavior. *Proc. Natl. Acad. Sci.* **108**, 3047–3052 (2011).
 35. Barrett, E., Ross, R. P., O'Toole, P. W., Fitzgerald, G. F. & Stanton, C. Gamma-Aminobutyric acid production by culturable bacteria from the human intestine. *J. Appl. Microbiol.* **113**, 411–417 (2012).
 36. Yano, J. M. *et al.* Indigenous bacteria from the gut microbiota regulate host serotonin biosynthesis. *Cell* **161**, 264–276 (2015).
 37. Wikoff, W. R. *et al.* Metabolomics analysis reveals large effects of gut microflora on mammalian blood metabolites. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3698–703 (2009).
 38. Goehler, L. E. *et al.* Activation in vagal afferents and central autonomic pathways: Early responses to intestinal infection with *Campylobacter jejuni*. *Brain. Behav. Immun.* **19**, 334–344 (2005).
 39. Borovikova, L. *et al.* Vagus nerve stimulation attenuates the systemic inflammatory response to endotoxin. *Nature* **405**, 458–462 (2000).
 40. Bercik, P. *et al.* The anxiolytic effect of *Bifidobacterium longum* NCC3001 involves

- vagal pathways for gut-brain communication. *Neurogastroenterol. Motil.* **23**, 1132–1139 (2011).
41. Bravo, J. A. *et al.* Ingestion of Lactobacillus strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 16050–5 (2011).
 42. Poutahidis, T. *et al.* Microbial symbionts accelerate wound healing via the neuropeptide hormone oxytocin. *PLoS One* **8**, (2013).
 43. Kleij, H. van der, O'Mahony, C., Shanahan, F., O'Mahony, L. & Bienenstock, J. Protective effects of Lactobacillus reuteri and Bifidobacterium infantis in Murine Models for Colitis Do Not Involve the Vagus Nerve. *Am. J. Physiol. - Regul. Integr. Comp. Physiol.* **295**, R1131–R1137 (2008).
 44. Bercik, P. *et al.* Chronic gastrointestinal inflammation induces anxiety-like behavior and alters central nervous system biochemistry in mice. *Gastroenterology* **139**, 2102–2112 (2010).
 45. Bailey, M. T. *et al.* Exposure to a social stressor alters the structure of the intestinal microbiota: Implications for stressor-induced immunomodulation. *Brain. Behav. Immun.* **25**, 397–407 (2011).
 46. Bharwani, A. *et al.* Structural & functional consequences of chronic psychosocial stress on the microbiome & host. *Psychoneuroendocrinology* **63**, 217–227 (2016).
 47. Collins, S. M., Surette, M. & Bercik, P. The interplay between the intestinal microbiota and the brain. *Nat. Rev. Microbiol.* **10**, 735–42 (2012).
 48. Cryan, J. F. & Dinan, T. G. Mind-altering microorganisms: the impact of the gut microbiota on brain and behaviour. *Nat. Rev. Neurosci.* **13**, 701–12 (2012).
 49. Jeffery, I. B. *et al.* An irritable bowel syndrome subtype defined by species-specific alterations in faecal microbiota. *Gut* **61**, 997–1006 (2012).
 50. Lydiard, R. B. Irritable bowel syndrome, anxiety, and depression: what are the links? *J. Clin. Psychiatry* **62 Suppl 8**, 38-45–7 (2001).
 51. Berer, K. *et al.* Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination - with comments. *Nature* **479**, 538–541 (2011).
 52. Bercik, P. *et al.* The intestinal microbiota affect central levels of brain-derived neurotropic factor and behavior in mice. *Gastroenterology* **141**, 599–609 (2011).
 53. Messaoudi, Michael, Lalonde Robert, Violle, Nicolas, Javelot, Herve, Desor, Didier, Nejd, Amine, Bisson, Jean-Francois, Rougeot, Cathrine, Pichelin, Mattieu, Cazaubiel, Murielle, Cazaubiel, J.-M. Assessment of psychotropic-like properties of a probiotic formulation (Lactobacillus helveticus R0052 and Bifidobacterium longum R0175) in rats and human subjects. *Br. J. Nutr.* **105**, 755–764 (2011).
 54. Tillisch Kirsten, L. J. *et al.* Consumption of Fermented Milk Product with Probiotics Modulates Brain Activity. *Gastroenterology* **144**, 1–15 (2014).
 55. McElhanon, B. O., McCracken, C., Karpen, S. & Sharp, W. G. Gastrointestinal Symptoms in Autism Spectrum Disorder: A Meta-analysis. *Pediatrics* peds.2013-3995 (2014). doi:10.1542/peds.2013-3995
 56. Finegold, S. M. *et al.* Pyrosequencing study of fecal microflora of autistic and control children. *Anaerobe* **16**, 444–53 (2010).
