GUT MICROBIOME IN IMMUNE-MEDIATED INFLAMMATORY DISEASE

by

JESSICA DAWN FORBES

A thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Medical Microbiology and Infectious Diseases

University of Manitoba

ABSTRACT

Immune-mediated inflammatory diseases (IMID) represent a group of ostensibly unrelated, chronic and highly disabling diseases that preferentially affect different organ systems. IMID are assumed to manifest as a result of the accumulation of genetic, environmental and immunological factors. A fundamental commonality between IMID is the idiopathic nature of disease, and moreover, substantial similarities are apparent in disease etiopathogenesis. The complex assemblage of microbes and their genes that exists within and on the human body, collectively known as the microbiome has emerged as a critical factor in human health and, altered microbial populations within the gastrointestinal tract lumen and mucosa have been linked to several IMID. Accordingly, we conducted several studies investigating the association of the gut microbiome with IMID. Our main study investigated differences in the microbial profile and functional potential of multiple IMID utilizing 16S rDNA amplicon sequencing and analysis of stool. We also investigated the mucosal-associated microbiome in IBD to characterize the microbial populations and their functions residing in distinct gastrointestinal compartments from inflamed and noninflamed mucosa. We also explored a potential environmental factor; specifically assessing whether microbes present in drinking water in low or high incidence areas of IBD might contribute to disease etiology. The findings of these studies are manifold. First, we show important differences of the stool microbial profile in IMID. In doing so, we were able to identify distinct states of gut dysbiosis and have revealed numerous microbes that are consistently or uniquely disproportionate between IMID. Second, we have shown the microbial profile associated with inflamed and noninflamed mucosa and have reported that a localized dysbiosis is not observed in the presence of inflammation. Third, we have

revealed that distinct gastrointestinal compartments are comprised of similar microbial communities. Lastly, we have reported the drinking water microbiome to differ between low and high incidence areas of IBD, thus suggesting a potential role in IBD etiology. Understanding the role of the gut microbiome in human disease will enable the development and application of more appropriate therapeutic strategies that specifically target microbes within the gut.

ACKNOWLEDGEMENTS

First and foremost, I'd like to express my sincerest gratitude to my supervisor Gary Van Domselaar for taking me on as a Ph.D. student, for being a stellar mentor and for introducing me to the world of computational biology of which I have grown to love! Secondly, a huge thanks to my committee: Charles Bernstein, Michelle Alfa and Teri De Kievit. Charles, I'll never understand how you can be *so* busy but still review and edit an 8,000-word literature review within hours of receiving it. Michelle, you have been so influential in guiding me to the right career path and it's an honour to be your last student (enjoy retirement!). And Teri, you taught me my first microbiology course 9 years ago – I truly believe I wouldn't have pursued a microbiology undergrad degree if it weren't for your enthusiasm and passion for teaching.

Massive thanks to everyone in the bioinformatics section, past and present, at the NML. If not just for putting up with some periodic yelling or crying on my part, but also for the help, the conversation, the jokes, and even the making fun of me for not having adequate figure creating skills (Adrian – please never get rid of the autographed figure I made). Natalie, you've been a fantastic mentor, and more importantly a friend. I'll miss our 'vent sessions' and I hope wherever I end up, I have someone like you to work with! And Eric, you've been a ton of help with *trying* to teach me to code, but our conversations will definitely be more memorable; for the incredibly random discussions and also your choice vocabulary that I have never, nor will probably ever hear again! And to everyone else: Phil, Eric, Shane, Deep, Cam, Tom, Aaron, Josh, Franklin, Kelly, Paul, Jen, Shivani and Mariam, it's been a blast! I'd also like to thank the genomics

section at the NML: Morag, Shaun, Vanessa, Brynn, Shari, Christine and everyone else, for their help with sequencing.

Thank you to Dionne and Kristina for our ever-eventful lunches in Elgin; you guys have been some of the best co-workers anyone could hope for! Amrit – the coffee trips and chats have kept me sane, especially in the last while. Thanks to everyone in Medical Microbiology, but specifically Angie: your help and answering (annoying) questions has been awesome. I would also like to thank the Mindel and Tom Olenick Research Studentship in Medicine for providing me with partial financial support for the last few years. And additionally, this research could not have been done without the patients – huge thanks for participating in these studies.

Lastly, thank you to my family (yes, including my cat Molly) for their support and encouragement throughout my PhD. And finally, the moment has come: I don't have to freak out anymore when someone asks *'how much longer until you're done?'*

DEDICATION

I dedicate this thesis

to my Oma

...my favourite person who ever lived.

FOREWARD

Sections of this dissertation have been published in the form of three peer-reviewed articles.

i. **Forbes JD**, Van Domselaar G, Bernstein CN. 2016. The gut microbiota in immune-mediated inflammatory diseases. Front Microbiol;7:1081. doi: 10.3389/fmicb.2016.01081.

ii. **Forbes JD**, Van Domselaar G, Bernstein CN. 2016. Microbiome survey of the inflamed and noninflamed gut at different compartments within the gastrointestinal tract of inflammatory bowel disease patients. Inflamm Bowel Dis;22(4):817-825. doi: 10.1097/MIB.0000000000000684.

iii. **Forbes JD**, Van Domselaar G, Sargent M, Green C, Springthorpe S, Krause DO, Bernstein CN. 2016. Microbiome profiling of drinking water in relation to incidence of inflammatory bowel disease. Can J Microbiol;62(9):781-793. doi: 10.1139/cjm-2016-0219.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF COPYRIGHTED MATERIAL FOR WHICH PERMISSION WAS OBTAINED

Forbes JD, Van Domselaar G, Bernstein CN. 2016. The gut microbiota in immune-mediated inflammatory diseases. Front Microbiol;7:1081. doi: 10.3389/fmicb.2016.01081.

Copyright permission in appendix VI

Forbes JD, Van Domselaar G, Bernstein CN. 2016. Microbiome survey of the inflamed and noninflamed gut at different compartments within the gastrointestinal tract of inflammatory bowel disease patients. Inflamm Bowel Dis;22(4):817-825. doi: 10.1097/MIB.0000000000000684.

Copyright permission in appendix VI

Forbes JD, Van Domselaar G, Sargent M, Green C, Springthorpe S, Krause DO, Bernstein CN. 2016. Microbiome profiling of drinking water in relation to incidence of inflammatory bowel disease. Can J Microbiol;62(9):781-793. doi: 10.1139/cjm-2016-0219.

Copyright permission in appendix VI

Figure 1. Layers of the gastrointestinal tract taken from Johnstone et al., (2016) published in: Johnstone, C., Hendry, C., Farley, A., and McLafferty, E. (2016). The digestive system: part 1. *Nurs. Stand.* 28, 37–45.

Copyright permission in appendix VI

study of inflammatory bowel disease: Searching for etiologic clues. *Am. J. Epidemiol.* 164, 615–623. doi:10.1093/aje/kwj260.

Copyright permission in appendix VI

LIST OF ABBREVIATIONS

ACE	abundance-based coverage estimator
ACPA	anti-citrullinated protein antibody
AIEC	adherent-invasive *Escherichia coli*
ANS	autonomic nervous system
anti-CCP	anticyclic citrullinated peptide
ARISA	automated ribosomal intergenic spacer analysis
bp	base pair
CD	Crohn's disease
CIS	clinically isolated syndrome
CNS	central nervous system
CTL	cytotoxic T lymphocyte
DC	dendritic cell
ddNTP	dideoxynucleoside triphosphate
dNTP	deoxynucleoside triphosphate
DNA	deoxyribonucleic acid
DMARD	disease modifying anti-rheumatic drug
EAE	experimental autoimmune encephalomyelitis
ENS	enteric nervous system
FDR	false discovery rate
FISH	fluorescent in situ hybridization
FMT	fecal microbiota transplantation

GALT	gut-associated lymphoid tissue
GBA	gut-brain axis
gDNA	genomic deoxyribonucleic acid
HC	healthy control
HIA	high incidence area
HLA	human leukocyte antigen
HPA	hypothalamic pituitary adrenal
IBD	inflammatory bowel disease
IEL	intraepithelial lymphocyte
IFN	interferon
Ig	immunoglobulin
IL	interleukin
IMID	immune-mediated inflammatory disease
KEGG	Kyoto encyclopedia of genes and genomes
KO	Kyoto encyclopedia of genes and genomes ortholog
LDA	linear discriminant analysis
LIA	low incidence area
LEfSe	linear discriminant analysis effect size
MALT	mucosa-associated lymphoid tissue
MaAsLin	multivariate association with linear models
mLN	mesenteric lymph node
MS	multiple sclerosis
NGS	next-generation sequencing

NK	natural killer
NOD	nucleotide-binding oligomerization domain
OTU	operational taxonomic unit
PAMP	pathogen-associated molecular pattern
PCoA	principal coordinate analysis
PCR	polymerase chain reaction
PICRUSt	phylogenetic investigation of communities by reconstruction of unobserved states
PPMS	primary progressive multiple sclerosis
PRMS	progressive relapsing multiple sclerosis
PRR	pathogen recognition receptor
RA	rheumatoid arthritis
RDP	ribosomal database project
Reg	regenerating islet-derived protein
RF	rheumatoid factor
RIS	radiologically isolated syndrome
RNA	ribonucleic acid
RRMS	relapsing remitting multiple sclerosis
rDNA	ribosomal deoxyribonucleic acid
rRNA	ribosomal ribonucleic acid
SCFA	short chain fatty acids
SLE	systemic lupus erythematosus
SPINGO	species level identification of metagenomic amplicons

SPMS	secondary progressive multiple sclerosis
T-RFLP	terminal restriction fragment length polymorphism
TGF	transforming growth factor
Th	T-helper
TLR	toll-like receptor
Treg	regulatory T
UC	ulcerative colitis
ZR	Zymo Research

CHAPTER ONE: GENERAL INTRODUCTION

The recognition of the complex assemblage of microbes and their genes that exists within and on the human body, collectively known as the microbiome has emerged as a principal factor in human health and disease. Humans and microbes have established a symbiotic association and perturbations in this association have been linked to several human disorders including those defined as immune-mediated inflammatory diseases (IMID). IMID is a term used to describe a group of chronic, highly disabling diseases that affect different organ systems. Though a cornerstone commonality between IMID is the idiopathic nature of disease, a considerable portion of their pathobiology overlaps including epidemiological co-occurrence, genetic susceptibility loci and environmental risk factors. At present, it is clear that persons with IMID are at an increased risk for developing comorbidities, including additional IMID, however, detailed mechanisms for the development of these comorbidities remain unknown.

Clinical, epidemiological, genetic and environmental links between IMID combined with an increased incidence of gastrointestinal disorders and with the preeminent yet incompletely defined association between inflammatory bowel disease (IBD) and the gut microbiome has collectively laid the foundation for studies evaluating the gut microbiome in other IMID. Advancements in sequencing technologies catalyzing an explosion of 16S rDNA and metagenomics community profiling studies have allowed for the characterization of microbiomes throughout the human body including the gut, in health and for myriad human diseases.

Intriguingly, recent research suggests that there are striking findings to support the gut microbiome as being intricately involved in the etiopathogenesis of several IMID including IBD (Naftali et al., 2016), multiple sclerosis (MS; Tremlett et al., 2016b) and rheumatoid arthritis (RA; Chen et al., 2016b). A fundamental challenge now is to determine if alterations of gut microbes are common between IMID or, if particular changes in the gut community are in fact specific to a single disease. Investigation of these microbial alterations may provide insights into particular microbes or groups of microbes that are consistently or uniquely disproportionate between IMID and ultimately lead to an understanding of how these microbes might influence the disease course.

While the gut microbiome may be involved in IMID onset, which is characterized by an overactive immune response, and in most cases preferentially targeting a particular organ, microbes alone cannot account for the epidemiological variability observed in IMID. Genetic susceptibility loci explain only part of disease risk; hence the remaining element should be environmental factors. A multitude of environmental risk factors have been extensively investigated in recent years, many of which are common between numerous IMID, but whether any of these factors are causal has yet to be established and raises the possibility that perhaps undefined environmental factors contribute to disease etiopathogenesis.

CHAPTER TWO: LITERATURE REVIEW

2.1 Overview of immune-mediated inflammatory diseases

Immune-mediated inflammatory disease (IMID) is a term used to define a group of clinically heterogeneous, ostensibly unrelated disorders that are recognized to share common pathogenic mechanisms. Most IMID are highly prevalent in well-developed industrialized countries. In Western populations the prevalence of IMID is approximately 5-8% (Bayry and Radstake, 2013) and encompasses over 100 different clinical diseases including inflammatory bowel disease (IBD), multiple sclerosis (MS), rheumatoid arthritis (RA), ankylosing spondylitis, systemic lupus erythematosus (SLE) and psoriasis/psoriatic arthritis.

Knowledge of the etiopathogenic mechanisms of IMID remains limited but these diseases are thought to arise secondary to a complex interplay between environmental and genetic factors. The rise in IMID prevalence in developed countries in the latter half of the 20th century implies that environmental determinants play a significant role in clinical disease onset. This view is corroborated by improved public health, antibiotic usage and dietary changes. Epidemiological data suggest that the increase of IMID is analogous to allergic and cancerous conditions whereas the frequencies of infections are decreasing thus establishing the basis for the hygiene hypothesis (Roduit et al., 2016). Geoepidemiologically, the linkage of IMID to their rising prevalence, socioeconomic factors, migratory influence, geographical gradients and relatively low concordance rate between monozygotic twins for most IMID collectively imply a dominant environmental impact rather than simply long-term genetic factors (Brant, 2011; Hedstrom et al., 2016; Howard et al., 2016; Sonnenberg et al., 2016; Szilagyi and Xue, 2016). Numerous additional environmental factors have been proposed to be important in the emergence of IMID

including cigarette smoking, diet, antibiotic usage, vitamin D, urbanization, hormones, appendectomy, excess alcohol, high coffee consumption and microbial exposure (Ananthakrishnan, 2015; Belbasis et al., 2015; Hedstrom et al., 2016; Pinto et al., 2015). Whether any of these factors are causal in IMID has yet to be established and certainly raises the possibility that perhaps undefined environmental factors contribute to disease etiopathogenesis.

IMID also share a considerable portion of their heritable etiology. This is evident through familial clustering of multiple IMID (referred to as kaleidoscope of autoimmunity), epidemiological co-occurrence and a similar efficacy of therapeutics directed at specific biologic loci across some diseases (Diaz-Gallo and Martin, 2012; Somers et al., 2006), which taken together, suggest that genetic factors predispose persons to IMID. The best-known genetic factor of IMID are the human leukocyte antigen (HLA) haplotypes (Wu et al., 2015). Moreover, genome-wide association studies have identified associations with more than 200 rare or common non-HLA variants (Wu et al., 2015) although known allelic variants account for a relatively low risk explaining 20-50% of disease heritability (Parkes et al., 2013; Sorrentino, 2014). Taken together, shared environmental, genetic and epidemiological characteristics of IMID suggest that microbial influences might have a role in etiology of disease.

2.1.1 Inflammatory bowel disease

IBD is a chronic inflammatory condition of the gastrointestinal tract. The two main clinical phenotypes of IBD are Crohn's disease (CD) and ulcerative colitis (UC), both of which present in a relapsing-remitting manner. Persons with IBD present with symptoms such as bloody diarrhea, diffuse abdominal pain and cramping, anemia, fever and fatigue, weight loss, tenesmus, mucousy and frequent stool. CD is a transmural disease characterized by deep ulcerations

distributed in a skipped pattern that can affect the gastrointestinal tract anywhere from the mouth to the anus, though the distal ileum and cecum are most frequently affected. UC, in contrast, demonstrates continuous, superficial inflammation (of the mucosa and submucosa) involving the rectum and extending proximally to the cecum. The presence of inflammation, ulceration, fistula and abscesses are often seen in CD whereas disease generally does not lead to fistulae nor abscesses in UC. Current therapeutic management is focused on reducing the inflammatory burden in persons with active disease and maintaining remission in persons with inactive disease.

Descriptive epidemiology

When describing population characteristics, by definition the term incidence describes the number of new occurrences per population within a specified time period whereas the term prevalence refers to the proportion of occurrences per population within a specified time period. Hence, incidence measures risk and prevalence measures how widespread a particular disease is. The incidence of UC varies from 0–19.2 and 0.6–24.3 per 100,000 in North America and Europe, respectively (Molodecky et al., 2012). This corresponds to a prevalence of 37.5–248.6 and 4.9–505 per 100,000. The incidence of CD is similar: 0–20.2 per 100,000 in North America and 0.3–12.7 per 100,000 in Europe. In the past several decades, the incidence of IBD has changed considerably; a North American population-based study in Olmsted County, MN, USA, reported the 1940–1943 UC incidence at 0.6 per 100,000 and 1984–1993 at 8.3 per 100,000 (Loftus Jr et al., 2000) with similar trends observed in CD (Loftus Jr et al., 1998). Studies of both CD and UC are reporting that the rate of increase might have decelerated although a decline as of yet has not been reported (Loftus et al., 2007). Trends are also variable within a particular geographical area. For example, urban populations report higher incidence than rural populations

(Soon et al., 2012). A latitudinal north-south gradient is also evident with higher incidences reported in northern latitudes (Holmes et al., 2015) and analogously; an inverse relationship might be present in the southern hemisphere, as New Zealand has reported a high incidence of IBD (Su et al., 2016). Lastly, particular populations within an area of high incidence have reported a significantly reduced incidence as seen in the First Nations population of Canada (Bernstein et al., 2011). The peak incidence of IBD occurs between the ages of 20 and 40 and some cohorts have reported a second more moderate incidence peak between 60 and 70 years particularly in UC (Molodecky et al., 2012). Incidence is comparable between male and females but race and ethnicity are influential. Jewish populations have a threefold higher risk of IBD than non-Jewish populations (Bernstein et al., 2006): among Jewish populations the risk of developing IBD is increased in Ashkenazi versus Sephardim populations and similarly, the risk is also greater in American and European Jewish populations compared to Israeli populations. African-American and Hispanic populations were once thought to have markedly lower prevalence of IBD, however, studies suggest that the difference in incidence between whites and nonwhites is not as sizeable as once thought (Nguyen et al., 2014).

2.1.2 Multiple sclerosis

MS is a chronic inflammatory demyelinating disorder of the CNS. The cause of disease is not entirely understood, but both genetic and environmental components are involved in disease susceptibility (Dendrou et al., 2015). Early in disease, patients present with fatigue, depression, pain and impairments related to sensory, vision, cognition; mobility, hand function, fatigue, bowel and bladder dysfunction, and spasticity are more common with longer disease duration (Fox et al., 2015). Though the CNS was originally regarded as an immune-privileged compartment, it is now recognized that immune cells survey the CNS regularly (Zhang et al.,

1994); in MS and related neurological diseases there is an enrichment of autoreactive immune cells targeting elements of the CNS (myelin in MS). The discovery that gut microbes can produce neurotransmitters that can in turn influence the enteric nervous system (ENS) and ultimately the CNS has led to the notion of the microbiota-gut-brain axis (Wang and Kasper, 2014).

The clinical pathophysiology of MS is classically defined as one of four phenotypes: relapsing-remitting (RRMS), secondary progressive (SPMS), primary progressive (PPMS) and progressive relapsing (PRMS). Further classification can also include the lesser common clinically isolated syndrome (CIS) and radiologically isolated syndrome (RIS; Rodriguez et al., 2016). RRMS is the most common phenotype of MS occurring in approximately 80% of MS presentations. The clinical characterizations of RRMS are unpredictable relapses that are followed by a variable length of remission. Functional deficits accrued over the course of a relapse can either be resolved or leave impairments, particularly in later stages of disease. In roughly 65% of RRMS cases, the disease eventually progresses (typically 20 years) to SPMS described as a progressive neurological and functional decline with no defined phases of remission. Roughly 15% of initial MS episodes are defined as PPMS that present with no apparent periods of remission after initial symptoms. Persons with PRMS present with a steady neurological and functional decline and no indication of attack. The classification of CIS is generally used as the first episode of MS where an attack is suggestive of demyelination but does not meet MS criteria in entirety. It is worth noting that many studies investigating gut involvement in MS are based on RRMS patients.

Descriptive epidemiology

The prevalence of MS is highest in temperate regions including North America, Europe, New Zealand and southern Australia and lowest in Asia, the tropics and subtropics (Ascherio and Munger, 2016). Following a similar trend as IBD, the incidence and prevalence of MS is reportedly increased with increasing latitude in the northern hemisphere (Kurtzke, 1995) and in the southern hemisphere, an inverse relationship is suspected (Mason et al., 2016). However, in recent years, the latitudinal gradient in the northern hemisphere has seemingly attenuated, effectively disappearing in the US (Koch-Henriksen and Sorensen, 2011). The incidence occurring in urbanized populations is also reportedly increased (Kotzamani et al., 2012). In the majority of established populations the female-to-male ratio ranges from 1.5:1 to 2.5:1 (Koch-Henriksen and Sørensen, 2010), which is likely attributed to environmental risk factors such as hormones (Tan et al., 2014) or smoking patterns (Palacios et al., 2011). The peak incidence of MS occurs between 25 and 35 years (Ascherio and Munger, 2016). Race and ethnicity might have an effect on incidence: whites, specifically of northern European descent appear to be the highest at-risk population, although recent studies indicate that the MS incidence in blacks is currently minimally different from MS incidence in whites (Wallin et al., 2015). This apparent change in MS demographics might be a result of increased incidence related to environmental factors such as the Western lifestyle or vitamin D associated with geographical location but could also be related to prior studies underrepresenting specific racial/ethnic populations.

2.1.3 Rheumatoid arthritis

RA is a common systemic disease that primarily manifests in the joints and if untreated can lead to chronic joint deformity, disability and increased mortality owing to cardiovascular and other concurrent, systemic complications. Despite recent advances in understanding the pathogenesis of RA (McInnes and Schett, 2011), the etiology remains undefined.

RA is often described as an autoimmune disease historically due to the presence of rheumatoid factor (RF), an antibody directed against the Fc region of immunoglobulin G (IgG). The presence of RF however, is not specific for RA but can be characterized as a general outcome of immune activation. Further, there is an absence of experimental studies indicating any form of proarthritogenic effect inherent in RF. Anti-citrullinated protein antibodies (ACPA) as discussed in section 2.5.1.3 demonstrate proarthritogenic abilities and are highly sensitive and specific serological markers for RA occurring in approximately 60% of RA patients and in normal populations at 2% (Nishimura et al., 2007). Citrullination is a term used to describe post-translational modification of a charged peptidyl arginine to a neutral peptidyl citrulline (Szekanecz et al., 2008). Citrullination physiologically occurs during processes of apoptosis, inflammation or keratinization. Preeminent evidence suggests a preclinical period (absence of synovitis) in the development of seropositive RA during which the abundance of autoantibodies including RF and ACPA are elevated often years preceding the onset of detectable joint manifestations and hence diagnosis of disease. As discussed in section 2.5.1.3 emerging data suggest the instigation of RA autoimmunity might originate in the intestinal mucosa. Furthermore, numerous genetic susceptibility risk alleles have been identified, but are insufficient to account for disease incidence (40–65% for seropositive RA; 20% for seronegative; Frisell et al., 2016). That said, RA is considered a multifactorial disease requiring both genetic and environmental factors for disease onset and/or progression.

Descriptive epidemiology

In North America and Europe, the incidence of RA is 25–50 per 100,000 with a corresponding prevalence of 0.5–1% (Minichiello et al., 2016). A recent systematic review could not define current frequency trends, as data was highly conflicting: cohorts have reported declining, increasing or stable incidence (Minichiello et al., 2016). Analogous to trends observed in IBD and MS, a latitudinal gradient in the northern hemisphere (higher incidence in the North) is apparent and in several established populations a higher incidence is associated with urbanization (Alamanos et al., 2006; Zeng et al., 2008). Moreover, within particular geographically defined areas, the risk of RA differs between distinct ethnicities, suggestive of a genetic impact. First Nations for example have a markedly increased risk of RA (El-Gabalawy et al., 2011). RA is exceedingly more common in females with a female-to-male ratio of 3:1 (Tobon et al., 2010). The peak age of onset is in the fifth decade of life, though recent studies suggest that this is shifting toward an older age (Tobon et al., 2010).

2.1.4 Immune-mediated inflammatory disease comorbidity

IMID are defined by the principle organ system that is affected (the gastrointestinal tract in IBD, the synovium in RA and the central nervous system (CNS) in MS); however, they are associated with comorbidities that extend beyond the primary target organ. In this regard, comorbidities that present in persons with IMID greatly contribute to the burden of disease and quality of life. IBD for example, frequently manifests with extra-intestinal complications in up to 50% of cases and covers a broad clinical spectrum affecting nearly every organ system (Harbord et al., 2016). Erythema nodosum and pyoderma gangrenosum of skin, ocular uveitis and episcleritis, primary sclerosing cholangitis and arthritides such as pauciarticular and polyarticular peripheral arthritis and seronegative spondyloarthropathy are among the more common extra-intestinal diseases that may coexist in IBD (Harbord et al., 2016). Rarely, a number of neurological complications

including peripheral neuropathy, myopathy and demyelinating disorders may occur in persons with IBD. Comorbidities are not specific to IBD; persons with other IMID often present with a number of additional clinical disorders including risks for osteoporosis, venous thromboembolic disease and ischemic heart disease (Bernstein et al., 2000, 2001a, 2008).

It is well established that in comparison to the general population, IMID patients are at greater risk for the development of additional IMID-related conditions. Weng et al., (2007) examined the co-occurrence of IBD with other IMID (asthma, psoriasis, type 1 diabetes (T1D), RA, MS, SLE, vitiligo, autoimmune thyroiditis and chronic glomerulonephritis). The authors reported 17% of IBD patients and 10% of persons without IBD were diagnosed with at least one additional IMID. Persons with IBD had significantly increased odds pertaining to the development of asthma (odds ratio; OR 1.5, 95% confidence interval (CI) 1.4-1.6), psoriasis (OR 1.7, 95% CI 1.5-2.0), RA (OR 1.9, 95% CI 1.5-2.3) and MS (OR 2.3 95% CI 1.6-3.3). These results have been corroborated by other studies that add further credence to the validity of the phenomenon of IMID co-occurrence (Bernstein et al., 2001b, 2005; Cohen et al., 2008; Cucino and Sonnenberg, 2001; Marin-Jiminez et al., 2014; Vanaclocha et al., 2015).

2.2 The gastrointestinal tract

The gastrointestinal tract, also known as the alimentary canal or digestive tract is a part of the organ system (digestive system) fundamental to the consumption and digestion of food, absorption of nutrients and expulsion of waste products. The gastrointestinal tract comprises a long, hollow tube that also includes a muscular layer, which allows for organ walls to move in a process referred to as peristalsis. The hollow centre of the gastrointestinal tract is referred to as the gut lumen. Peristalsis moves ingested food and liquids through the gastrointestinal tract,

mixing the contents within each organ. The gastrointestinal tract extends from the oral cavity where food and liquids are introduced and continue via the esophagus to the stomach and small intestine where food is absorbed and digested; muscles of the small intestine propel waste products to the large intestine where stool is formed and expelled through the rectum in a process that takes an estimated 18-24 hours. The following section focuses on the lower gastrointestinal tract, as this compartment harbours the vast majority of gut microbes and is thus directly relevant to studies conducted in this thesis.

The small intestine is roughly 6-7 metres in length and has a particular structure to maximize surface area to over 30 m^2 (Helander and Fandriks, 2014) and thus acquires the ability to maximize nutrient absorption: large folds of mucosa are covered in villi (long fingerlike projects) and microvilli (microscopic protrusions) that project into the lumen. Compartments of the small intestine include the duodenum, jejunum and of relevance to this thesis, the ileum. Subsequent to the digestion of food that occurs in the duodenum, nutrients are absorbed into enterocytes throughout the jejunum and ileum by a variety of mechanisms including simple or facilitated diffusion, active transport, endocytosis or by paracellular passage (Goodman, 2010). The ileum occupies the central and right lower abdomen and pelvis, constituting approximately three-fifths (2-4 metres) of the small intestine; the small intestine extends to the ileocecal valve, which in turn moves contents to the large intestine. The large intestine is approximately 1.5 metres in length and is considerably wider in diameter (6-7 cm) than the small intestine. The cecum forms the beginning segment of the large intestine, followed by the ascending (right) colon, which ascends on the right side of the abdomen, to the transverse colon, which passes across the abdominal cavity, to the descending (left) colon, which descends on the left side of the abdomen

into the sigmoid colon and ultimately, the rectum. Also a section of the large intestine is an appendage connected to the cecum, the vermiform appendix, was until recently thought vestigial, but is now thought to act as a 'safe-house' for normal gut bacteria (Laurin et al., 2011). The appendix contains the same layers as the rest of the large intestine and also has a more pronounced population of lymphoid tissue. Large populations of bacteria inhabit the large intestine (discussed in section 2.5); these commensals play a pivotal role in host health including functions related to food digestion and metabolism. Furthermore, nutrients and byproducts produced by commensals can be absorbed into the host body or utilized proximally by intestinal epithelial cells. The large intestine secretes mucus and is fundamentally responsible for the absorption of water from the lumen and stores stool in the rectum and sigmoid colon until it is eliminated. Stool is comprised of water, cellulose fibre, fatty acids, microbes, epithelial cells, mucus and cellular detritus (Waugh and Grant, 2010).

2.2.1 Layers of the gastrointestinal tract

The gastrointestinal tract contains four core layers: the mucosa, submucosa, muscularis externa and the serosa (Figure 1). The mucosa is the innermost layer and is made up of three sublayers: the epithelium, lamina propria and muscularis mucosae. The intestinal epithelium represents the innermost layer found in a cross section of the gastrointestinal tract; it lines the intestinal lumen and is thus in direct contact with the external environment including commensals, pathogens, food, chemicals, pollutants and other antigens. Four major types of epithelial cells comprise the intestinal epithelium: enterocytes, goblet cells, enteroendocrine cells and present only in the small intestine, Paneth cells (Birchenough et al., 2015; Höfer et al., 1998). Enterocytes are the most abundant cells of the epithelium; they provide defense against luminal microbes via the formation of a physical barrier and production of antimicrobial peptides such as β-defensins,

cathelicidins, and regenerating islet-derived protein 3β (RegIIIβ) and RegIIIγ; they also have a primary role in nutrient absorption. The lamina propria is an underlying layer consisting of connective tissue, small blood and lymphatic vessels and lymphoid tissue that provide vascular support for the epithelium (Pearson et al., 2012). The muscularis mucosa consists of a thin double layer of smooth muscle for movement of the mucosa. The submucosa, the second layer of the gastrointestinal tract, is comprised of a dense layer of connective tissue, lymphoid tissue (mucosa-associated lymphoid tissue; MALT) and nerves (Seeley et al., 2007). The nerves form a complex – the submucosal plexus (plexus of Meissner) that is controlled by the ANS. The role of the submucosa is to control mucosal gland secretions, mucosal movement and blood flow (Tortora and Derrickson, 2013). The third layer of the gastrointestinal tract is the muscularis externa, which typically includes a circular and longitudinal smooth muscle layer with the myenteric plexus (plexus of Auerbach) found between the layers, all of which are critical for peristalsis (Tortora and Derrickson, 2009). The outermost layer, the serosa, consists of connective tissue and is covered by the visceral peritoneum.

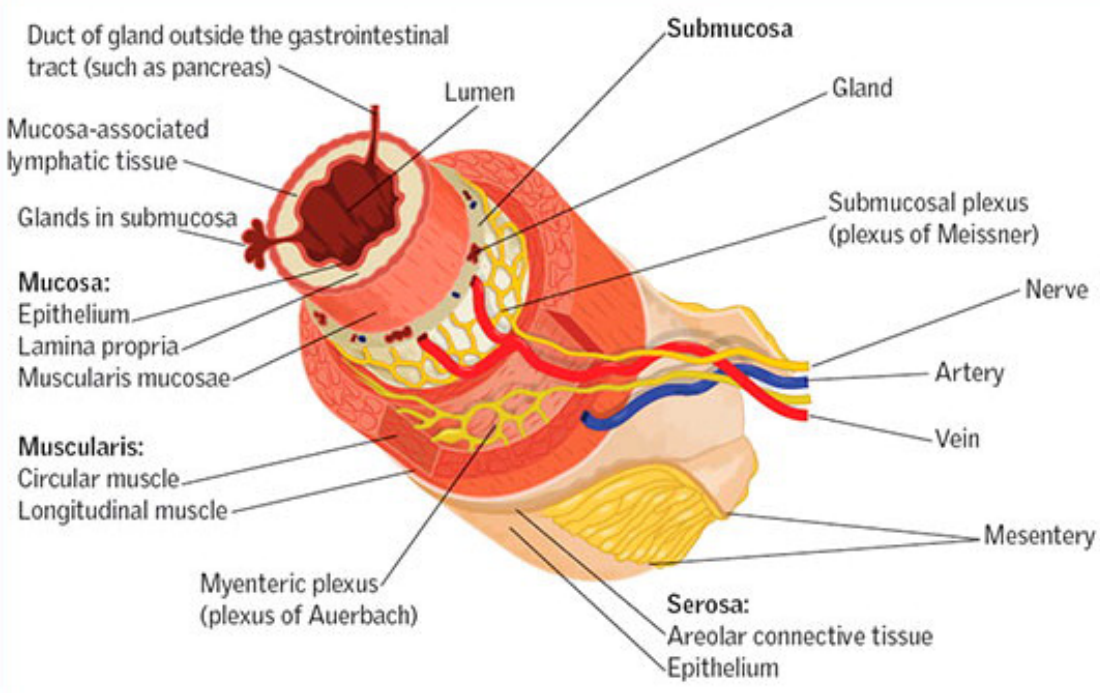

Figure 1. Layers of the gastrointestinal tract. Used with permission from Johnstone et al., (2016).

2.2.2 Intestinal barrier

The adult human lower gastrointestinal tract consists of a vast surface area (approximately 30 m^2) representing the largest area of the human body continuously in contact with the external environment (Hao and Lee, 2004). The intestinal barrier resembles a polymeric impermeable membrane and is composed of a thick secreted mucus layer, a single layer of epithelial cells and the lamina propria. Optimized for the uptake of nutrients, water and electrolytes, the intestinal barrier also functions as the primary defense separating the host from the luminal environment – a potentially harmful milieu of bacteria, fungi, viruses, food allergens, chemicals, pollutants and other antigens. This dynamic barrier is continually renewed and functions to maintain a state of

mucosal homeostasis via interaction with luminal antigens through innate receptors (i.e. Toll-like receptor; TLR) and host immune cells via secreted chemokines and others. Tight junctions apically connect intestinal epithelial cells; over 50 tight junction proteins that regulate paracellular permeability constitute these complex structures including occludin, claudin, zonulin-1 and tricellulin. Additional factors such as immune cells, the gut microbiota and antimicrobial peptides also have a supportive role in intestinal barrier function.

Maintaining epithelial barrier function is essential for health. Disruption of this barrier results in increased intestinal permeability commonly referred to as a "leaky gut" which has been shown to contribute to the pathophysiology of numerous gastrointestinal disorders including but not limited to IBD, celiac disease and acute microbial infections with pathogenic *E. coli*, *C. difficile* and *Vibrio vulnificus* and others (Dheer et al., 2016; Hollon et al., 2015; Jafari et al., 2016; Lee et al., 2016; Morampudi et al., 2016). A rising number of factors have been identified to enhance gut permeability such as dietary fats, stress/cortisol, and alcohol (Bala et al., 2014; Paula et al., 2012; Vanuytsel et al., 2014). A disrupted epithelial barrier allows for the indiscriminate passage of intraluminal compounds including macromolecules, toxins, and both commensal and pathogenic bacterial species to the vascular compartment that could in turn potentiate systemic inflammation. This phenomenon has the ability to lead to sepsis, organ damage or failure and death (Doig et al., 1998). Importantly, their presence in systemic circulation is thought to trigger chronic immune-mediated diseases (i.e. MS, RA, type I diabetes and others) in different systems distant from the gastrointestinal tract. As an example, it was recently shown that RRMS patients have a significantly higher proportion of intestinal permeability compared to healthy controls (Buscarinu et al., 2016). An increasing number of studies are exploring the association of

intestinal permeability and IMID, under the notion that translocation of intraluminal contents could either directly or indirectly affect host function elsewhere (the CNS in MS and synovium in RA; Buscarinu et al., 2016; Ho et al., 2016; Terato et al., 2015). What remains to be elucidated, however, is whether increased intestinal permeability is a cause or consequence of these diseases. Understanding how the host regulates intestinal epithelial barrier integrity is critical for understanding the underlying pathophysiological mechanisms of gastrointestinal disorders and other IMID.

2.3 Immune response

The host is continuously exposed to antigens that may be innocuous or potentially harmful. Under normal circumstances the immune system exists in a finely tuned equilibrium; a disruption of this equilibrium triggers an immune response whereby the function is twofold – protection of the host against pathogenic antigens and maintains tolerance to innocuous antigens. An appropriate immune response is a critical aspect to host health: an inadequate immune response can increase susceptibility to infection whereas an overactive immune response can lead to chronic inflammation and other autoimmune diseases. The following section provides a brief overview of the host immune response.

The immune response is divided into two arms: innate and adaptive immunity. The innate immune response represents the first line of defense. It is an unspecific and rapid response with the primary function to limit early infections by preventing the entry and systemic migration of microbes utilizing both molecular and cellular defense mechanisms. Innate immunity is comprised of three barriers: i) physical – mucosal/skin barrier, ii) chemical – secreted cytokines,

complement proteins, antimicrobial peptides and enzymes and iii) cellular – neutrophils, macrophages dendritic cells (DC) and natural-killer (NK) cells.

The second arm of immunity, and second line of defense, is the adaptive immune response, which is characterized by immunological memory, antigen specificity and recognition of non-self antigens. The principal cellular components are the (T and B) lymphocytes that play pivotal roles in humoral and cell-mediated immunity, respectively. Lymphocytes identify microbes by their antigen specific receptors, T cell receptor (T lymphocytes) and B cell receptor (B lymphocytes). Subsequent to activation, T lymphocytes proliferate and differentiate into either effector T cells that secrete signal molecules (cytokines, chemokines) or into memory T cells. Activated B-lymphocytes proliferate and differentiate into plasma cells that secrete antibodies and memory B cells.

T cells

While a comprehensive review of the host immune system components are outside the scope of this review, T cells and their secreted cytokines are highly relevant to IMID and therefore discussion here is thus warranted. Briefly, T helper (Th) cells are a subpopulation of effector T cells that express the CD4 co-receptor. Th cells have numerous immune functions including cytokine secretion and activation of B and T cells. Dependent on the cytokine milieu present in which they are activated, Th cells can be polarized into different subclasses: Th1, Th2, Th9 and Th17 (Kaplan, 2013; Zhu and Paul, 2008). Each subclass secretes specific cytokines to support different biological functions. Th1 cells are primarily involved in cell-mediated immunity; they secrete IFN-γ that stimulates macrophages to eliminate intracellular microbes, trigger an

increased toxicity of cytotoxic T lymphocyte (CTL; CD8+) and NK-cell and can induce inflammation (Zhu and Paul, 2008). Th2 cells secrete IL-4, IL-5 and IL-13 and are involved in allergic reactions, clearing extracellular parasitic infections and aid in B cell differentiation (Zhu and Paul, 2008). Th9 cells produce IL-9, IL-21 and IL-10 and are primarily involved in allergic or intestinal inflammation (Kaplan, 2013). Th17 cells produce pro-inflammatory cytokines (IL-17A, IL-17F, IL-22 and IL-26) and as such, play a critical role in the host immune response. Th17 cells are involved in defense against bacterial and fungal infections, inflammation (Torchinsky and Blander, 2010) and interestingly, a correlation between increased production of IL-17A and multiple IMID (including IBD, MS and RA) has been reported (Hamburg et al., 2016; Tremlett et al., 2016a; Wu et al., 2016). Naïve CD4+ Th cells can also differentiate into regulatory T (Treg) cells that demonstrate regulatory and inhibitor functions. Treg cells secrete IL-10 and transforming growth factor (TGF)-β1 and are mostly involved in suppressing the immune response and controlling inflammation (Joller et al., 2014).

2.3.1 Immune defense of the gut

Host defense in the intestinal mucosa is first provided by the intestinal epithelium: the intestinal epithelium comprises a single layer of cells situated at the intestinal lumen derived from epithelial stem cells within the crypt. Participating in immunologic surveillance and facilitation of the gut immune response, epithelial cells are responsible for mucosal barrier function. Epithelial cells produce mucus and antimicrobial peptides including β-defensins, cathelicidins, RegIIIβ and RegIIIγ to prevent pathogen entry. Epithelial (and other innate immune) cells express pattern recognition receptors (PRR); PRRs are germline-encoded receptors including TLR and nucleotide-binding oligomerization domain (NOD) like receptor that recognizes microbial components termed pathogen-associated molecular patterns (PAMP). In this regard,

TLR and/or NOD recognition and interaction of microbial products such as bacterial DNA, lipopolysaccharides, peptidoglycan, flagellin and metabolites can initiate a response that stimulates immune cells (i.e. cytokines) to secrete signal molecules to recruit cells of the adaptive immune response.

The gastrointestinal tract is the largest immune organ in the body containing up to 70% of the body's immunocytes; it is the gut-associated lymphoid tissue (GALT), a component of the MALT, which is responsible for mounting appropriate immune responses to produce microbe-specific IgA and to maintain gut homeostasis. IgA is one of the most critical molecules in regulating intestinal homeostasis and is the most abundantly produced antibody representing 70% of host antibody production (Macpherson et al., 2008).

2.4 Techniques for studying the gut microbiome

Outside the laboratory, microbes almost never exist in isolation but are instead most frequently observed as complex community assemblages or biofilms whereby various microbes share a similar ecological niche. Understanding the structure, function and mechanisms of microbial communities exemplifies simple yet inadequately understood questions with substantial ecological importance.

2.4.1 Conventional methods for microbial identification

Culture-dependent methods require isolation of a bacterium by growth in a medium and identification of the bacterium by analysis of cultural and cellular morphology, metabolism, growth and other factors. Culturing remains a relevant technique, particularly in the clinical laboratory, when studying interactions (i.e. antibiotic resistance), and remains the gold standard approach for taxonomically defining new species. However, traditional culture based

microbiology is contingent upon the ability to cultivate viable microbes outside of their natural habitat which poses a challenge for the gut as many inhabiting microbes are unable to survive elsewhere. Early studies attempting to investigate the microbes located in the gut have largely underestimated the complexity of gut populations as more than 80% of species could not be cultured (Eckburg et al., 2005).

2.4.2 Studying microbial community structure via culture-independent techniques

To overcome the limitations inherent of culture-based approaches, culture-independent methods were developed. Many culture-independent techniques function on the basis of nucleotide sequencing whereby the order of nucleotides in a DNA molecule will result in sequencing reads that can be taxonomically informative. Early methods have transformed our understanding of microbes, as made evident by the characterization of the three phylogenetic domains of life – Bacteria, Archaea and Eukarya (Woese and Fox, 1977).

Sanger DNA sequencing was developed nearly four decades ago (Sanger et al., 1977) and is still routinely in use based on the same principle applied when it was first introduced. Briefly, millions of copies of the sequence to be determined are amplified (or purified). Reverse strand synthesis is performed on these copies utilizing an upstream known priming sequence and a mixture of deoxynucleoside triphosphate (dNTP) and dideoxynucleoside triphosphate (ddNTP). dNTP and ddNTP trigger random, irreversible termination of the extension reaction resulting in molecules extended to different lengths (due to copy number). Thereafter, resulting molecules are sorted by their size via capillary electrophoresis, which corresponds to the point of termination, and the label attached to the ddNTP is read out sequentially. This method, however, is laborious and expensive relative to more recently introduced sequencing technologies.

2.4.2.1 Next-generation sequencing technologies

Next-generation sequencing (NGS) has revolutionized the discipline of microbial ecology: NGS technologies combine miniaturization with massive parallelization allowing several thousand to millions of sequence reads per sample. Pyrosequencing was the first NGS technology to be commercially introduced (Margulies et al., 2005). This technique requires the production of PCR by DNA amplification with fusion primers that contain sequencing adaptor sequences. The resulting DNA fragments are immobilized on DNA capture beads decorated with PCR primers complementary to the sequencing adaptors and emulsified with amplification reagents in a water-oil amalgam at a dilution that favors a single DNA fragment per bead per water-in-oil droplet. An emulsion PCR step is performed, amplifying the single fragment such that millions of copies of that fragment are bound to the DNA capture bead. The beads are loaded onto a PicoTiterPlate and placed into the instrument. Next, sequencing-by-synthesis occurs whereby reagents are flowed across the PicoTiterPlate and the incorporation of dNTP complementary to the template strand results in the release of pyrophosphate indirectly detected as a chemiluminescent light signal by luciferase. Because the added dNTP is known the template sequence can be determined. Pyrosequencing, while historically popular, is no longer supported. A number of technologies currently exist, though two in particular are in common use for studying the microbiome: reversible dye-terminator chemistry commercialized by Illumina and ion-semiconductor sequencing incorporated into the Life Technologies line of sequencers.

Reversible dye-terminator chemistry

Illumina's sequencing technology is based on a sequencing-by-synthesis approach (Bentley et al., 2008). In this approach, templates are immobilized on a flow cell surface that is specifically

designed to present DNA such that access to enzymes, stability of the template and minimal non-specific binding occurs. Illumina's technology utilizes fluorescently labeled dNTP where the label functions as a polymerization terminator. During each cycle the labeled dNTP is added to the nucleic acid chain, the fluorescent dye is imaged to identify the nucleotide and the label is subsequently cleaved to allow incorporation of the following dNTP.

Ion-semiconductor sequencing

Ion-semiconducting sequencing (Ion Torrent) uses an ion detection method for DNA sequencing (Rothberg et al., 2011). Sequencing is performed in a micro-well array; the Ion Torrent semiconductor-sequencing chip comprises roughly one million copies of a DNA molecule and nucleotides continuously flow across the chip. If a nucleotide complements the DNA sequence in a specific micro-well, DNA polymerase incorporates it and a proton (hydrogen ion) is released. This results in a local change in pH in that micro-well that is detected by an ion sensor.

Additional technologies exist including PacBio and MinIon; however, Illumina currently is the most widely used platform for microbiome analyses, primarily because of its inherent lower error rate and higher read numbers (25 million to 6 billion dependent on the instrument). Ion Torrent is presently the closest competitor to Illumina.

2.4.2.2 Targeted amplicon sequencing

The vast majority of microbial community profiling studies target genes that encode bacterial ribosomal RNA– an essential component of protein synthesis. Ribosomes are comprised of two-subunit ribosomal RNA (rRNA) complexes. In prokaryotes the large subunit includes 5S and 23S rRNA and the small subunit includes the 16S rRNA. The genes are named after their

sedimentation rates measured in Svedberg (S). The 16S rRNA gene is present in all bacteria and archaea; however, the genomic copy number varies from one to fifteen in bacteria (Kembel et al., 2012). The 16S rRNA is an approximately 1500 nucleotide sequence, which is characterized by highly conserved regions flanked by hypervariable; regions (V1–V9), which are distinct among particular groups of bacteria and archaea (Figure 2). Combined with curated databases such as Greengenes (DeSantis et al., 2006b), SILVA (Pruesse et al., 2007) and Ribosomal Database Project (RDP; Wang et al., 2007), the sequence information of hypervariable regions can be used to identify microbes; however, the 16S and corresponding hypervariable regions are often insufficient to provide species level resolution. Although the 16S rRNA represents the gold standard phylogenetic marker, alternative markers have been suggested including 5S rRNA (Pei et al., 2012), 23S rRNA (Taylor et al., 1997), cpn60 (Katyal et al., 2016) and rpoB (Volokhov et al., 2012).

Amplification of 16S rRNA or other markers requires universal oligonucleotide primers. These primers utilize the conserved regions of the gene, which in turn are appropriate annealing targets for PCR amplification. Degenerate primers may also be used which have different nucleotides at degenerate positions. The use of degenerate primers helps to increase diversity on sequencing runs. PCR is unable to achieve true universality (i.e. the possibility to amplify rRNA from all known microbes) and accordingly, no primers are able to cover taxa from all Bacteria, Archaea and Eukarya. Primers are designed to amplify one or more hypervariable regions of the 16S rRNA. Numerous primer combinations exist and are generally dependent on targeting specific microbes or sequencing technology. Moreover, multiplexing is commonly used. This technique allows for numerous amplicon libraries to be sequenced in a single sequencing reaction by

applying primers with attached, unique barcodes (i.e. multiplex identifiers). Sequence reads from individual samples are identified by their barcodes are subsequently demultiplexed for downstream analysis. The advantage of 16S sequencing is the depth of investigation. Specifically, microbes that represent less than 0.01% of microbial populations can be detected.

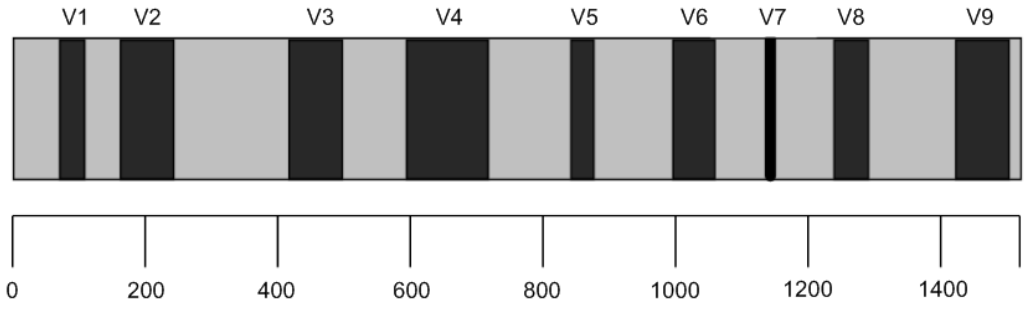

Figure 2. 16S rRNA gene showing conserved and hypervariable regions.

2.4.2.3 Shotgun sequencing

Whole-genome shotgun sequencing represents an approach to sequence large amounts of DNA (i.e. an entire genome). Conversely, shotgun metagenomics sequencing is a technique that has the capacity to partially sequence the majority of community genomes within an environmental sample. In either method, genomic DNA (gDNA) is randomly sheared into smaller fragments and is used directly as a template for sequencing. Fragmented sequences obtained from either sequencing method are reassembled based on overlaps into their original order ultimately yielding the partial or complete original sequence or collection of sequences, depending on the depth of sequencing performed and the diversity of the metagenome. The outcome of shotgun

metagenomics sequencing is a community composition profile that can be associated with the functional structure of the microbiome.

2.4.3 Analysis of sequencing data

Irrespective of the sequencing method or instrument utilized, bioinformatics analysis of sequencing data is required to provide any meaningful interpretation of the sequences (Figure 3).

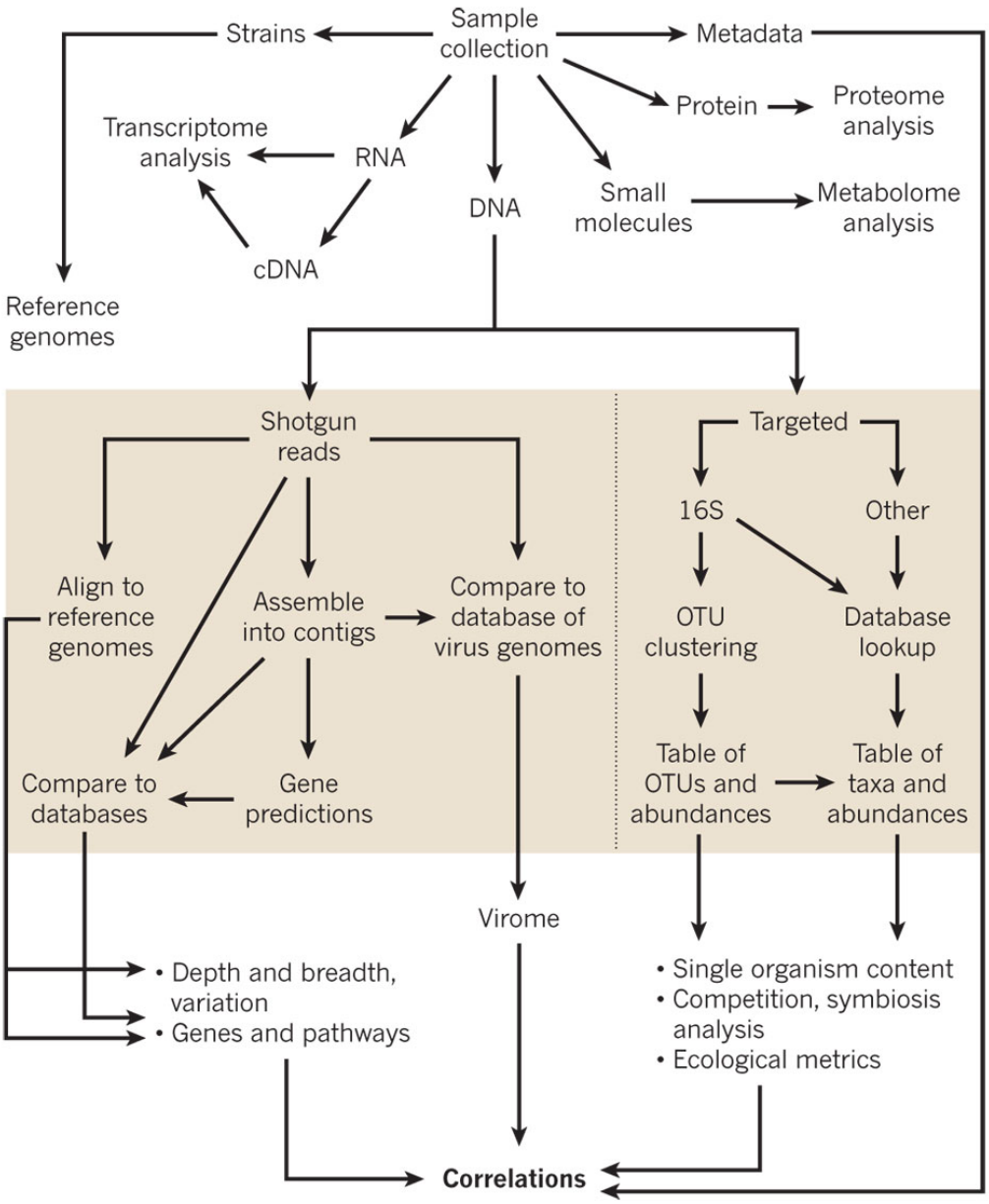

Figure 3. Schematic representation of a microbiome analysis study. Used with permission from Weinstock, (2012).

Targeted amplicon sequencing analysis

Numerous computational tools or software to perform 16S data analysis are available, though QIIME (Caporaso et al., 2010) and mothur (Schloss et al., 2009) are most commonly used. The

analysis of 16S data can be divided into three phases (Weinstock, 2012). The first phase of processing sequencing data is quality control of the assembled reads (contigs); this includes filtering contigs based on quality scores, contig length, chimaerism, presence of homopolymers and ambiguous base calls.

The second phase is operational taxonomic unit (OTU) picking and taxonomic assignment. Contigs are clustered into OTUs depending on their sequence identity where a 97% similarity cutoff is thought to correspond to the same species (Stackebrandt and Goebel, 1994). There are a number of OTU picking strategies including de novo, closed or open reference. De novo OTU picking involves clustering contigs (into OTUs) based on shared sequence similarity. Closed reference utilizes a reference database and contigs not matched are removed whereas open reference similarly utilizes a reference database, however, contigs not matched are then clustered de novo. Thereafter, OTUs are taxonomically classified using one of several curated reference databases including greengenes (DeSantis et al., 2006b), SILVA (Pruesse et al., 2007) or RDP (Wang et al., 2007). Taxonomic classification is the method of predicting the taxonomical positions from a particular community of interest. Classification can be applied to sequences and/or OTUs. Most classification approaches apply one of three strategies to identify similar sequences: i) pairwise alignment (i.e. BLAST) to a reference database with known taxonomies (Altschul et al., 1997); ii) incorporation into an existing multiple alignment of reference sequences (DeSantis et al., 2006a); or iii) using *n*-mer frequencies compared to reference sequences (Rosen et al., 2008). Classification is next based on sequence similarities in a subset of sequences via a number of algorithms including nearest neighbour (Arya et al., 1998), lowest common ancestor (Huson et al., 2007) or Naïve Bayes Classifier (Wang et al., 2007). An

alternative to analyzing OTUs is to analyze sequences that are the same genus, or other taxonomic rank.

The third phase of analysis uses the above data to produce representations of community similarities, trees, abundance curves, other statistical descriptors of community structure or inferred functionality. Such analyses can be executed in a number of additional software packages including phyloseq (McMurdie and Holmes, 2013), vegan (Dixon, 2003), PICRUSt (Langille et al., 2013) and others.

Data from targeted-amplicon sequencing presents with intrinsic limitations to ecological and statistical interpretation; library sizes are quite variable and microbiota data are inflated with many zeros. Generally, microbiota analyses function to compare relative abundances of taxa in the ecological niche of two or more groups and as such, we can only measure taxon abundances obtained from specimens of such niches. As the comparison between specimen taxon abundances is not equivalent to taxon abundances in the ecological niche we are faced with difficulties in interpretation. Moreover, as microbiota analyses are based upon compositional data (i.e. sum to 1) many standard analyses methods are not appropriate. Thus, microbiota analyses are constrained by normalization and differential abundance testing methods.

Shotgun metagenomics sequencing analysis

Subsequent to quality control filtering, three general analytical strategies are used to determine the taxonomic profile, phylogenetic lineage or genomes present in a metagenome (Sharpton, 2014). First, marker gene analyses compare each read to a reference database containing

taxonomically or phylogenetically informative sequences. Classification algorithms are employed to determine if reads are homologous to a marker gene and are annotated based on similarity. Second, binning metagenomes can be performed via compositional binning (sequence composition are used to classify and/or cluster metagenomes into taxonomic groups), similarity binning (reads are classified based on similarity to previously identified genes) or fragment recruitment (alignment of reads to nearly identical genome sequences to produce estimates of metagenomic coverage). Third, assembly of reads can be performed whereby overlapping reads are merged to construct contigs. A functional profile can also be developed through application of gene prediction and functional annotation pipelines (Sharpton, 2014). Briefly, subsequences in a metagenomic read that encodes proteins can be identified via utilization of several algorithms and each predicted protein can be functionally annotated by comparison to curated databases of proteins and protein families. Peptides that are found to be homologs of a protein or protein family are annotated with the corresponding function. Thus, this approach allows for the characterization of both taxonomic and functional structure of the metagenome.

Measures of biodiversity

Several useful statistical indices have been adapted for use on sequencing data to assist in extracting instructive information from massive ecological datasets. Indices measure an aspect of biological diversity and can be categorized into alpha or beta diversity measures.

In community profiling studies, alpha diversity indices measure a number of different aspects of biodiversity including richness, evenness, diversity or coverage of a sample. Richness is the simplest and most intuitive concept for characterizing ecological diversity; it is defined as the count of the different taxa, or OTUs, present in a community. Chao1 (Chao, 1984) and ACE

(abundance-based coverage estimator; Chao and Lee, 1992) nonparametric indices are commonly used and measure slightly different dynamics of species richness – Chao1 incorporates the frequencies of singletons and doubletons (i.e. taxa observed once and twice) to estimate minimum richness whereas ACE takes abundant taxa into consideration. These indices have inherent limitations, specifically pertaining to underestimating richness: they do not account for unobserved OTUs, bias from 16S rRNA copy number, sequencing artefacts and outliers. Diversity indices take both richness and evenness (i.e. how equal OTUs are distributed in a community) into consideration. Shannon (Shannon, 1948) and Simpson (Simpson, 1949) diversity indices are commonly used. A key dissimilarity between these diversity indices is that the Shannon index puts more weight on rare taxa versus the Simpson index, which puts more weight on dominant species. Often, the inverse of Simpson is reported such that an increased measure indicates more even taxon abundances.

In contrast to alpha diversity that measures the ecological profile within a sample, beta diversity is defined as the degree of community differentiation between two or more microbial communities or samples (Whittaker, 1960). Beta diversity measurements can take into consideration microbial membership (i.e. OTUs present/absent) and microbial structure (i.e. abundance of OTUs and/or their phylogenetic relationship). A number of beta diversity measures are in routine use including Jaccard (Jaccard, 1912), Sørenson (Sørensen, 1948), Morisita-Horn (Morisita, 1959), UniFrac (Lozupone and Knight, 2005) and, the Bray-Curtis index (Bray and Curtis, 1957). These indices can be implemented in multivariate statistics such as principal coordinate analysis (PCoA) through multidimensional scaling (and ordination techniques) as a means of visualizing the similarity or dissimilarity among communities.

2.4.4 Importance of standardization

To achieve an accurate representation of microbial community profiles sequenced, a number of factors must be considered. In the last few years many studies have investigated the effect that study design factors may have on 16S amplicon data. These include type, collection and storage of samples, DNA extraction, primer selection and sequencing platform.

Sample considerations

Gut microbiome studies in particular are highly influenced by the sample type utilized. Stool is often used as a proxy to investigate the microbial structure and function of the gastrointestinal tract, the advantage being the simple, effective and non-invasive route of collection. Though the abundance of most mucosal bacteria are highly similar throughout the ileum to the rectum in health and disease (Eckburg et al., 2005; Lepage et al., 2005) bacteria inhabiting mucosal surfaces of the gastrointestinal tract are extensively different from those present in the lumen. Eckburg et al., (2005) compared the fecal and mucosal microbiotas; the fecal microbiota was comprised of less Bacteroidetes and more Firmicutes and Proteobacteria. While these differences between the microbiota structures might be a result of the flushing action of beverages and gastric juices, it's likely an outcome of antimicrobial peptides secreted by the intestinal mucosa. Distinct bacterial species are known to preferentially colonize the mucosa including some belonging to the genera *Escherichia* (Darfeuille-Michaud et al., 2004)*, Fusobacterium* (Li et al., 2016) and *Akkermansia* (Reunanen et al., 2015). Stool sample collection in itself can introduce bias to sequencing data primarily as a result of transporting the sample to the laboratory for processing (distance, temperature, transport regulations). Accordingly, the storage conditions

(temperature, preservatives, length) of stool samples can affect the integrity of nucleic acids (Choo et al., 2015).

DNA extraction

Numerous studies have evaluated the influence of nucleic acid extraction methods. Several DNA extraction kits are available that contain premade buffers, materials and protocols for the disruption of cellular membranes, protein denaturation and nucleic acid purification. Though commercial kits generally contain similar components such as guanidine-based chaotropic salts or silica-adsorption spin-columns, variability exists in buffer or enzyme composition and method of cell lysis (i.e. mechanical lysis such as bead-beating or chemical lysis). In general, the lysis step receives the most scrutiny, as intensity can bias the microbial community towards a particular taxonomic group. Commercial kits can be used with modifications to the protocol or manual extraction methods can be developed in-house further enhancing methods of nucleic acid extraction variability. Nonetheless, the choice of extraction method can profoundly impact on both the yield and microbial community profile via failure to recover specific taxa (Knudsen et al., 2016).

Primer selection

As discussed in section 2.4.2.2 the vast majority of amplicon-based sequencing targets the 16S rRNA gene for amplification. The choice of primer pair and thus the selected variable region(s) to be analyzed is in part dictated by the sequencing platform; 2×800 bp reads can cover the entire length of the 16S rRNA gene by Sanger sequencing, while NGS platforms are more apt for shorter read lengths (i.e. 2×300 bp for Illumina) and hence typically one to three variable regions

are sequenced. Selected PCR primers should be taxonomically and phylogenetically informative and generate appropriate specificity to taxonomic groups of interest. As an example, specific primers (and variable regions) are more apt to amplify particular taxonomic groups than others; F27–R338 is highly specific to bacteria but poorly amplifies Verrucomicrobia and *Bifidobacterium* (Hayashi et al., 2004). Currently, amplification of the V4 region and Illumina sequencing is considered the gold standard for many environmental sample types.

2.5 Gut microbiome

Knowledge of the role of microbes that inhabit mucosal surfaces of the human body and their role in health and disease is increasing at a remarkable rate. The term "microbiota" refers to the population of microbes at a particular anatomical niche and "microbiome" refers to the collective genes encoded by all microbes of that particular niche. The human gastrointestinal tract comprises approximately 10^{14} bacterial microbes and amounts to a biomass of approximately 2 kg (Qin et al., 2010). It has been assumed for some time that there are roughly 10 times as many microbes as there are eukaryotic cells in the human body, but recent support proposes the ratio for microbial to human cells is 1.3:1 (Abbott, 2016). The bacteria of the gut belong to more than 1000 different species encoding more than 3 million bacterial genes (metagenome) exceeding the number encoded by the human genome by 150–fold (Qin et al., 2010).

The study of the gut microbiome cannot be separated from its environmental context; host genetics, nutrition, the environment, geographical location, early microbial exposures and other factors profoundly impact the microbiome of the healthy human gut. Extensive sampling of the human microbiome has revealed that although the diversity and abundance of microbial communities of particular niches vary widely both within and among individuals (Huttenhower

et al., 2012), the functional repertoire of the healthy microbiome, including the gut, is relatively stable regardless of the microbe composition. A microbiome becomes more stable in adulthood, which subsequently changes in the elderly (Claesson et al., 2012). However, there are temporal variations within a healthy individual's microbiome that can occur even over a short period of time (David et al., 2014).

The healthy human gut is dominated by the presence of four bacterial phyla: Bacteroidetes, Firmicutes, Actinobacteria and Proteobacteria, with Bacteroidetes and Firmicutes accounting for a large majority of endemic bacteria in the gut (Tap et al., 2009). Within the healthy human gut, the phylum Firmicutes are divided into two major classes of Gram-positive bacteria: Bacilli and Clostridia (primarily *Clostridium* cluster IV and *Clostridium* XIVa). The Bacteroidetes are Gram-negative bacteria, of which the *Bacteroides* represents one of the most abundant genera in the gut (Huttenhower et al., 2012). Considerable diversity of bacterial species exist within a normal gut; recent studies have recognized common core subsets within the microbiome that are relatively stable throughout large populations and that can even persist in an individuals' gut for their entire adult life (Faith et al., 2013).

There is a clear spatial distribution of microbes within the gastrointestinal tract with diversity increasing from the stomach to the colon, and it is at the terminal ileum where prevalent species change from aerobes to anaerobes (Sekirov et al., 2010). Within the gut itself, there is a significant difference in microbial populations on mucosal surfaces compared to within the lumen (Li et al., 2015). Microbes at the mucosal surface are in closer proximity to the intestinal epithelium and may have a greater influence on the immune system whereas luminal/fecal

microbes might be more essential for energy and metabolic interactions. This is relevant since many studies of the gut microbiota use fecal material for community profiling and thus may not adequately reflect the totality of viable microbes within the gut.

Under normal physiological conditions, the human gut microbiota is a homeostatic ecosystem with several vital functions and interrelationships important to host health including food digestion, development of the host immune system and intestinal epithelial barrier and protection against pathogens (Shreiner et al., 2015). Disruption of this equilibrium can result in dysbiosis and increase risk of disease. Gut dysbiosis refers to an altered composition of intestinal microbial populations and is thought to provide continuous immunological stimulation leading to immune response anomalies in numerous IMID. Altered community composition has been established in a number of gastrointestinal diseases: IBD, celiac disease, irritable bowel syndrome, functional dyspepsia, antibiotic-associated diarrhea, tropical enteropathy and others (Brown et al., 2015; D'Argenio et al., 2016; Distrutti et al., 2016; Keely et al., 2015; Larcombe et al., 2016). Accumulating evidence proposes that dysbiosis of the intestinal microbiota is not limited to gastrointestinal diseases thereby suggesting that gut bacteria can affect the systemic immunological response. A number of studies have investigated gut dysbiosis in relation to obesity, diabetes, chronic periodontitis, vaginosis, atopic diseases, nonalcoholic steatohepatitis, Alzheimer's disease, and others (Adams and Morrison, 2016; Blasco-baque et al., 2016; Daulatzai, 2014; Johnson and Ownby, 2016). While a breakdown in the equilibrium of the intestinal milieu may be widely recognized, perturbations surrounding biological mechanisms driving dysbiosis exist and it is unclear whether dysbiosis manifests as a cause or consequence of disease.

2.5.1 Gut microbiome in immune-mediated inflammatory disease

2.5.1.1 Inflammatory bowel disease

At present, not only have numerous candidate pathogens (adherent invasive *E. coli* (AIEC), *Bacteroides fragilis, Clostridium difficile, Mycobacterium avium paratuberculosis, Yersinia enterocolitica, Listeria monocytogenes, Salmonella, Campylobacter*) been implicated in disease onset and/or perpetuation (Becker et al., 2015), but over 400 complex microbiome surveys have been conducted to identify the totality of microbes in the IBD gut and precisely how the IBD gut differs from that of a healthy person. Indeed, advances in culture-independent sequencing and bioinformatics analyses in recent years have provided tremendous insight into the structure and function of the gut microbiota. Though patterns of gut dysbiosis are inconsistent among published studies, partly owing to specimen type (stool versus biopsy) or analysis methods, studies surveying the gut microbiome in persons with IBD have consistently demonstrated perturbations of the structure and function of the microbiome.

Studies frequently document an overall reduction of diversity, the total number of species in a community. In fact, data from the MetaHIT consortium suggest that persons with IBD harbour on average 25% fewer microbial genes than healthy persons (Qin et al., 2010). Diversity is reduced in the fecal and mucosal microbiomes of IBD (Lepage et al., 2011; Manichanh et al., 2006) and has also been documented among monozygotic twins discordant for CD (Dicksved et al., 2008). Decreased diversity has been attributed to shifts in the abundance of the Firmicutes, and more specifically the *C. leptum* and *C. coccoides* group (Manichanh et al., 2006). Likewise, utilizing a custom phylogenetic microarray Kang et al., (2010) reported some bacteria belonging to the Firmicutes phylum including *Eubacterium rectale* of the Lachnospiraceae and

Ruminococcus albus, *R. callidus*, *R. bromii* and *F. prausnitzii* of the Ruminococcaceae were 5-10–fold more abundant in healthy persons compared to CD.

The gut microbiome was recently characterized according to the three most dominant phenotypes of CD: inflammation of the colon, the terminal ileum, or both (Naftali et al., 2016). The authors demonstrated ileal and colonic CD sustain distinct microbiome patterns. For example, ileal CD samples were richer in *Escherichia* (Enterobacteriaceae) and disease activity correlated with *Fusobacterium* abundance whereas colonic CD had higher levels of *Faecalibacterium* and two unidentified genera of the Clostridiales and Ruminococcaceae. The variance in CD microbiome patterns suggests that different mechanisms might underlie the two major clinical manifestations of CD. Recently, Meij et al., (2015) described the gut microbiome composition in active new-onset pediatric CD compared with 3 months following initiation of treatment. At baseline, a greater diversity of Proteobacteria and a lower abundance of Bacteroidetes were observed in CD compared to controls. Upon clinical remission, the microbiota profile of CD seemed to shift to a profile similar to the control population.

The largest study to date related to the mucosal-associated microbiota and new-onset CD conducted by Gevers et al., (2014) included 468 children and adolescents (<17 years) with newly diagnosed CD and a control group of 229 persons with non-inflammatory conditions of the gastrointestinal tract. Mucosal and fecal samples were obtained prior to the initiation of treatment in subjects with a spectrum of disease phenotypes based on severity, location (terminal ileum and rectum) and behaviour of disease. Overall, disease status strongly correlated to increases in the abundance of the Enterobacteriaceae, Pasteurellaceae, Veillonellaceae and Fusobacteriaceae and

a decrease in the abundance of Erysipelotrichales, Bacteroidales and Clostridiales. Importantly, the alterations described were more strongly observed in tissue samples with weaker correlations observed in stool, implying a less dramatic shift in the luminal microbiota regardless of disease presence. The authors also demonstrated that microbiota profiles at time of diagnosis could be predictive for future disease outcomes; abundant levels of *Fusobacterium* and *Haemophilus* positively correlated with the pediatric CD activity index whereas levels of Enterobacteriaceae were negatively correlated. Furthermore, the microbiome composition between terminal ileal and rectal mucosal specimens were determined to be less different than differences observed between tissue and stool. Our group has recently reported similar observations of a somewhat homogenous gastrointestinal tract (using mucosal biopsies from the ileum, cecum, mid-colon and rectum) in terms of microbe content and abundance within CD and UC cohorts (Forbes et al., 2016). We also explored the presence and absence of inflammation as an influencing factor on microbiome composition; shifts in microbe abundance comparing the inflamed mucosa between CD and UC were observed but these shifts were more dramatic in the non-inflamed mucosa between CD and UC (see section 3.2).

Bacteroides is the most dominant genus in Western microbiotas and can both positively and negatively affect the host (Wexler, 2007). Generally, while the overall abundance of the order Bacteroidales is increased in IBD, in certain circumstances particular species may be reduced; *Parabacteroides distasonis* is significantly decreased in inflamed IBD mucosa (Zitomersky et al., 2013). Pathogenic bacteria including *E. coli, and Shigella, and others such as Rhodococcus* and *Stenotrophomonas maltophilia* are increasingly observed in IBD (Guerrero et al., 2015; Jensen et al., 2015; Knösel et al., 2009). Other pathobionts with potential roles in the disease

course include *Prevotellaceae, C. difficile, Klebsiella pneumoniae, Proteus mirabilis* and *Helicobacter hepaticus*.

The relative abundance of the Enterobacteriaceae in persons with IBD (Kolho et al., 2015) and mouse models of IBD (Nagao-Kitamoto et al., 2016) is increased. Much research has focused on the role of *E. coli*, specifically AIEC, in IBD etiology; AIEC strains have been isolated from ileal-involving CD tissue (Darfeuille-Michaud et al., 2004) and the genera *Escherichia* and *Shigella* (indistinguishable by 16S analysis) are highly enriched in patients with IBD. the enrichment of this phylotype in IBD is more pronounced in mucosal samples versus stool samples (Petersen et al., 2015). It has also recently been demonstrated that a Western diet induces a shift in microbiome composition increasing the susceptibility to AIEC infection (Agus et al., 2016). AIEC possess the capability to invade enterocytes, replicate within macrophages and induce granuloma formation in vitro (Martinez-Medina and Garcia-Gil, 2014), and though AIEC strains may be present within a healthy gut, they do not have the ability to adhere to ileal enterocytes. Therefore, colonization of the inflamed ileum provides a preferential environment for these microbes to influence the disease course. It has also been shown that the anti-inflammatory drug, mesalamine, used for IBD therapy, decreases gut inflammation and correlates with a reduction of *Escherichia/Shigella* abundance (Morgan et al., 2012a).

Fusobacterium species are members of the normal oral and gut microbiota in humans; however, particular species (adherent, invasive, pro-inflammatory) are recognized as opportunistic pathogens in CD and UC. The relative abundance of *Fusobacterium* is increased in mouse models and the colonic mucosa of IBD; Strauss et al., (2011) isolated *Fusobacterium* species

from 63.6% of patients with IBD versus 26.5% of healthy controls. Moreover, 69% of all *Fusobacterium* isolated from IBD patients were identified as *F. nucleatum* and strains isolated from the IBD inflamed mucosa were more invasive in a Caco-2 cell invasion assay than strains isolated from the healthy mucosa of IBD or the control group (Strauss et al., 2011). *Fusobacterium* species have also been linked to colorectal carcinoma (Mima et al., 2015) and interestingly, IBD is a recognized risk factor for colorectal carcinoma (Lutgens et al., 2015).

The reduction of particular commensal microbes and a concomitant loss of their protective function possibly have a substantial impact on the course of disease. The ability of commensal bacteria to produce SCFA including butyrate, acetate and propionate via dietary fiber fermentation is a key benefit to the human gut. SCFA are a primary source of energy for colonic epithelial cells and have recently been shown to mediate homeostasis of colonic regulatory T cell populations (Smith et al., 2013). The abundance of *Faecalibacterium* of the Ruminococcaceae family is reduced in IBD (Lopez-Siles et al., 2015). *F. prausnitzii* is an anti-inflammatory commensal bacterium and produces the SCFA butyrate. *F. prausnitzii* administration in mouse models has been shown to reduce inflammation and secondly, a reduction in the abundance of *F. prausnitzii* has been linked to postoperative disease recurrence in ileal CD (Sokol et al., 2008). The abundance of other SCFA producers including the Leuconostocaceae, *Odoribacter splanchus*, *Phascolarctobacterium* and *Roseburia* are reduced in IBD (Machiels et al., 2013; Morgan et al., 2012a). *Roseburia hominis*, in particular, inversely correlates with UC disease activity and duration (Machiels et al., 2013). This species has been described as an acetate user and butyrate producer and is therefore dependent on other SCFA producers (Duncan et al., 2002). Alterations of these bacterial populations and concurrent variability of SCFA production

may have profound consequences on host regulatory mechanisms of inflammation, though studies haven't established causality.

2.5.1.2 Multiple sclerosis

Recent research exploring the role of the gut microbiome in humans and experimental autoimmune encephalomyelitis (EAE), an animal model of MS, suggests a relationship between the gut microbes and the development of MS.

2.5.1.2.1 Gut-brain axis

Insights into gut-brain crosstalk have revealed a multifaceted communication network (Figure 4). The gut-brain axis (GBA) ensures the proper maintenance of gastrointestinal homeostasis but also likely demonstrates manifold consequences on affect, motivation and higher cognitive functions as mediated through a bidirectional interaction (Rhee et al., 2009). The role of the GBA is to connect the emotional and cognitive centers of the brain to peripheral intestinal functions and mechanisms including intestinal barrier function, immune activation and entero-endocrine signaling and additionally, to monitor and integrate gastrointestinal functions. GBA communication is facilitated via the involvement of neuro-immuno-endocrine mediators (Rhee et al., 2009).

The complex communication network of the GBA comprises the CNS, autonomic nervous system (ANS), ENS and the hypothalamic pituitary adrenal (HPA) axis. The sympathetic and parasympathetic limbs of the ANS facilitate both afferent (from lumen via enteric, spinal and vagal pathways to the CNS) and efferent (from CNS to intestinal wall) signals. The HPA axis is a part of the limbic system that also includes the amygdala, hippocampus and hypothalamus. In particular, the limbic system is an essential region of the brain intricate to memory and emotional

responses and the HPA axis is believed to function as the core stress efferent axis that directs adaptive responses to stressors (Tsigos and Chrousos, 2002). The limbic system is activated through external stressors and increased systemic pro-inflammatory cytokines. Corticotropin-releasing factor (CRF) is secreted from the hypothalamus, which in turn stimulates the pituitary gland to secrete adrenocorticotropic hormone (ACTH) resulting in the release of a major stress hormone, cortisol, from the adrenal glands. The CNS communicates along both afferent and efferent autonomic pathways in parallel, with several intestinal targets including the ENS, muscle layers and gut mucosa ultimately influencing motility, immunity, permeability and mucus production. Recently, the role of the gut microbiota has been recognized as being a part of the GBA as the gut microbiota also demonstrates a bidirectional interaction with the same intestinal targets (Mayer et al., 2014); therefore, the gut microbiota is being modulated by brain-gut interaction and can itself influence gastrointestinal function. In fact, not only does clinical and experimental evidence support the role of the gut microbiota to heavily impact the GBA through localized interaction with intestinal epithelial cells and the ENS, but evidence also suggests that the gut microbiota interacts with the CNS via metabolic and neuroendocrine pathways.

The current perception of the GBA implies that the gastrointestinal tract and associated factors (for example, ingested food) are likely to be involved in neurological and/or neuropsychiatric disorders. A seminal observation from more than two decades ago provided compelling evidence for the association between the gut microbiota and brain. Specifically, following administration of oral antibiotics to persons with hepatic encephalopathy, dramatic clinical improvements were detected (Morgan, 1991). Evidence of gut microbiota involvement is beginning to unfold in

conditions such as MS, Parkinson's disease, schizophrenia, autism, stress, migraines and others (Sasdelli et al., 2016). Moreover, it is recognized that many gastrointestinal ailments are associated with a risk to develop neurological comorbidities and vice versa.

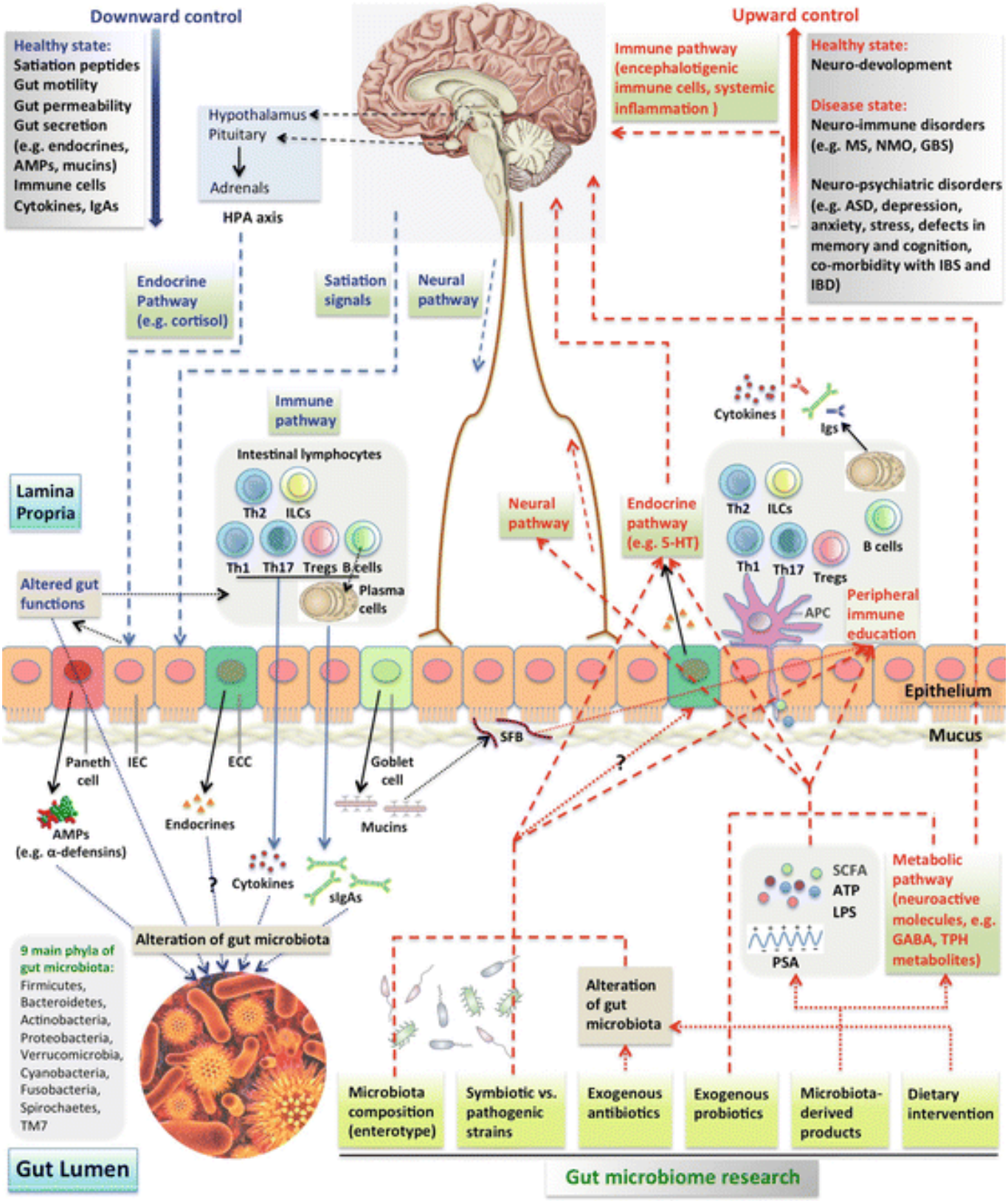
Downward control
Healthy state:
Satiation peptides
Gut motility
Gut permeability
Gut secretion
(e.g. endocrines, AMPs, mucins)
Immune cells
Cytokines, IgAs
Hypothalamus
Pituitary
Adrenals
HPA axis
Endocrine Pathway (e.g. cortisol)
Satiation signals
Neural pathway
Upward control
Immune pathway (encephalotigenic immune cells, systemic inflammation)
Healthy state:
Neuro-devolopment
Disease state:
Neuro-immune disorders (e.g. MS, NMO, GBS)
Neuro-psychiatric disorders (e.g. ASD, depression, anxiety, stress, defects in memory and cognition, co-morbidity with IBS and IBD)
Immune pathway
Intestinal lymphocytes
Th2
ILCs
Th1
Th17
Tregs
B cells
Plasma cells
Lamina Propria
Altered gut functions
Neural pathway
Endocrine pathway (e.g. 5-HT)
Cytokines
Igs
Th2
ILCs
Th1
Th17
Tregs
B cells
APC
Peripheral immune education
Epithelium
Mucus
SFB
Paneth cell
IEC
ECC
Goblet cell
AMPs (e.g. α-defensins)
Endocrines
Mucins
Cytokines
sIgAs
Alteration of gut microbiota
9 main phyla of gut microbiota:
Firmicutes,
Bacteroidetes,
Actinobacteria,
Proteobacteria,
Verrucomicrobia,
Cyanobacteria,
Fusobacteria,
Spirochaetes,
TM7
Gut Lumen
SCFA
ATP
LPS
PSA
Metabolic pathway (neuroactive molecules, e.g. GABA, TPH metabolites)
Alteration of gut microbiota
Microbiota composition (enterotype)
Symbiotic vs. pathogenic strains
Exogenous antibiotics
Exogenous probiotics
Microbiota-derived products
Dietary intervention
Gut microbiome research

Figure 4. Microbiota–gut–brain axis in relation to CNS disorders. Multiple pathways guide the downward and upward directions of the microbiota–gut–brain axis in the contexts of health and disease. On the left (*blue arrows*): downwardly, CNS controls gut microbiota composition through satiation signaling peptides that affect nutrient availability and endocrines that affect gut functions and neural pathways. HPA axis release of cortisol regulates gut movement and integrity. Immune (cells, cytokines, and IgAs) pathways can be turned on in response to altered gut functions. Endocrine and neural pathways can also regulate the secretion from specialized gut epithelial cells, including paneth cells, enteroendocrine cells (ECC), and goblet cells. Their secretory products affect the survival and the resident environment of microbiota. On the right (*red arrows*): Upwardly, gut microbiota controls CNS activities through neural (direct activation of neurons by microbiota), endocrine (e.g. ECC release of 5-HT), metabolic (microbiota synthesis of neuroactive molecules), and immune (CNS infiltrating immune cells and systemic inflammation) pathways. Microbiota influences CNS at healthy (neurodevelopment) and disease (a range of neuro-immune and neuro-psychiatric disorders) states. Gut luminal microbiota, their products sampled by APCs, and epithelium-attaching SFBs mediate peripheral immune education. *Abbreviations*: *AMPs* antimicrobial peptides, *TPH* tryptophan, *5-HT* 5-hydroxytryptamine, *SFB* segmented filamentous bacteria, *PSA* polysaccharide A from *B. fragilis*, *ATP* adenosine-5′-triphosphate, *SCFA* short-chain fatty acid, *IEC* intestinal epithelial cell, *ILCs* innate lymphoid cells, *APC* antigen-presenting cell, *MS* multiple sclerosis, *NMO* neuromyelitis optica, *GBS* Guillain-Barré syndrome, *ASD* autism spectrum disorder, *IBS* irritable bowel syndrome, *IBD* inflammatory bowel disease. Used with permission from Wang and Kasper, (2014).