 57. Parracho, H. M. R. T., Bingham, M. O., Gibson, G. R. & McCartney, A. L. Differences between the gut microflora of children with autistic spectrum disorders and that of healthy children. *J. Med. Microbiol.* **54**, 987–991 (2005).

58. Tomova, A. *et al.* Gastrointestinal microbiota in children with autism in Slovakia. *Physiol. Behav.* **138**, 179–187 (2015).
59. Sandler, R. H. *et al.* Short-Term Benefit From Oral Vancomycin Treatment of Regressive-Onset Autism. *J. Child Neurol.* **15**, 429–435 (2000).
60. Kang, D.-W. *et al.* Microbiota Transfer Therapy alters gut ecosystem and improves gastrointestinal and autism symptoms: an open-label study. *Microbiome* **5**, 10 (2017).
61. Naseribafrouei, A. *et al.* Correlation between the human fecal microbiota and depression. *Neurogastroenterol. Motil.* **26**, 1155–62 (2014).
62. Jiang, H. *et al.* Altered fecal microbiota composition in patients with major depressive disorder. *Brain. Behav. Immun.* **48**, 186–194 (2015).
63. Maes, M., Kubera, M., Leunis, J. C. & Berk, M. Increased IgA and IgM responses against gut commensals in chronic depression: Further evidence for increased bacterial translocation or leaky gut. *J. Affect. Disord.* **141**, 55–62 (2012).
64. Consortium, H. M. P. Structure, Function and Diversity of the Healthy Human Microbiome. *Nature* **486**, 207–214 (2013).
65. Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–31 (2006).
66. Wexler, H. M. Bacteroides: The good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* **20**, 593–621 (2007).
67. Wexler, A. G., Goodman, A. L., Voigt, C. A., Lu, T. K. & Hennet, T. An insider's perspective: Bacteroides as a window into the microbiome. *Nat. Microbiol.* **2**, 17026 (2017).
68. Ndeh, D. *et al.* Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* **544**, 65–70 (2017).
69. Li, M., Shang, Q., Li, G., Wang, X. & Yu, G. Degradation of Marine Algae-Derived Carbohydrates by Bacteroidetes Isolated from Human Gut Microbiota. *Mar. Drugs* **15**, 92 (2017).
70. Roelofs, K. G. & Comstock, L. E. Molecules Necessary for Gut Colonization and Mediate Competition In Vivo. *MBio* **7**, 1–10 (2016).
71. Russell, A. B. *et al.* A type VI secretion-related pathway in Bacteroidetes mediates interbacterial antagonism. *Cell Host Microbe* **16**, 227–236 (2014).
72. Desai, M. S. *et al.* A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell* **167**, 1339–1353.e21 (2016).
73. Cordonnier, C. *et al.* Vitamin B12 uptake by the gut commensal bacteria bacteroides thetaiotaomicron limits the production of shiga toxin by enterohemorrhagic escherichia coli. *Toxins (Basel)*. **8**, (2016).
74. Mazmanian, S. K., Round, J. L. & Kasper, D. L. A microbial symbiosis factor prevents intestinal inflammatory disease. *Nature* **453**, 620–5 (2008).
75. Neff, C. P. *et al.* Diverse Intestinal Bacteria Contain Putative Zwitterionic Capsular Polysaccharides with Anti-inflammatory Properties. *Cell Host Microbe* **20**, 535–547 (2016).
76. Vatanen, T. *et al.* Variation in Microbiome LPS Immunogenicity Contributes to Autoimmunity in Humans. *Cell* **165**, 842–853 (2016).
77. Ivanov, I. I. *et al.* Specific microbiota direct the differentiation of Th17 cells in the mucosa of the small intestine. **4**, 337–349 (2008).
78. Sheehan, D. V. *et al.* The Mini-International Neuropsychiatric Interview (M.I.N.I.): The

- development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* **59**, 22–33 (1998).
79. Lovibond, S.H. & Lovibond, P. F. *Manual for the Depression Anxiety Stress Scales. (2nd. Ed.)*. (Sydney Psychology Foundation, 1995).
 80. Bernstein, D. P., Ahluwalia, T., Pogge, D. & Handelsman, L. Validity of the Childhood Trauma Questionnaire in an Adolescent Psychiatric Population. *J. Am. Acad. Child Adolesc. Psychiatry* **36**, 340–348 (1997).