2.5.1.2.2 Dysbiosis evidence and characteristics

Research investigating the role of the gut microbiome in EAE suggests a relationship between the gut microbes and development of MS. Ochoa-Repáraz et al., (2009) utilized a cocktail of oral antibiotics (ampicillin, vancomycin, neomycin sulfate and metronidazole) to disrupt the gut microbiome of mice prior to induction of EAE. The authors reported that oral antibiotic treatment significantly decreased gut bacterial populations and reduced the onset and severity of EAE. The reduced severity of EAE was ascribed to decreased levels of pro-inflammatory cytokines and chemokines and elevated levels of anti-inflammatory cytokines including IL-10

and IL-13; IL-10-producing Foxp3$^+$ Treg cells were shown to accumulate in mesenteric and cervical lymph nodes and seemed to mediate the severity. Additionally, adoptive transfer of Foxp3$^+$ Treg cells to naïve recipients conferred protection against disease. Furthermore, disease was observed among antibiotic treated mice with reduced CD25$^+$ T cells. In contrast, in the absence of antibiotics, disease was exacerbated in CD25$^+$-deficient mice. These findings imply that regulatory populations other than Th1, Th17 or Treg cells may confer protection against disease. An additional study in mice treated with oral antibiotics suggests a role for IL-10 producing CD19$^+$ B cells (Ochoa-Repáraz et al., 2010a). Adoptive transfer from antibiotic treated mice to naïve recipient mice reduced severity of disease; this protective effect was attributed to shifting the immune response from a pro-inflammatory Th1/Th17 response to an anti-inflammatory Th2 response.

Lee et al., (2011) demonstrated that disease protection in EAE is associated with reduced levels of the pro-inflammatory cytokines IFN-γ and IL-17A, both of which play critical roles in the development of EAE. Increased Foxp3$^+$ Treg cells in the peripheral lymphoid tissues and the CNS were observed in germ-free mice. The authors further reported that SFB monocolonization of mice restored disease and examination of CNS tissues showed enhanced invasion by Th1 and Th17 cell populations. Using a relapsing-remitting mouse model (SJL/J) of spontaneously developing EAE, Berer et al., (2011) reported that the commensal gut microbiota is essential in triggering immunological processes that lead to relapsing-remitting autoimmune disease and that myelin-specific CD4+ T cells were driving factors of autoimmunity. Additionally, the activation of autoantibody-producing B cells was shown to be dependent upon the availability of the myelin autoantigen (specifically, MOG) and the gut microbiota.

The ability of probiotics to modify the clinical outcome of murine EAE has also been investigated. Prophylactic treatment with *Lactobacillus paracasei* and *L. plantarum* DSM 15312 was shown to induce Treg cells in mesenteric lymph nodes and enhance production of TGFβ1 whereas increased serum IL-27 levels were observed with *L. plantarum* DSM 15313 (Lavasani et al., 2010). Interestingly, administration of each strain independently was therapeutically unsuccessful whereas combinatorial administration suppressed disease perpetuation and reversed clinical signs of EAE. Disease amelioration was attributed to attenuation of the pro-inflammatory Th1/Th17 response and induction of Treg cells through elevated IL-10 levels. Comparably, an additional probiotic, *Bifidobacterium animalis* demonstrated similar suppressive effects (Ezendam et al., 2008). It has also been reported that some probiotic bacteria, notably *L. casei* Shirota, enhanced the duration of EAE (Ezendam and van Loveren, 2008); however, this was refuted in a subsequent study that determined neither *L. casei* Shirota nor *B. brevis* strain Yakult exacerbates EAE (Kobayashi et al., 2010). Prophylactic administration of five probiotics (IRT5; *L. casei*, *L. acidophilus*, *L. reuteni*, *B. bifidum*, and *Streptococcus thermophilus*) prior to disease induction significantly suppressed development of EAE (Kwon et al., 2013) whereas treatment with IRT5 to EAE delayed disease onset. Prophylactic and therapeutic efficacy of IRT5 probiotics was ascribed to inhibition of Th1 and Th17 responses and induction of IL-10-producing and/or $Foxp3^+$ Treg cells.

To further explore how the microbiome affects EAE susceptibility and severity, the role of particular commensals or their products was considered. Ochoa-Repáraz et al., (2010b) investigated oral treatment of mice with polysaccharide A (PSA) of *B. fragilis.* In PSA treated mice, EAE severity was decreased relative to untreated mice and decreased numbers of Th1 and

Th17 cells were reported in the EAE brains of PSA-treated mice. Moreover, lymphocytes isolated from the cervical lymph nodes of PSA-treated mice reportedly produced more IL-10 and less IL-17 and IFN-γ when stimulated with myelin oligodendrocyte glycoprotein (induction of EAE). Phosphorylated dihydroceramides, lipids derived from the oral commensal *P. gingivalis*, demonstrated enhanced EAE severity through induction of DC IL-6 secretion in a TLR2-dependent manner and a subsequent decrease in $Foxp3^+$ Treg cell populations (Nichols et al., 2009). A recent study demonstrated oral administration of a *Lactococcus lactis* strain that produces heat shock protein-65 prevented development of murine EAE; the suppressive effect was ascribed to elevated IL-10 and reduced IL-17 production (Rezende et al., 2013).

Little is known about what impact the gut microbiome may have on human MS etiopathogenesis; gut microbiome studies in persons with MS are mostly limited to case-control examinations. Mowry et al., (2012) reported RRMS patients to have differing levels of Firmicutes, Bacteroidetes and Proteobacteria, and that shifts in the abundance of Firmicutes was more dramatic in patients treated with glatiramer acetate, an immunomodulator drug. Comparatively, vitamin D treatment was associated with changes in Firmicutes, Actinobacteria and Proteobacteria in RRMS patients, and increases in Enterobacteria in both RRMS and healthy controls. Vitamin D has been a main suspect in MS etiology for some time (Holmøy and Torkildsen, 2016). The beneficial aspects of vitamin D in the treatment of MS including exerting anti-inflammatory effects, possible influence on remyelination and its link to disease activity is widely accepted and is actively used as a therapeutic (Jarrett et al., 2014). The mechanism of benefit is unknown but it is plausible that it has an impact on the gut microbiome. Additional analysis by the same group revealed that although overlap of bacterial communities has been

observed. Particular taxa including *Faecalibacterium* were reported to be lower in persons with MS (Cantarel et al., 2015). Of note, *Faecalibacterium* is considered to be immunosuppressive through production of the SCFA butyrate. Glatiramer acetate-treated patients showed differences in microbe composition including Bacteroidaceae, *Faecalibacterium, Ruminococcus,* Lactobacillaceae, *Clostridium* and additional Clostridiales compared to untreated persons. After vitamin D supplementation, untreated persons with MS had an increase in *Akkermansia*, *Faecalibacterium* and *Coprococcus.* Although these studies are exploratory and future studies in larger cohorts are needed to confirm these findings, these results suggest that the microbial communities of the MS gut are subject to modulation by some drugs commonly used in the treatment of MS. Jhangi et al., (2014) found a significant increase in the highly immunogenic Archaea *Methanobrevibacter smithii*, which might play a role in inflammation due to their structural components. *Butyricimonas* and Lachnospiraceae, which are butyrate producers with anti-inflammatory properties, were lower in untreated MS. Tremlett et al., (2015) explored gut microbiome profiles in pediatric MS; increases in *Escherichia, Shigella* and *Clostridium* were observed, and decreases in *Eubacterium rectale* and *Corynebacterium.* Recently, the association between gut community profiles in 15 pediatric RRMS patients and host immunological markers such as IFN-γ, IL-17, IL-10, $CD4^+CD25^+Foxp3^+$ Treg were explored (Tremlett et al., 2016a). The authors reported immune markers did not differ between RRMS and controls but did observe divergence in the association between the gut microbiome and immunological markers. $IL\text{-}17^+$ T cells positively correlated with overall richness and evenness whereas an inverse correlation between $IL\text{-}17^+$ T cells and Bacteroidetes abundance in MS was reported. Another study reporting on the association between the gut microbiota composition and relapse risk in pediatric MS showed a decrease in Fusobacteria, increase in Firmicutes and the presence of the

Archaea Euryarchaeota was associated with a shorter time to relapse (Tremlett et al., 2016b). Thus far, current studies indicate that gut dysbiosis is linked with MS. Common findings include a reduction in the phyla Bacteroidetes and Firmicutes, which are important for immunoregulation and production of SCFA.

2.5.1.3 Rheumatoid arthritis

Gut microbes have been recognized to be an autoimmune trigger in RA for several decades. In recent years, research has investigated the role of the gut microbiome in humans and experimental animal models of RA and thus suggests a critical role in disease.

2.5.1.3.1 Gut-joint axis

Many clinical and experimental observations strengthen the existence of the gut-joint axis in rheumatologic and gastrointestinal ailments. Autoimmune inflammation is believed to initiate in the gut mucosa and often many years preceding any clinical sign of joint inflammation. Microscopic gut inflammation is associated with age, sex, disease activity, comorbidities, degree of MRI inflammation on sacroiliac joints and is predictive for disease course, therapeutic efficacy and prognosis (Gill et al., 2015; Wilde et al., 2015). Gut dysbiosis similarities are evident between IBD and spondyloarthritis: decreased abundance of the Firmicutes, and particularly *F. prausnitzii* and *Clostridium leptum* have been reported and may have an important role in the gut-joint axis (Clemente et al., 2012).

Numerous underlying mechanisms are potentially involved in the immune response at the gastrointestinal mucosa thereby initiating rheumatologic disease or rheumatologic symptoms in concurrent gastrointestinal disease. Molecular mimicry may occur between mucosal antigens

(microbial or other) and self-proteins generating autoantibodies and leading to inflammation (Catrina et al., 2016). A notable example of this cross-reactivity is observed in rheumatic fever; pharyngeal infection with group A *Streptococcus* bacteria generates antibodies that cross-react with host self-antigens of the heart and other tissues (Cunningham, 2012). Molecular mimicry in the context of RA might be influenced by food intake. In RA, production of cross-reactive food antibodies from serum and perfusion fluid from the jejunum is reportedly increased (Hvatum et al., 2006). Food-related issues might theoretically be an adverse additive effect of multiple moderate hypersensitivity immune reactions. An alternative mechanism may involve microbial-mediated respiratory burst to facilitate mucosal tissue citrullination through processes of peptidylarginine deiminase activation or carbamylation (Makrygiannakis et al., 2006; Shi et al., 2014). The role of TLR in autoimmune development is relevant as they are activated by endogenous nuclear material aberrantly released during cell death (Lerner et al., 2016). Anti-tissue transglutaminase antibodies are associated with reduced apoptosis clearance and apoptosis produces citrullinated endogenous proteins (Roth et al., 2006). Moreover, activated neutrophil extracellular traps can also trigger citrullination through release of citrullinated peptides (Kleyer et al., 2014).

The premise that antibodies are released in the peripheral blood and circulates throughout the host many years preceding the onset of detectable joint manifestations strongly indicates that antibody production occurs at extra-articular anatomical locations. Albeit, antibodies produced in the joints in the absence of macro- and/or microscopic indication of inflammation remains a possibility. Persons with antibodies are devoid of clinical complaints and inflammation suggesting that the antibodies are passive bystanders of disease. That said, ACPA inherently

possess multiple effector pathogenic functions. ACPA have been shown to exacerbate minor joint disease by passive transfer in mice (Kuhn et al., 2006), activate the complement system (Trouw et al., 2009), trigger macrophage activation (Sokolove et al., 2011), promote the release of neutrophil extracellular traps by neutrophils and lastly, to stimulate inflammation (Kleyer et al., 2014).

While noteworthy advancements have been made to understand the perpetuating role of ACPA to joint inflammation the trigger of the delayed clinical manifestations remains unclear, though several scenarios are conceivable. First, a 'second hit' such as a trauma, transient infection or dysbiosis might stimulate expression of citrullinated proteins in a healthy joint (Makrygiannakis et al., 2006). Newly expressed antigens would thereafter be targeted by pre-existing circulating ACPA and thus lead to joint inflammation. The principle behind this mechanism is that circulating ACPA generated in response to mucosal antigens might target similar antigens in the joint. In fact, a study utilizing mass spectrometry techniques have identified the same citrullinated peptides in both the joint and lungs of RA patients (Joshua et al., 2014). Second, anatomical sites other than the synovial membrane might be the primary joint constituent to be affected with synovial involvement a mere secondary effect; evidence of bone destruction preceding synovial inflammation has been reported (Kleyer et al., 2014). Third, progressive epitope spreading (Brink et al., 2013) and subclinical inflammation (Hughes-Austin et al., 2013) might be required for ACPA to acquire effector functions suggesting that antibodies might begin as nonpathogenic and thereafter acquire proarthritogenic properties.

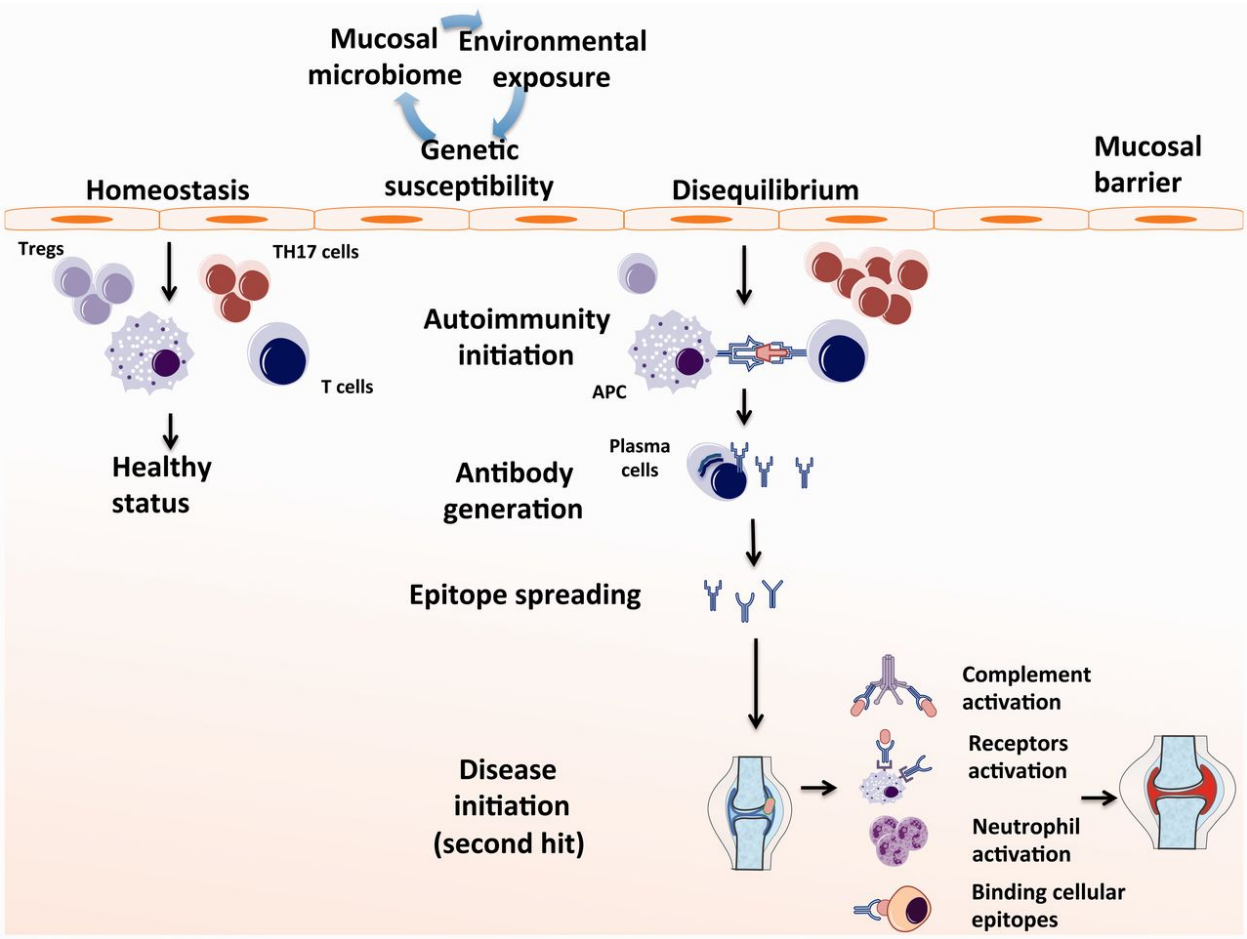

Figure 5. A schematic representation of how mucosal disequilibrium might lead to generation of autoimmunity and later to joint disease development. Complex mechanisms dependent on environmental exposure and the host microbiome are responsible for maintaining homeostasis at mucosal surfaces (such as the respiratory and gastrointestinal tract). In genetically susceptible hosts, failure of these mechanisms may lead to mucosal disequilibrium and molecular changes such as post-translational modifications (citrullination), with subsequent antigen presentation by professional antigen presenting cells, activation of immune effector T cells (such as Th17) and relative deficiency of Tregs. These changes lead to activation of B cells and generation of antibodies (such as ACPAs) by plasma cells. These antibodies undergo somatic hypermutation and epitope spreading, leading to joint disease initiation and perpetuation through several mechanisms (complement activation, cell surface receptor activation, ligation of cell surface components and neutrophil activation). Used with permission from Catrina et al., (2016).

2.5.1.3.2 Dysbiosis evidence and characteristics

Gut bacteria as an autoimmune trigger in RA has been recognized for decades. Mansson and Colldahl, (1965) reported the majority of RA patients to have an intestinal microbiota with increased amounts of *Clostridium perfringens* type A in their feces. The patients also showed immune responses with elevated antibody titers to *C. perfringens* alphatoxin and positive skin tests. Ultimately, the elevation of *C. perfringens* was not specific to RA but was instead identified in a number of diseases, in particular, other chronic arthritides. It was later hypothesized that there are many bacteria involved in the pathogenesis of RA and not just one bacterium (Gullberg, 1978). An accumulating body of literature exists to support a connection between gut bacteria and arthritis or more commonly referred to as the 'gut–joint axis' hypothesis.

The groundbreaking study to identify the proarthritogenic role played by intestinal bacteria was reported by Kohashi et al., (1979) who discovered germ-free rats developed severe joint inflammation with 100% incidence in an adjuvant-induced arthritis model as compared to conventionally raised controls that developed less severe arthritis and at a modest incidence. This observation suggests that the gut microbiota may have important immunosuppressive effects. Other notable early findings supporting gut involvement in RA include studies reporting either protective or proarthritogenic roles of *E. coli* and *Bacteroides* species when introduced into germ-free arthritis-prone rats (Kohashi et al., 1985, 1986; Rath et al., 1996); germ-free *HLA-B27* transgenic rats remained unaffected from inflammatory (intestinal and peripheral) joint disease (Taurog et al., 1994), and germ-free rats become susceptible to arthritis –attributed to a loss of T-cell tolerance– in a streptococcal cell wall-induced rat arthritis model (van den Broek et al.,

1992). More recently, Abdollahi-Roodsaz et al., (2008) determined arthritis is attenuated in IL-1 receptor antagonist-knockout (*Il1rn*$^{-/-}$) mice (a model whereby mice spontaneously develop an autoimmune T-cell mediated arthritis). Applying gnotobiological methods, the introduction of the commensal species *Lactobacillus bifidus* resulted in disease onset. Development of arthritis in this model was attributed to an imbalance of Treg-Th17 homeostasis. Additional evidence comes from a study using the K/BxN T-cell receptor transgenic model of inflammatory arthritis (Wu et al., 2010). Germ-free mice did not develop disease, implying the gut microbiota is critical to the development of arthritis; protection of arthritis was ascribed to an absence of peripheral Th17 cells. Interestingly, monocolonization of germ-free K/Bxn mice with SFB effectively stimulated an autoimmune response thereby resulting in arthritis; autoimmunity developed as a result of promoting Th17 cells. Collectively, these findings derived from germ-free and gnotobiotic experimental animal models imply that the gut microbiota – particularly in a state of dysbiosis – is deemed necessary to trigger autoimmunity and consequent inflammatory arthritis. Moreover, the transition from a 'normal' gut microbiota to a state of dysbiosis likely requires a genetic predisposition.

One of the initial microbiome surveys in RA reported that persons early in disease course (disease duration ≤ 6 months) harbor significantly less bifidobacteria, *Bacteroides fragilis* subgroup, bacteria of the *Eubacterium rectale – Clostridium coccoides* subgroup and *Bacteroides-Porphyromonas-Prevotella* groups in comparison to persons with non-inflammatory fibromyalgia (Vaahtovuo et al., 2008). Another study specifically examined the relationship of *Lactobacillus* to RA (Liu et al., 2013); the authors identified significantly more *Lactobacillus* (10.62 ± 1.72 16S amplicon copies/gram of feces) in the RA gut compared to healthy controls

(8.93 ± 1.60 16S amplicon copies/gram of feces). Furthermore, an increase in diversity and bacterial abundance was reported suggesting a possible relationship between *Lactobacillus* populations and disease onset and/or progression of RA. The abundance of *Lactobacillus* communities in RA, and other IMID, is relevant as these microorganisms and others are considered probiotic microbes that may confer a health benefit to the host. Sher et al., (2013) identified a strong association between *Prevotella copri* with new-onset untreated RA (NORA) and increases in the abundance of *Prevotella* corresponded with a reduction of several beneficial microbes, including the *Bacteroides*. Moreover, a negative correlation between the relative abundance of *P. copri* and the presence of shared epitope risk alleles was identified, indicating that the role of *P. copri* may play a more important role in persons with a lower load of genetic susceptibility. Zhang et al., (2015) recently analyzed the fecal, dental and salivary microbiome from treatment-naïve RA and unrelated healthy controls, as well as treatment-naïve RA paired with healthy relatives and disease-modifying anti-rheumatic drugs (DMARD) treated RA patients, representing the largest RA cohort and most comprehensive metagenomics analysis to date. The fecal analysis showed that *Haemophilus* spp. were depleted and *Lactobacillus salivarius* was over-represented in RA, and that more dramatic changes in abundance were observed in cases of active disease. In contrast, the abundance of *P. copri* was detected as a function of RA duration within the first year. Metagenomic linkage groups comprised of *Clostridium asparagiforme*, *Gordonibacter pamelaeae*, *Eggerthella lengta* and *Lachnospiraceae bacterium* as well as *B. dentium, Lactobacillus sp.* and *Ruminococcus lactaris* were enriched in the RA gut. The investigators also showed that the altered gut microbiome could be used to identify RA patients, correlated with clinical indices [for example, titers of IgA, major serum IgG, platelet count, anticyclic citrullinated peptide (anti-CCP) and RF], and could be utilized to

stratify individuals based on their therapeutic response with DMARDs. And functionally, the redox environment and transport and metabolism of iron, zinc and sulfur were altered in persons with RA. A newly published study examined the microbial and metabolite profile in RA (Chen et al., 2016b). RA patients were found to demonstrate a reduced microbial diversity and this correlated to disease duration and antibody levels. The gut in persons with RA was characterized by an expansion of the typically low-abundant phylum Actinobacteria and a decrease in more dominant phyla. The genera *Eggerthella* and *Actinomyces* of the Actinobacteria and *Turicibacter* and *Streptococcus* of the Firmicutes were significantly expanded in RA whereas the abundance of *Faecalibacterium* was significantly reduced. Moreover, the abundance of *Collinsella* correlated with high levels of alpha-aminoadipic acid and asparagine and IL-17A production. The study also confirmed the role of *Collinsella* in altering gut permeability and disease severity in experimental arthritis.

A list of some of the most current and/or influential microbiome profiling studies in IBD, MS and RA are presented in Table 1.

Table 1. Current studies investigating the role of the gut microbiota in IMID patients.

IMID	Subjects	Comments	Reference
IBD			
	31 patients with ileal, ileocolic and colon-restricted CD	Distinct microbiota profiles observed between ileal and colonic CD: ileal CD richer in *Escherichia* and disease activity correlated with abundance of *Fusobacterium*; increased abundance of *Faecalibacterium,* Clostridiales and Ruminococcaceae in colonic CD.	(Naftali et al., 2016)
	60 newly diagnosed pediatric CD	Increased diversity of Proteobacteria and decreased abundance of Bacteroidetes in CD at baseline. Microbiota profile of CD resembled profile of controls upon clinical remission.	(Meij et al., 2015)
	468 newly diagnosed pediatric CD	Increased abundance of Enterobacteriaceae, Pasteurellaceae, Veillonellaceae, Fusobacteriaceae and decreased abundance of Erysipelotrichales, Bacteroidales and Clostridiales correlates with disease status. Shifts more strongly observed in tissue versus stool.	(Gevers et al., 2014)
	15 patients with CD, 21 patients with UC	Bacteroidetes and Fusobacteria more abundant in inflamed CD mucosa versus inflamed UC. Proteobacteria and Firmicutes more abundant in inflamed UC mucosa. *Faecalibacterium, Bacteroides* and *Pseudomonas* levels differ between inflamed CD and UC; 13 genera significantly differed within the noninflamed mucosa. Noninflamed UC mucosa most different from non-IBD mucosa. No variation in taxa when comparing different gut compartments within CD or UC.	(Forbes et al., 2016)
MS			
	7 RRMS patients	Differences in Firmicutes, Bacteroidetes and Proteobacteria. *Faecalibacterium* decreased in MS. Differences in Bacteroidaceae, *Faecalibacterium, Ruminococcus,* Lactobacillaceae, *Clostridium* and other Clostridiales in glatiramer acetate-treated patients compared to untreated. Vitamin D associated with changes in Firmicutes, Actinobacteria and Proteobacteria. Untreated patients had increase in *Akkermansia, Faecalibacterium* and *Coprococcus* after vitamin D treatment.	(Mowry et al., 2012) (Cantarel et al., 2015)

	53 MS patients: 22 untreated; 13 glatiramer acetate treated; 18 IFN-β treated	Increased *Methanobrevibacter smithii* in MS. Decreased *Butyricimonas* and Lachnospiraceae in untreated MS.	(Jhangi et al., 2014a)
	20 pediatric RRMS	Increased *Escherichia, Shigella, Clostridium*. Decreased *Eubacterium rectale*, *Corynebacterium*.	(Tremlett et al., 2015)
	15 pediatric RRMS	IL-17^{+} T cells positively correlated with overall richness and evenness; inverse correlation between IL-17^{+} T cells and Bacteroidetes	(Tremlett et al., 2016a)
	17 pediatric RRMS	Decreased Fusobacteria, increased Firmicutes and Euryarchaeota linked to shorter relapse time.	(Tremlett et al.,2016b)
RA			
	51 early (≤6 mo) RA	Decreased bifidobacteria, *Bacteroides fragilis* subgroup, *Eubacterium rectale – Clostridium coccoides* subgroup and *Bacteroides-Porphyromonas-Prevotella* groups	(Vaahtovuo et al., 2008)
	15 early (≤6 mo) RA	Increased *Lactobacillus* and diversity.	(Liu et al., 2013)
	44 new-onset RA; 26 chronic, treated RA; 16 PsA	Increased *Prevotella copri* in NORA; decreased *Bacteroides* and beneficial microbes.	(Sher et al., 2013)
	77 treatment naïve RA; 17 treatment naïve RA paired; 21 DMARD-treated RA	Increased *Lactobacillus salivarius*; decreased *Haemophilus*. Metagenomic linkage groups containing *Clostridium asparagiforme*, *Gordonibacter pamelaeae*, *Eggerthella lengta, Lachnospiraceae, Bifidobacterium dentium, Lactobacillus sp.* and *Ruminococcus lactaris* enriched.	(Zhang et al., 2015)
	40 RA	Expansion of Actinobacteria (*Eggerthella, Actinomyces*), *Turicibacter, Streptococcus*. Reduced *Faecalibacterium*. *Collinsella* correlated with high levels of alpha-aminoadipic acid and asparagine and IL-17A production. Role of *Collinsella* in altering gut permeability and disease severity confirmed in experimental arthritis.	(Chen et al., 2016b)

2.5.2 Gut virome in immune-mediated inflammatory disease

To date, human microbiome studies have largely focused on bacterial components of the gut (microbiome), though emerging data suggest that the viral components of the human gut (virome) can have a profound impact on the host. Advances in sequencing technology and bioinformatics analysis have led to the discovery of a diverse human gut virome comprised of eukaryotic viruses and bacteriophages (viruses that infect bacteria) that outnumber human cells by 100-fold (Mokili et al., 2012). The functionality of the virome is not well defined although recent studies suggest the virome does in fact play an important role. It has been shown that a common enteric RNA virus, murine norovirus, can replace the beneficial function of commensal bacteria in germ-free or antibiotic-treated mice (Kernbauer et al., 2014). Importantly, eukaryotic viruses have demonstrated an interaction with IBD susceptibility genes to alter gut microbial populations as demonstrated in studies of mice carrying mutations in *Il-10* (Basic et al., 2014) or *Atg16l1* (Cadwell et al., 2010).

In the absence of disease, gut bacteriophage populations exhibit significant diversity and are predominated by members of the Caudovirales or Microviridae (Castro-Mejía et al., 2015). A number of studies have investigated the role of bacteriophages with respect to IBD. The abundance of bacteriophages in mucosal biopsies are significantly increased in CD versus healthy controls (Lepage et al., 2008), and are also more diverse in pediatric CD compared to healthy controls (Wagner et al., 2013). New evidence demonstrates that disease-specific alterations exist in the gut virome in IBD (Norman et al., 2015). This study showed that the IBD gut virome is abnormal; the gut virome richness is increased and a significant expansion of Caudovirales was observed in both CD and UC, but particular viromes of CD and UC were

disease and cohort specific. Importantly, the expansion and diversification of the gut virome was found to be independent from alterations in bacterial populations.

As virome profiling in health and disease is a relatively new field, data are currently limited to investigations of IBD. With numerous viral infections as potential environmental triggers in IMID, particularly Epstein-Barr virus or cytomegalovirus in MS (Belbasis et al., 2015) and several other lines of evidence that indicates a role for viruses, we anticipate future studies will target the gut virome in other IMID.

2.5.3 Gut mycobiome in immune-mediated inflammatory disease

The human fungal microbial populations (mycobiome) are now appreciated as complex and important for human health. Research of the mycobiome has only recently begun, particularly in health and in IBD. Fungal DNA accounts for approximately 0.2% of the mucosa-associated microbiota and 0.3% of the fecal microbiota (Ott et al., 2008). The fungal constituents of the healthy human gut have been characterized and were found to contain *Aspergillus, Candida, Cryptococcus, Penicillum, Pneumocystis* and *Saccharomyces* (Dollive et al., 2012). Moreover, age and gender have been identified to affect the fungal populations of the human gastrointestinal tract (Strati et al., 2016). In IBD, the fecal mycobiome was recently reported to be skewed (Sokol et al., 2016). Specifically, the authors observed an increased Basidiomycota/Ascomycota ratio, a reduced abundance of *Saccharomyces cerevisiae* and an increased abundance of *Candida albicans* relative to healthy controls. The study also explored the relationship between fungal and bacterial populations and suggests that CD may favour fungi at the expense of bacteria and that disease-specific inter-kingdom alterations exist. Mukhopadhya et al., (2014) studied the mycobiome in newly diagnosed pediatric IBD patients and found the

phylum Basidiomycota to dominate in IBD whereas Ascomycota dominated in healthy controls. A higher fungal diversity has also been reported in CD but no disease-specific fungal species were reported in either CD or UC (Ott et al., 2008). Fungal abundance has also been linked to diet (Hoffmann et al., 2013); *Candida* positively associated with carbohydrate consumption and negatively associated with saturated fatty acid levels whereas *Aspergillus* negatively associated with SCFA levels in a carbohydrate-rich diet. Interestingly, anti-*Saccharomyces cerevisiae* antibodies (ASCA) have a role in diagnosis, disease phenotype and prognosis in CD (Russell et al., 2009).

Other domains of microorganisms including the archaea, fungi or viruses are increasingly being recognized to have important functions in host health. However, in terms of studying the microbiota, research has been very bacteria-centric, as only a limited number of studies have investigated the micro-eukaryotes or viral components in IMID.

2.5.4 Therapeutic manipulation of the gut microbiome

Antibiotics

The gut microbiota is increasingly being recognized as an attractive target for therapeutic intervention. Exposure to particular antibiotics is a known risk factor for development of some IMID, however, antibiotics may also impact on IMID disease course by reducing concentrations of luminal gut bacteria and subsequently alter the gut microbiota composition in a way that is beneficial to the host, resulting in the induction of remission and/or the prevention of relapse. For example, even in the absence of robust clinical evidence antibiotics have been considered by some as an effective therapy in IBD and routinely used as a therapeutic strategy. A diverse group of antibiotics including macrolides, fluoroquinolones, 5-nitroimidazoles, rifaximin and

antimycobacterial therapy either administered alone or in combination have been evaluated in clinical trials for IBD treatment. In a meta-analysis antibiotics were found to be superior to placebo (Khan et al., 2011). However, the antibiotics had such a disparate spectrum of activity the results of this meta analysis raises the question as to whether it is important to attack a specific spectrum of organisms or simply important to alter the gut microbiome in any way. Antibiotic treatment for RA has been utilized without a firm rationale as early as the 1930s with sulfasalazine, a combination of sulphapyridine and 5-aminosalicylic acid (Svartz, 1948), and later with tetracycline derivatives (Brown et al., 1971). A number of clinical trials have evaluated the use of minocycline (Kloppenburg et al., 1994; O'Dell et al., 1997, 2001; Tilley et al., 1995), macrolides (Ogrendik, 2007a, 2009; Ogrendik and Karagoz, 2011; Saviola et al., 2002, 2013) and levofloxacin (Ogrendik, 2007b) for the treatment of RA. Evidence concerning therapeutic use of antibiotics in MS is limited; use of minocycline for clinically isolated syndrome and RRMS is currently in phase III of clinical trials (Metz et al., 2009).

Probiotics

Repopulating the gut with a healthy community is an alternative approach to perturbing the existing microbiome, but so far, outcomes from these methods are controversial or not efficacious. Probiotics are dietary supplements that contain live (or dead) microorganisms that when administered are thought to strengthen the existing gut microbiome. For IBD, the efficacy of probiotics in clinical outcomes demonstrates variable success. A recent meta-analysis of randomized controlled trials evaluating the effectiveness of probiotics reported probiotic supplementation is linked to inducing remission/response in active UC but not CD (Shen et al., 2014). And, only VSL#3 (a combination of *L. casei, L. plantarum, L. acidophilus, L. delbruekii*

subsp. *bulgaricus*, *B. longim, B. breve, B. infantis* and *S. salivarus* subsp. *thermophilus*) significantly increased remission/response rates. Studies indicate that probiotics cannot treat CD and UC with equal effectiveness suggesting altering a single parameter might not be enough to cure disease. There is clinical evidence that probiotics (*L. casei* 01) as adjunctive therapy improves disease activity and inflammatory status in patients with RA (Vaghef-Mehrabany et al., 2014). In contrast, a probiotic preparation containing *S. salivarius, B. lactis* and *L. acidophilus* demonstrated no significant benefit in patients with spondylarthritis (Jenks et al., 2010). There are currently no published clinical trials to date of probiotic administration in MS, though studies utilizing mice suggest a potential benefit in MS (discussed in section 2.5.1.2.2).

Prebiotics

Prebiotic refers to a food substance that may be fermented but not digested. The fermentation of prebiotics stimulates growth and activity of gut microbes potentially benefiting the host via production of energy and metabolic substrates. Studies evaluating the clinical use of prebiotics in specific diseases are limited, with the exception of CD. The administration of 15 g of fructo-oligosaccharides (inulin) in patients with active ileocolonic CD has shown both promising results (Lindsay et al., 2006) and no clinical benefit (Benjamin et al., 2011). A 10 g lactulose administered to active CD patients demonstrated no significant improvement (Hafer et al., 2007). Prebiotics may be bloating and may in themselves lead to gastrointestinal symptoms.

Fecal microbiota transplantation

Fecal microbiota transplantation (FMT) is another approach that can be used in diseases linked to gut dysbiosis. FMT is a novel technique in which the gut microbiota are transferred from a

healthy donor to the patient with the overall goal to introduce a stable microbial community in the gut. Until recently, FMT has primarily been utilized to successfully treat recurrent antibiotic-resistant *C. difficile* infection (van Nood et al., 2013). The clinical efficacy of FMT in UC is promising as FMT has been shown to induce remission in a greater percentage of patients than placebo and no difference in adverse events were reported (Moayyedi et al., 2015). FMT has been investigated in few other diseases including CD (Gordon and Harbord, 2014; Quera et al., 2014) as case reports, albeit data are too limited to determine clinical usefulness. More randomized controlled trials are needed to evaluate donor selection and the frequency of FMT administration.

2.6 Research rationale and hypotheses

This research is divided into three major aims each with their own respective hypotheses. The congruity of these aims allows us to explore several aspects of IMID etiopathogenesis.

i. Gain insights into the immune-mediated inflammatory disease gut microbiome

Rationale

IMID are multifaceted disorders that share many common underlying dynamics including epidemiological co-occurrence, genetic susceptibility and numerous environmental factors. But perhaps more importantly, IMID are highly influenced by the structural and functional characteristics of the gut microbiome. Many microbiome profiling studies have recently elucidated particular dysbiosis patterns in IBD, and more recently, MS and RA; however, by including multiple disease cohorts in a single sample population we are able to characterize the gut microbial populations simultaneously to accurately examine how particular microbial groups are altered or common between IMID. The identification of microbial taxa common, unique or consistently disproportionate to each IMID will provide further credence to overlapping IMID pathophysiology.

Hypotheses

a. The gut microbiota will differ between persons with different IMID and healthy controls.

b. The inferred functional capacity of the gut microbiome will differ between persons with different IMID and healthy controls.

ii. Understand the mucosal microbiome in inflammatory bowel disease

Rationale

Relatively few studies have specifically investigated the compositional differences among inflamed and non-inflamed mucosa of the lower gastrointestinal tract in CD or UC. Likewise, there is a paucity of studies profiling the taxonomic shifts between distinct anatomical locations of the gastrointestinal tract. Given that different sections of the gastrointestinal tract perform vastly distinct host functions or that particular immunological reactions are occurring in inflammation versus in the absence of inflammation, these anatomical areas should be populated by a different group of microbes.

Hypotheses

a. The mucosal microbiome will be different in IBD and healthy controls.
b. Gastrointestinal compartments will exhibit distinct microbiome profiles and functions in IBD.
c. The inflamed mucosa in IBD will be characterized by a different microbiome profile and function than non-inflamed mucosa.

iii. Investigate a potential inflammatory bowel disease environmental etiology

Rationale

Despite exhaustive scrutiny and comprehensive investigations into the etiopathogenic mechanisms of IBD a definitive cause of disease is unknown. It is widely accepted that IBD (and IMID in general) is the result of an interaction between a genetic predisposition and environmental factors combined with gut dysbiosis and an overactive immune response. A pivotal question shaping how we view IBD remains: is gut dysbiosis a trigger or consequence of disease? If it is a trigger of disease what environmental factors can profoundly impact on the host and gut microbiome to initiate an inflammatory cascade? A plethora of environmental risk factors for IBD are recognized (see section 2.1), however to date, a causal association has yet to be established, suggesting that perhaps undefined environmental factors contribute to disease etiopathogenesis. The concept of drinking water as an etiological source of IBD is novel and may provide rationale to the varied geographical distribution of incidence rates observed globally.

The average Canadian adult consumes 1.5 litres of water daily (http://www.hc-sc.gc.ca). Studies have suggested a risk of developing IBD to be associated with iron content of drinking water (Aamodt et al. 2008) as iron may impact the microbial ecology of the human gut (Dostal et al. 2012). At a microbial level, drinking water systems harbor a vast diversity of microbes (Hammes et al. 2008; Navarro-Noya et al. 2013); bacterial concentrations of drinking water are estimated at 10^6-10^8 cells per liter (Hammes et al. 2008; Navarro-Noya et al. 2013; Lautenschlager et al. 2010). Modern water treatment systems apply filtration and disinfection methods to purify drinking water and inactivate bacteria though it is not possible to extricate all microbes from drinking water distribution systems due to 1) microbes' survival and persistence including biofilm formation and nitrification and 2) inadequate disinfection strategies and a lack of

understanding of microbial ecology (Berry et al. 2006). Bacterial communities of drinking water systems may be innocuous, may benefit human health or may be detrimental if harbouring potential pathogens (Berry et al. 2006; Thomas and Ashbolt 2011), it is therefore critical to determine the identity of bacteria present in water distribution systems. Further, microorganisms that are not considered pathogenic may be injurious in specific hosts whose intestinal immune system may be genetically programmed to react aberrantly to them. That said, microbes specific to either low or high incidence areas of IBD might be instructive in regards to IBD etiology.

Hypothesis

a. The water microbiota of low and high incidence areas of IBD will be characterized by different microbial profiles.

CHAPTER THREE: MATERIALS AND METHODS

3.1 Insights into the immune-mediated inflammatory disease gut microbiome

3.1.1 Patients and acquisition of biological specimen

Research and ethics approval was acquired from the University of Manitoba Research Ethics Board (see APPENDIX V). Patients were recruited in clinic. Healthy controls (HC) were recruited at the University of Manitoba and elsewhere; HC who had not taken antibiotics in the previous two months and with no history of gastrointestinal, neurological or joint manifestations were included in this study. Participants self-collected two stool samples approximately two months apart. Stool was transported to the laboratory and stored at -80 °C until processing.

Table 2. IMID patient data at time of sample procurement.

Disease	Average age, yrs[a]	Female/Male[a]
Crohn's disease	49.9	14/5
Ulcerative colitis	51.2	11/8
Multiple sclerosis	47.3	14/4
Rheumatoid arthritis	62.3	14/7
Healthy controls	32.4	12/11

[a] tabulated meta-data does not include information from patients whose meta-data was not available

3.1.2 DNA extraction

Stool samples were divided into three aliquots. One aliquot of each sample was thawed at 4 °C and genomic DNA was isolated in 96-well plate purification format using the ZR-96 Fecal DNA

Kit (Zymo Research, Irvine, CA) with modifications. Briefly, 80 mg of stool was added to tubes (containing 0.1 and 0.5 mm beads) of the ZR BashingBead™ Lysis Rack, and 400 µl lysis solution was added. Samples were processed at 1500 rpm for 20 min (2 × 10 min with 5 min rest in between) in a MiniG® SpexSamplePrep (Metuchen, NJ, USA) and centrifuged at 3,500 × g for 5 min. 150 µl of supernatant was added to each well of a deep well block, in triplicate. 350 µl Fecal DNA binding buffer with beta-mercaptoethanol (0.5% v/v) and 0.5X isopropanol (250 µl) was added to each well. Each block was mixed by placing on a plate shaker at 1800 rpm for 2 min and centrifuged at 3,500 × g for 5 min. 500 µl of each mixture was added to the wells of the ZR Silicon-A™ plate on a collection plate and centrifuged at 3,500 × g for 5 min. Flow through was discarded and this step was repeated until all mixture was applied to the column. DNA Pre-Wash buffer (200 µl) was added to the wells of the ZR Silicon-A™ plate, centrifuged at 3,500 × g for 5 min followed by the addition of Fecal DNA Wash Buffer (500 µl) and centrifugation at 3,500 × g for 5 min. DNA elution buffer was heated to 60 °C and added directly to the matrices in the wells of the ZR Silicon-A™ plate on a clean elution plate and was centrifuged at 3,500 × g for 5 min after waiting 3 min. The eluent was then added back onto the same column and centrifuged at 3,500 × g for 5 min after waiting 3 min. The ZR Silicon-A™ –HRC Plate was prepared according to manufacturer's guidelines. Eluted DNA was transferred to the ZR Silicon-A™ –HRC Plate on a clean elution plate and centrifuged at 3,500 × g for 3 min. DNA was quantitated with PicoGreen.

DNA extraction optimization

Several methods for isolating genomic DNA were investigated in order to optimize the microbial profile of the stool samples. First, on a subset of samples, we tested whether a compositional

difference in the microbial profile exists between solid stool and the use of a fecal/water slurry in a 1:4 dilution utilizing the above-mentioned DNA extraction procedure. Next, we implemented several modifications to the manufacturer's protocol in addition to the above-mentioned protocol utilizing the ZR-96 Fecal DNA Kit (Zymo Research, Irvine, CA). These modifications included varying times of physical lysis via bead-beating (1 min, 5 min or 10 min) and different amounts of lysis solution (300 µl and 400 µl). We also tested a different DNA extraction kit – MoBio PowerSoil DNA Isolation Kit according to the manufacturer's protocol. These troubleshooting techniques were worked up on IMID samples, in addition to fresh, never frozen, healthy stool.

3.1.3 Library preparation and genomic sequencing

Templates were prepared following the "Metagenomic Sequencing Library Preparation: Preparing 16S Ribosomal RNA Gene Amplicons for the Illumina MiSeq System - Part # 15044223 Rev. B" with modifications to the manufacturer's guidelines. Amplicon PCR reactions (25 µl) contained 2x KAPA HiFi HotStart Ready Mix (Kapa Biosciences, Wilmington, MA, USA), 2.5 µl microbial DNA and 5 µl of 1 µM primers 515fXT (GTGBCAGCMGCCGCGGTAA) and 806rXT (GGACTACHVGGGTWTCTAAT) to target the 16S V4 region. PCR parameters consisted of 1 cycle at 95 °C for 3 min, 25 cycles of 95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s and a final extension cycle at 72 °C for 5 min. 16S V4 amplicons were cleaned utilizing AMPure XP (Beckman Counter Canada, LP, Mississauga, ON) beads with modifications to the manufacturer's guidelines: 18 µl of 10 mM Tris pH 8.5 was added to each well of the amplicons and 15 µl of the supernatant was recovered. Index PCR reactions (50 µl) contained 15 µl DNA, 2x KAPA HiFi HotStart Ready Mix (Kapa Biosciences, Wilmington, MA, USA), and 5 µl of each dual-indexed (Nextera® XT Index Kit v2) barcodes. PCR parameters consisted of 1 cycle at 95 °C for 3 min, 8 cycles of 95 °C for 30 s, 55 °C for 30

s, 72 °C for 30 s and a final extension cycle at 72 °C for 5 min. The final libraries were cleaned with no modifications to the guidelines. The final library size was approximately 420 bp. Libraries were quantitated with PicoGreen and pooled in equimolar volumes. Next, BluePippin (Sage Science Inc., Beverly, MA) was utilized with 1.5% cassettes to gate and select for 250-650 bp fragments and quantified by Qubit™ 2.0 fluorometric quantitation (Thermo Fisher Scientific Inc., Waltham, MA). The pools were cleaned with a 0.6X AMPure XP (Beckman Counter Canada, LP, Mississauga, ON) bead cleanup. The size and concentration of the pooled libraries were determined by an Agilent Tapestation analyzer (Agilent Technologies Canada Inc., Mississauga, ON). Approximately 11 pM of the pools and 37.5% PhiX spike-in control DNA were used for sequencing on the Illumina MiSeq instrument generating 2×300 bp. The final dataset included 4 MiSeq runs with 56 samples multiplexed for a total of 224 samples including technical replicates, mock communities of known composition (HM-782D; BEI Resources, Manassas, VA) and negative control templates.

3.1.4 Bioinformatics and statistical analyses

3.1.4.1 Quality control

16S amplicon data was analyzed by the mothur analysis software package (v.1.38.0; Schloss et al., 2009) with customized batch scripts. Paired-end reads (2×300 bp) were assembled into contigs and V4 amplicon primers were removed. Contigs were subjected to quality control whereby contigs > 275 bp, containing homopolymers > 8 nt or containing any ambiguous base calls were screened and filtered from the dataset. Trimming sequence limits of the curated 16S rDNA SILVA reference alignment (Pruesse et al., 2007) created a custom reference alignment database specific to the V4 region; contigs were aligned to the custom database. Specifically, alignment of contigs to the database uses kmer (8mer) searching and the Needleman-Wunsch

pairwise alignment approach. Contigs aligning outside of the 16S V4 region were removed. Sequencing noise was reduced by (pre)clustering contigs that differed by a maximum of 2 bp. Chimeric artefacts were identified utilizing mothur's implementation of UCHIME (Edgar et al., 2011) and removed. Mock community groups were selected from the data to identify sequence error rates from each MiSeq run and were removed from downstream analyses.

3.1.4.2 Operational taxonomic unit picking and taxonomic assignment

Sequences were taxonomically classified using the RDP reference database (Wang et al., 2007) with a 60% bootstrap and those identified as unwanted lineages (chloroplast, mitochondria, archaea, eukaryota or unknown) were removed. The classification algorithm implemented in mothur is a naïve Bayes classifier that utilizes the reference database as a training set whereby sequence taxonomies are assigned according to the training set. The remaining contigs were binned into species-level (≥97% sequence similarity) OTUs using the average neighbour algorithm, and taxonomic classification of OTUs was performed using the RDP reference database with a 60% minimum bootstrap (Wang et al., 2007). Resultant and relevant files included an abundance matrix that contains the number of OTUs per sample and a list of taxonomies for corresponding OTUs. These files were manipulated for compatibility in downstream software packages.

3.1.4.3 Statistical analysis of operational taxonomic unit data

Processing of OTU data were executed using the R package (http://www.r-project.org) phyloseq (McMurdie and Holmes, 2013). OTUs were normalized to relative abundance and were filtered to include only OTUs with a mean abundance of $> 1\times10^{-3}$. Filtered OTU and taxonomy data in addition to sample meta-data were exported from R and merged into a single file containing all of the above. OTUs with identical taxonomies were aggregated using a customized python script.

Genus data was analyzed via LEfSe (Segata et al., 2011) as a means of biomarker discovery combining both statistical and biological relevance. Default parameters were applied (i.e. LDA > 2 and all-against-all multi-class analyses). In addition, customized Kruskal-Wallis one-way analysis of variance R-scripts was applied to OTU data to compare median similarities. *P*-values were corrected for multiple comparisons utilizing false discovery rate (FDR) and were considered significant at $P < 0.05$. Furthermore, where the presence of available meta-data allowed, we applied a multivariate statistical framework to identify relationships between clinical meta-data to microbial community abundance utilizing MaAsLin (multivariate association with linear models; Morgan et al., 2012b).

3.1.4.4 Species-level taxonomic assignment

Sequences were degapped and were split into groups (i.e. samples). Degapped fasta files belonging to the same disease cohort were concatenated and served as input for SPINGO (species level identification of metagenomic amplicons; Allard et al., 2015). Briefly, SPINGO is an accurate, flexible and rapid classifier for low-level taxonomic assignment that uses a species database prepared from RDP taxonomies. SPINGO was executed with default parameters: *k*-mer size of 8, bootstrap samples of 10 and subsample of size 8. Customized python scripts were applied to combine SPINGO output from each job and additionally, to tabulate the count of species per group.

3.1.4.5 Diversity analyses

Alpha diversity estimates including richness (ACE, Chao1) and diversity (Shannon, Simpson and inverse-Simpson) were calculated in mothur (Schloss et al., 2009) and plotted in phyloseq (McMurdie and Holmes, 2013). The ACE estimator utilized 10 as the discriminating cutoff between rare and abundant. For example, if an OTU was present in more than 10 samples, then it

was considered abundant. Kruskal-Wallis R-scripts were used to determine statistical significance. *P*-values were corrected with FDR and considered significant at $P<0.05$

Beta diversity analyses were conducted using phyloseq (McMurdie and Holmes, 2013). The Bray-Curtis index was used and distances visualized via PCoA. A multivariate t-distribution was assumed with an 80% confidence level.

3.1.4.6 Functional inference of the metagenome

Sequences were additionally taxonomically classified using mothur (Schloss et al., 2009) according to the Greengenes database (DeSantis et al., 2006b), version 13.5, allowing for the construction of a biological observation matrix (biom) file compatible with downstream applications. The biom file served as input for PICRUSt (phylogenetic investigation of communities by reconstruction of unobserved states; Langille et al., 2013) to impute the functional potential of the metagenome. PICRUSt was applied with default parameters; KEGG (Kyoto encyclopedia of genes and genomes; Kanehisa et al., 2004) ortholog (KEGG Ortholog; KO) functional predictions were made and categorized according to function at a KEGG hierarchy level of 3. The metagenomes were then analyzed with LEfSe (linear discriminant analysis effect size; Segata et al., 2011) with an LDA score > 1.5 and a one-against-all multi-class analysis to determine genes significantly enriched.

3.2 Understanding the mucosal microbiome in inflammatory bowel disease

3.2.1 Patients and acquisition of biological specimen

Research and ethics approval was acquired from the University of Manitoba Research Ethics Board (see APPENDIX V). Patients undergoing endoscopy for other reasons were recruited to

enroll in this study. In all, 3-4 three mm biopsies were collected from each of CD, UC and non-IBD participants at colonoscopy from the ileum, cecum, mid-colon and/or rectum for a total of 166 specimens. The specimens were histologically defined as inflamed or non-inflamed as per the clinical pathologist. Non-IBD participants were undergoing colonoscopy either for screening or for other gastrointestinal complaints unrelated to IBD. Biopsies were immediately snap frozen (i.e. not rinsed) in cryovials without preservative in liquid nitrogen and transferred to a -70°C freezer, hence it is possible there were some adherent lumen contents. Patient characteristics are described in Table 3.

Table 3. Patient and biopsy characteristics at time of sample procurement.

Disease (N)	Average Disease Duration, yrs	Male/Female	Biopsy Location (N)	Inflamed/Noninflamed
CD (15)	10.24 ± 6.9	8/7	Ileum (13)	9/4
			Cecum (5)	2/3
			Mid-colon (24)	6/18
			Rectum (15)	1/14
UC (21)	10.14 ± 8.8	8/13	Ileum (19)	1/18
			Cecum (20)	1/19
			Mid-colon (22)	10/12
			Rectum (21)	14/7
non-IBD (7)		3/4	Ileum (6)	-
			Cecum (6)	-
			Mid-colon (8)	-
			Rectum (7)	-

3.2.2 DNA extraction and genomic sequencing

Genomic DNA was extracted from tissue samples using a DNeasy Blood and Tissue Kit (Qiagen, Valencia CA) according to manufacturer's instructions. The DNA concentration and purity were assessed by microspectrophotometry (Beckman DU/800; Beckman Coulter, Inc., Fullerton, CA). Genomic DNA was diluted to a concentration of 20 ng/µl.

The bacterial 16S rRNA genes were amplified with primers 967–985 (CAACGCGARGAACCTTACC) and 1078–1061 (ACAACACGAGCTGACGAC) targeting the V6 hypervariable region (58 bp) as described elsewhere (Gloor et al., 2010). The reads were generated as part of a multiplexed sequencing run (Hamady et al., 2008). Sequencing was conducted at Research and Testing Laboratory (Lubbock, TX; http://Researchandtesting.com) using the Illumina HiSeq platform.

3.2.3 Bioinformatics and statistical analyses

3.2.3.1 Quality control

Sequence reads were demultiplexed and multiplex identifiers allowed for assigning reads to their respective sample. Paired-end reads (2×100 bp) were assembled with the FLASH assembler (Magoč and Salzberg, 2011) within the Galaxy framework (Goecks et al., 2010). Assembly of the reads generated 23,789,945 raw contigs. Quality control and taxonomic profiling was conducted using mothur (v.1.34.0; Schloss et al., 2009), a suite of tools for microbial community investigation. Contigs were screened and filtered based on an average quality score < 20, having read lengths > 80 bp, containing hompolymers > 8 nt or containing any ambiguous base calls. Contigs were aligned against the 16S rDNA SILVA reference database (Pruesse et al., 2007) via

kmer searching and the Needleman-Wunsch pairwise alignment method. Sequencing noise was reduced by (pre)clustering contigs that differed by a maximum of 1 bp. Chimeric artefacts were identified via mothur's implementation of UCHIME (Edgar et al., 2011) and removed.

3.2.3.2 Operational taxonomic unit picking and taxonomic classification

Sequences were taxonomically classified using the RDP reference database (Wang et al., 2007) with a 50% bootstrap and those identified as unwanted lineages (chloroplast, mitochondria, archaea, eukaryota or unknown) were removed. The remaining contigs were binned into species-level (≥97% sequence similarity) OTUs using the average neighbour algorithm, and taxonomic classification of OTUs was performed using the RDP reference database (Wang et al., 2007) with a 50% minimum bootstrap. Resultant and relevant files included an abundance matrix that contains the number of OTUs per sample and a list of taxonomies for corresponding OTUs. These files were manipulated for compatibility in downstream software packages.

3.2.3.3 Statistical analysis of operational taxonomic unit data

Analyses were executed using the R package (http://www.r-project.org) phyloseq (McMurdie and Holmes, 2013). Samples with <1000 reads and OTUs present in <5 samples were removed. OTUs were normalized to relative abundance. The data were filtered to include only taxa present with an overall mean abundance of $>10^{-3}$. Statistical significance of OTU data was tested using customized R-scripts; the Kruskal-Wallis one-way analysis of variance was applied to compare median similarities. *P*-values were corrected for multiple comparisons with FDR and were considered significant at $P<0.05$.

3.2.3.4 Species-level taxonomic assignment

Sequences were degapped and were split into groups (i.e. samples). Degapped fasta files belonging to the same disease cohort were concatenated and served as input for SPINGO (Allard et al., 2015). SPINGO was executed utilizing default parameters: *k*-mer size of 8, bootstrap samples of 10 and subsample of size 8. Customized python scripts were applied to combine SPINGO output from each job and additionally, to tabulate the count of species per group.

3.2.3.5 Diversity analyses

Alpha diversity estimates including richness (ACE, Chao1) and diversity (Shannon, Simpson and inverse-Simpson) were calculated in mothur (Schloss et al., 2009) and visualized in phyloseq (McMurdie and Holmes, 2013). The discriminating cutoff between rare and abundant taxa for ACE was 10. Kruskal-Wallis R-scripts were used to determine statistical significance. *P*-values were corrected with FDR and considered significant at $P<0.05$

Beta diversity analyses were conducted using phyloseq (McMurdie and Holmes, 2013) The Bray-Curtis index was used and distances visualized via PCoA. A multivariate t-distribution was assumed with an 80% confidence level.

3.2.3.6 Functional inference of the metagenome

Sequences were additionally taxonomically classified using mothur (Schloss et al., 2009) according to the greengenes database (DeSantis et al., 2006b), version 13.5, to create a biom matrix compatible with PICRUSt (Langille et al., 2013). The software PICRUSt was used to predict the functional gene content. PICRUSt was used within the galaxy framework (Goecks et al., 2010). The biom matrix was normalized to correct for multiple 16S copy number. KO functional metagenome predictions were made and categorized according to function at a KEGG

pathway hierarchy level of 3. The metagenomes were then analyzed with LEfSe (Segata et al., 2011) to determine functional genes associated with particular sample types. For defining functional biomarkers specific to gut compartments, LEfSe was executed with an LDA score > 2.5 and a one-against-all multi-class analysis. Functional biomarkers of distinct inflammatory states where identified with an LDA score > 1.5 and one-against all testing. Default settings were used for the remainder of parameters.

3.3 A potential inflammatory bowel disease environmental etiology

3.3.1 Sample collection

The first population based incidence rates from Canada were reported from the province of Manitoba, a central province with a population of 1.27 million in 2013 (Bernstein et al. 1999). Capitalizing on universal health care provided within each province and the administrative health databases established within each province, investigators have shown that these high rates in Manitoba are comparably elevated in at least 5 of the other 9 provinces (Bitton et al. 2014; Bernstein et al. 2006).

Our group previously identified high incidence areas (HIA) and low incidence areas (LIA) of IBD (Green et al., 2006; Figure 6) using the University of Manitoba IBD Epidemiology Database (UMIBDED; Bernstein et al., 1999).The incidence of HIA was ≥ 2-fold when compared to LIA. The UMIBDED is a population-based archive dating back to 1984 as described in detail elsewhere (Bernstein et al. 2006); it was created from the Manitoba Health databases and includes all Manitobans who meet a validated administrative case definition for IBD.

Incidence rates of IBD were calculated for 2001 for the purpose of this study. We have reassessed small area incidence rates for 2009-2010 and they are minimally changed from 2001 (data not shown). Hence, we anticipate HIA and LIA remained as such in 2005 when water samples were collected. It is worth noting that based on epidemiological surveillance, LIA of Crohn's disease (CD) correlate highly with LIA of ulcerative colitis (UC), with a similar trend observed in HIA.

The city of Winnipeg receives its water from Shoal Lake, a large isolated lake in southeastern Manitoba (http://winnipeg.ca/waterandwaste/). It is 137 kilometers from Winnipeg and approximately 92 metres higher. Water flows via gravity through the aqueduct to Deacon Reservoir, a four-cell open reservoir on the eastern side of Winnipeg. The City of Winnipeg's water treatment process includes dissolved air flotation, ozonation, filtration and ultraviolet light disinfection. Fluoride and orthophosphate are added then water flows from the water treatment plant to one of three regional reservoirs and pumping stations. Chlorine is added and water flows through the distribution system. Brandon, Manitoba a city of approximately 50,000 is 199 kilometers west of Winnipeg, uses water from the Assiniboine River as a source for its treatment plant; Steinbach, Manitoba, a city of approximately 14,000 is 58 kilometers southeast of Winnipeg uses groundwater as its drinking source. Both rural communities employ water disinfection strategies similar to those just described; however, the key difference is the source of the drinking water.

Public buildings were used as sample locations as they were a source of high consumption of water. Verbal permission was obtained from personnel at each location. Each location was

equipped with a water filtration system and three pieces of PVC piping. The PVC piping was used to collect biofilm that naturally occurs in drinking water pipes. The water filters were changed every 3 months and the PVC replaced every 4 months over the course of 1 year, providing replicates for each sampling location. Samples were collected in 2005 from HIA (n=20) and LIA (n=20) across Winnipeg, Brandon and Steinbach. We collected samples from 3 water distribution networks (Shoal Lake (n=12), Brandon (n=4) and Steinbach (n=4)). Of the water distributed to Winnipeg areas, samples were obtained from 5 reservoirs – Shoal Lake Intake (n=3), Deacon (n=3), McLean (n=5), McPhillips (n=14) and Wilkes (n=7). Sample types used in this analysis include, bulk drinking water (n=8), filter (n= 15) and pipe wall material (n=17). Samples were removed from the installed PVC system one section at a time. They were placed in zip locked sealed bags, shipped to the laboratory and stored at 4°C until processing.

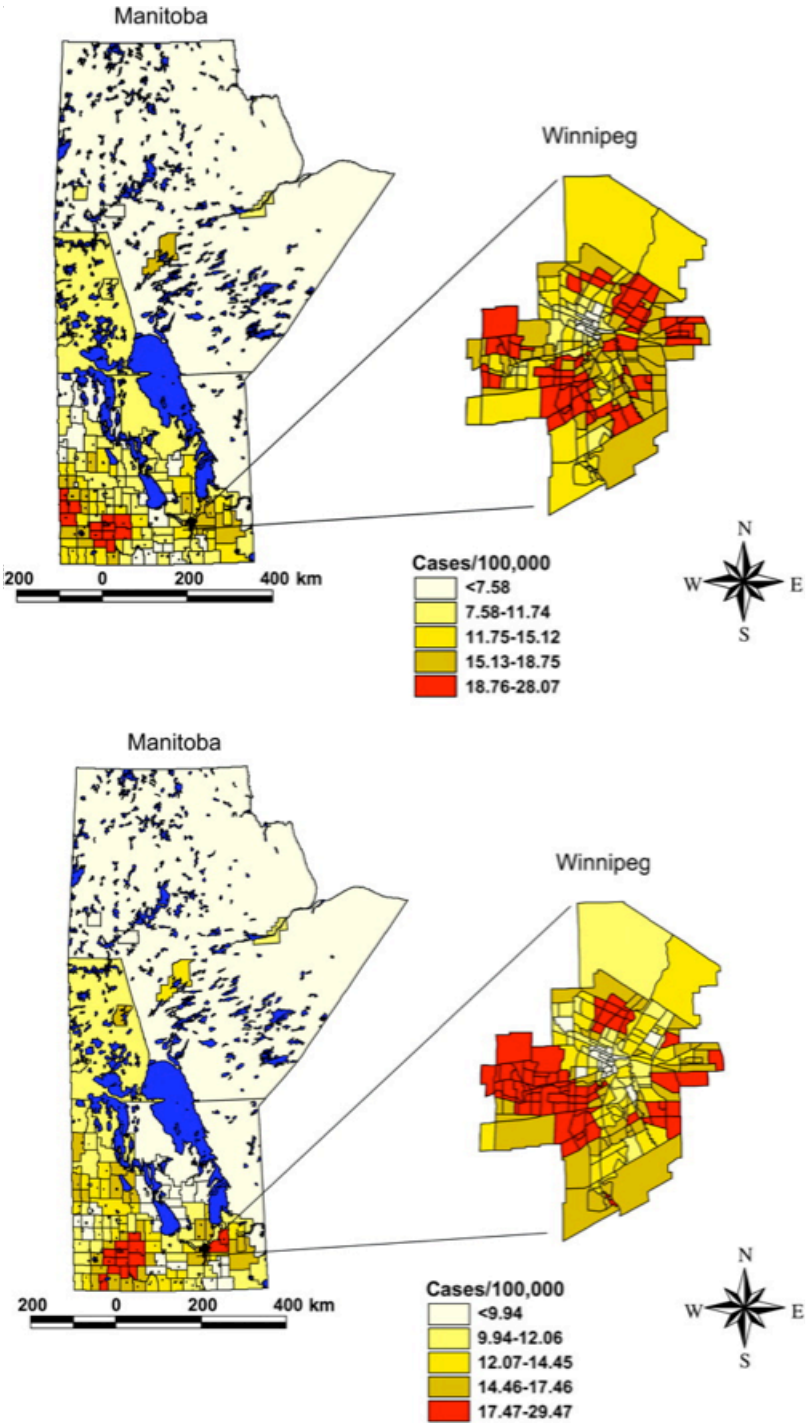

Figure 6. Map of Manitoba, Canada, illustrating incidence of inflammatory bowel disease (IBD). Incidence (number of cases per 100 000 population) of Crohn's disease (top) and ulcerative colitis (bottom) in the province of Manitoba and the City of Winnipeg, Manitoba Canada, 1990-2001. Results were age-standardized to the 1996 Manitoba Population. Used with permission from Green et al., (2006).

3.3.2 DNA extraction and genomic sequencing

The following nucleic acid extraction protocol was implemented for optimal DNA recovery as the biomass of water was expected to be low: 1 ml of each sample was inoculated in 9 ml 2.5% LB broth and incubated aerobically at 37 °C until slightly turbid. The culture suspension was used for genomic DNA extraction with Zymo Research Fecal DNA Kits. The DNA concentration and purity were assessed by microspectrophotometry (Beckman DU/800; Beckman Coulter, Inc., Fullerton, CA). Genomic DNA samples were diluted to a concentration of 20 ng/µl.

The bacterial 16S rRNA genes were amplified with primers 28F (5'-GAGTTTGATCNTGGCTCAG-3') and 519R (5'-GTNTTACNGCGGCKGCTG-3') targeting the V1-V3 hypervariable region (approximately 510 bp). The V1-V3 region has been shown to be suitable for identifying and distinguishing several microbes including but not limited to *Streptococcus, Staphylococcus, Mycobacterium* and *Haemophilus* spp. and is less suitable for distinguishing *Escherichia* spp. and other closely related Enterobacteriaceae (Chakravorty et al., 2007). Pyrosequencing was conducted at Research and Testing Laboratory (Lubbock, TX; http://Researchandtesting.com) using a 454 GS FLX (454 Life Sciences, a Roche Company, Brandford, CT) and titanium technology. Sequencing generated 110,760 raw reads.

3.3.3 Bioinformatics analysis

Quality control and taxonomic profiling was conducted using mothur (v.1.34.0; Schloss et al. 2009), a microbial community software program. Barcodes and primers were removed and low quality reads were filtered based on an average quality score < 20, having read lengths < 210 bp, containing homopolymers > 8 nucleotides and containing any ambiguous base calls. Contigs

were aligned against the 16S rDNA SILVA database (Pruesse et al. 2007). Sequencing noise was reduced by clustering reads that differ by only 1 bp. Chimeric sequences were detected using mothur's implementation of UCHIME and removed (Edgar et al. 2011). The remaining reads were binned into species-level (≥97% sequence similarity) operational taxonomic units (OTUs) using the average neighbour algorithm, and taxonomic classification was performed using the SILVA database with Ribosomal Database Project taxonomy (Wang et al. 2007) and a 70% minimum bootstrap. Species richness estimates and diversity indices were calculated by algorithms implemented in mothur (Schloss et al. 2009).

3.3.4 Statistical analysis

Analyses were executed using the R package (http://www.r-project.org) phyloseq (McMurdie and Holmes 2013). Samples with <500 reads were removed. OTUs were normalized to relative abundance and filtered to include only OTUs with a mean abundance of $> 1\times10^{-4}$. Statistical significance of community composition differences was tested using customized R-scripts: the Kruskal-Wallis one-way analysis of variance was applied to microbiota data and community estimates to compare similarities. *P*-values were considered significant at $P < 0.05$. Visualization of OTU-based community composition differences was employed using the Bray-Curtis dissimilarity index.

CHAPTER FOUR: RESULTS

4.1 Insights into the immune-mediated inflammatory disease gut microbiome

The average base calling error rate assessed by the co-sequenced mock communities was 0.0145%. A total of 31,257,335 raw contigs were assembled. Filtering out low quality, chimeric and non-bacterial contigs generated 26,461,364 high quality sequences with an average amplicon length of 253 bp. Sequences were clustered into 67,694 OTUs based on their shared sequence similarity at a 97% threshold (3% sequence divergence assumed to define a species). The average number of sequences per sample was 120,472 ± 37,420 (range: 20,160 – 296,587). The average coverage for OTU characterization based on Good's coverage was 99.7%. All samples had sufficient depth therefore; all samples were retained in downstream analysis.

4.1.1 Microbial diversity of the stool microbiota

Alpha diversity

We observed significant differences in community richness, as estimated by Chao1 and ACE between disease cohorts and healthy (Figure 7; Table 4). The lowest community richness was reported in CD and the highest in HC and between the two cohorts significant differences were reported: P = 3.72e-05 and P = 2.70e-03 for Chao1 and ACE, respectively. Significant differences were also reported between HC and RA (P = 2.86e-03, P = 9.40e-03). Chao1 reported the community richness profile to differ between CD and UC (P = 1.49e-02) and similarly between CD and MS (P = 3.33e-02). Both richness indices followed the same trend (HC > UC > MS > RA > CD) indicating similar observations between rare (Chao1) and abundant (ACE) taxa.

We used several different diversity indices (Shannon, Simpson, inverse-Simpson) to survey the microbial diversity within samples and found that overall, significant differences between cohorts were apparent (Figure 7; Table 4). Analogous to richness trends, diversity was lowest in CD and highest in HC. Shannon, Simpson and inverse-Simpson indices reported significant differences between CD and HC: P = 3.74e-09, P = 1.14e-08 and P = 1.14e-08. We also observed diversity dissimilarities between HC and RA (P = 6.31e-05, P = 5.85e-05, P = 5.85e-05) and between HC and UC (P = 9.96e-03, P = 1.25e-02, P = 1.25e-02). Moreover, each diversity index between CD and UC (P = 7.38e-05, P = 7.75e-05, P = 7.75e-05) and between CD and MS (P = 1.60e-05, P = 1.69e-04, P = 1.69e-04) reported significant differences whereas only the Shannon diversity index reported diversity dissimilarities between CD and RA (P = 8.87e-03) indicating that the microbial community diversity between these cohorts are due to rare taxa rather than abundant. The microbial diversity between UC, MS and RA were comparable both in the context of rare and abundant taxa.

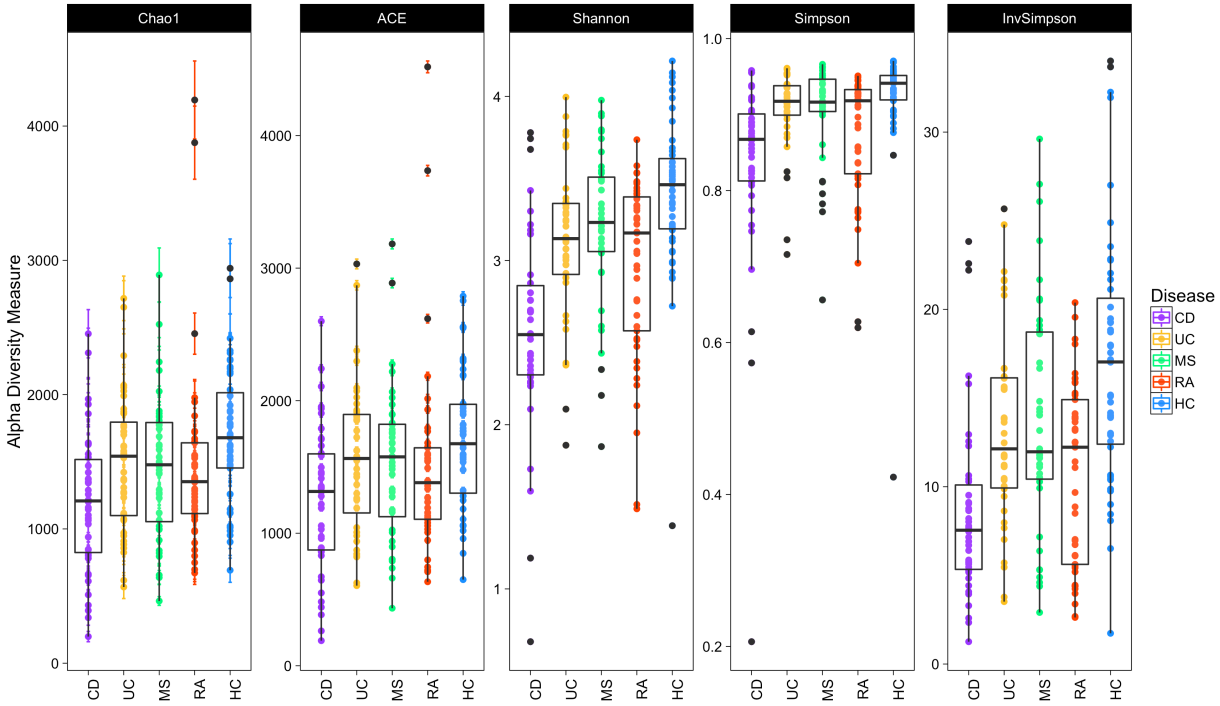

Figure 7. Alpha-diversity plots between disease cohorts. Community richness estimated by Chao1 and ACE; diversity estimated by Shannon, Simpson and inverse-Simpson.

Table 4. Alpha diversity indices with significant differences displayed between cohorts. Community richness estimated by Chao1 and ACE; diversity estimated by Shannon, Simpson and inverse-Simpson.

	HC/CD	HC/UC	HC/MS	HC/RA	CD/UC	CD/MS	CD/RA	UC/MS	UC/RA	MS/RA
Chao1	***			**	*	*				
ACE	**			**						
Shannon	***	**		***	***	***	**			
Simpson	***	*		***	***	***				
InvSimpson	***	*		***	***	***				

* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$

Beta diversity

To compare the overall structure of the fecal microbiota between IMID and in health, beta diversity analyses were executed. The PCoA results between all cohorts reveal an intriguing phenomenon (Figure 8). While the community structure of each cohort to some extent overlaps, HC form the tightest cluster. The communities encompassed at the 80% confidence level for MS and UC are nearly identical indicating strong community similarities, and furthermore, of IMID, MS and UC overlap the most with HC. The community structure of RA also overlaps with MS, UC and HC, though the variability in the confidence ellipse suggests community dissimilarities. Lastly, the CD cohort forms the most variable cluster and is most similar to RA. Figure 8 also shows that samples obtained from the same patient are for the most part, highly similar which adds validity to gut microbiome resiliency within individuals.

PCoA plots were also constructed to view IMID clusters in the context of HC. CD (Figure 9A), showing large community variability, minimally overlaps with a tight HC cluster. Nearly the entire HC cluster overlaps with UC (Figure 9B), however, many UC samples do not overlap with HC. MS (Figure 9C) and HC clusters indicate a high degree of similarity although several MS samples do not overlap with HC. RA (Figure 9D) also shows high variability through a large confidence ellipse, of which the HC samples form a tight cluster within the RA confidence level.

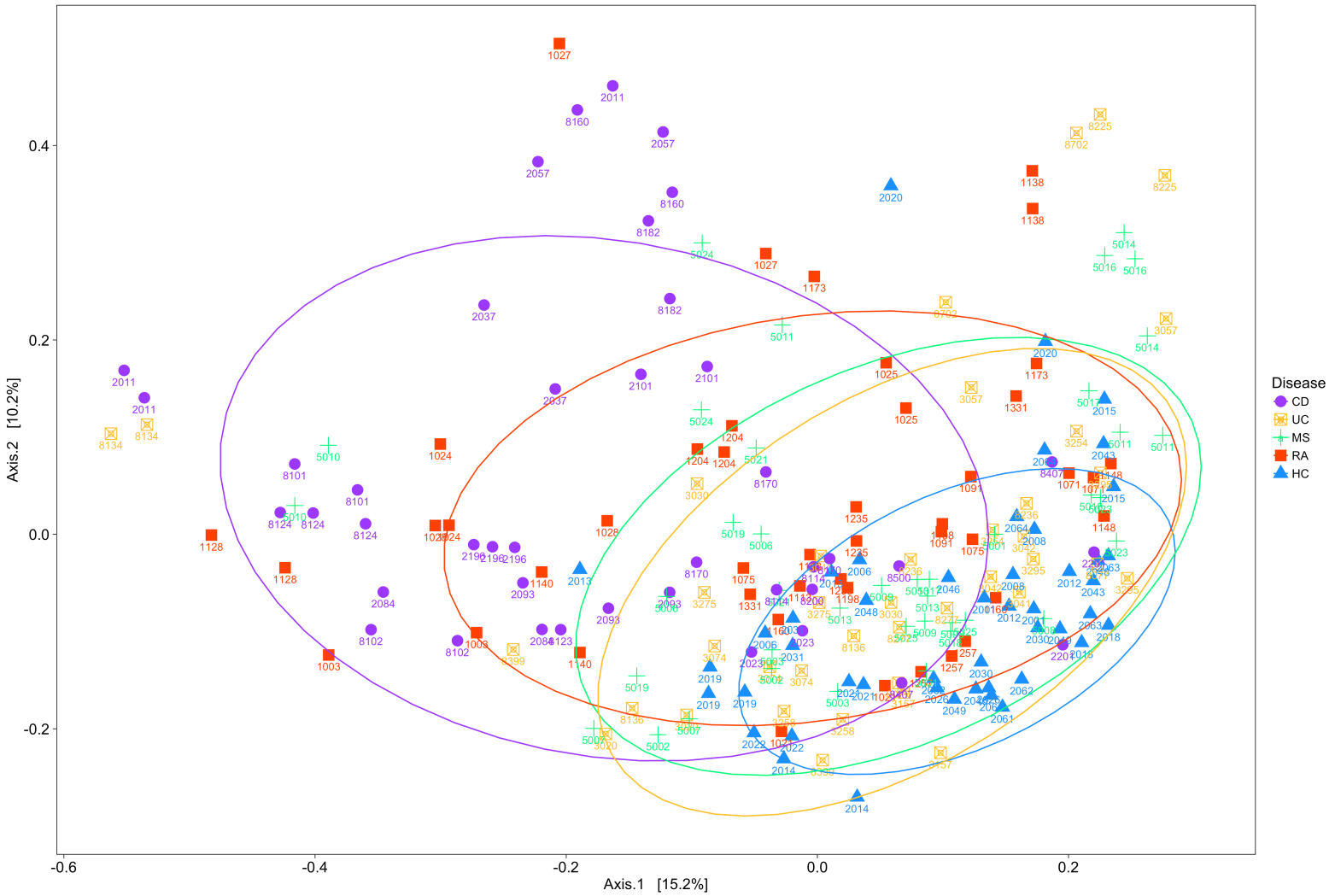

Figure 8. Principal coordinate analysis based on the overall structure of the stool microbiota in all samples. Each data point represents an individual sample. Principal coordinate analysis was calculated using Bray-Curtis distances with a multivariate t-distribution. Ellipses represent an 80% confidence level. Colour/symbol is indicative of disease cohort. Data points are labeled by subject.

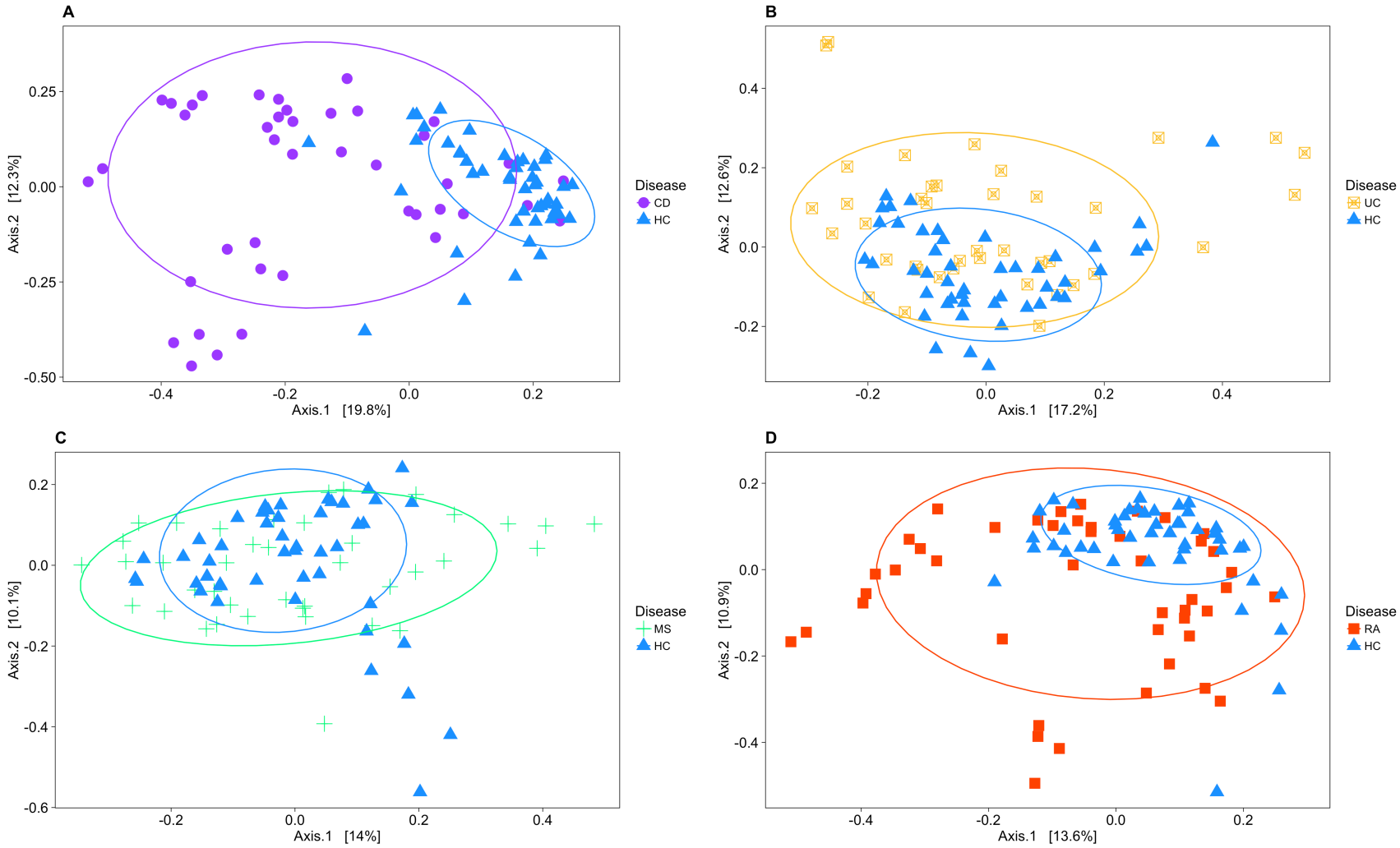

Figure 9. Principal coordinate analysis based on the overall structure of the stool microbiota. A) CD, B) UC, C) MS and D) RA compared to healthy controls. Each data point represents an individual sample. Principal coordinate analysis was calculated using Bray-Curtis distances with a multivariate t-distribution. Ellipses represent an 80% confidence level. Colour/symbol is indicative of disease cohort.