 81. Meyer, T. J., Miller, M. L., Metzger, R. L. & Borkovec, T. D. Development and validation of the penn state worry questionnaire. *Behav. Res. Ther.* **28**, 487–495 (1990).
 82. Hamilton, M. A rating scale for depression. *J. Neurol. Neurosurg. Psychiatry* **23**, 56–62 (1960).
 83. Williams, J. B. W. & Kobak, K. A. Development and reliability of a structured interview guide for the Montgomery-??sberg Depression Rating Scale (SIGMA). *Br. J. Psychiatry* **192**, 52–58 (2008).
 84. Fraser, A., Delaney, B. C., Ford, A. C., Qume, M. & Moayyedi, P. The Short-Form Leeds Dyspepsia Questionnaire validation study. *Aliment. Pharmacol. Ther.* **25**, 477–486 (2007).
 85. Mostafa, R. Rome III: The functional gastrointestinal disorders, third edition, 2006. *World J. Gastroenterol.* **14**, 2124 (2008).
 86. Busner, J. & Targum, S. D. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont)*. **4**, 28–37 (2007).
 87. Hebden, L., Kostan, E., O’Leary, F., Hodge, A. & Allman-Farinelli, M. Validity and Reproducibility of a Food Frequency Questionnaire as a Measure of Recent Dietary Intake in Young Adults. *PLoS One* **8**, (2013).
 88. Bartram, A. K., Lynch, M. D. J., Stearns, J. C., Moreno-Hagelsieb, G. & Neufeld, J. D. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end Illumina reads. *Appl. Environ. Microbiol.* **77**, 3846–3852 (2011).
 89. Joshi, N. & FASS, J. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files. [Software]. Available at <https://github.com/najoshi/sickle>. Version 1.33 (2011).
 90. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10–12 (2011).
 91. Masella, A. P., Bartram, A. K., Truszkowski, J. M., Brown, D. G. & Neufeld, J. D. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**, 31 (2012).
 92. Ye, Y. Identification and quantification of abundant species from pyrosequences of 16S rRNA by consensus alignment. *Proc. - 2010 IEEE Int. Conf. Bioinforma. Biomed. BIBM 2010* 153–157 (2010). doi:10.1109/BIBM.2010.5706555
 93. Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335–336 (2011).
 94. McDonald, D. *et al.* An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–8 (2012).
 95. R Core Team. R Software. (2015). at <<http://www.r-project.org/>>
 96. Mandal, S. *et al.* Analysis of Composition of Microbiomes (ANCOM): A novel method for studying microbial composition. *Microb. Ecol. Health Dis.* **26**, 1–22 (2015).
 97. Schliep, K. P. phangorn: Phylogenetic analysis in R. *Bioinformatics* **27**, 592–593 (2011).

98. McMurdie, P. J. & Holmes, S. Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* **8**, (2013).
99. Oksanen, J., Blanchet, F., Kindt, R., Legendre, P. & O'Hara, R. Vegan: community ecology package. *R Packag.* 2.3-3 Available at: <https://cran.r-project.org/web/packa> (2016). at <<http://cran.r-project.org/package=vegan>>
100. Ahlman-Eltze, C. Ggsignif: Significance Bars for 'ggplot2'. R package version 0.2.0 (2017).
101. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlong, 2009).
102. Harrell Jr., F. E. & Dupont, C. Hmisc: Harrell Miscellaneous. R package version 4.0-3 (2017).
103. Auguie, B. GridExtra: Miscellaneous Functions for 'Grid' Graphics. R package version 2.2.1 (2016).
104. Wilke, C. O. Cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 0.7.0 (2016).
105. Menopause. *Canadian Women's Health Network* 1 (2012).
106. Berry, M. Microbial Community Diversity Analysis Tutorial with Phyloseq. (2016). at <http://denefflab.github.io/MicrobeMiseq/demos/mothur_2_phyloseq.html>
107. Hughes, J. B. *et al.* Counting the Uncountable : Statistical Approaches to Estimating Microbial Diversity MINIREVIEW Counting the Uncountable : Statistical Approaches to Estimating Microbial Diversity. *Appl. Environ. Microbiol.* **10**, 4399–4406 (2016).
108. Luna, R. A. & Foster, J. A. Gut brain axis: diet microbiota interactions and implications for modulation of anxiety and depression. *Curr. Opin. Biotechnol.* **32**, 35–41 (2015).
109. Muegge, B. D. *et al.* Diet drives convergence in gut microbiome functions across mammalian phylogeny and within humans. *Science* **332**, 970–4 (2011).