4.1.2 Stool microbiota of immune-mediated inflammatory diseases relative to health

A taxonomic comparative analysis was performed to accurately determine if the gut microbial populations in IMID demonstrate a similar or dissimilar pattern of dysbiosis relative to the healthy gut (Table 5; Figure 10).

Crohn's disease

Phylum level analyses between CD and HC cohorts revealed an increased abundance of Firmicutes and Proteobacteria and a parallel decrease in Actinobacteria and Bacteroidetes in CD. Of the Firmicutes, several genera were significantly increased in CD including *Blautia, Clostridium*_XIVa, *Enterococcus, Streptococcus* and *Veillonella*. *Blautia* and *Streptococcus* genera were among the most abundant taxa in CD. Other taxa including *Catenibacterium*, unclassified Clostridiales, *Clostridium* groups IV, *Clostridium* XI, *Clostridium* sensu stricto, *Coprococcus*, unclassified Erysipelotrichaceae, *Faecalibacterium*, unclassified Firmicutes, *Gemmiger,* Ruminococcaceae, *Ruminococcus* genera belonging to both the Ruminococcaceae and Lachnospiraceae, *Subdoligranulum* and *Turicibacter* were decreased in CD. Moreover, the underrepresentation of Actinobacteria in CD was largely driven by decreases in the genera *Asaccharobacter*, *Bifidobacterium, Collinsella,* unclassified Coriobacteriaceae, *Gordonibacter* and *Slackia*, although an increase of *Eggerthella* was observed. The decrease in Bacteroidetes was driven by shifts in the *Bacteroides* whereas the increase in Proteobacteria was due to increased *Escherichia/Shigella* and unclassified Enterobacteriaceae abundance.

Table 5. Distribution of taxa in disease cohorts relative to healthy controls. Taxa unable to be classified to the genus level were assigned to a lower level of taxonomy.

	Crohn's disease	Ulcerative colitis	Rheumatoid arthritis	Multiple sclerosis
Actinobacteria	↓***			
Asaccharobacter	↓***			
Bifidobacterium	↓**			
Collinsella	↓**			
Coriobacteriaceae	↓***	↓**	↓***	↓*
Eggerthella	↑**	↑**	↑**	↑***
Gordonibacter	↓**			
Slackia	↓*			
Bacteroidetes	↓*			
Bacteroides	↓*			
Firmicutes	↑*			
Anaerostipes				
Blautia	↑*			
Catenibacterium	↓*			
Clostridiales_unclassified	↓**			
*Clostridium*_III		↑*	↑**	↑***
*Clostridium*_IV	↓***			
*Clostridium*_sensu_stricto	↓***		↓*	
*Clostridium*_XI	↓**		↓**	
*Clostridium*_XlVa	↑**	↑**	↑***	↑***
*Clostridium*_XVIII				
Coprococcus	↓***			
Dialister				↓**
Dorea			↓***	
Enterococcus	↑***	↓**		
Erysipelotrichaceae	↓*			
Faecalibacterium	↓***		↓**	↓***
Firmicutes_unclassified	↓***		↓**	
Gemmiger	↓***	↓***	↓**	
Lachnospiraceae		↓*	↓**	
Lactobacillus			↑*	
Lactococcus				
Peptococcus				
Roseburia			↓*	
Ruminococcaceae	↓**		↓*	
Ruminococcus[a]	↓***		↓*	
Ruminococcus[b]	↓***			
Streptococcus	↑**		↑***	↑*
Subdoligranulum	↓*	↑*	↓**	
Turicibacter	↓***			
Veillonella	↑*			

Proteobacteria	↑***	↑*
Escherichia/Shigella	↑***	
Enterobacteriaceae	↑*	
Verrucomicrobia		
Akkermansia		

↑↓ relative to healthy control

* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$

[a] Lachnospiraceae

[b] Ruminococcaceae

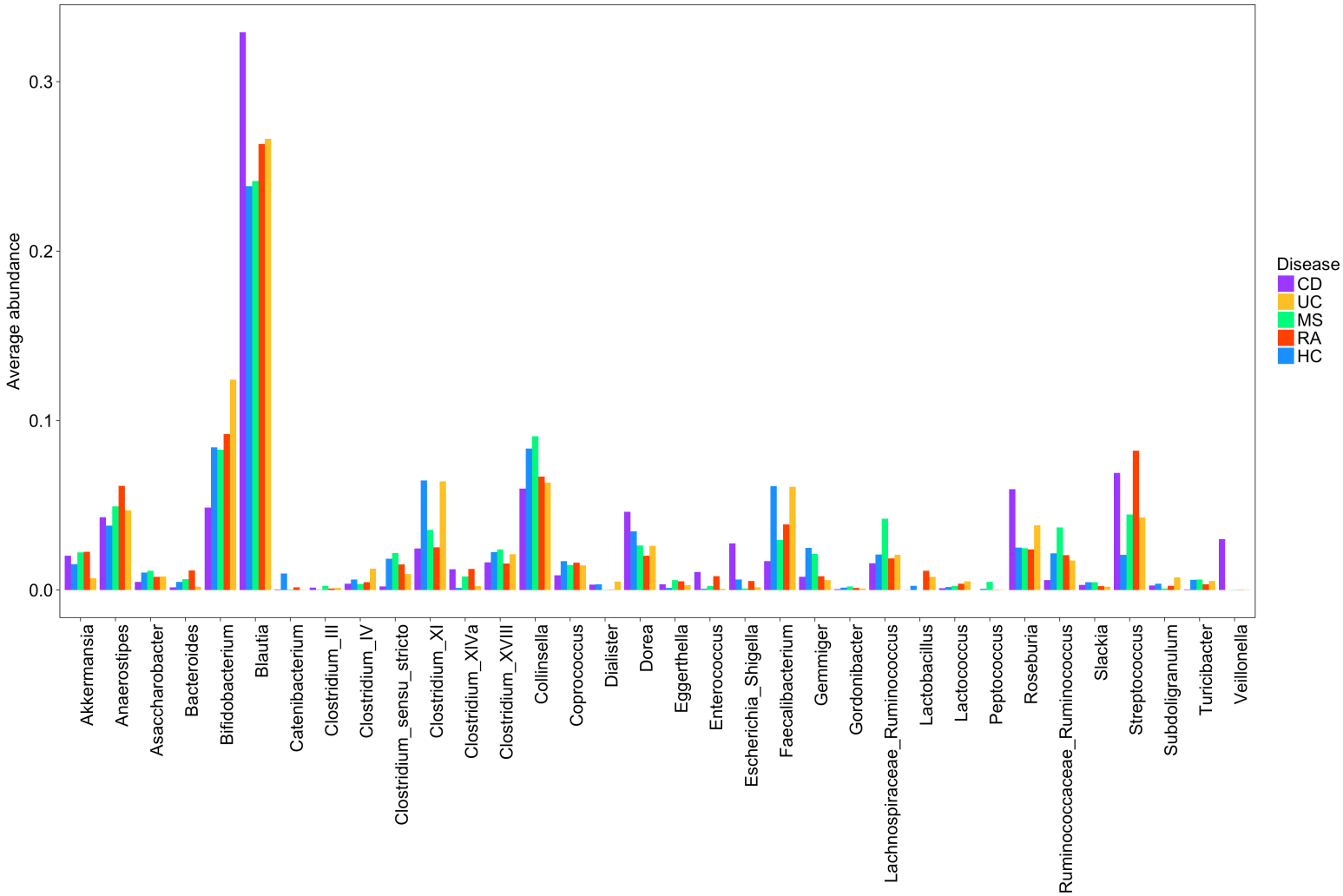

Figure 10. Average abundances for taxa classified to the genus level of taxonomy.

Ulcerative colitis

Significant differences in phyla were not observed between UC and HC. The abundance of Proteobacteria and Verrucomicrobia nearly reached statistical significance prior to multiple correction testing ($P = 0.059$). At a nominal significance ($P \leq 0.05$) the abundance of 18 genera were different, with 8 maintaining significance after FDR correction for multiple comparisons. Genera significantly increased in UC include *Eggerthella, Clostridium* III*, Clostridium* XIVa and *Subdoligranulum* whereas genera decreased include *Enterococcus, Gemmiger* and unclassified Coriobacteriaceae and Lachnospiraceae.

Rheumatoid arthritis

Several taxonomic groups demonstrated varied abundance between RA and HC. At the phylum level, Proteobacteria was increased in RA. *Eggerthella*, *Clostridium* III*, Clostridium* XIVa, *Lactobacillus* and *Streptococcus* were increased in RA; unclassified Coriobacteriaceae, *Clostridium* sensu strico*, Clostridium* XI, *Dorea, Faecalibacterium,* unclassified Firmicutes, *Gemmiger*, unclassified Lachnospiraceae, *Roseburia*, Ruminococcaceae, *Ruminococcus* (Ruminococcaceae) and *Subdoligranulum* were decreased in RA relative to HC. We identified one species belonging to the *Subdoligranulum* genus – *S. variabile*.

Multiple sclerosis

Similar to the relationship of UC to HC, we did not observe phylum level shifts between MS and HC. The abundance of 14 genera was considered significant at a nominal significance ($P \leq 0.05$), however, after FDR correction for multiple comparisons, only *Eggerthella*, *Clostridium* III*, Clostridium* XIVa, *Streptococcus*, unclassified Coriobacteriaceae, *Dialster*, and

Faecalibacterium maintained significance. Utilizing SPINGO analysis, *E. lenta* was the most commonly detected *Eggerthella* species. Several *Streptococcus* species were identified, however, *S. mutans* and *S. thermophilus* were the most common.

Overall trends observed among immune-mediated inflammatory diseases relative to health

Taxonomic differences in IMID relative to health confirm beta-diversity analyses. Specifically, taxonomic shifts were most dramatic in CD, followed by RA. Genus level changes were also apparent in UC and MS but the numbers of distinct taxa significantly variable from HC were considerably less than those observed within CD or RA.

Several dysbiosis patterns were observed between more than one IMID relative to HC. Coriobacteriaceae was decreased in all IMID whereas the abundance of *Eggerthella* and *Clostridium* XIVa was increased in all IMID. *Clostridium* III was increased in UC, MS and RA and similarly, *Streptococcus* was increased in CD, RA and MS. Characteristic decreases were observed with *Faecalibacterium* (*F. prausnitzii*) in CD, RA and MS, and with *Gemmiger* (*G. formicilis*) in CD, UC and RA.

4.1.3 Stool microbiota comparisons between immune-mediated inflammatory diseases

We next aimed to survey microbiota patterns of dysbiosis between IMID to determine similarity or dissimilarity in gut communities (Table 6).

Table 6. Distribution of taxa significantly different between disease cohorts. Taxa unable to be classified to the genus level were assigned to a lower level of taxonomy.

	CD/UC	CD/MS	CD/RA	UC/MS	UC/RA	MS/RA
Actinobacteria	*	*				
Asaccharobacter	**	***	**			
Bifidobacterium	***					
Coriobacteriaceae	**	***				
Gordonibacter		**				
Bacteroidetes	*					
Bacteroides	*					
Firmicutes						
Clostridiales_unclassified		**	*			
*Clostridium*_III		*				
*Clostridium*_IV	**	***	**			
*Clostridium*_sensu_stricto	*	***				*
*Clostridium*_XI		*				
*Clostridium*_XVIII	*	*				
Coprococcus	**	**	*			
Enterococcus		*				
Erysipelotrichaceae						
Faecalibacterium	***	**	**			
Firmicutes_unclassified	***	***	**			*
Gemmiger	*	***	**			
Ruminococcaceae	**				*	
Ruminococcus[a]		***				**
Ruminococcus[b]		***	*			
Turicibacter	***	***	**			
Veillonella		*				
Proteobacteria	*	**				
Escherichia/Shigella	*	**				
Verrucomicrobia					**	
Akkermansia					*	

* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$

[a] Lachnospiraceae

[b] Ruminococcaceae

Crohn's disease and ulcerative colitis

Several taxonomic differences were observed between CD and UC (Figure 11). Specifically, CD was associated with increased abundances of *Escherichia-Shigella* and *Gemmiger*. *Escherichia-*

Shigella was largely defined by the presence of *Escherichia-Shigella coli* as only a few sequences were identified as other species (i.e. spp. *flexneri, fergusonii*). A single *Gemmiger* species was identified via SPINGO, *G. formicilis*. An associated between UC and several genera were reported: *Bifidobacterium*, *Faecalibacterium (*spp. *prausnitzii), Turicibacter* (spp. *sanguinis*)*, Bacteroides, Ruminococcus* of the Lachnospiraceae*, Coprococcus, Asaccharobacter* and *Clostridium* XI, *Clostridium* IV, *Clostridium* sensu strico and *Clostridium* XVIII. Several *Bifidobacterium* species were identified including *B. bifidum* and *B. dentium*. Lachnospiraceae *Ruminococcus* included *R. obeum, R. torques, R. lactaris* and *R. gnavus* among others. Many of the above-listed genera retained statistical significance following multiple comparison correction (Table 6).

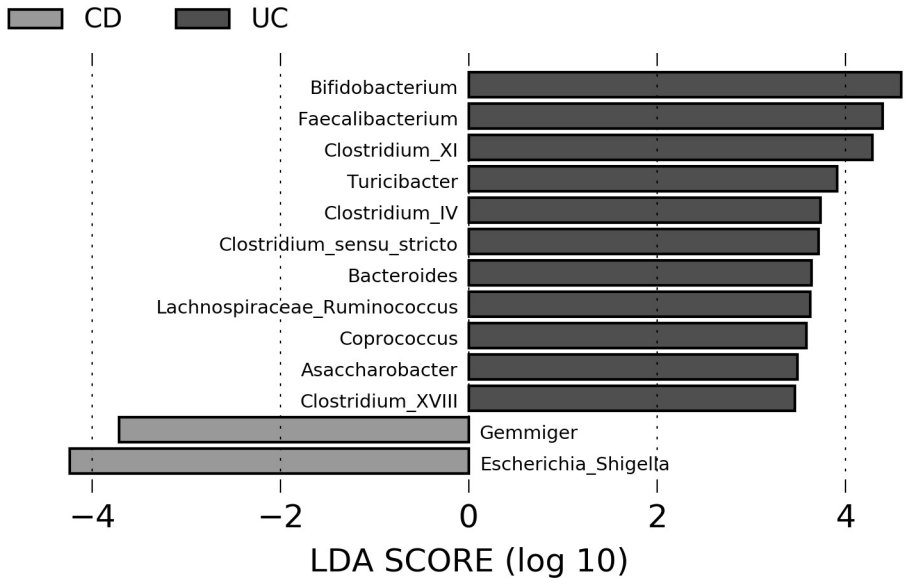

Figure 11. Genera identified as discriminating features between Crohn's disease and ulcerative colitis.

Crohn's disease and multiple sclerosis

Figure 12 shows the relationship between genera enriched in CD and MS and Table 6 shows significance levels following multiple comparison correction. MS was associated with increased abundances of *Ruminococcus* belonging to both the Ruminococcaceae and Lachnospiraceae families, *Faecalibacterium, Gemmiger, Turicibacter, Asaccharobacter, Gordonibacter, Coprococcus, Lactococcus* and *Clostridium* sensu stricto, *Clostridium* XVIII, *Clostridium* XI, *Clostridium* IV and *Clostridium* III. *Blautia, Escherichia-Shigella, Veillonella, Enterococcus* and *Dialster*, in contrast, were associated with CD. *Blautia* in general was the most abundant genus reported amongst all cohorts; several species were detected, however, *B. luti* and *B. faecis* were the most abundant. Moreover, the differential relative abundance of *Veillonella* was largely driven by *V. dispar*.

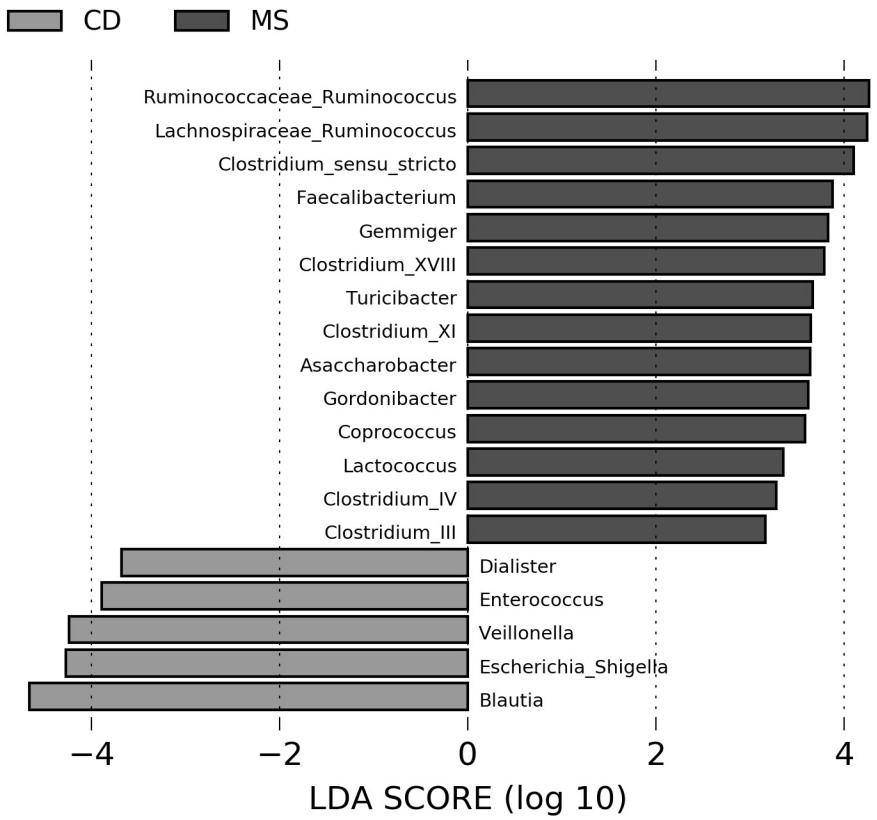

Figure 12. Genera identified as discriminating features between Crohn's disease and multiple sclerosis.

Crohn's disease and rheumatoid arthritis

Compared to RA, CD again demonstrated increasingly abundant populations of *Escherichia-Shigella, Veillonella* and *Enterococcus*. Several *Enterococcus* species were identified, albeit at considerably low abundances; *E. faecium* was detected most frequently. RA was associated with higher abundances of *Faecalibacterium, Ruminococcus* of the Ruminococcaceae and Lachnospiraceae families, *Akkermansia, Turicibacter, Coprococcus, Gemmiger, Gordonibacter, Asaccharobacter* and *Clostridium* sensu stricto and *Clostridium* IV. In our species taxonomic classification, *Akkermansia muciniphila* was detected and additionally, *Gordonibacter*

pamelaeae. Subsequent to multiple comparison correction, many of these genera retained statistical significance (Table 6).

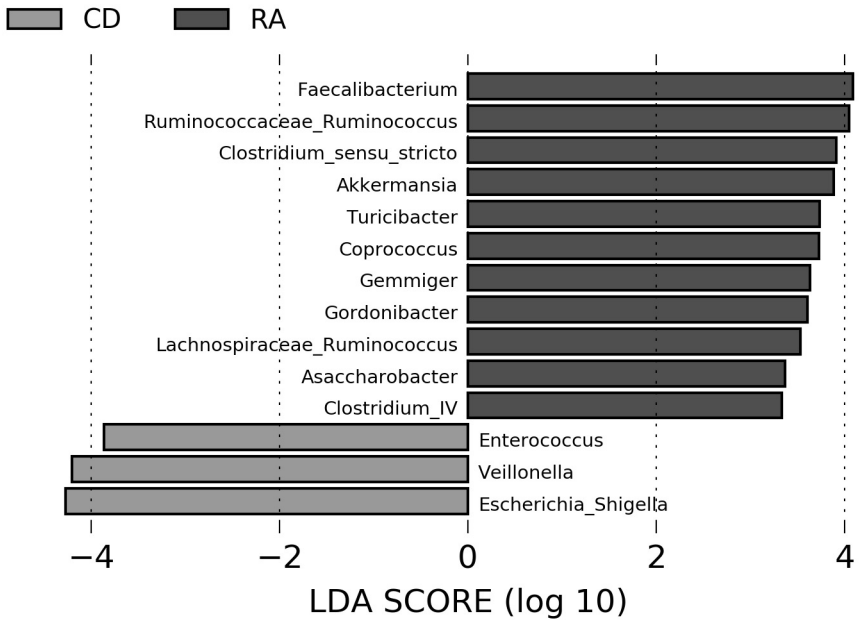

Figure 13. Genera identified as discriminating features between Crohn's disease and rheumatoid arthritis.

Ulcerative colitis and multiple sclerosis

LEfSe analysis of the gut microbial communities of UC and MS reveals differentially abundant genera (Figure 14). These include *Bifidobacterium, Faecalibacterium, Veillonella, Subdoligranulum and Dialster* that are increased in UC and *Ruminococcus* (Ruminococcaceae and Lachnospiraceae), *Akkermansia, Gemmiger, Clostridium* sensu stricto*, Peptococcus* and Gordonibacter in MS. However, as characterized in Table 6 the varied abundance of none of these genera reached statistical significance after FDR multiple corrections.

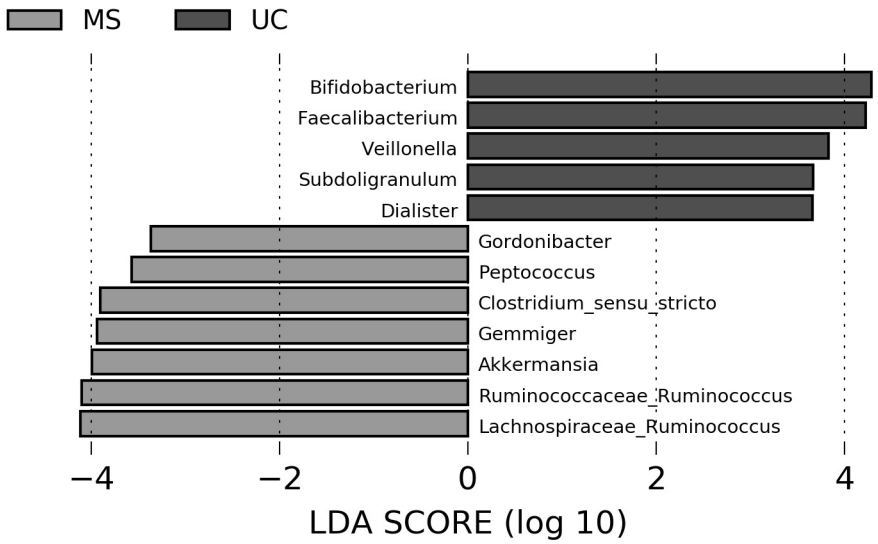

Figure 14. Genera identified as discriminating features between ulcerative colitis and multiple sclerosis.

Ulcerative colitis and rheumatoid arthritis

By LeFSe analysis, *Clostridium* XI and *Clostridium* XVIII, *Bifidobacterium, Faecalibacterium* and *Roseburia* were reportedly increased in UC relative to RA (Figure 15). Increased abundances of *Streptococcus, Akkermansia, Peptococcus* and *Clostridium* XIVa were associated with RA. However, as Table 6 indicates, of the genera listed here, only *Akkermansia* retained significance following FDR correction. Additionally, unclassified Ruminococcaceae and the phylum Verrucomicrobia were increased in RA.

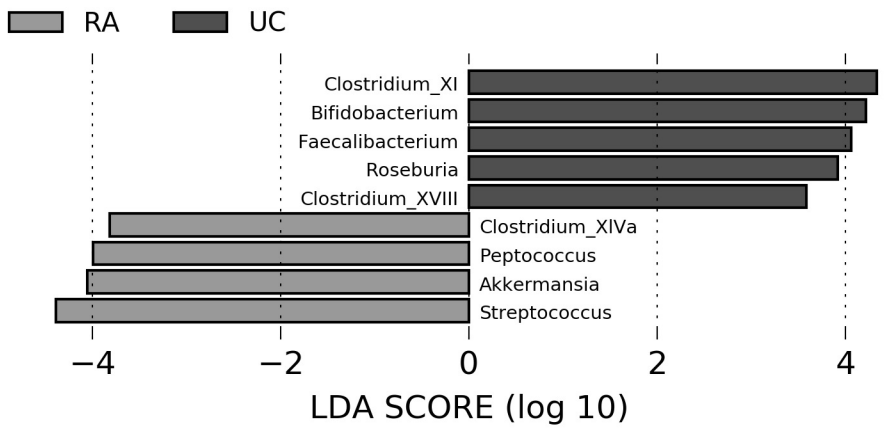

Figure 15. Genera identified as discriminating features between ulcerative colitis and rheumatoid arthritis.

Multiple sclerosis and rheumatoid arthritis

Comparative analysis between MS and RA identified increased abundances of *Ruminococcus* (Lachnospiraceae), *Turicibacter* and *Clostridium* XI, *Clostridium* XVIII and *Clostridium* sensu strico to associate with MS whereas increases in *Streptococcus* and *Subdoligranulum* were characteristic of RA (Figure 16). Following FDR correction, *Ruminococcus* and *Clostridium* sensu stricto remained significant (Table 6).

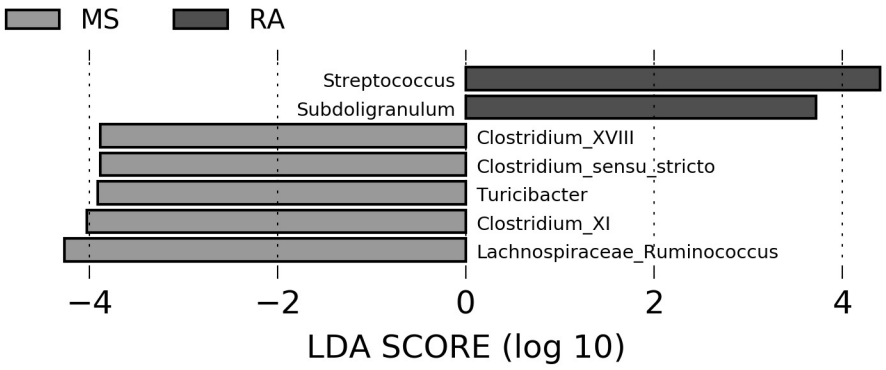

Figure 16. Genera identified as discriminating features between multiple sclerosis and rheumatoid arthritis.

Effect of clinical meta-data on rheumatoid arthritis

For the RA cohort, we also applied analyses of multivariate linear associations between meta- and taxonomic data (see APPENDIX I). Briefly, the relationship between clinical phenotypes (i.e. joints affected), comorbidities and antibiotic usage to microbial clades were tested. We found a positive association between lung comorbidities with *Subdoligranulum* ($P = 0.02$) and *Streptococcus* ($P = 0.02$) abundances. Moreover, phenotypes involving miscellaneous joints (i.e. ankles, back, neck and jaw) negatively associated with *Subdoligranulum* abundance ($P = 0.02$). Finally, we observed an association between *Clostridium* XVIII and prostate cancer in one patient ($P = 0.03$).

4.1.4 Microbial profile optimization

The stool samples displayed an unusual ratio of Gram-negative to Gram-positive bacteria (compared to published literature), which suggests possible bias in DNA extraction methodology. We tested several troubleshooting techniques in order to optimize the microbial profile on select samples, and alongside mock communities (not shown). First, Figure 17 shows

the average abundance of phyla relative to i) original aliquots and secondary aliquots worked up as; ii) fecal slurries and iii) solid stool. Though length of time in -80 °C were the same among aliquots, secondary aliquots had fewer freeze-thaws. These results indicate no difference between primary and secondary aliquots or between fecal slurries and solid stool.

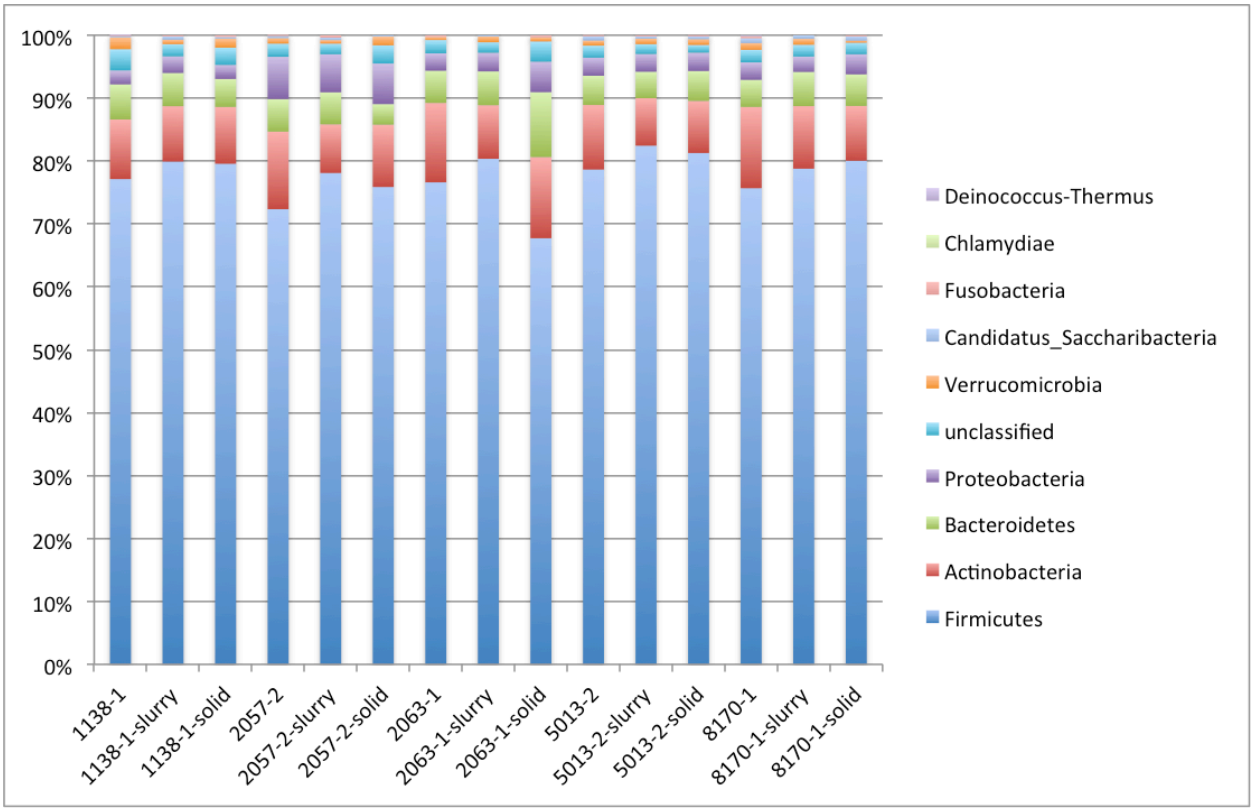

Figure 17. Microbial profile of stool samples utilizing original aliquots, fecal slurries and solid stool.

Modifications to the manufacturer's guidelines and to the protocol outlined in section 3.1.2 were next conducted to attempt optimization of the stool samples microbial profile (Figure 18). Minimal variability was observed between IMID samples with any troubleshooting modification.

However, we can clearly see that the ratio of Gram-positives to Gram-negatives in the healthy fresh sample represents an expected community profile whereas the microbial profile of healthy frozen stool did not show an expected ratio of Gram-positive to Gram-negatives as indicated by a substantial depletion of Bacteroidetes and Proteobacteria.

Figure 18. Microbial profile of stool samples utilizing several troubleshooting modifications. (st = standard ZR protocol, mod = modified ZR protocol, mobio = MoBio PowerSoil DNA Isolation Kit; 1, 5, 10 = bead beating in minutes; 300, 400 = lysis solution in µl).

4.1.5 Inferred functionality of the stool microbiome in immune-mediated inflammatory disease and in health

A total of 274 KEGG pathways were identified. Several differences in the abundance of KOs belonging to functional gene categories were identified between IMID and HC (APPENDIX III). In each cohort we observed statistically significant increases in specific genes and utilizing LEfSe analysis these genes were ranked into their ability to act as biomarkers of disease (or health). All genes found to significantly associate with any cohort are illustrated in Figure 19; briefly, the strongest associations were reported between CD and ABC transports, UC and metabolism of nicotinate and nicotinamide, MS and ribosomes, RA and several metabolism pathways (notably amino and nucleotide sugar metabolism) and HC with electron transfer carriers. See APPENDIX III for median abundances, *P*-values and FDR corrected *P*-values.

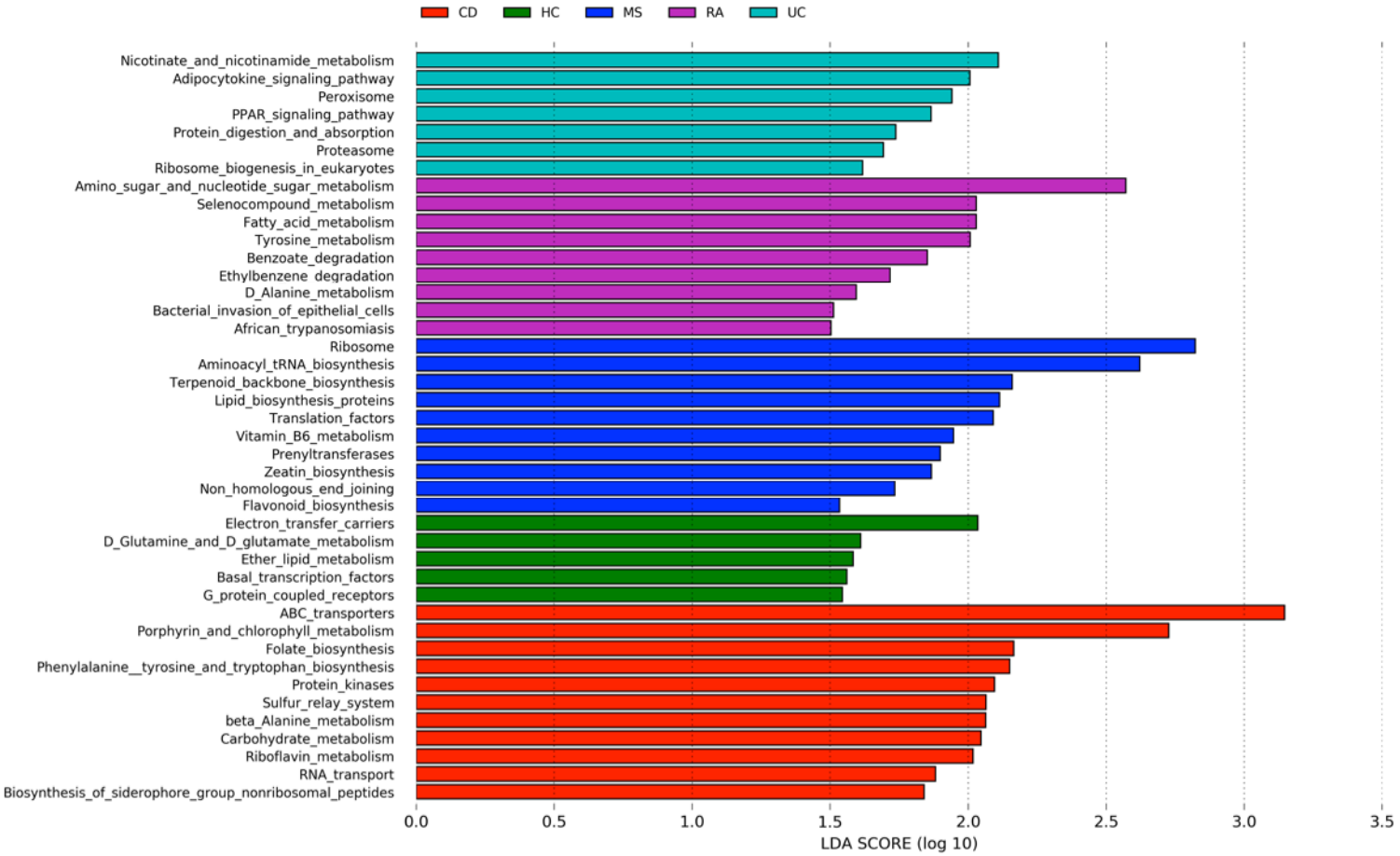

Figure 19. Predicted functional composition of metagenomes associated with stool in immune-mediated inflammatory diseases and in health.

4.2 Understanding the mucosal microbiome in inflammatory bowel disease

Filtering out low quality, chimeric and non-bacterial reads generated 22,913,986 high quality 16S rDNA reads with an average amplicon length of 74 bp. Sequences were clustered into 13,040 OTUs based on their shared sequence similarity at a 97% threshold (3% sequence divergence). The average number of sequences per sample was 181,806; 40 of the 166 samples were excluded from statistical analysis due to insufficient sequence numbers (<1000). The average coverage for OTU characterization based on Good's coverage was 99.5%.

4.2.1 Microbial diversity of the mucosal microbiota in inflammatory bowel disease and in health

Alpha diversity

Using a number of different measures (Shannon, Simpson, inverse-Simpson) to survey bacterial diversity we found that overall there were significant differences between particular inflammatory states within each disease group using each diversity index (Figure 20). Significant differences (P<0.05) were observed when comparing non-inflamed tissues from CD to UC and likewise between inflamed tissues of UC to non-IBD. In general, the diversity of inflamed tissues of UC (or non-IBD) was higher than inflamed tissues of CD though these trends did not reach statistical significance. Comparable diversity was observed between anatomical sites of a specific disease (CD or UC) regardless of inflammatory state and anatomical sites when compared between diseases (Figure 21). However, microbial diversity was noticeably lowest in the cecum when comparing gastrointestinal tract sites of CD. Trends were observed in cecal (UC > non-IBD > CD) and mid-colonic diversity (non-IBD > UC > CD).

Significant differences ($P < 0.05$) in community richness (Chao1 and ACE) when considering the inflammatory state of tissue were observed and presented a noteworthy trend (CD non-inflamed > CD inflamed > UC non-inflamed > non-IBD > UC inflamed). Significant differences ($P<0.05$) of community richness were observed between non-inflamed tissues of CD and UC, and inflamed UC and non-IBD tissues. We observed an interesting pattern in richness within the CD cohort: richness tended to be lowest in the ileum and increased distally to the mid-colon. Furthermore, within each gut compartment (ileum, cecum, mid-colon, rectum) richness was higher in IBD compared to controls and greater in CD than UC.

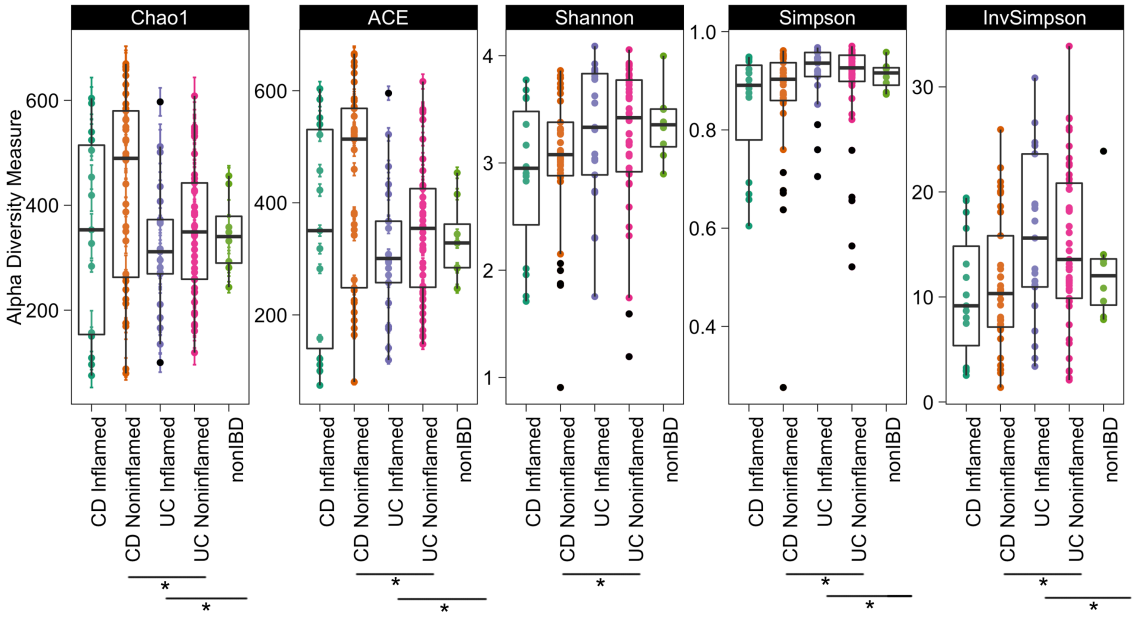

Figure 20. Alpha-diversity plots between disease cohort inflammatory states. Community richness estimated by ACE and Chao1; diversity estimated by Shannon, Simpson and inverse-Simpson. Differences considered significant at *$P < 0.05$.

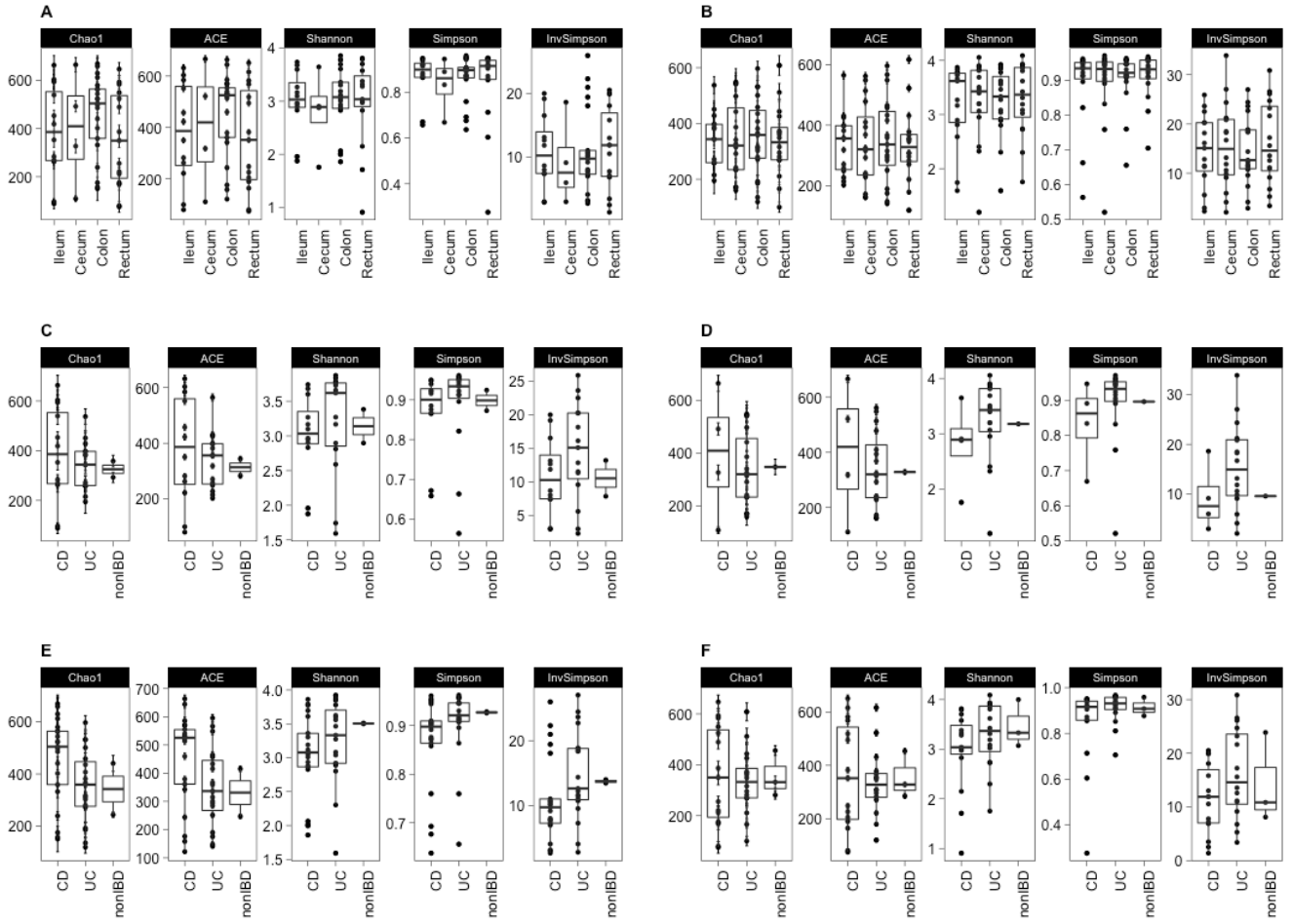

Figure 21. Alpha-diversity plots between gastrointestinal compartments in (A) CD and (B) UC; and between disease cohorts in the (C) ileum; (D) cecum; (E) mid-colon and (F) rectum. Community richness estimated by ACE and Chao1; diversity estimated by Shannon, Simpson and inverse-Simpson.

Beta diversity

To compare the overall structure of the mucosal microbiota between (i) CD, UC and non-IBD, (ii) inflammatory states, (iii) anatomical sites of the gastrointestinal tract and (iv) patients, we visualized Bray-Curtis distances between samples using PCoA. The results of PCoA indicate a difference in structure of the microbiota between CD, UC and non-IBD (Figure 22). Although to

an extent, overlap of all disease groups was observed, samples tended to cluster together based on disease. PCoA of inflammatory states overlapped and could not be well separated (Figure 23) however, cluster analysis of only inflamed tissues from CD and UC shows clear separation as did the non-inflamed tissues (Figure 24). Inflamed and non-inflamed biopsies of CD or UC share the characteristic that there are no structural patterns apparent. There was also no separation based on the anatomical site of the biopsy. Cluster analysis indicates that the overall gut microbiota composition are more dependent on inter-individual variation rather than disease. Our results indicate tissues from the same participant are most similar regardless of the inflammatory state and/or sampling location of the specimen.

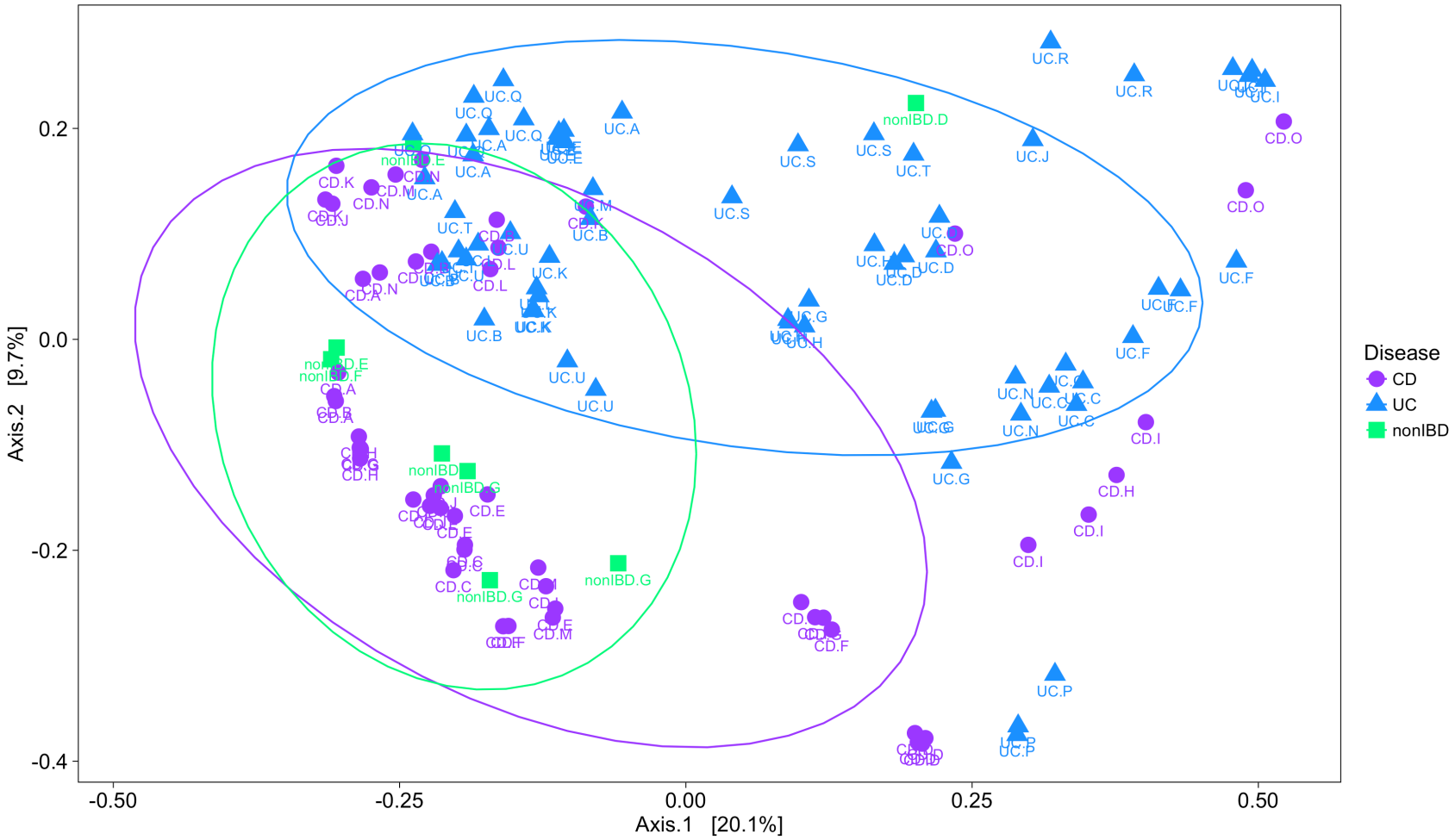

Figure 22. Principal coordinate analysis based on the overall structure of the mucosal microbiota in all biopsy samples. Each data point represents an individual sample. Principal coordinate analysis was calculated using Bray-Curtis distances with a multivariate t-distribution. Ellipses represent an 80% confidence level. Colour/symbol is indicative of disease cohort. Data points are labeled by subject disease type.

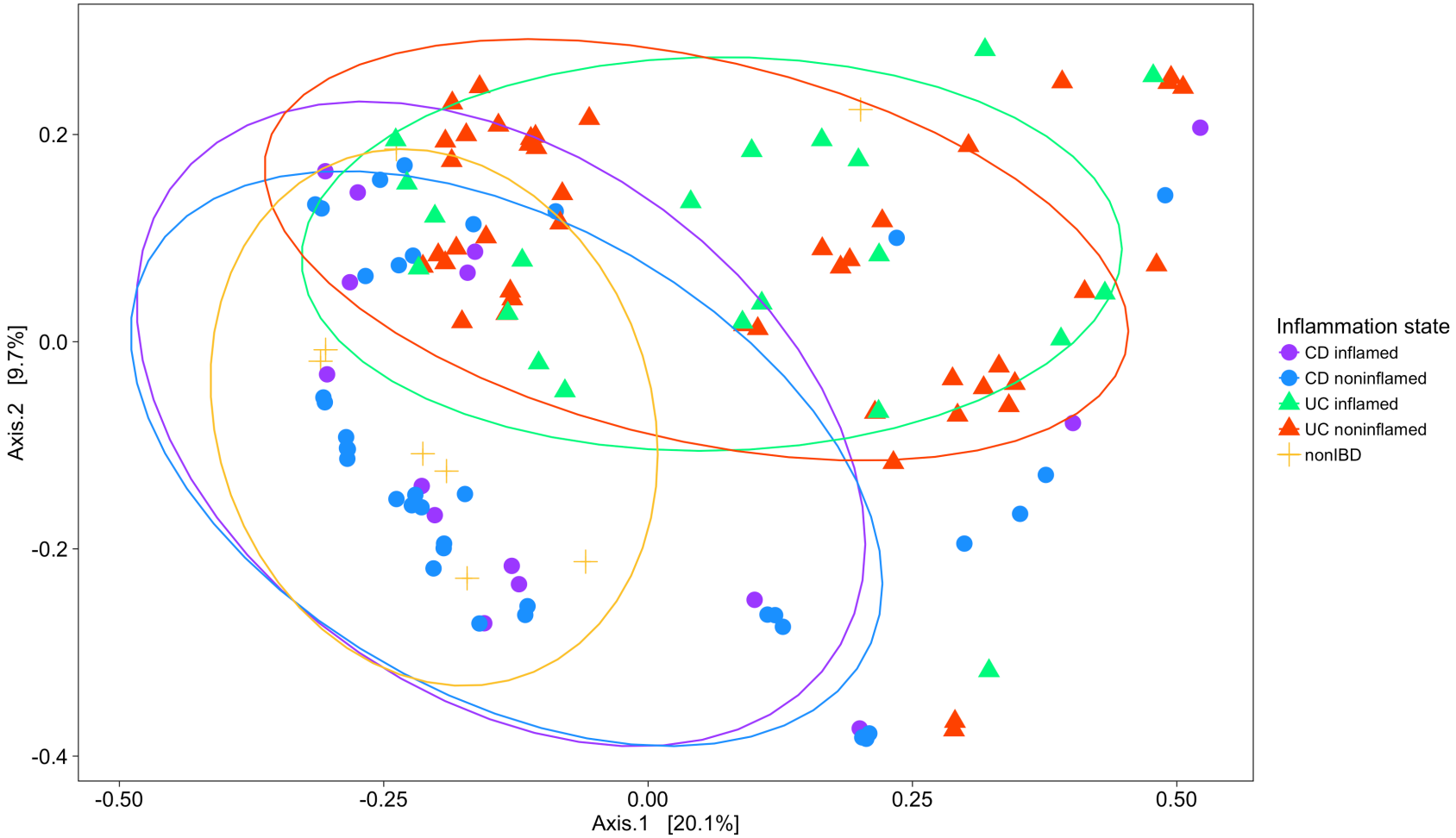

Figure 23. Principal coordinate analysis based on the overall structure of the mucosal microbiota in all biopsy samples with corresponding inflammatory state. Each data point represents an individual sample. Principal coordinate analysis was calculated using Bray-Curtis distances with a multivariate t-distribution. Ellipses represent an 80% confidence level. Colour is indicative of inflammatory cohort and symbol is indicative of disease cohort.

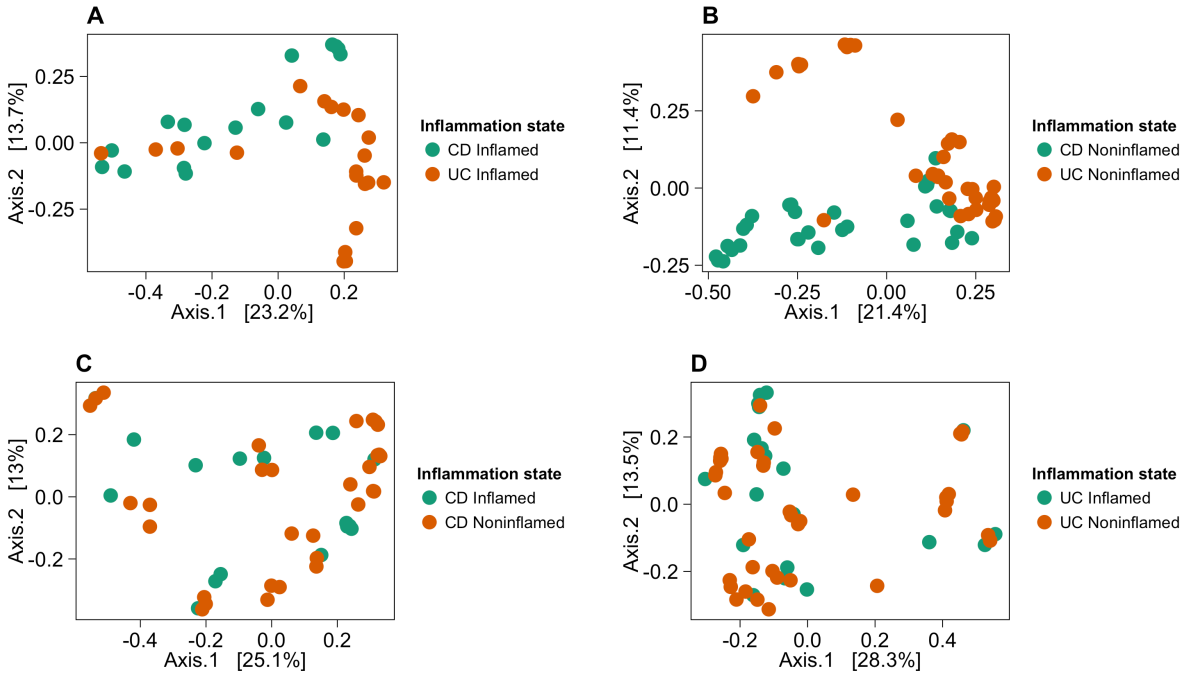

Figure 24. Principal coordinate analysis based on the overall structure of the mucosal microbiota in (A) inflamed mucosa, (B) noninflamed mucosa, (C) inflamed and non-inflamed mucosa of CD and (D) inflamed and non-inflamed mucosa of UC. Principal coordinate analysis was calculated using Bray-Curtis distances.

1.1.1 Overall taxonomic structure of the mucosal microbiota

The filtered data included 131 OTUs (clustered at >97% sequence identity) that were assigned to 6 bacterial phyla. The Firmicutes (47.0%) and Bacteroidetes (36.0%) were the most diverse phyla as the bulk of genera identified belonged to one of these groups. We identified the genus *Bacteroides* to be the most abundant bacterial genus accounting for 28.7% of reads. Further, 31 *Bacteroides* species were identified, the most abundant being *B. vulgatus, B. thetaiotomicron* and *B. fragilis*. Proteobacteria represented 13.1% of sequences, of which *Escherichia-Shigella* (which are indistinguishable as a 16S-based phylotype) accounted for 5.6%. *Escherichia-*

Table 7. Microbial distribution of taxa differentially abundant among inflammatory states.

	Inflamed	Non-inflamed	CD			UC		
	CD/UC	CD/UC	Inflamed	Non-inflamed	Inflamed/ Non-inflamed	Inflamed	Non-inflamed	Inflamed/ Non-inflamed
Firmicutes	↓↓*	↓↓***					↑**	
Clostridium		↓↓*						
Coprococcus				↓*				
Dorea							↓*	
Enterococcus		↓↓*						
Faecalibacterium	↓↓*	↓↓**						
Oscillibacter				↓*			↓*	
Phascolarctobacterium		↑↑*					↓*	
Sporacetigenium		↓↓*						
Streptococcus		↓↓***						
Turicibacter		↓↓*						
Bacteroidetes	↑↑**	↑↑***				↓**	↓***	
Alistipes		↑↑**				↓*	↓*	
Bacteroides	↑↑**	↑↑***					↓*	
Parabacteroides						↓*	↓*	
Prevotella							↓*	
Proteobacteria	↓↓*	↓↓***				↑*	↑**	
Haemophilus		↓↓**						
Klebsiella		↓↓**						
Marinobacter							↑*	
Pseudomonas	↓↓***	↓↓***				↑*	↑*	
Verrucomicrobia								
Actinobacteria								
Actinomyces		↓↓***					↑*	
Fusobacteria	↑↑*							

↓↓ and ↑↑ observed in CD relative to UC; ↑ and ↓ relative to nonIBD
* = $P<0.05$, ** = $P<0.01$, *** = $P<0.001$

4.2.5 Inferred functionality of the mucosal microbiome in distinct inflammatory states of inflammatory bowel disease and in health

We have identified 284 KEGG pathways. A number of differences in the abundance of KOs belonging to functional gene categories were identified among distinct inflammatory states in IBD and nonIBD (Figure 25) and as such, identified as potential biomarkers. nonIBD was characterized by increased genes for energy and taurine/hyptotaurine metabolism, primary and secondary bile acid biosynthesis, carbon fixation pathways in prokaryotes, RNA degradation, peroxisome, adipocytokine signaling pathway and glutamatergic synapse. In UC, inflamed mucosa had significantly increased genes for atrazine degradation and carotenoid biosynthesis whereas noninflamed mucosa had increased genes pertaining to fluorobenzoate degradation. Inflamed CD mucosa was associated with increased biotin metabolism genes and no genes were increased in noninflamed CD mucosa. Of these pathways identified as potential biomarkers, multiple correction testing was performed and is presented in Table 8.

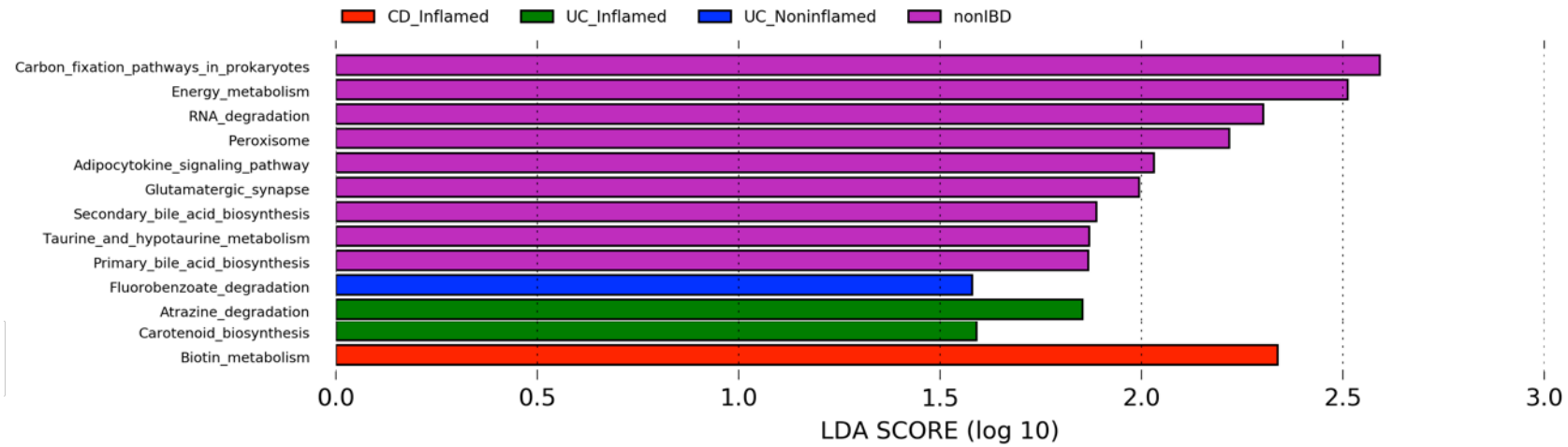

Figure 25. Predicted functional composition of metagenomes associated with distinct inflammatory states in inflammatory bowel disease and in health.

Table 8. Enriched pathways in inflammatory bowel disease inflammatory states identified as biomarkers.

	CD[a]		UC[a]		nonIBD[a]		
KEGG pathway	Inflamed	Noninflamed	Inflamed	Noninflamed		*P*-value[b]	*P*-corr[c]
Adipocytokine signaling pathway	6.7E-02	7.3E-02	5.6E-02	5.3E-02	7.7E-02	3.5E-04	1.6E-03
Atrazine degradation	1.8E-02	1.7E-02	2.5E-02	2.3E-02	1.3E-02	3.2E-03	9.7E-03
Biotin metabolism	2.1E-01	2.0E-01	1.7E-01	1.6E-01	1.9E-01	2.6E-07	9.3E-06
Carbon fixation pathways in prokaryotes	9.5E-01	9.7E-01	8.8E-01	8.8E-01	9.8E-01	4.1E-04	1.8E-03
Carotenoid biosynthesis	2.6E-03	1.7E-03	3.8E-03	3.4E-03	0.0E+00	3.6E-06	3.1E-05
Energy metabolism	8.6E-01	8.9E-01	8.5E-01	8.5E-01	9.2E-01	1.2E-03	4.3E-03
Fluorobenzoate degradation	2.7E-03	1.8E-03	9.8E-03	7.1E-03	7.1E-04	1.8E-04	8.7E-04
Glutamatergic synapse	1.2E-01	1.2E-01	1.0E-01	1.0E-01	1.1E-01	7.6E-10	2.2E-07
Peroxisome	1.7E-01	1.8E-01	1.6E-01	1.6E-01	2.0E-01	3.7E-06	3.1E-05
Primary bile acid biosynthesis	3.3E-02	3.8E-02	2.7E-02	2.3E-02	3.9E-02	3.2E-06	3.1E-05
RNA degradation	4.6E-01	4.8E-01	4.6E-01	4.5E-01	4.8E-01	4.0E-03	1.1E-02
Secondary bile acid biosynthesis	3.3E-02	3.8E-02	2.7E-02	2.3E-02	3.9E-02	2.7E-06	3.1E-05
Taurine and hypotaurine metabolism	1.1E-01	1.2E-01	1.1E-01	1.1E-01	1.3E-01	3.2E-06	3.1E-05

[a] median abundance

[b] *P*-values calculated via Kruskal-Wallis

[c] FDR corrected *P*-values; indicate a significantly different median abundance of pathways between inflammatory states

4.2.6 Comparative microbiota analysis of gut compartments within Crohn's disease and ulcerative colitis

In analyses comparing genera from distinct anatomical gastrointestinal tract sites within each disease group, there were no significant differences in the average abundances. Figure 26 shows that regardless of the genera in either CD or UC, their abundance was uniform across all sites with a few exceptions: *Escherichia-Shigella* in CD appears to be numerically highest in the cecum and *Pseudomonas* in CD also appears to be numerically highest in the rectum.

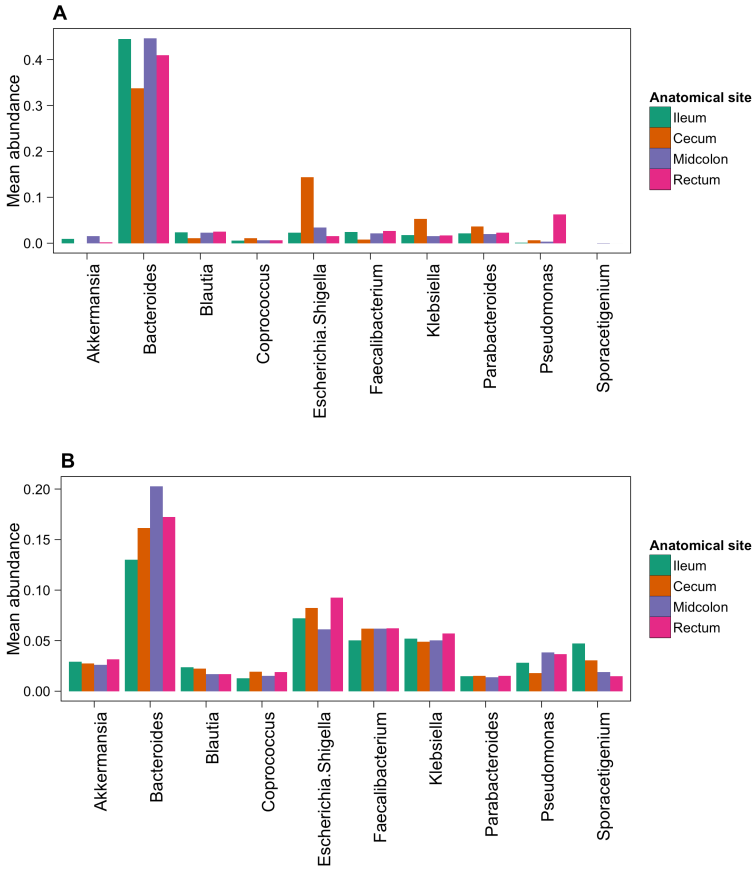

Figure 26. Average abundance of the 10 most abundant genera in the CD and UC mucosal microbiota within specific gastrointestinal compartments: ileum, cecum, mid-colon, and rectum in (A) CD and (B) UC.

4.2.7 Comparative microbiota analysis of specific gut compartments between Crohn's disease, ulcerative colitis and in health

We also used a taxonomy-based analysis to recognize particular phyla (Figure 27) or genera (Figure 28) found to be differentially represented in tissues of the ileum, cecum, mid-colon or rectum between distinct disease groups. We have demonstrated that the abundance of Bacteroidetes significantly varies between disease groups and that their degree of variability is dependent on particular compartments of the gastrointestinal tract. For example, the disproportional representation of Bacteroidetes is most clear in ileal and mid-colonic tissue, followed by rectal tissue. We have also shown the abundance of Firmicutes to vary between disease groups in ileal and mid-colonic tissue. Similarly, the Proteobacteria were disproportionally abundant within ileal, mid-colonic and rectal tissue. Of importance, no significant variability was observed between disease groups within the cecal compartment. Additionally, we have identified significant differences between genera including *Bacteroides* and *Pseudomonas* among the ileal, mid-colonic and rectal tissues.

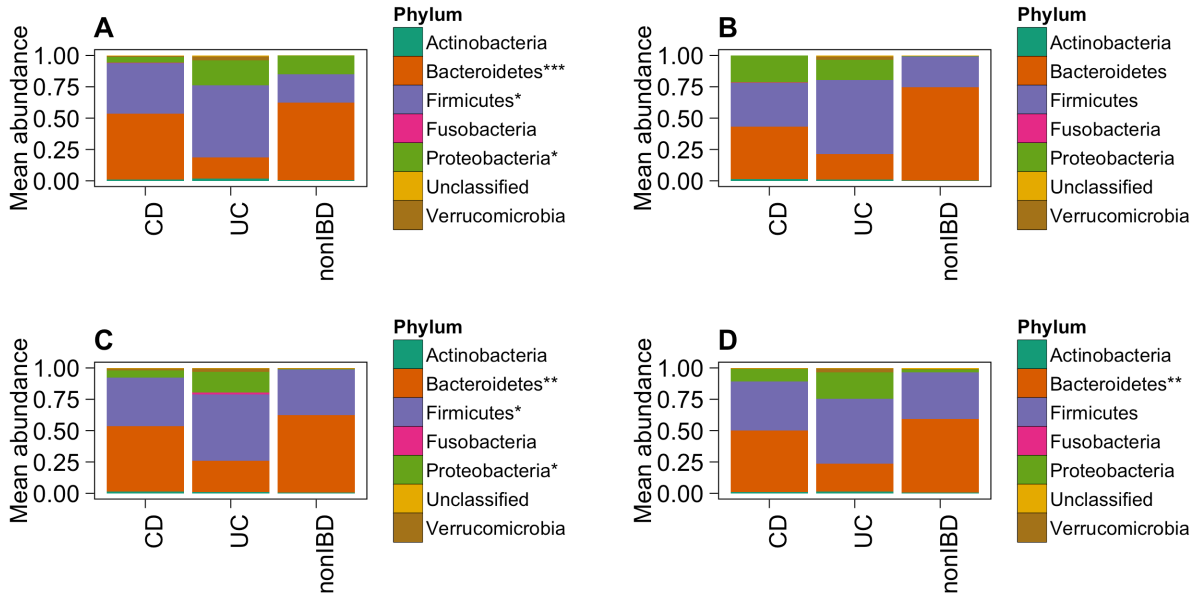

Figure 27. Phylum-level comparisons between CD, UC and non-IBD within the (A) ileum (B) cecum (C) mid-colon and (D) rectum. Differences considered significant at *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

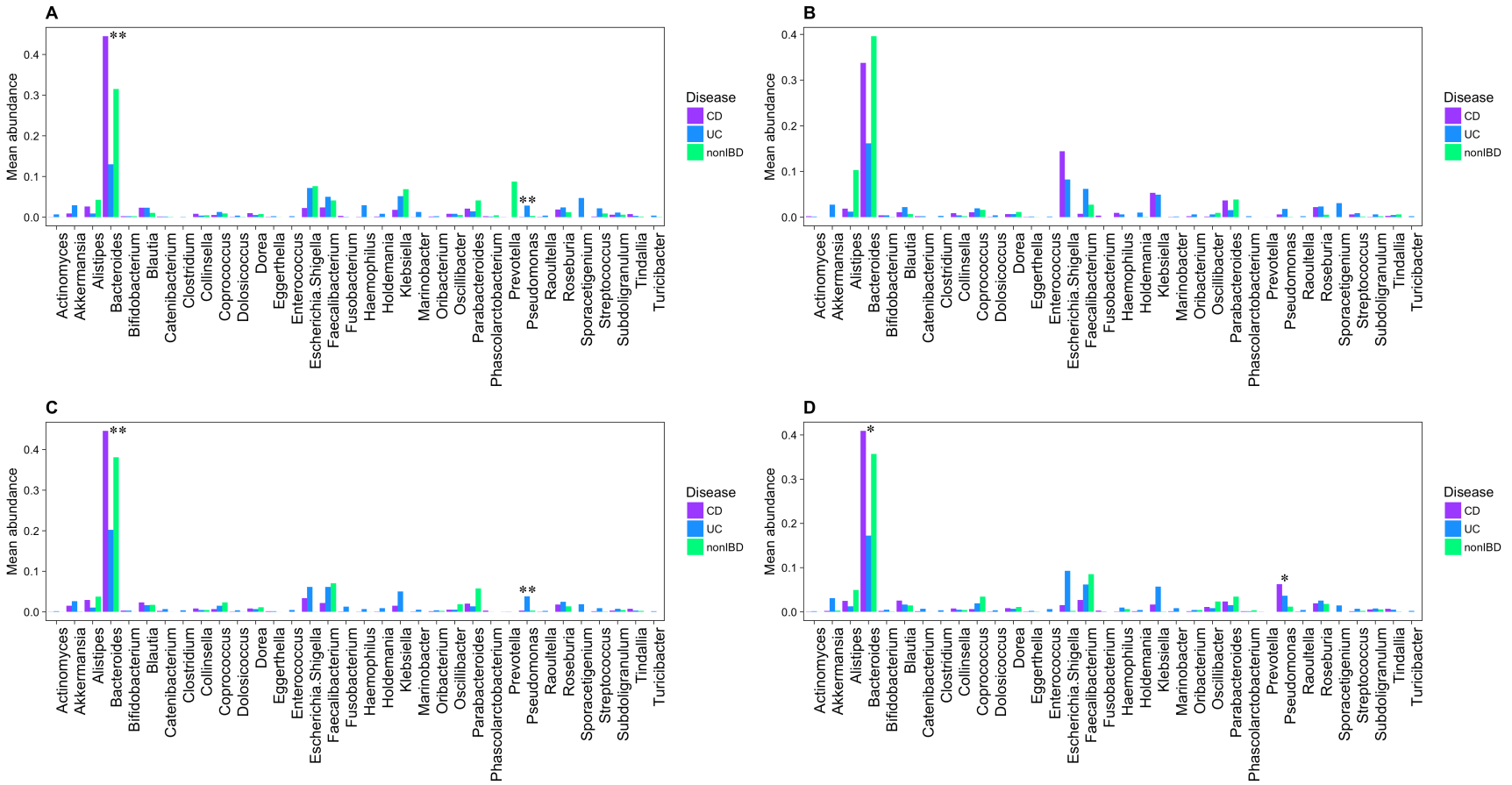

Figure 28. Genus-level comparisons between CD, UC and non-IBD within the (A) ileum (B) cecum (C) mid-colon and (D) rectum. Differences considered significant at *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.

4.2.8 Inferred functionality of the mucosal microbiome in distinct gut compartments in inflammatory bowel disease and in health

We performed an inferred functionality analysis to determine genes identified at a significantly different abundance in CD or UC gut compartments compared to in health. Several differences in the abundance of KOs belonging to functional gene categories were identified among gut compartments in IBD and nonIBD (Figure 29) and were identified as potential biomarkers. Specifically, numerous genes were associated with nonIBD and hence identified at a lower abundance in CD and UC compartments including: butanoate and tryptophan metabolism and carbon fixation pathways in prokaryotes in the ileum; sphingolipid metabolism, lysosome, glycosaminoglycan and other glycan degradation in the cecum; arginine and proline metabolism, lysine biosynthesis, and general function prediction in the colon and; histidine, thiamine and alanine, aspartate and glutamate metabolism, transcription machinery, homologous recombination, amino acid related enzymes and chaperones and folding catalysts in the rectum. In UC, genes for ABC transporters and benzoate degradation were associated with the ileum, and secretion system genes with the rectum. No genes could be identified as biomarkers of disease in the cecum or colon of UC. Of note, no genes were associated with any CD gut compartment. Of these pathways identified as potential biomarkers, multiple correction testing was performed and is characterized in Table 9.

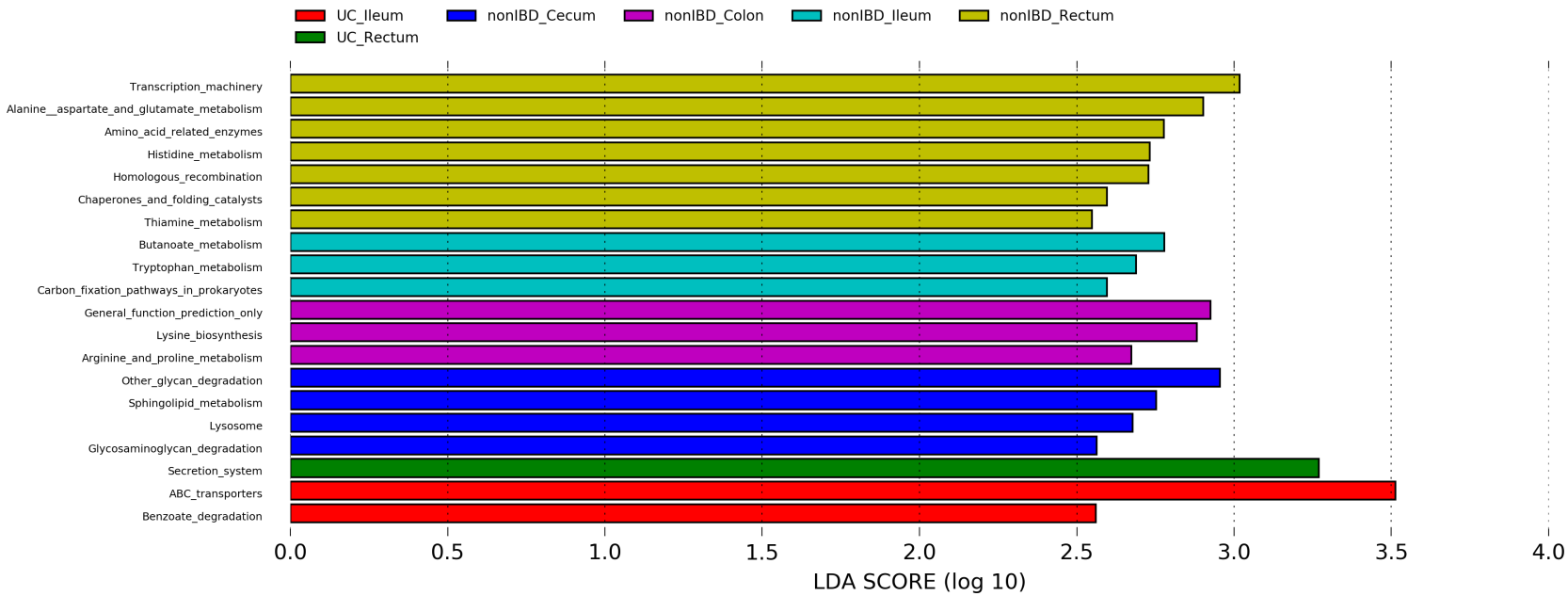

Figure 29. Predicted functional composition of metagenomes associated with gut compartments in inflammatory bowel disease and in health.

Table 9. Enriched pathways in inflammatory bowel disease gut compartments identified as biomarkers.

	CD[a]				UC[a]				nonIBD[a]					
Kegg Pathway	Ileum	Cecum	Colon	Rectum	Ileum	Cecum	Colon	Rectum	Ileum	Cecum	Colon	Rectum	*P*-value[b]	*P*-corr[c]
ABC transporters	3.008	3.076	2.982	3.141	3.613	3.533	3.474	3.493	3.068	2.992	2.987	2.981	2.6E-02	9.9E-02
Alanine aspartate and glutamate metabolism	1.185	1.191	1.173	1.163	1.079	1.090	1.105	1.092	1.138	1.186	1.199	1.166	1.1E-04	3.0E-03
Amino acid related enzymes	1.472	1.447	1.473	1.465	1.396	1.420	1.409	1.442	1.429	1.456	1.510	1.484	5.0E-02	1.6E-01
Arginine and proline metabolism	1.364	1.344	1.348	1.372	1.244	1.274	1.256	1.272	1.291	1.359	1.379	1.329	1.4E-03	9.7E-03
Benzoate degradation	0.179	0.166	0.168	0.177	0.211	0.210	0.203	0.200	0.236	0.187	0.180	0.177	3.1E-02	1.1E-01
Butanoate metabolism	0.532	0.520	0.521	0.528	0.575	0.588	0.581	0.591	0.678	0.545	0.536	0.534	2.2E-04	3.5E-03
Carbon fixation pathways in prokaryotes	0.967	0.904	0.972	0.957	0.879	0.899	0.891	0.883	0.994	0.952	0.983	1.000	2.3E-02	8.8E-02
Chaperones and folding catalysts	1.046	1.048	1.061	1.034	0.993	0.990	0.999	1.007	1.039	1.077	1.037	1.082	2.2E-02	8.6E-02
General function prediction only	3.653	3.647	3.625	3.613	3.497	3.531	3.537	3.511	3.636	3.643	3.652	3.657	2.7E-04	3.8E-03
Glycosaminoglycan degradation	0.086	0.121	0.114	0.089	0.044	0.076	0.066	0.061	0.104	0.130	0.105	0.128	7.8E-04	6.0E-03
Histidine metabolism	0.663	0.656	0.671	0.668	0.614	0.631	0.625	0.618	0.662	0.675	0.700	0.668	1.3E-04	3.3E-03
Homologous recombination	0.902	0.896	0.890	0.895	0.837	0.861	0.856	0.865	0.850	0.862	0.918	0.931	3.6E-02	1.2E-01
Lysine biosynthesis	0.941	0.957	0.936	0.928	0.904	0.884	0.872	0.866	0.801	0.905	0.962	0.945	2.9E-02	1.0E-01
Lysosome	0.148	0.168	0.168	0.135	0.069	0.104	0.092	0.099	0.154	0.191	0.143	0.186	2.3E-04	3.5E-03
Other glycan degradation	0.357	0.357	0.376	0.327	0.265	0.268	0.248	0.276	0.370	0.396	0.359	0.379	3.9E-04	4.9E-03
Secretion system	1.019	1.133	1.078	1.052	1.171	1.199	1.240	1.221	1.448	1.108	0.981	1.025	2.2E-02	8.6E-02
Sphingolipid metabolism	0.289	0.299	0.305	0.264	0.242	0.252	0.219	0.253	0.246	0.321	0.297	0.286	1.8E-03	1.2E-02
Thiamine metabolism	0.538	0.537	0.536	0.537	0.479	0.481	0.491	0.496	0.489	0.530	0.547	0.551	6.3E-04	5.4E-03
Transcription machinery	1.086	1.072	1.083	1.071	0.977	0.972	0.945	0.992	0.886	1.065	1.108	1.087	2.6E-03	1.6E-02
Tryptophan metabolism	0.125	0.133	0.119	0.119	0.155	0.153	0.167	0.157	0.243	0.144	0.114	0.128	5.0E-02	1.6E-01

[a] median abundance
[b] *P*-values calculated via Kruskal-Wallis
[c] FDR corrected *P*-values

4.3 A potential inflammatory bowel disease environmental etiology

Filtering out low quality, chimeric and non-bacterial sequences generated 83,392 high quality 16S rDNA sequences with an average amplicon length of 456 bp. The sequencing depth was comparable among groups with 2323 ± 770 in HIA and 2253 ± 693 in LIA. Sequences were clustered into 749 OTUs based on their shared sequence similarity at a 97% threshold (3% sequence divergence). Four samples were excluded from statistical analysis due to insufficient sequence numbers (<500).

4.3.1 Bacterial community composition characteristics

The average coverage for OTU characterization based on Good's coverage was 99.1% (min-90.0%; max-99.8%). Richness (Chao1 and ACE) and diversity (Shannon and Simpson) were comparable among incidence areas (Figure 30A) and sample type (Figure 30B). Investigation of the 7 sampling locations identified differences of richness as estimated by ACE (P=0.046; Figure 30C). To further investigate richness estimates and diversity indices, sample locations were divided into three respective water distribution systems: (i) all Winnipeg locations receiving its water from Shoal lake, (ii) Brandon (Assiniboine River) and (iii) Steinbach (ground-water). Both Chao (P=0.04) and ACE (P=0.01) richness estimators indicate species richness to vary among the different water systems (Figure 30D). For example, species richness was highest in Steinbach followed by Winnipeg (Shoal Lake) and Brandon. In terms of IBD incidence, species richness was lowest in Brandon and highest in Steinbach which were selected communities of HIA and LIA, respectively.