110. Ze, X., Duncan, S. H., Louis, P. & Flint, H. J. Ruminococcus bromii is a keystone species for the degradation of resistant starch in the human colon. *ISME J.* **6**, 1535–1543 (2012).
111. Yolken, R. *et al.* Individuals hospitalized with acute mania have increased exposure to antimicrobial medications. 404–409 (2016). doi:10.1111/bdi.12416
112. Falkow, S., Rosenberg, E., Schleifer, K.-H. & Stackebrandt, E. *The Prokaryotes: Vol. 4: Bacteria: Firmicutes, Cyanobacteria*. (Springer Science and Buisness Media, 2006).
113. Phillips, R. W. *et al.* Kineococcus radiotolerans sp. nov., a radiation- resistant, Gram-positive bacterium. *Int. J. Syst. Evol. Microbiol.* 933–938 (2002).
114. Goodsir, J. XXIII.-History of a Case in which a Fluid periodically ejected from the stomach contained vegetable organisms of an undescribed form. *Ann. Mag. Nat. Hist.* **11**, 125 (1843).
115. Haroon Al Rasheed, M. R. & Senseng, C. G. Sarcina ventriculi: Review of the literature. *Arch. Pathol. Lab. Med.* **140**, 1441–1445 (2016).
116. Ratuapli, S. K., Lam-Himlin, D. M. & Heigh, R. I. Sarcina ventriculi of the stomach: A case report. *World J. Gastroenterol.* **19**, 2282–2285 (2013).
117. Thirkell, D. & Hunter, M. I. . Cell Division in Sarcina flava. *Proc. Biochem. Soc.* **15**, 490 (1967).
118. Crowther, J. S. Sarcina ventriculi in Human faeces. **4**, 343–350 (1970).
119. Glick-Bauer, M. & Yeh, M. C. The health advantage of a vegan diet: Exploring the gut microbiota connection. *Nutrients* **6**, 4822–4838 (2014).
120. Kageyama, A., Sakamoto, M. & Benno, Y. Rapid identification and quantification of Collinsella aerofaciens using PCR. *FEMS Microbiol. Lett.* **183**, 43–47 (2000).

121. Calcagno, C., Lobatto, M. E., Robson, P. M. & Millon, A. Developmental dynamics of the preterm infant gut microbiota and antibiotic resistome. *28*, 1304–1314 (2016).
122. Gomez-Arango, L. F. *et al.* Connections between the gut microbiome and metabolic hormones in early pregnancy in overweight and obese women. *Diabetes* **65**, 2214–2223 (2016).
123. Miragoli, F. *et al.* Impact of cystic fibrosis disease on archaea and bacteria composition of gut microbiota. *FEMS Microbiol. Ecol.* **93**, fiw230 (2017).
124. Kassinen, A. *et al.* The Fecal Microbiota of Irritable Bowel Syndrome Patients Differs Significantly From That of Healthy Subjects. *Gastroenterology* **133**, 24–33 (2007).
125. Opstelten, J. L. *et al.* Gut Microbial Diversity Is Reduced in Smokers with Crohn's Disease. *Inflamm. Bowel Dis.* **22**, 2070–2077 (2016).
126. Rienzi, S. C. Di *et al.* The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *Elife* 1–25 (2013). doi:10.7554/eLife.01102
127. Stewart, I., Schluter, P. J. & Shaw, G. R. Cyanobacterial lipopolysaccharides and human health – a review. *Environ. Heal.* **5**, 7 (2006).
128. McIntyre, a, Gibson, P. R. & Young, G. P. Butyrate production from dietary fibre and protection against large bowel cancer in a rat model. *Gut* **34**, 386–391 (1993).
129. Zhou, L. & Foster, J. A. Psychobiotics and the gut-brain axis: in the pursuit of happiness. *Neuropsychiatr. Dis. Treat.* **11**, 715–723 (2015).
130. Sherwin, E., Sandhu, K. V., Dinan, T. G. & Cryan, J. F. May the Force Be With You: The Light and Dark Sides of the Microbiota–Gut–Brain Axis in Neuropsychiatry. *CNS Drugs* (2016). doi:10.1007/s40263-016-0370-3
131. Eren, A. M. *et al.* Oligotyping: Differentiating between closely related microbial taxa using 16S rRNA gene data. *Methods Ecol. Evol.* **4**, 1111–1119 (2013).
132. Coil, D., Jospin, G. & Darling, A. E. A5-miseq: An updated pipeline to assemble microbial genomes from Illumina MiSeq data. *Bioinformatics* **31**, 587–589 (2015).
133. Seemann, T. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–2069 (2014).
134. Ellner, P., Stoessel, C., Drakeford, E. & Vasi, F. A new culture medium for medical bacteriology. *Am J Clin Pathol* **45**, 502–504 (1966).