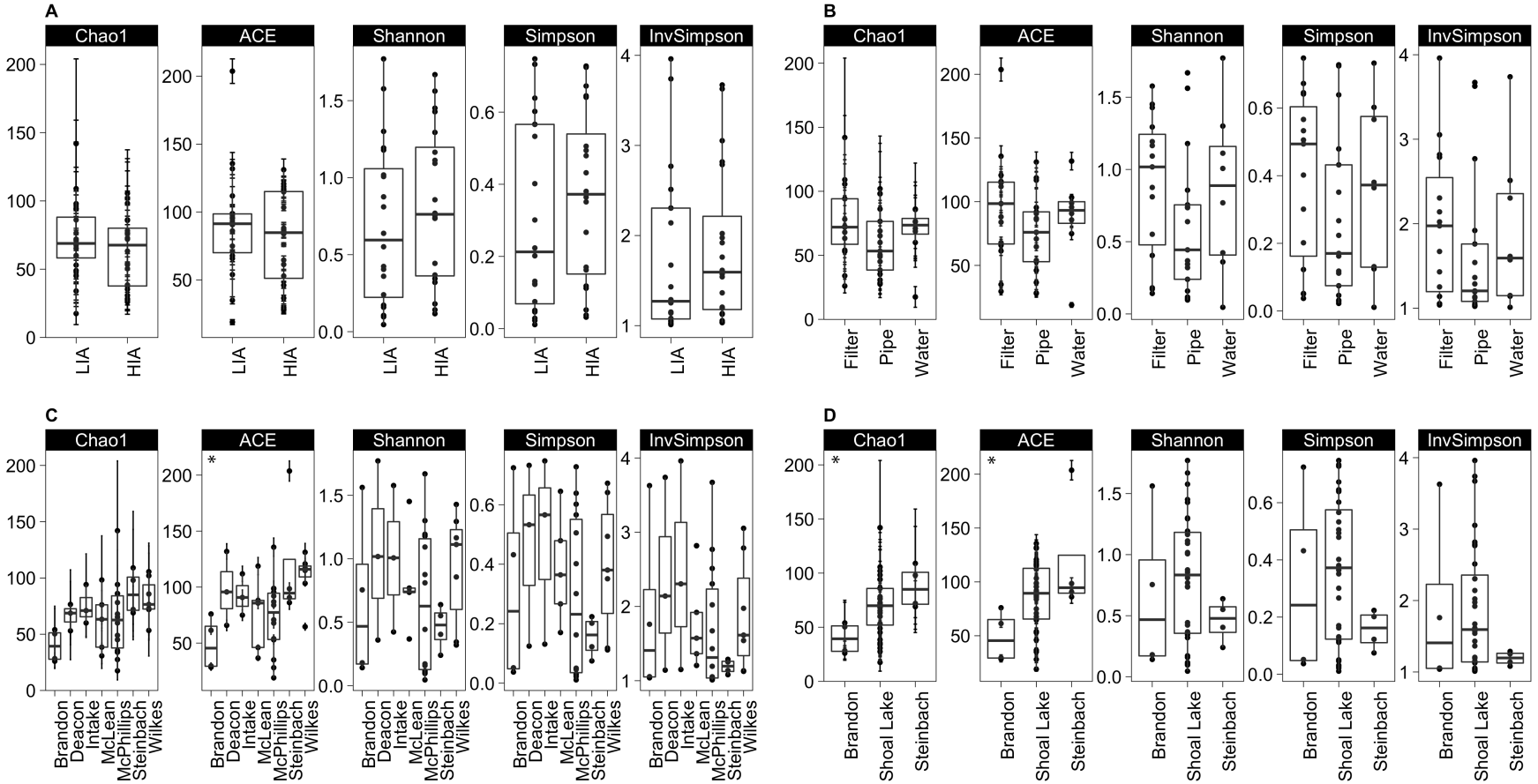

Figure 30. Alpha-diversity plots between (A) low and high incidence areas, (B) sample type, (C) sample location and (D) water distribution networks. Community richness estimated by ACE and Chao1; diversity estimated by Shannon, Simpson and inverse-Simpson. Differences considered significant at $*P < 0.05$.

4.3.2 Bacterial population comparisons between low and high incidence areas of inflammatory bowel disease

Sequences were assigned to 10 phyla, of which phylotypes belonging to the Proteobacteria dominated the overall microbial population. The average relative abundance of Proteobacteria was 92%. The abundance of Proteobacteria was higher in LIA (94% ± 21) in contrast to HIA (91% ± 22; Figure 31A). Furthermore, while there was a lack of significant variation of Proteobacteria between sample type (Figure 31B) and locations (Figure 31C), there was a significant distinction in abundance between water treatment systems (P=0.046; Steinbach (100%) > Shoal Lake (93% ± 20) > Brandon (76% ± 41) Figure 31D). Actinobacteria, Firmicutes and Bacteroidetes also represent common phyla. Phyla present at very low abundances include Acidobacteria, Chloroflexi, Fusobacteria, Spirochaetes, TM7 and Verrucomicrobia.

In our analysis, Gammaproteobacteria showed an overall dominance of 54%. The abundance of Gammaproteobacteria was found to differ between IBD incidence areas (P=0.006). Specifically, Gammaproteobacteria was elevated in LIA (77% ± 41) when compared to HIA (28% ± 45). We observed moderate changes of Alphaproteobacteria (P=0.07), which were highest in LIA. Investigation of sample type presented an interesting dynamic: the abundance of Alphaproteobacteria (P=0.055) and Betaproteobacteria (P=0.091) seemed to shift based on the sample type. A moderate variation of Betaproteobacteria was observed among sample locations (P=0.056).

Few genera were found to correlate with either incidence area; *Pseudomonas* (Gammaproteobacteria; *P*=0.016) was highest in LIA whereas *Bradyrhizobium* (Alphaproteobacteria; *P*=0.02) was highest in HIA (Figure 32A). *Pseudomonas* comprised approximately 48% of the microbial population. To a minor extent, the abundance of *Mycobacterium* (*P*=0.06) was higher in HIA. We identified *Mycobacterium* in only 3 samples, all of which were HIA and identified in pipe wall material. Though many of the taxa identified in our study do not reach statistical significance with respect to abundance, application of a principal coordinate analysis (Figure 33A) shows that many HIA samples cluster separately from LIA thereby suggesting that the overall microbial population structure varies. The HIA samples form a tighter cluster compared to LIA samples; many HIA samples were comprised of a relatively low number of different OTUs reflecting a more homogenous microbiota. Conversely, LIA were more heterogeneous typically including several different OTUs.

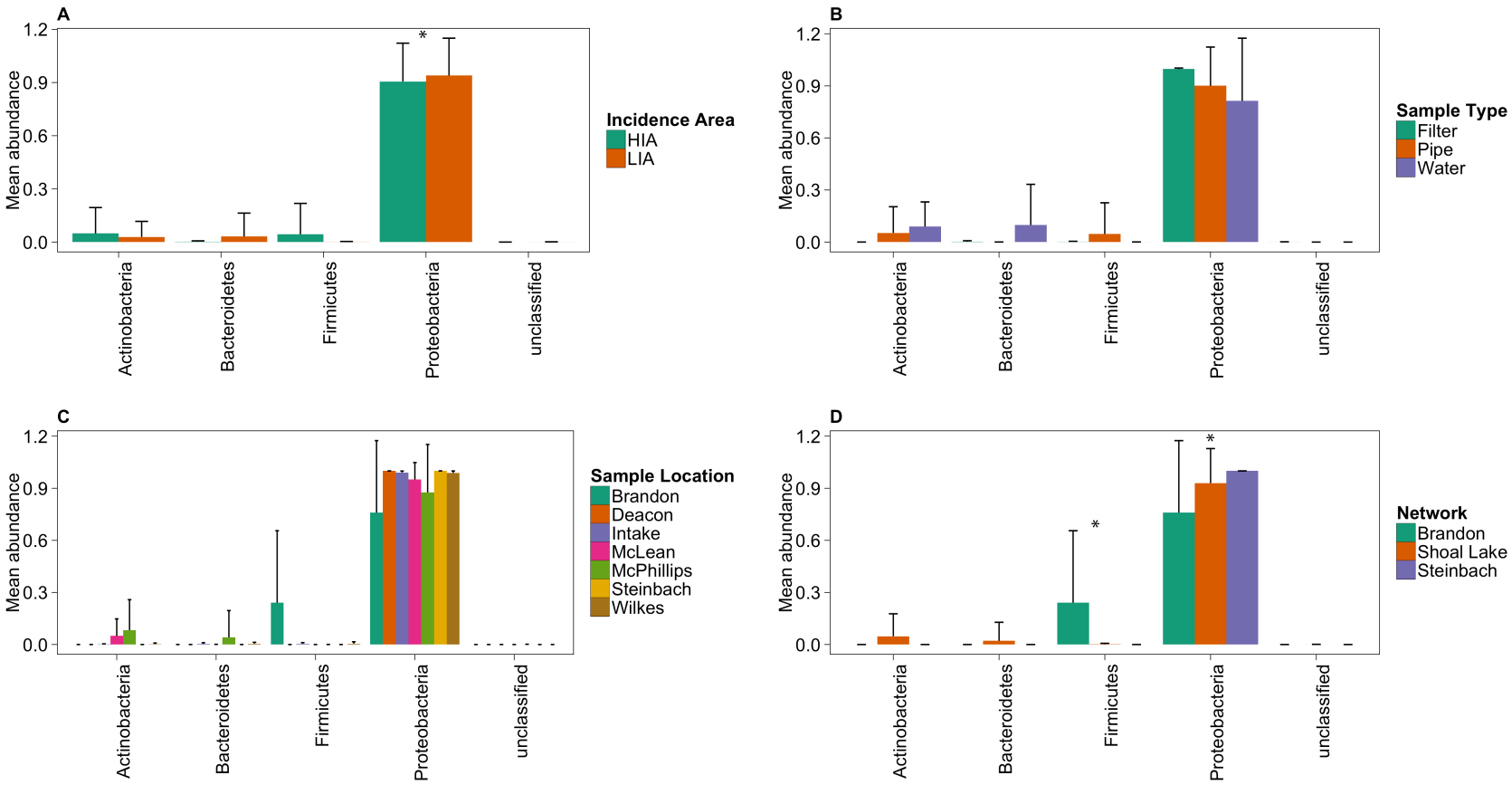

Figure 31. Phylum-level abundances of the water microbiota. Comparison of phyla abundance documented between (A) low and high incidence areas, (B) sample type, (C) sample location and (D) water distribution networks. Phyla are expressed as average abundance with standard deviations. Differences considered significant at $*P < 0.05$.

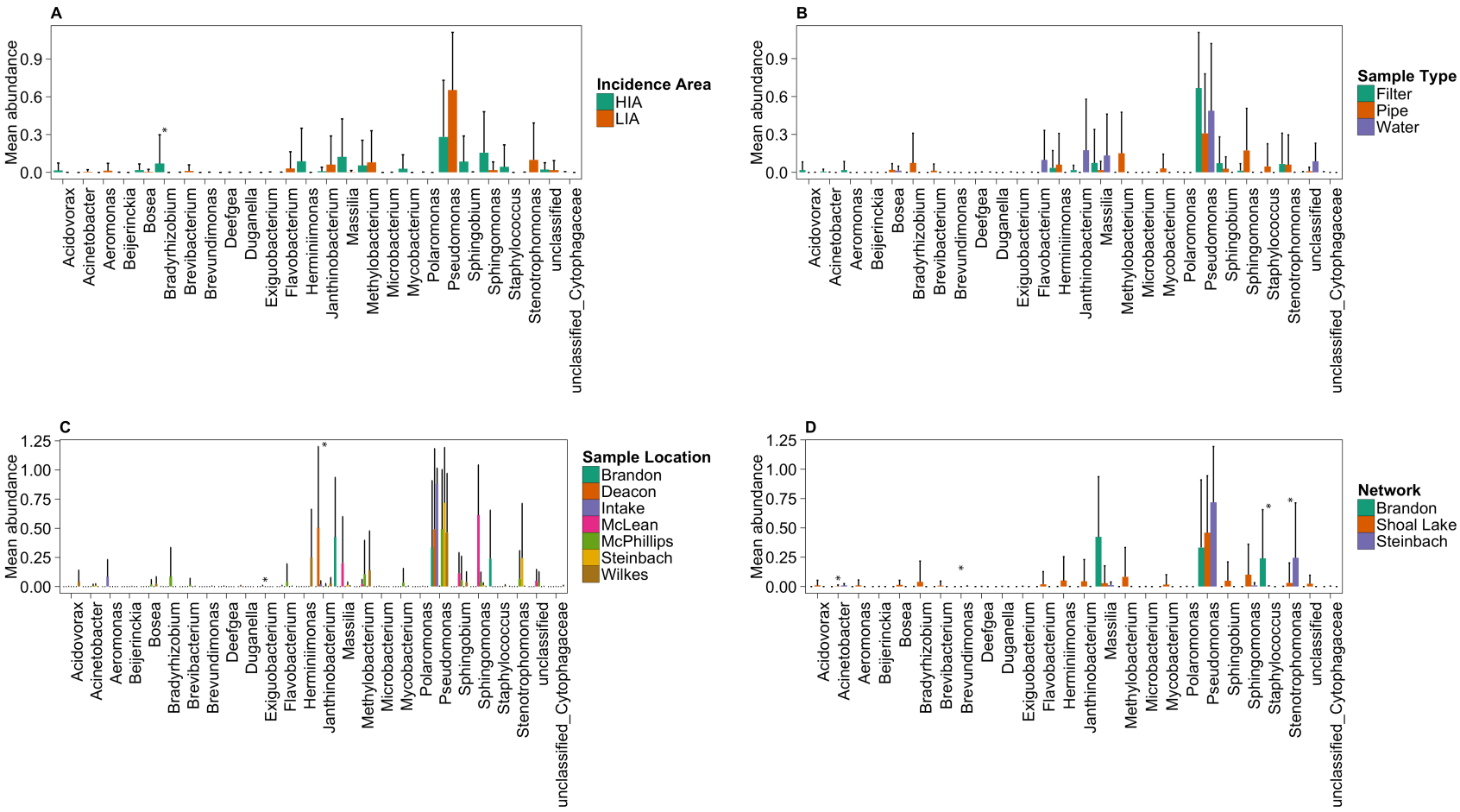

Figure 32. Genus-level abundances of the water microbiota. Comparison of genera abundance documented between (A) low and high incidence area, (B) sample type, (C) sample location and (D) water distribution networks. Genera are expressed as average abundance with standard deviations. Differences considered significant at $*P < 0.05$.

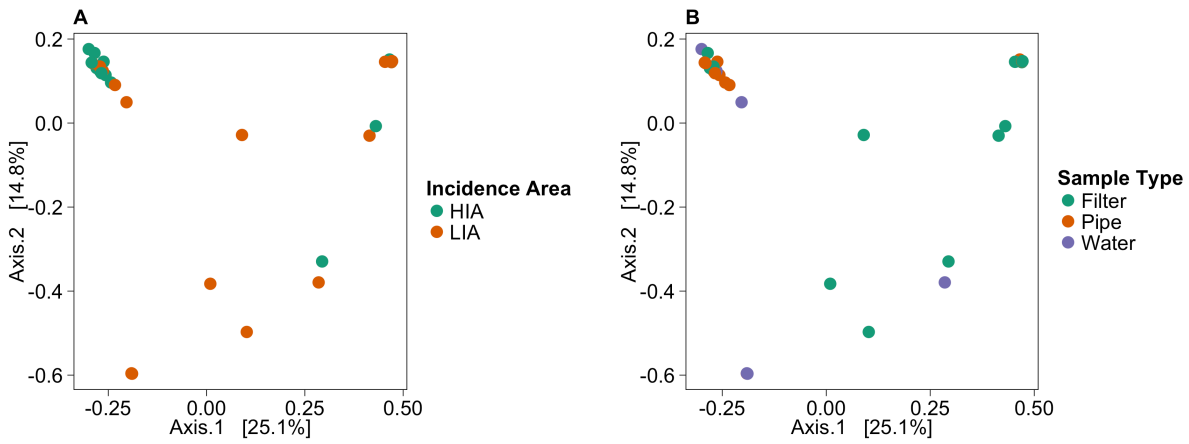

Figure 33. Principal coordinate analysis based on the overall structure of the microbiota of all samples in terms of (A) incidence area and (B) sample type. PCoA was calculated using Bray-Curtis distances. Each data point represents an individual sample.

Sample type

In bulk water, Proteobacteria represented 81% of OTUs, Bacteroidetes represented 10% and Actinobacteria accounted for 9% (P=0.05; Figure 31B). Dominant genera (Figure 32B) of bulk water included *Pseudomonas* (49%), *Janthinobacterium* (18%) and *Massilia* (13%). In terms of pipe wall (biofilm), Proteobacteria accounted for 90% of OTUs, and Actinobacteria and Firmicutes represented 5% of OTUs individually (Supplementary Table 5). Within pipe material, *Pseudomonas* (31%), *Sphingomonas* (17%) and *Methylobacterium* (15%) dominated the microbiota. Proteobacteria represented 99.7% of OTUs associated with filter material (Supplementary Table 5). Intriguingly, Actinobacteria accounted for 9% of OTUs in bulk water yet in filter material their abundance was negligible. The main genera of filter material included *Pseudomonas* (67%), *Massilia* (8%) and *Sphingobium* (7%). PCoA implies that the overall

microbial populations vary; most of the pipe wall samples clustered together and likewise, a trend can be observed among filter samples (Figure 33B).

Sample locations

The microbial community of sample locations indicated minor variability. The abundance of Firmicutes (P=0.06) demonstrated minimal shifts based on location (Figure 31C); in the context of genera, the abundance of *Exiguobacterium* (P=0.0009) and *Janthinobacterium* (P=0.01) were statistically significant among sample locations (Figure 32C).

Water distribution networks

As mentioned, in this study there are three water distribution networks. The first receives its water from Shoal Lake intake and includes the additional locations of Deacon, McPhillips, McLean and Wilkes Reservoirs. The second includes the location Brandon, Manitoba and lastly, Steinbach, Manitoba that utilizes water distribution systems previously described. In this regard, the abundance of Proteobacteria (P=0.05) was highest in LIA Steinbach and lowest in HIA Brandon (Figure 31D). *Brevundimonas* (0.17% ± 0.3; P=0.02), *Stenotrophomonas* (25% ± 47; P=0.02) and *Acinetobacter* (0.98% ± 1; P=0.03) of Proteobacteria were highest in Steinbach (Figure 32D). The abundance of Firmicutes (P=0.05) was significantly different among distribution systems, likely due to Staphylococcus, which was highest in Brandon (24% ± 41; P=0.01).

CHAPTER FIVE: DISCUSSION

5.1 Insights into the immune-mediated inflammatory disease gut microbiome

Host-microbe interactions have progressively been shown to play a fundamental role in the induction of immune-mediated inflammation in both experimental animal models and human diseases including IBD, MS and RA. Though the etiology of these particular IMID are in large part undefined and combined with relatively low disease concordance rates observed in monozygotic twins, it can be assumed environmental elements greatly contribute to disease etiopathogenesis. The gut microbiome and its associated dysbiosis generally observed in IMID is likely critically involved. Particular characteristics of gut dysbiosis are widely accepted for some IMID (i.e. IBD), however, evidence has only begun to unfold in recent years in IMID not classically defined as gastrointestinal disease, such as MS and RA.

We undertook studies to determine the analytical differences in the microbial profile of multiple IMID utilizing 16S rDNA amplicon sequencing and analysis of stool. To our knowledge, our study represents the first investigative analysis to characterize the gut microbiome in numerous IMID simultaneously that preferentially affects different body systems. This study aimed to accurately determine patterns of dysbiosis and the inferred functional potential of the metagenome relative to health, and to other IMID. The resultant findings of this study are two-fold. First, we show important differences in the stool microbial profile in IMID relative to health. In doing so, we were able to identify consistent patterns of dysbiosis between multiple IMID in addition to dysbiosis characteristics that are specific to one disease. Second, we show how microbial populations and their functional potential are altered between IMID. A limitation

of our study, however, is the lack of available patient meta-data; thus we are faced with difficulties interpreting compositional shifts associated with the heterogeneous, IMID microbiome.

In the present study, we have shown that significant variability is present between species richness observed in HC compared to both CD and RA, which is in agreement with previously reported data (Chen et al., 2016b; Manichanh et al., 2006). Interestingly, Chen et al., (2016b) reported decreased species richness in RA correlated with an elevated body mass index, RF levels, and disease duration whereas treatment regimens that included methotrexate and hydrochloroquinone correlated with increased species richness (and diversity). Further, we detected no significant differences between HC with neither MS nor UC, though we did detect decreasing trends. Our MS findings agree with earlier reports (Chen et al., 2016a), which have also shown that RRMS patients with active disease tended to have lower richness than those in remission. Microbiome profiling studies in UC often report significantly decreased richness (Duranti et al., 2016), and although we noticed a decreasing trend, our indices did not reach statistical significance. Species richness analyses utilizing the Chao1 index also revealed significant differences between CD compared to UC and MS. Combined with the finding that significant variability was not observed with the ACE index, we can conclude that while these IMID present with distinct states of gut dysbiosis, their difference (CD compared to UC and CD compared to MS) is in large part attributed to the presence of rare or uncommon species. Analogous to previous reports (Chen et al., 2016a, 2016b; Hoarau et al., 2016; Michail et al., 2012), relative to HC, microbial diversity (richness and evenness) was decreased in all IMID utilizing Shannon and Simpson indices, and significant differences were observed with CD, UC

and RA. Furthermore, we also report significantly decreased diversity in CD compared with UC, MS and RA. While both rare and dominant taxa explain diversity variability observed with UC and MS, only the Shannon index reported significant differences between CD and RA suggesting that the microbiota diversity dissimilarities are due to rare taxa rather than dominant. To our knowledge, this study represents the first to concurrently analyze gut populations from multiple IMID; hence our findings indicate that there are varying degrees of microbial richness and diversity loss relative to HC between particular IMID.

Our results on the composition of the IMID microbiome have revealed that the key constituents of the complex assemblage of gut microbes that define an IMID in the context of health are variably different. Unsurprisingly, gut populations of CD were dramatically different from HC. The gut profile of RA was additionally strikingly different from HC. Further, while UC and MS both demonstrated a clear dysbiosis, observed microbial shifts were minimal compared to those observed within CD or RA. This represents one of the most interesting findings of this study suggesting that even phenotypes of IBD are dramatically different and that other IMID demonstrate more similar microbial patterns to particular forms of IBD. We propose overlapping etiopathogenesis including genetic susceptibilities and environmental factors (i.e. diet, cigarette smoking, urbanization and others) play a key role in this finding. Further study is needed to elucidate parameters that might affect the IMID microbiome.

In the current study, we have identified several microbes consistently or uniquely disproportionate in abundance in IMID relative to health and to other IMID; many taxa found to be disproportionately abundant belonged to the Actinobacteria and Firmicutes phyla. Of taxa

belonging to the Actinobacteria phylum, the relative abundance of unclassified Coriobacteriaceae was decreased whereas *Eggerthella* was increased in all IMID relative to health. The observed increase of *Eggerthella* was similar across all IMID, but shifts were slightly more pronounced in MS. *E. lenta* was the most commonly detected *Eggerthella* species in our data; the abundance of *E. lenta* is reportedly increased in type II diabetes (Qin et al., 2012), and has recently been described in a Japanese cohort of RRMS patients (Miyake et al., 2015) and identified in high abundance in an RA cohort (Chen et al., 2016b). While generally considered both a commensal and an opportunistic pathogen, this species has been implicated in CD bacteremia subsequent to ileocecal resection (Thota et al., 2011) and other disseminated infections (Salameh et al., 2012). Additional genera belonging to the Actinobacteria phylum – *Asaccharobacter, Collinsella, Gordonibacter, Slackia* and *Bifidobacterium* – were uniquely decreased in abundance in CD relative to health. Several species belonging to these taxa exhibit beneficial, protective or anti-inflammatory effects on the host (Rivière et al., 2016; Selma et al., 2014); thus, their preferential decrease in CD might indicate a more perturbed state of gut dysbiosis or potential thereof. Reductions of particular Coriobacteriaceae members have been reported in relatives of persons with CD (Joossens et al., 2011). Of note, a recent study characterizing the MS gut microbiome has reported decreased abundances of *Collinsella* and *Slackia* in untreated patients (Jangi et al., 2016). This suggests that therapeutic intervention might impact these gut populations that in turn benefits host health. Within IMID, and specifically with respect to CD, we also observed significant variability with Coriobacteriaceae and particular genera within this family (i.e. *Asaccharobacter* and *Gordonibacter*) and *Bifidobacterium* (Bifidobacteriaceae). *Asaccharobacter* was significantly decreased in CD relative to UC, MS and RA. The relative abundance of Coriobacteriaceae was similarly reduced

in CD compared to UC and MS. Lastly, *Bifidobacterium* and *Gordonibacter* were significantly decreased in CD relative to UC and MS, respectively. The reduced abundance of *Bifidobacterium* (a probiotic bacterium) in CD relative to in health and UC, warrants further study to elucidate its role in disease. These findings highlight key abundance similarities of Actinobacteria in IMID relative to in health; moreover, when comparative analyses are executed between IMID, critical dissimilarities of microbe abundance are observed.

We observed similar increasing shifts in *Clostridium* XIVa (Lachnospiraceae) across all IMID. Species belonging to Clostridial cluster XIVa are among the most dominant in the healthy distal human gut (Maukonen et al., 2006) and have been implicated in maintaining mucosal homeostasis and stimulation of anti-inflammatory responses (Atarashi et al., 2011). This cluster however, can also be characterized as opportunistic; some species can cause systemic infections (Huh et al., 2010), and furthermore, factors related to their flagellins have been implicated in IBD (Duck et al., 2007) and irritable bowel syndrome (Schoepfer et al., 2008). An enrichment of *Clostridium* XIVa has also been reported in *C. difficile* infection (Pérez-Cobas et al., 2014). Thus, considering their opportunistic function, the overrepresentation of *Clostridium* XIVa might be a consequence of microbial imbalance in IMID. The overrepresentation of this genus was similar between all IMID, which suggests a pivotal role in IMID etiopathogenesis. We have also shown that the *Clostridium* III genus (Ruminococcaceae) was increased in UC, MS and RA relative to HC and that the abundance in MS was significantly greater than that observed in CD. The genus *Streptococcus* was increased in CD, MS and RA; increases in *Streptococcus* have recently been documented in a Japanese cohort of RRMS patients (Miyake et al., 2015). Analogous to our findings, this group identified *S. thermophilus* as a highly abundant species. *S.*

*pneumonia*e has also been investigated as a candidate pathogen in MS etiology; subsequent to EAE induction, systemic infection of mice with *S. pneumoniae* has resulted in increased disease severity (Herrmann et al., 2006). The increased abundance of *Streptococcus* in many IMID compared to in health warrants further study to determine what, if any, role these microbes play as causative agents of IMID.

Clostridium IV (Ruminococcaceae) are known to exhibit anti-inflammatory effects. We have shown that the abundance of *Clostridium* IV are significantly decreased in CD in comparison to HC and other IMID. Studies that have characterized the UC and MS microbiome have however, reported an underrepresentation of Clostridial cluster IV (Kabeerdoss et al., 2013; Miyake et al., 2015). The Clostridial cluster IV is constituted by a highly diverse group of microbes (i.e. including *Eubacterium, Ruminococcus* and *F. prausnitzii*), some of which are characterized by the ability to produce SCFA. SCFA (i.e. butyrate) in turn, are implicated in colonic epithelium homeostasis, stimulation of anti-inflammatory responses and the concurrent induction of colonic Treg cells (Atarashi et al., 2013). We have shown that *Ruminococcus* (Ruminococcaceae) is significantly decreased in CD relative to HC, MS and RA. The abundance of *Faecalibacterium* was underrepresented in CD, MS and RA and differences were observed amongst CD with other IMID. Additional members of the Ruminococcaceae (*Gemmiger* and *Subdoligranulum*) associated with multiple IMID, and in the case of *Subdoligranulum*, both decreased and increased representation was observed. Lachnospiraceae is a dominant (butyrate producing) family in the healthy human gut and reductions of genera within this taxon are consistently reported in several IMID (Balzola et al., 2011; Jhangi et al., 2014b; Tremlett et al., 2016c). We have detected several Lachnospiraceae members that are uniquely disproportionate to particular

IMID. Interestingly, *Blautia* was the only Lachnospiraceae genus to be increased in IMID, specifically in CD. A primary function of *Blautia* is the removal of end-products (such as gasses) via processes of acetate formation. Overrepresented abundances of *Blautia spp.* has been reported in irritable bowel syndrome (Rajilić-Stojanović et al., 2011) and MS (Chen et al., 2016a). Our data indicates that the CD microbiome, more so than other IMID, are constituted by abundant groups of intestinal microbes that produce gas (i.e. *Dorea spp.*) In concordance with previous reports, *Coprococcus* was decreased in CD. The underrepresentation of this taxon has recently been characterized in CD patients with penetrating disease (Shahir et al., 2016). Furthermore, analysis of the stool microbiome in CD patients following ileocolonic resection has identified *C. catus* L8 in addition to a *Butyricicoccus spp.* to significantly associate with remission at six months following surgery (Mondot et al., 2016). This finding indicates an increased abundance of these taxa might be involved in establishing early remission. Significant differences were also reported in the abundance of *Coprococcus* in CD to other IMID. Both *Dorea* and *Roseburia* were decreased in RA relative to health. The families Erysipelotrichaceaeare and Veillonellaceae are reportedly decreased in RRMS patients (Chen et al., 2016a). Erysipelotrichaceae are involved in bile acid metabolism and in turn induce anti-inflammatory properties (Vavassori et al., 2009). An underrepresentation of Erysipelotrichaceae species has recently been observed in the fecal and mucosal microbiome in CD patients (Gevers et al., 2014). We observed a significant decrease in *Turicibacter* (Erysipelotrichaceae) in CD relative to health, but a decrease was not observed in MS. Veillonellaceae, specifically *Dialister,* was significantly decreased in MS. This taxon is closely related to *Clostridium* species that have been shown to induce Treg cells (Atarashi et al., 2011). The decreased abundance of anti-inflammatory microbes is a consistent finding of the IMID microbiome and accordingly it is

conceivable that depletion of beneficial microbes contributes to disease. Since we report different taxa to be preferentially underrepresented and to varying degrees, we propose that the depletion of particular anti-inflammatory microbes might be disease-specific and warrants further investigation into their effect on gut homeostasis.

In the present study we have revealed interesting findings related to taxonomic alterations between IMID and relative to health. Investigation of functional analyses may provide more relevance in annotating clinical significance. In recent years, numerous studies in persons with IMID or experimental animal models have identified an altered metagenome through either shotgun metagenomics or 16S amplicon sequencing (Chen et al., 2016b; Gevers et al., 2014; Jin et al., 2015; Tremlett et al., 2016c). Herein, we used the 16S rRNA gene profiles to predict the functional potential of the stool putative metagenomes and as such have identified numerous functional gene categories as differentially expressed between IMID and in health. A principle observation in the present study is that specific pathways characterize distinct IMID, thus, microbes of the IMID microbiome perform differential functions. Prior studies characterizing the IBD putative metagenome have identified numerous microbial functional alterations including decreased SCFA production, butanoate and propanoate metabolism, amino acid biosynthesis and increased auxotrophy, amino acid and sulfate transport, oxidative stress and type II secretion systems (reviewed in Kostic et al., 2014). In concordance, we have shown CD to be associated with several metabolic and transport systems and UC with metabolic and signaling pathways. We observed several pathways significantly overrepresented in MS and thus could act as a biomarker of disease. Genes pertaining to ribosomes represented the most associative factor of MS. Tremlett et al., (2016c) did not observe significant differences in the abundance of

ribosomal metabolic pathways between MS cases and controls but did report significant variability among MS patients treated with and without immunomodulatory drugs. We also observed an increase in several biosynthesis pathways, many of which are in agreement with Tremlett et al., (2016c), which is suggestive of an important role in MS. Chen et al., (2016a) has also recently characterized the functional potential of the MS metagenome; COG (clusters of orthologous group) categories related to signal transduction mechanisms, defense mechanisms, transcription and carbohydrate transport and metabolism to be significantly increased and lipid transport/metabolism and intracellular trafficking and others to be significantly decreased compared to controls. Moreover, at a nominal significance the authors also identified 34 differentially abundant KEGG pathways including an increased abundance of nitrotoluene degradation, sporulation and porphyrin and chlorophyll metabolism. Though we did not observe an association between porphyrin and chlorophyll metabolism and MS, we did find this pathway to be significantly increased which suggests these pathways are similar in CD and MS. We also observed a significant association with prenyltransferases. Interestingly, prenyltransferases are involved in the production of dolichol, which is a major lipid component of human substantia nigra neuromelanin (a dark pigment in the brain; Fedorow et al., 2005). Studies concerning the functional role of the metagenome in MS are limited and data that is available is derived from relatively small sample populations. Hence, we suggest that functional studies conducted in larger cohorts would provide more meaningful associations and would allow for the establishment of general trends. Functions of the RA metagenome were largely associated with various metabolism and degradation pathways. The pathway most characteristic of RA was related to amino and nucleotide sugar metabolism. However, in a recent functional analysis, a

decrease in the amino acid metabolism pathway was reported (Chen et al., 2016b). Further analyses in larger cohorts are needed to establish metabolic pathway trends in RA.

As discussed in section 2.4.4, numerous studies in recent years have illustrated the importance of standardization when conducting microbiome analyses to achieve an accurate representation of microbial populations. These include type, collection and storage of samples, DNA extraction methodology, appropriate primer selection and sequencing platform. At the time that this study was initiated, though study design factors were acknowledged, the importance of them were not as well recognized. We have conducted a series of troubleshooting experiments on select samples and alongside mock communities; our results indicate that either the length of time in storage or the amount of freeze-thaws contributed to our unusual ratio of Gram-negative to Gram-positive bacteria. Since all samples were subjected to the same (degradation) conditions and analyzed comparatively, our results are meaningful and add substantially to current literature. Importantly, these conditions affect particular groups of microbes differently; as microbial abundances vary between samples, results may be skewed differently dependent on the composition. As discussed previously, our samples were characterized by an unusually low abundance of Gram-negative populations (i.e. Bacteroidetes and Proteobacteria); many taxa belonging to these phyla exhibit potentially detrimental effects (i.e. pathogens) on the host. Thus, while we identified significant microbial abundance shifts of Proteobacteria in CD and RA as well as the Bacteroidetes in CD we feel as though based on the limitations (i.e. sample quality) that this study is more apt to focus our findings on the Gram-positive populations (i.e. Actionbacteria and Firmicutes), which were overall, observed in expected abundances. We have reported that many Gram-positive taxa identified in our data may demonstrate beneficial effects

on the host, many of which, were observed at a reduced abundance in IMID, compared to in health. While we do not suggest that an underrepresentation of beneficial microbes represents the complete characterization of gut dysbiosis observed in IMID, we do highlight its importance on host health. We feel as though if additional studies were conducted utilizing pristine stool samples, that Gram-negative species would be present in expected abundances and that disproportional abundances would be observed relative to in health and between IMID. These findings would add considerable depth of information to current literature regarding potential pathogens that might be consistently (or uniquely) increased between IMID. This study was also hampered with another pivotal disadvantage: a lack of patient meta-data. It is now widely accepted that factors related to disease including disease duration, activity state, phenotype and treatment, epidemiology including age, sex, ethnicity and geographical location, environment including lifestyle, diet or other pertinent known risk factors, and genetic susceptibility including the presence or absence of susceptibility risk alleles profoundly impact the microbiome. In the absence of this information, we are able to draw general conclusions from our data but are unable to stratify the microbial abundances or functions to particular meta-data characteristics.

In summary, this study presents a comprehensive analysis of the fecal microbiome in IMID. Although we conclude that the composition and function of the gut microbiome is altered in CD, UC, MS and RA, we have reported that varying degrees of gut dysbiosis are evident. We have determined gut dysbiosis to be most dramatic in CD, followed by RA and less so in UC and MS. Moreover, we also show that within IMID, the gut communities in UC and MS are most similar, which inherently suggests that mechanisms of etiopathogenesis of these diseases are closely related. We have uncovered several microbes that are consistently or uniquely disproportionate

in IMID relative to health and suggest further research into these microbes and their associated functions are needed in order to establish any sort of causality in disease pathogenesis.

5.2 Understanding the mucosal microbiome in inflammatory bowel disease

An alteration of the microbiome composition and its interaction with the host immunological response likely plays a critical role in the pathogenesis of IBD, however this relationship is poorly understood. This study shows important differences in the bacterial profiles of the mucosal microbiota between participants with CD, UC or non-IBD controls based on analysis of mucosal tissue using a 16S rDNA amplicon-based analysis. To our knowledge, our study represents one of the largest 16S rDNA amplicon-based studies investigating the mucosal microbiota composition in distinct inflammatory states of particular compartments within the gastrointestinal tract of persons with IBD.

Our principal and most surprising finding was that the mucosal-associated microbiota of inflamed and non-inflamed regions of the gastrointestinal tract in CD or UC, respectively, were indistinguishable as virtually no taxa demonstrated disproportional abundances at a significant threshold nor were there significant diversity differences observed. Thus, as we were unable to recognize a specific microbe or group of microbes to consistently associate with the inflamed (or non-inflamed) tissue in CD or in UC, our results suggest that important localized changes in the mucosal-associated microbiota do not exist between the inflamed and non-inflamed gut. These findings argue against the hypothesis that inflammation drives gut dysbiosis and instead promotes the notion that gut dysbiosis is a prerequisite for inflammation. Although, it would be particularly useful to know if inflammation is present outside of the lesions (i.e. via elevated antimicrobial peptide expression or reactive oxygen species levels). This is further supported by

the clinical findings that resecting CD tissue will lead to recurrences in the previously un-inflamed tissue (De Cruz et al., 2015). This would also occur in UC although segmental resections in UC have long been abandoned because of the very high recurrence in the tissue remaining behind. While these microbial patterns may not be causal they seem to be signature patterns for each disease. As the participants all had IBD for a lengthy duration and hence findings are associations and could not be construed as causal; it is noteworthy that within an individual, patterns were similar regardless of inflammatory state. Since CD participants had consistent patterns as a group, the implication is that these really are unique to the entire extent of the distal gastrointestinal tract of a person with CD (or UC). Even if not causal it is plausible that altering this pattern could impact on disease course.

Few studies have reported a localized dysbiosis within the inflamed gut (Mylonaki et al., 2005; Sepehri et al., 2007). Mylonaki et al., (2005) noted a reduction of *Bifidobacterium* and an increase of *E. coli* in the epithelial cell surface of inflamed UC colorectal mucosa compared to non-inflamed UC mucosa. *Lactobacillus* and *Bacteroides* were found in similar abundances between the two groups. Similar observations were reported for inflamed CD colorectal mucosa. And, though Sepehri et al., (2007) described an increase of Firmicutes and parallel reduction of Bacteroidetes in addition to significant differences in community richness and biodiversity, inflamed (and non-inflamed) biopsies were not separated on the basis of disease type (i.e. CD and UC). These studies used molecular methods that include fluorescent in situ hybridization (FISH), automated ribosomal intergenic spacer analysis (ARISA) and terminal restriction fragment length polymorphism (T-RFLP) which are less discriminatory methods compared to our approach yielding a greater sequencing depth and taxonomic resolution. Several additional

studies have taxonomically surveyed the inflamed and non-inflamed gut in IBD, most of which agree with our data indicating an absence of a localized dysbiosis (Bibiloni et al., 2006; Gophna et al., 2006; Gupta et al., 2014; Seksik et al., 2005; Vasquez et al., 2007).

While our study strengthens the notion that dysbiosis of the overall gut microbiota in CD or UC is testament to either disease, we also performed comparative analyses to determine the dysbiosis profiles between the inflamed gut of CD to UC and similarly with the non-inflamed gut. Our data indicated that both the inflamed and non-inflamed gut has unique taxonomic profiles when compared across disease types and that importantly, those differences are more apparent within the non-inflamed gut. In active disease, CD and UC clinically differ based on the formation of strictures and fistulas in CD, hence it's theoretically compelling that the microbiotas of the inflamed mucosa are less different than the non-inflamed mucosa demonstrating no evidence of active disease. Since we observed greater community differences among the non-inflamed or inflamed IBD gastrointestinal tracts than between the inflamed and non-inflamed CD (or UC) mucosa, there may be additional systemic perturbations of host-bacteria interactions in CD (or UC) that is independent of the inflammation status of the mucosa.

A seminal study performed by Eckburg et al., (2005) demonstrated microbial homogeneity between mucosal sites of the large intestine of healthy individuals which has been replicated in both health and IBD (Lepage et al., 2005). Our data corroborates previously reported findings of gut mucosa microbial homogeneity in CD and UC regardless of whether the mucosa is inflamed or non-inflamed. Though all microbial populations demonstrated fairly consistent abundances across gut compartments, in CD, *Escherichia-Shigella* was notably highest in the cecum, and

Pseudomonas in the rectum. We also compared the microbiota composition of gut anatomical sites between disease groups. Our main findings include the significant difference of *Bacteroides* and *Pseudomonas* among ileal, mid-colonic and rectal specimens between CD, UC and nonIBD.

This study confirms previously noted findings regarding the presence of particular bacteria. For example, a number of the microbes identified in this study are recognized to have the ability to adhere to the mucosa, invade intestinal epithelial cells and possibly exacerbate inflammation including *Escherichia* (Krause et al., 2011) and *Fusobacterium* (Ohkusa et al., 2002). The average abundance of both of these taxa was lowest in non-IBD participants. In fact, the presence of *Fusobacterium* was only identified in rectal tissue of one non-IBD sample. Several groups of microbes including *Bifidobacterium, Lactobacillus* and *Faecalibacterium* have been shown to exhibit protective effects via inflammatory cytokine regulation (Llopis et al., 2009) or conversely, stimulation of IL-10 (Sokol et al., 2008). We also detected a few butyrate-producing bacteria in our data (Vital et al., 2013); *Coprococcus* and *Oscillibacter* were reduced relative to non-IBD. Butyrate and other SCFA are the result of bacterial metabolism (Kelly et al., 2015). SCFA are essential sources of energy for gut epithelial cells and their reduction in IBD may perturb epithelial barrier integrity and modulate the gut immunological response thereby exacerbating disease (Frank et al., 2007; Kelly et al., 2015).