135. Brazier, J. S. Yellow fluorescence of fusobacteria. *Lett. Appl. Microbiol.* **2**, 125–126 (1986).
136. Lech, K. & Brent, R. Techniques for bacterial cell culture: media preparation and bacteriological tools. *Curr. Protoc. Mol. Biol.* **Appendix 3**, Appendix 3E (2001).
137. Livingston, S. J., Kominos, S. D. & Yee, R. B. New medium for selection and presumptive identification of the *Bacteroides fragilis* group. *J. Clin. Microbiol.* **7**, 448–453 (1978).
138. Saito, M., Seki, M., Iida, K., Nakayama, H. & Yoshida, S.-I. A novel agar medium to detect hydrogen peroxide-producing bacteria based on the prussian blue-forming reaction. *Microbiol. Immunol.* **51**, 889–92 (2007).
139. Bauer, A. W., Kirby, W. M., Sherris, J. C. & Turck, M. Antibiotic susceptibility testing by a standardized single disk method. *Am J Clin Pathol* **45**, 493–496 (1966).
140. Bodner, S. J., Koenig, M. G., Treanor, L. L. & Goodman, J. S. Antibiotic susceptibility testing of *Bacteroides*. *Antimicrob. Agents Chemother.* **2**, 57–60 (1972).
141. Evers, B. M., Ishizuka, J., Townsend, C. M. & Thompson, J. C. The human carcinoid cell

- line, BON. A model system for the study of carcinoid tumours. *Ann. N. Y. Acad. Sci.* **733**, 393–406 (1994).
142. Larsen, M. V. *et al.* Benchmarking of methods for genomic taxonomy. *J. Clin. Microbiol.* **52**, 1529–1539 (2014).
 143. Darling, A. E., Mau, B. & Perna, N. T. Progressivemauve: Multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* **5**, (2010).
 144. Nikitina, A. S. *et al.* Complete Genome Sequence of an Enterotoxigenic *Bacteroides fragilis*. **3**, 4–5 (2015).
 145. Nayfach, S. & Pollard, K. S. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* **16**, 51 (2015).
 146. Ames, S. K. *et al.* Scalable metagenomic taxonomy classification using a reference genome database. *Bioinformatics* **29**, 2253–2260 (2013).
 147. Zhou, Y., Liang, Y., Lynch, K. H., Dennis, J. J. & Wishart, D. S. PHAST : A Fast Phage Search Tool. *Nucleic Acids Res.* **39**, 347–352 (2011).
 148. Arndt, D. *et al.* PHASTER : a better , faster version of the PHAST phage search tool. *Nucleic Acids Res.* **44**, 16–21 (2016).
 149. Dinh, D. M. *et al.* Intestinal Microbiota, microbial translocation, and systemic inflammation in chronic HIV infection. *J. Infect. Dis.* **211**, 19–27 (2015).
 150. Chen, W., Liu, F., Ling, Z., Tong, X. & Xiang, C. Human intestinal lumen and mucosa-associated microbiota in patients with colorectal cancer. *PLoS One* **7**, (2012).
 151. Sze, M. & Schloss, P. D. Looking for a Signal in the Noise: Revisiting Obesity and the Microbiome. *bioRxiv* (2016). doi:10.1101/057331
 152. McMurdie, P. J. & Holmes, S. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Comput. Biol.* **10**, (2014).
 153. Dhariwal, A. *et al.* MicrobiomeAnalyst: a web-based tool for comprehensive statistical, visual and meta-analysis of microbiome data. *Nucleic Acids Res.* 1–9 (2017). doi:10.1093/nar/gkx295
 154. Su, D. *et al.* Vitamin D signaling through induction of paneth cell defensins maintains gut microbiota and improves metabolic disorders and hepatic steatosis in animal models. *Front. Physiol.* **7**, 1–18 (2016).
 155. Chen, W., Zhang, C. K., Cheng, Y., Zhang, S. & Zhao, H. A Comparison of Methods for Clustering 16S rRNA Sequences into OTUs. *PLoS One* **8**, (2013).
 156. Franzén, O. *et al.* Improved OTU-picking using long-read 16S rRNA gene amplicon sequencing and generic hierarchical clustering. *Microbiome* **3**, 43 (2015).
 157. Ho, T. K. The Random Subspace Method for Constructing Decision Forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, 832–844 (1998).
 158. Stearns, J. C. *et al.* Ethnic and diet-related differences in the healthy infant microbiome. *Genome Med.* **9**, 32 (2017).