Taxonomic alterations revealed interesting findings, although, functional analyses may be more relevant in annotating clinical significance. Several studies in persons with IBD or mouse models of experimental colitis have identified altered microbiome function through either shotgun metagenomics sequencing or 16S amplicon sequencing (Gevers et al., 2014; Jin et al., 2015). In

the present study, we used the 16S rRNA gene profiles to predict the functional potential of the mucosal-associated putative metagenomes and as such have identified numerous functional gene categories as differentially abundant between (i) distinct inflammatory states of IBD and in health and; (ii) specific gastrointestinal compartments of IBD and in health. Concerning IBD, previous studies have identified numerous microbial functional alterations including decreased SCFA production, butanoate and propanoate metabolism, amino acid biosynthesis and increased auxotrophy, amino acid and sulfate transport, oxidative stress and type II secretion systems (reviewed in Kostic et al., 2014). Functional divergence between the inflamed and non-inflamed IBD mucosal-associated microbiota has recently been identified (Davenport et al., 2014). In UC, carbohydrate and nucleotide metabolism were increased in the non-inflamed mucosa whereas amino acid and lipid metabolism were decreased; in CD, an increased abundance of genes for energy metabolism and nervous system pathways were reported in the inflamed mucosa. Predictably, functional genes associated with non-IBD largely manifested in nutrient metabolism and synthesis pathways. We observed an association between fluorobenzoate degradation and the noninflamed UC mucosa and between benzoate degradation with the ileum of UC. Interestingly, benzoate metabolism intermediaries have the ability to influence dysbiosis as a stress response and promote the growth and virulence of Enterobacteriaceae (Freestone et al., 2007; Lyte et al., 2011). Increased genes for benzoate degradation have previously been reported in treatment-naïve new-onset CD patients (Gevers et al., 2014) in addition to models of experimental colitis (Rooks et al., 2014). Biotin metabolism, which was associated with the inflamed CD mucosa, has unclear metabolic implications though their metabolism is believed to be a complementary source to endogenous vitamins (Fiorucci and Distrutti, 2015). This observation, combined with the minimal genus level taxonomic differences detected between the inflamed CD mucosa and

non-IBD suggests that the inflamed CD mucosa still participates in benefiting host health and is not exclusively a pathogenic milieu. Similarly, pathways potentially beneficial to host health were associated with the inflamed UC mucosa including atrazine degradation and carotenoid biosynthesis. Further, genes pertaining to ABC transporters and secretion systems were associated with the UC ileum and rectum, respectively. An increase in secretion systems, particularly type II, are increased in IBD and are clinically relevant due to their involvement in secreting toxins (Kostic et al., 2014).

A limitation of this study is the paucity of patient meta-data. Factors related to disease including disease duration, activity state, phenotype and treatment, epidemiology including age, sex, ethnicity and geographical location, environmental parameters including lifestyle, diet and others, and genetic susceptibility factors including the presence or absence of susceptibility risk alleles profoundly impact the microbiome. In the absence of this information, we are able to draw general conclusions from our data but are unable to stratify the microbial abundances or functions to particular meta-data characteristics.

This study presents a comprehensive analysis of the mucosal-associated gut microbiome in IBD while investigating a possible localized dysbiosis in inflamed or non-inflamed mucosa at distinct gut compartments. Although we conclude that the composition of the mucosal-associated microbiota was not altered by the presence of inflammation we were able to identify an interesting phenomenon whereby the microbiota of the non-inflamed gut in IBD are more diverse than the microbiota of the inflamed gut. We have identified several microbes and

functional genes and suggest their disproportional abundances contribute to gut dysbiosis within the inflamed or non-inflamed mucosa.

5.3 A potential inflammatory bowel disease environmental etiology

Many acute illnesses may be secondary to the consumption of contaminated drinking water including legionellosis (Cohn et al. 2014), cholera (Alam et al. 2014) and gastroenteritis (Beaudeau et al. 2014). It is unknown if immunocompetent hosts can acquire chronic diseases like IBD through microbes present in drinking water. Linkages between the development of IBD and specific microorganisms have been investigated such as *Mycobacterium paratuberculosis* (Rhodes et al. 2013; Suwandi et al. 2014) or AIEC (Kotlowski et al. 2007; Agus et al. 2014). In the past decade studies have focused on exploring the gut microbiome using culture-independent techniques for microorganisms that may either be protective or injurious to the healthy gut and associated with one phenotype of IBD or another (Balzola et al. 2011; Frank et al. 2007), rather than environmental microbiomes.

Proteobacteria as a principal phylum in drinking water is well documented in the literature (Lautenschlager et al. 2014; Wang et al. 2014; Liu et al. 2014; Lührig et al. 2015). Our data corresponds with related population-based studies whereby the abundance of Proteobacteria can range anywhere from 44 to 98% (Wang et al. 2014; Liu et al. 2014; Lührig et al. 2015). Many population-based or environmental studies have reported either Beta- (Pinto et al. 2012; Lin et al. 2014) or Alphaproteobacteria (Gomez-Alvarez et al. 2012; Bai et al. 2013) to outweigh the Gammaproteobacteria population, though in our analysis Gammaproteobacteria was a dominant class. A noteworthy observation from this study is that the Alphaproteobacteria and Betaproteobacteria appeared to shift relative to sample type and that these populations are

inversely correlated. For example, Alphaproteobacteria (competitive under oligotrophic conditions and degrading complex organic compounds) are more prevalent within pipe walls whereas Betaproteobacteria (fast-grower and nutrient lovers; Newton et al. 2011) are more prevalent within water and filter material. Rudi et al. (2010) observed a similar phenomenon utilizing a ProteoQuant assay; the authors concluded that the Alphaproteobacteria and Betaproteobacteria are in competition, independent of the abundance of Gammaproteobacteria.

Pseudomonas constituted nearly half of the microbial population and was significantly increased in LIA. Such a predominance of *Pseudomonas* is of particular concern as many species belonging to this genus are biofilm formers (Fazli et al. 2014), which therefore lead to an increase in pathogen persistence. *Pseudomonas* has been implicated in the pathogenesis of IBD; the *Pseudomonas fluorescens*-associated sequence I2 that encodes for a T cell superantigen is prevalent among CD (38-60%) and UC (42%) (Prideaux et al. 2012; Wei et al. 2002). Huang et al. (2014) reported *Pseudomonas aeruginosa* to represent 11.92% of total sequence reads of filtered water and 11.16% in drinking water; hence, *Pseudomonas* was found to be resistant to particular disinfection methods due to unknown mechanisms. Many studies support these findings in that *P. aeruginosa* is frequently detected in chlorinated drinking water (Lee et al. 2011; Wingender and Flemming 2011). Conversely, *Bradyrhizobium* (increased abundance in HIA) has previously been identified to be a dominant member of chlorinated water (Gomez-Alvarez et al. 2012). The opportunistic pathogenic association, if any, of *Bradyrhizobium* species to human gastrointestinal diseases is not well defined. To our knowledge, *Bradyrhizobium* has not been implicated in the pathogenesis of IBD; however, *Bradyrhizobium enterica* is a newly identified bacterium and has been identified among persons with cord colitis following

umbilical-cord hematopoietic stem-cell transplantation (Bhatt et al. 2013). It should be mentioned that both *Pseudomonas* and *Bradyrhizobium* have been reported to be a common contaminant of NGS studies, particularly in samples containing low biomass (Salter et al. 2014). *Mycobacterium* was minimally increased in HIA. This genus, particularly, *M. paratuberculosis* is scrutinized in the possible etiology of CD as there are many conflicting reports regarding its relationship to CD (Suwandi et al. 2014, Nazareth et al. 2015). We identified *Mycobacterium* in only 3 samples, all of which were HIA and recovered from pipe wall material.

Liu et al. (2014) has previously characterized the bulk water microbiome; they reported *Polaromonas* (69%), *Sphingomonas* (13%), *Acidovorax* (5%) and *Janthinobacterium* (4%) to dominate the bulk water microbiota. In our data, *Polaromonas*, *Sphingomonas* and *Acidovorax* were identified, albeit at very low abundances. Navarro-Noya et al. (2013) investigated the bacterial communities of drinking water wells in Mexico and concluded similar findings with respect to the relative observed abundance of *Pseudomonas*, *Janthinobacterium* and *Massilia*. Interestingly, they also reported *Perlucidibaca* (13.9%), *Rheinheimera* (5.9%), *Alkanindiges* (2.2%) and *Psychrobacter* (2.1%) to be dominant members of the Gammaproteobacteria. These genera were not observed in our study; we speculate that environmental characteristics such as temperature, in addition to varied water treatment processes explain their absence. Moreover, while *Duganella* and *Flavobacterium* were present in our samples, their abundances were less than previously reported (Navarro-Noya et al. 2013).

Our findings of dominant biofilm genera supplements previous studies (Liu et al. 2014). Furthermore, Wang et al. (2014) detected *Legionella*, *Mycobacterium* and *P. aeruginosa* in

biofilms of pipe material. While our pipe material samples were devoid of *Legionella*, the presence (however paltry) of *Mycobacterium* and *Pseudomonas* is promising with respect to their known potential pathogenesis. The identification and dominance of *Pseudomonas* and *Sphingomonas* detected in pipe wall biofilms has been previously established (Berry et al. 2006; Simões et al. 2010). Both genera are able to form biofilms allowing persistence in otherwise harsh, oligotrophic (*Sphingomonas*) conditions. Moreover, *Sphingomonas* are able to transform between biofilm and planktonic modes that further supports their dominance in bulk water environments (Bereschenko et al. 2010).

Actinobacteria accounted for 9% of OTUs in bulk water yet in filter material their abundance was negligible. A possible explanation for their absence is their planktonic nature and hence do not preferentially attach to surfaces (Newton et al. 2011). *Nitrospirae* are frequently detected in filters but were absent in our samples (Lautenschlager et al. 2014). The variation and associated importance of microbial communities between sample types was surprisingly nominal. Previous population-based studies suggest sizeable structural and compositional variability among distinct sample material (Wang et al. 2014; Liu et al. 2014; Lin et al. 2014). In the present study, however, no significant microbial population differences were observed between sample types. Pinto et al. (2012) recently reported *Acidovorax*, *Hydrogenophaga* and *Denitratisoma* to be associated with filter effluent and hence less impacted by disinfection processes, though our samples were devoid of the latter two genera. It has been reported that different locations (within the same geographical region and subject to similar environmental exposures) will be comprised of diverse microbial communities (Roeselers et al. 2015). As water flows from one reservoir to another subjected to sequential disinfection methods it would theoretically be expected that

bacterial populations would shift in response. However, the shifts observed in our study were not so dramatic as to significantly affect the microbial populations.

We acknowledge our study has methodological limitations; we recognize that the absence of water quality characterization including microbe concentration, pH and levels of chlorine, minerals and trace elements and others is a major shortcoming of this paper. An additional limitation of this study is a lack of culture data and specifically, the ability to distinguish between presence of bacterial DNA and live bacteria. Future studies should incorporate the aforementioned data. Furthermore, the small sample size volume and methods we applied to enrich the samples may have also impacted our results. Enrichment methodologies chosen in particular were selected to help overcome low biomass issues. Culture enriched molecular profiling has to our knowledge not been performed specifically in water but has been explored in other low biomass samples, including the airway of cystic fibrosis (Sibley et al. 2011). This approach, particularly the incubation temperature may have biased the results to underrepresent some species and may partially account for the high abundance of Proteobacteria. Thus, these limitations should be considered when interpreting the microbial abundance distributions in all samples. We do however feel that as all samples were subjected to the same enrichment parameters that comparative analyses and interpretations can be made.

It is important to highlight that the drinking water examined did not harbor potentially unsafe levels of any known pathogenic bacteria and hence even differences in microbial ecology should not be misconstrued as any measure of lack of safety in the drinking water of Manitoba. However, we have uncovered that within the complexity of the microbial ecology of what is

considered safe drinking water there are differences in areas supplied by different water sources of varying incidence of IBD. That having been said, we have not proved any causal associations between microbes and IBD, but rather ecological associations. Our intent in this approach was not to be definitive, but instead, to look for trends that could be pursued in the future.

CHAPTER SIX: CONCLUSIONS AND FUTURE DIRECTIONS

The work presented herein examines the role of the gut microbiome in IMID. Studies within this thesis have examined the (i) stool microbial populations found in CD, UC, MS and RA; (ii) inferred functional potential of distinct IMID metagenomes; (iii) mucosal-associated microbiome associated with inflamed and non-inflamed mucosa in IBD; (iv) mucosal-associated microbiome in distinct gastrointestinal compartments in IBD; (v) inferred functionality of mucosal-associated microbiomes and; (vi) drinking water microbiome in low and high incidence areas of IBD.

Findings of this work suggest that while the composition and function of the gut microbiome in IMID may be perturbed, differential states of dysbiosis in specific diseases are evident. We have determined that the gut microbiome in CD exhibits substantial dysbiosis, followed by RA and less so in UC and MS. Moreover, we also show that within IMID, the gut communities in UC and MS are most similar, which inherently suggests that mechanisms of etiopathogenesis of these diseases are closely related. Many observations we have made are related to the underrepresentation of taxa that exhibit beneficial, anti-inflammatory mechanisms in the host and thus suggest that their depletion contributes to disease. We have found that these groups of microbes are disproportionally abundant across IMID and suggest particular species are likely implicated in specific, rather than all, diseases. Future studies should investigate the IMID microbiome in larger cohorts, perhaps with additional IMID that target different organs or are defined as systemic including SLE, ankylosing spondylitis and psoriasis. Further, studies should also consider examining the mycobiome and virome in IMID; these communities are gaining

increased attention, particularly in IBD and it would likewise be useful to see how these communities shift in different IMID. Future studies should also incorporate patient meta-data including those related to disease (i.e. duration, activity, treatment, phenotype), epidemiology (i.e. age, sex, ethnicity, geographical location), environment (i.e. lifestyle, diet or other pertinent known risk factors) and genetics (i.e. presence or absence of susceptibility risk alleles) as many of these influences have shown the ability to alter the microbiome. Incorporating meta-data will allow for a thorough examination of associations between clinical meta-data and microbial community abundance or function.

The abundance of most mucosal bacteria are highly similar throughout the ileum to the rectum in both health and disease. Our investigation of the IBD mucosal-associated microbiome has substantiated this finding. However, bacteria inhabiting mucosal surfaces of the gastrointestinal tract are extensively different from those present in the lumen and due to their proximity to the epithelial barrier likely play a more prominent role in disease. Our investigation of the IBD mucosal-associated microbiome revealed that a localized dysbiosis does not exist between the inflamed and non-inflamed mucosa and further has also revealed that the non-inflamed mucosa might be more important in disease pathogenesis. These findings suggest that gut dysbiosis is a prerequisite for inflammation rather than a trigger of inflammation. Future studies in characterizing the IMID microbiome should consider utilizing mucosal tissue alongside stool.

While the fecal and mucosal microbiome studies were designed independent of one another, and direct analytical comparisons cannot be conducted due to differences in sequencing technology, primers utilized and variances in data analysis pipelines, microbial structure patterns between the

two studies can loosely be inferred. Briefly, compared to in health, the abundance of *Coprococcus* was significantly decreased in the CD fecal microbiome and a similar trend was observed in the non-inflamed CD mucosa. The genus *Oscillibacter* was also reduced in the non-inflamed CD mucosa, however, was absent in the fecal microbiome. Interestingly, *Dorea* was decreased in the inflamed UC mucosa, though these changes were not observed in the UC fecal microbiome; *Dorea* was however reportedly decreased in the RA fecal microbiome compared to in health. In the UC mucosal microbiome, several genera belonging to the Bacteroidetes and Proteobacteria phyla were differentially abundant compared to in health, however these particular genera were not detected in the fecal microbiome. Comparisons between CD and UC also show several genera that are consistently disproportionate among the fecal and mucosal microbiomes including *Bacteroides, Clostridium, Faecalibacterium* and *Turicibacter*; interestingly, *Enterococcus* was significantly disproportionate between the non-inflamed CD and UC mucosa but was observed at a similar abundance in the fecal CD and UC microbiomes. Future studies that incorporate both stool and mucosa from IMID patients would add substantially to current research.

Lastly, in our exploration of a potential IBD environmental factor in drinking water, we have uncovered discrepant microbial ecologies in low and high incidence areas; particularly that *Bradyrhizobium* was observed in significantly higher abundance in HIA and *Pseudomonas* in LIA. While this study is independent from the fecal and mucosal microbiome studies in IMID and IBD respectively, it is worth noting that *Bradyrhizobium* was not observed in either gut microbiome study and that the abundance of *Pseudomonas* was significantly more abundant in the ileal, mid-colonic and rectal gastrointestinal compartments of the UC mucosa and in the

rectum of CD mucosa. Moreover, in the context of inflammation, *Pseudomonas* was similarly increased in the inflamed and non-inflamed UC mucosa relative to non-IBD. This is an interesting finding considering the association of *Pseudomonas* with LIA rather than HIA. This research lays the ground work for further exploration in other low and high incidence areas, of not only IBD elsewhere in Canada but also for other IMID, for similar associations as well as determination as to what extent these microorganisms appear in the human gut of affected and unaffected individuals. We conclude that drinking water as an etiological source warrants further study.

IMID are multifaceted disorders that share many common underlying dynamics including epidemiological co-occurrence, genetic susceptibility, and numerous environmental factors. IMID are highly influenced by the structural and functional characteristics of the gut microbiome. Recent technological advancements have facilitated comprehensive analyses through large cohort studies and have expanded our knowledge of the structure and function of the gut in health and in many diseases. Most available gut microbiome research is correlative and experimental models for testing hypotheses emergent from correlative associations are necessary. Investigation of the microbiome and relevant mechanisms poses a significant challenge because of the complexity of the relationships between the microbiota with host genetics and environmental factors. Nevertheless, many microbiome surveys have focused on only a single IMID; our fecal microbiome study, which includes multiple disease cohorts, has allowed for the simultaneous investigation of different diseases to accurately examine how particular microbial groups are altered between IMID. Furthermore, our mucosal microbiome

study has revealed interesting findings related to the non-inflamed mucosa and hence strongly encourages interrogation into the mucosal microbiome in other IMID.

CHAPTER SEVEN: REFERENCES

Abbott, A. (2016). Scientists bust myth that our bodies have more bacteria than human cells. *Nat. News*. doi:doi:10.1038/nature.2016.19136.

Abdollahi-Roodsaz, S., Joosten, L. A. B., Koenders, M. I., Devesa, I., Roelofs, M. F., Radstake, T. R. D. J., et al. (2008). Stimulation of TLR2 and TLR4 differentially skews the balance of T cells in a mouse model of arthritis. *J. Clin. Invest.* 118, 205–216. doi:10.1172/JCI32639.

Adams, L. A., and Morrison, M. (2016). The Microbiome in Obesity, Diabetes, and NAFLD: What is Your Gut Telling Us? *Curr. Hepatol. Reports*, 1–7. doi:10.1007/s11901-016-0299-5.

Agus, A., Denizot, J., Thévenot, J., Martinez-Medina, M., Massier, S., Sauvanet, P., et al. (2016). Western diet induces a shift in microbiota composition enhancing susceptibility to Adherent-Invasive E. coli infection and intestinal inflammation. *Sci. Rep.* 6, 19032. doi:10.1038/srep19032.

Alamanos, Y., Voulgari, P. V., and Drosos, A. A. (2006). Incidence and Prevalence of Rheumatoid Arthritis, Based on the 1987 American College of Rheumatology Criteria: A Systematic Review. *Semin. Arthritis Rheum.* 36, 182–188. doi:10.1016/j.semarthrit.2006.08.006.

Allard, G., Ryan, F. J., Jeffery, I. B., and Claesson, M. J. (2015). SPINGO: a rapid species-classifier for microbial amplicon sequences. *BMC Bioinformatics* 16, 324. doi:10.1186/s12859-015-0747-1.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST:a new generation of protein database search programs. *Nucleic Acids Res* 25, 3389–3402. doi:10.1093/nar/25.17.3389.

Ananthakrishnan, A. N. (2015). Environmental Risk Factors for Inflammatory Bowel Diseases: A Review. *Dig. Dis. Sci.* 60, 290–298. doi:10.1007/s10620-014-3350-9.

Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R., and Wu, A. Y. (1998). An optimal algorithm for approximate nearest neighbor searching in fixed dimensions. *J. ACM* 45, 891–923. doi:10.1145/293347.293348.

Ascherio, A., and Munger, K. L. (2016). Epidemiology of Multiple Sclerosis: From Risk Factors to Prevention—An Update. *Semin. Neurol.* 36, 103–114.

Atarashi, K., Tanoue, T., Oshima, K., Suda, W., Nagano, Y., Nishikawa, H., et al. (2013). Treg induction by a rationally selected mixture of Clostridia strains from the human microbiota. *Nature* 500, 232–6. doi:10.1038/nature12331.

Atarashi, K., Tanoue, T., Shima, T., Imaoka, A., Kuwahara, T., Momose, Y., et al. (2011). Induction of Colonic Regulatory T Cells by Indigenous Clostridium Species. *Science (80-.).* 331, 337–342. doi:10.1126/science.1198469.

Bai, Y., Liu, R., Liang, J., and Qu, J. (2013). Integrated Metagenomic and Physiochemical Analyses to Evaluate the Potential Role of Microbes in the Sand Filter of a Drinking Water Treatment System. *PLoS One* 8, e61011. doi:10.1371/journal.pone.0061011.

Bala, S., Marcos, M., Gattu, A., Catalano, D., and Szabo, G. (2014). Acute Binge Drinking Increases Serum Endotoxin and Bacterial DNA Levels in Healthy Individuals. *PLoS One* 9, e96864. doi:10.1371/journal.pone.0096864.

Balzola, F., Bernstein, C., and Ho, G. T. (2011). A pyrosequencing study in twins shows that gastrointestinal microbial profiles vary with inflammatory bowel disease phenotypes: Commentary. *Inflamm. Bowel Dis. Monit.* 11, 166. doi:10.1053/j.gastro.2010.08.049.

Basic, M., Keubler, L. M., Buettner, M., Achard, M., Breves, G., Schröder, B., et al. (2014). Norovirus triggered microbiota-driven mucosal inflammation in interleukin 10-deficient mice. *Inflamm. Bowel Dis.* 20, 431–43. doi:10.1097/01.MIB.0000441346.86827.ed.

Bayry, J., and Radstake, T. R. (2013). Immune-mediated inflammatory diseases: progress in molecular pathogenesis and therapeutic strategies. *Expert Rev. Clin. Immunol.* 9, 297–9. doi:10.1586/eci.13.10.

Becker, C., Neurath, M. F., and Wirtz, S. (2015). The Intestinal Microbiota in Inflammatory Bowel Disease. *Ilar J* 56, 192–204. doi:10.1093/ilar/ilv030.

Belbasis, L., Bellou, V., Evangelou, E., Ioannidis, J. P. A., and Tzoulaki, I. (2015). Environmental risk factors and multiple sclerosis: An umbrella review of systematic reviews and meta-analyses. *Lancet Neurol.* 14, 263–273. doi:10.1016/S1474-4422(14)70267-4.

Benjamin, J., Hedin, C., Koutsoumpa, A., Ng, S., McCarthy, N., Hart, A., et al. (2011). Randomised, double-blind, placebo-controlled trial of fructo- oligosaccharides in active Crohn's disease. *Gut* 60, 923–929. doi:10.1136/gut.2010.232025.

Bentley, D. R., Gormley, N. A., Balasubramanian, S., Baybayan, P. A., Benoit, V. A., Swerdlow, H. P., et al. (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456, 53–9. doi:10.1038/nature07517.

Berer, K., Mues, M., Koutrolos, M., Rasbi, Z. A., Boziki, M., Johner, C., et al. (2011). Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination - with comments. *Nature* 479, 538–541. doi:10.1038/nature10554.

Bernstein, C. N., Blanchard, J. F., Leslie, W., Wajda, A., and Yu, B. N. (2000). The Incidence of Fracture among Patients with Inflammatory Bowel Disease: A Population-Based Cohort Study. *Ann. Intern. Med.* 133, 795–799.

Bernstein, C. N., Blanchard, J. F., Rawsthorne, P., and Wajda, a (1999). Epidemiology of Crohn's disease and ulcerative colitis in a central Canadian province: a population-based study. *Am. J. Epidemiol.* 149, 916–924. doi:10.1093/oxfordjournals.aje.a009735.

Bernstein, C. N., Blanchard, J., Houston, D., and Wajda, A. (2001a). The incidence of venous thromboembolic disease among patients with IBD: A population-based study. *Thromb. Haemost.* 85, 430–434.

Bernstein, C. N., Blanchard, J., Rawsthorne, P., and Yu, N. (2001b). The prevalence of extraintestinal diseases in inflammatory bowel disease: a population-based study. *Am. J. Gastroenterol.* 96, 1116–1122. doi:10.1111/j.1572-0241.2001.03756.x.

Bernstein, C. N., Blanchard, J., and Wajda, A. (2008). The incidence of arterial thromboembolic diseases in inflammatory bowel disease: A population-based study. *Clin. Gastroenterol. Hepatol.* 6, 41–45.

Bernstein, C. N., El-gabalawy, H., Sargent, M., Landers, C. J., Rawsthorne, P., Elias, B., et al. (2011). Assessing inflammatory bowel disease-associated antibodies in Caucasian and First Nations cohorts. *Can. J. Gastroenterol.* 25, 269–274.

Bernstein, C. N., Rawsthorne, P., Cheang, M., Math, M., and Blanchard, J. F. (2006). A Population-Based Case Control Study of Potential Risk Factors for IBD. *Am. J. Gastroenterol.* 101, 993–1002. doi:10.1111/j.1572-0241.2006.00381.x.

Bernstein, C. N., Wajda, A., and Blanchard, J. F. (2005). The clustering of other chronic inflammatory diseases in inflammatory bowel disease: A population-based study. *Gastroenterology* 129, 827–836. doi:10.1053/j.gastro.2005.06.021.

Bibiloni, R., Mangold, M., Madsen, K. L., Fedorak, R. N., and Tannock, G. W. (2006). The

bacteriology of biopsies differs between newly diagnosed, untreated, Crohn's disease and ulcerative colitis patients. *J. Med. Microbiol.* 55, 1141–1149. doi:10.1099/jmm.0.46498-0.
Birchenough, G. M. H., Johansson, M. E., Gustafsson, J. K., Bergström, J. H., and Hansson, G. C. (2015). New developments in goblet cell mucus secretion and function. *Mucosal Immunol.* 8, 712–719. doi:10.1038/mi.2015.32.
Blasco-baque, V., Garidou, L., Pomié, C., Escoula, Q., Loubieres, P., Gall-david, S. Le, et al. (2016). Periodontitis induced by Porphyromonas gingivalis drives periodontal microbiota dysbiosis and insulin resistance via an impaired adaptive immune response. *Gut* 0, 1–14. doi:10.1136/gutjnl-2015-309897.
Brant, S. R. (2011). Update on the Heritability of Inflammatory Bowel Disease: The Importance of Twin Studies. *Inflamm. Bowel Dis.* 17, 1–5. doi:10.1002/ibd.21385.
Bray, J., and Curtis, J. (1957). An ordination of upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27, 325–349.
Brink, M., Hansson, M., Mathsson, L., Jakobsson, P. J., Holmdahl, R., Stenlund, H., et al. (2013). Multiplex analyses of antibodies against citrullinated peptides in individuals prior to development of rheumatoid arthritis. *Arthritis Rheum.* 65, 899–910. doi:10.1002/art.37835.
van den Broek, M. F., van Bruggen, M. C., Koopman, J. P., Hazenberg, M. P., and van den Berg, W. B. (1992). Gut flora induces and maintains resistance against streptococcal cell wall-induced arthritis in F344 rats. *Clin Exp Immunol* 88, 313–317. Available at: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=1572097.
Brown, E. M., Wlodarska, M., Willing, B. P., Vonaesch, P., Han, J., Reynolds, L. A., et al. (2015). Diet and specific microbial exposure trigger features of environmental enteropathy in a novel murine model. *Nat. Commun.* 6, 7806. doi:doi:10.1038/ncomms8806.
Brown, T. M., Clark, H. W., Bailey, J. S., and Gray, C. W. (1971). A mechanistic approach to treatment of rheumatoid type arthritis naturally occurring in a gorilla. *Trans. Am. Clin. Climatol. Assoc.* 82, 227–247.
Buscarinu, M. C., Cerasoli, B., Annibali, V., Policano, C., Lionetto, L., Capi, M., et al. (2016). Altered intestinal permeability in patients with relapsing – remitting multiple sclerosis : A pilot study. *Mult. Scler. J.*, 1–5. doi:10.1177/1352458516652498.
Cadwell, K., Patel, K. K., Maloney, N. S., Liu, T. C., Ng, A. C. Y., Storer, C. E., et al. (2010). Virus-Plus-Susceptibility Gene Interaction Determines Crohn's Disease Gene Atg16L1 Phenotypes in Intestine. *Cell* 141, 1135–1145. doi:10.1016/j.cell.2010.05.009.
Cantarel, B. L., Waubant, E., Chehoud, C., Kuczynski, J., DeSantis, T. Z., Warrington, J., et al. (2015). Gut microbiota in multiple sclerosis. *J. Investig. Med.* 63, 729–734.
Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., et al. (2010). QIIME allows analysis of high- throughput community sequencing data. *Nat. Publ. Gr.* 7, 335–336. doi:10.1038/nmeth0510-335.
Castro-Mejía, J. L., Muhammed, M. K., Kot, W., Neve, H., Franz, C. M. A. P., Hansen, L. H., et al. (2015). Optimizing protocols for extraction of bacteriophages prior to metagenomic analyses of phage communities in the human gut. *Microbiome* 3, 64. doi:10.1186/s40168-015-0131-4.
Catrina, A. I., Deane, K. D., and Scher, J. U. (2016). Gene, environment, microbiome and mucosal immune tolerance in rheumatoid arthritis. *Rheumatology* 55, 391–402. doi:10.1093/rheumatology/keu469.
Chakravorty, S., Helb, D., Burday, M., and Connell, N. (2007). A detailed analysis of 16S

ribosomal RNA gene segments for the diagnosis of pathogenic bacteria. *J Microbiol Methods* 69, 330–339. doi:10.1016/j.mimet.2007.02.005.A.

Chao, A. (1984). Nonparametric Estimation of the Number of Classes in a Population. *Scand. J. Stat.* 11, 265–270.

Chao, A., and Lee, S.-M. (1992). Estimating the Number of Classes via Sample Coverage. *J. Am. Stat. Assoc.* 87, 210–217.

Chen, J., Chia, N., Kalari, K. R., Yao, J. Z., Novotna, M., Soldan, M. M. P., et al. (2016a). Multiple sclerosis patients have a distinct gut microbiota compared to healthy controls. *Sci. Rep.* 6, 28484. doi:10.1038/srep28484.

Chen, J., Wright, K., Davis, J. M., Jeraldo, P., Marietta, E. V., Murray, J., et al. (2016b). An expansion of rare lineage intestinal microbes characterizes rheumatoid arthritis. *Genome Med.* 8, 43. doi:10.1186/s13073-016-0299-7.

Choo, J. M., Leong, L. E., and Rogers, G. B. (2015). Sample storage conditions significantly influence faecal microbiome profiles. *Sci. Rep.* 5, 16350. doi:10.1038/srep16350.

Claesson, M. J., Jeffery, I. B., Conde, S., Power, S. E., O'Connor, E. M., Cusack, S., et al. (2012). Gut microbiota composition correlates with diet and health in the elderly. *Nature* 488, 178. doi:10.1038/nature11319.

Clemente, J. C., Ursell, L. K., Parfrey, L. W., and Knight, R. (2012). The Impact of the Gut Microbiota on Human Health : An Integrative View. *Cell* 148, 1258–1270. doi:10.1016/j.cell.2012.01.035.

Cohen, R., Robinson, D., Paramore, C., Fraeman, K., Renahan, K., and Bala, M. (2008). Autoimmune disease concomitance among inflammatory bowel disease patients in the United States, 2001-2002. *Inflamm. Bowel Dis.* 14, 738–743. doi:10.1002/ibd.20406.

De Cruz, P., Kang, S., Wagner, J., Buckley, M., Sim, W. H., Prideaux, L., et al. (2015). Association between specific mucosa-associated microbiota in Crohn's disease at the time of resection and subsequent disease recurrence: A pilot study. *J. Gastroenterol. Hepatol.* 30, 268–278. doi:10.1111/jgh.12694.

Cucino, C., and Sonnenberg, a (2001). The comorbid occurrence of other diagnoses in patients with ulcerative colitis and Crohn's disease. *Am. J. Gastroenterol.* 96, 2107–2112. doi:10.1111/j.1572-0241.2001.03943.x.

Cunningham, M. (2012). Streptococcus and rheumatic fever. *Curr. Opin. Rheumatol.* 24, 408–416.

D'Argenio, V., Casaburi, G., Precone, V., Pagliuca, C., Colicchio, R., Sarnataro, D., et al. (2016). Metagenomics Reveals Dysbiosis and a Potentially Pathogenic N. flavescens Strain in Duodenum of Adult Celiac Patients. *Am. J. Gastroenterol.*, 1–12. doi:10.1038/ajg.2016.95.

Darfeuille-Michaud, A., Boudeau, J., Bulois, P., Neut, C., Glasser, A. L., Barnich, N., et al. (2004). High prevalence of adherent-invasive Escherichia coli associated with ileal mucosa in Crohn's disease. *Gastroenterology* 127, 412–421. doi:10.1053/j.gastro.2004.04.061.

Daulatzai, M. A. (2014). Obesity and Gut's Dysbiosis Promote Neuroinflammation , Cognitive Impairment , and Vulnerability to Alzheimer ' s disease: New Directions and Therapeutic Implications. *Mol. Genet. Med.* S1. doi:10.4172/1747-0862.S1-005.

Davenport, M., Poles, J., Leung, J. M., Wolff, M. J., Abidi, W. M., Ullman, T., et al. (2014). Metabolic alterations to the mucosal microbiota in inflammatory bowel disease. *Inflamm. bowel Dis.* 29, 997–1003. doi:10.1016/j.biotechadv.2011.08.021.Secreted.

David, L. A., Materna, A. C., Friedman, J., Campos-Baptista, M. I., Blackburn, M. C., Perrotta,

A., et al. (2014). Host lifestyle affects human microbiota on daily timescales. *Genome Biol.* 15, R89. doi:10.1186/gb-2014-15-7-r89.
Dendrou, C. a., Fugger, L., and Friese, M. a. (2015). Immunopathology of multiple sclerosis. *Nat. Rev. Immunol.* 15, 545–558. doi:10.1038/nri3871.
DeSantis, T. Z., Hugenholtz, P., Keller, K., Brodie, E. L., Larsen, N., Piceno, Y. M., et al. (2006a). NAST: A multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res.* 34, 394–399. doi:10.1093/nar/gkl244.
DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006b). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi:10.1128/AEM.03006-05.
Dheer, R., Davies, J., Quintero, M. A., Santaolalla, R., Deshpande, A., Kerman, D., et al. (2016). Mucosal Microbiota that Breach the Intestinal Epithelial Barrier—A Step Towards Identifying Potential Pathogens in IBD. *Inflamm Bowel Dis* 22, 2016. doi:doi: 10.1097/01.MIB.0000480312.66811.0f.
Diaz-Gallo, L.-M., and Martin, J. (2012). Common genes in autoimmune diseases: a link between immune-mediated diseases. *Expert Rev. Clin. Immunol.* 8, 107–9. doi:10.1586/eci.11.90.
Dicksved, J., Halfvarson, J., Rosenquist, M., Jarnerot, G., Tysk, C., Apajalahti, J., et al. (2008). Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *ISME J.* 2, 716–727. doi:10.1038/ismej.2008.37.
Distrutti, E., Monaldi, L., Ricci, P., Fiorucci, S., Distrutti, E., Monaldi, L., et al. (2016). Gut microbiota role in irritable bowel syndrome: New therapeutic strategies. *World J. Gastroenterol.* 22, 2219–2241. doi:10.3748/wjg.v22.i7.2219.
Dixon, P. (2003). VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* 14, 927–930. doi:10.1111/j.1654-1103.2003.tb02228.x.
Doig, C. J., Sutherland, L. R., Sandham, J. D., Fick, G. H., Verhoef, M., and Meddings, J. O. N. B. (1998). Increased Intestinal Permeability Is Associated with the Development of Multiple Organ Dysfunction Syndrome in Critically Ill ICU Patients. *Am. J. Respir. Clin. Care Med.* 158, 444–451.
Dollive, S., Peterfreund, G. L., Sherrill-Mix, S., Bittinger, K., Sinha, R., Hoffmann, C., et al. (2012). A tool kit for quantifying eukaryotic rRNA gene sequences from human microbiome samples. *Genome Biol* 13, R60. doi:gb-2012-13-7-r60 [pii]\r10.1186/gb-2012-13-7-r60.
Duck, L. W., Walter, M. R., Novak, J., Kelly, D., Tomasi, M., Cong, Y., et al. (2007). Isolation of flagellated bacteria implicated in Crohn's disease. *Inflamm. Bowel Dis.* 13, 1191–1201. doi:10.1002/ibd.20237.
Duncan, S. H., Hold, G. L., Barcenilla, A., Stewart, C. S., and Flint, H. J. (2002). Roseburia intestinalis sp. nov., a novel saccharolytic, butyrate-producing bacterium from human faeces. *Int. J. Syst. Evol. Microbiol.* 52, 1615–1620. doi:10.1099/ijs.0.02143-0.
Duranti, S., Gaiani, F., Mancabelli, L., Milani, C., Grandi, A., Bolchi, A., et al. (2016). Elucidating the gut microbiome of ulcerative colitis: bifidobacteria as novel microbial biomarkers. *FEMS Microbiol. Ecol.* 92, fiw191. doi:10.1093/femsec/fiw191.
Eckburg, P. B., Bik, E. M., Berstein, C. N., Purdom, E., Dethlefsen, L., Sargent, M., et al. (2005). Diversity of the human intestinal microbial flora. *Science (80-.).* 308, 1635–1638. doi:10.2307/3841624.
Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves

sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi:10.1093/bioinformatics/btr381.
El-Gabalawy, H. S., Robinson, D. B., Daha, N. A., Oen, K. G., Smolik, I., Elias, B., et al. (2011). Non-HLA genes modulate the risk of rheumatoid arthritis associated with HLA-DRB1 in a susceptible North American Native population. *Genes Immun.* 12, 568–574. doi:10.1038/gene.2011.30.
Ezendam, J., De Klerk, A., Gremmer, E. R., and Van Loveren, H. (2008). Effects of Bifidobacterium animalis administered during lactation on allergic and autoimmune responses in rodents. *Clin. Exp. Immunol.* 154, 424–431. doi:10.1111/j.1365-2249.2008.03788.x.
Ezendam, J., and van Loveren, H. (2008). Lactobacillus casei Shirota administered during lactation increases the duration of autoimmunity in rats and enhances lung inflammation in mice. *Br. J. Nutr.* 99, 83–90. doi:10.1017/S0007114507803412.
Faith, J. J., Guruge, J. L., Charbonneau, M., Subramanian, S., Seedorf, H., Goodman, A. L., et al. (2013). The long-term stability of the human gut microbiota. *Science (80-.).* 341, 1237439. doi:10.1126/science.1237439.
Fedorow, H., Pickford, R., Hook, J. M., Double, K. L., Halliday, G. M., Gerlach, M., et al. (2005). Dolichol is the major lipid component of human substantia nigra neuromelanin. *J. Neurochem.* 92, 990–995. doi:10.1111/j.1471-4159.2004.02975.x.
Fiorucci, S., and Distrutti, E. (2015). Bile Acid-Activated Receptors, Intestinal Microbiota, and the Treatment of Metabolic Disorders. *Trends Mol. Med.* 21, 702–714. doi:10.1016/j.molmed.2015.09.001.
Forbes, J. D., Van Domselaar, G., and Bernstein, C. N. (2016). Microbiome Survey of the Inflamed and Noninflamed Gut at Different Compartments Within the Gastrointestinal Tract of Inflammatory Bowel Disease Patients. *Inflamm. Bowel Dis.* 22, 817–825. doi:10.1097/MIB.0000000000000684.
Fox, R., Bacon, T., Chamot, E., Salter, A., Cutter, G., Kalina, J., et al. (2015). Prevalence of multiple sclerosis symptoms across lifespan: data from the NARCOMS Registry. *Neurodegener. Dis. Manag.* 5, 3–10.
Frank, D. N., St Amand, A. L., Feldman, R. a, Boedeker, E. C., Harpaz, N., and Pace, N. R. (2007). Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc. Natl. Acad. Sci. U. S. A.* 104, 13780–13785. doi:10.1073/pnas.0706625104.
Freestone, P. P. E., Walton, N. J., Haigh, R. D., and Lyte, M. (2007). Influence of dietary catechols on the growth of enteropathogenic bacteria. *Int. J. Food Microbiol.* 119, 159–169. doi:10.1016/j.ijfoodmicro.2007.07.039.
Frisell, T., Hellgren, K., Alfredsson, L., Raychaudhuri, S., Klareskog, L., and Askling, J. (2016). Familial aggregation of arthritis-related diseases in seropositive and seronegative rheumatoid arthritis: a register-based case-control study in Sweden. *Ann. Rheum. Dis.* 75, 183–189.
Gevers, D., Kugathasan, S., Denson, L. a., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., et al. (2014). The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* 15, 382–392. doi:10.1016/j.chom.2014.02.005.
Gill, T., Asquith, M., Rosenbaum, J. T., and Colbert, R. A. (2015). The intestinal microbiome in spondyloarthritis. *Curr. Opin. Rheumatol.* 27, 319–325. doi:10.1097/BOR.0000000000000187.

Gloor, G. B., Hummelen, R., Macklaim, J. M., Dickson, R. J., Fernandes, A. D., MacPhee, R., et al. (2010). Microbiome profiling by illumina sequencing of combinatorial sequence-tagged PCR products. *PLoS One* 5. doi:10.1371/journal.pone.0015406.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11, R86. doi:10.1186/gb-2010-11-8-r86.

Gomez-Alvarez, V., Revetta, R. P., and Domingo, J. W. S. (2012). Metagenomic analyses of drinking water receiving different disinfection treatments. *Appl. Environ. Microbiol.* 78, 6095–6102. doi:10.1128/AEM.01018-12.

Goodman, B. E. (2010). Insights into digestion and absorption of major nutrients in humans. *Adv. Physiol. Educ.* 34, 44–53. doi:10.1152/advan.00094.2009.

Gophna, U., Sommerfeld, K., Gophna, S., Doolittle, W. F., and Veldhuyzen Van Zanten, S. J. O. (2006). Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *J. Clin. Microbiol.* 44, 4136–4141. doi:10.1128/JCM.01004-06.

Gordon, H., and Harbord, M. (2014). A patient with severe Crohn's colitis responds to Faecal Microbiota Transplantation. *J. Crohn's Colitis* 8, 256–257. doi:10.1016/j.crohns.2013.10.007.

Green, C., Elliott, L., Beaudoin, C., and Bernstein, C. N. (2006). A population-based ecologic study of inflammatory bowel disease: Searching for etiologic clues. *Am. J. Epidemiol.* 164, 615–623. doi:10.1093/aje/kwj260.

Guerrero, C., Tort, J., Pérez, J., Andrés, M., and Espejo, E. (2015). Rhodococcus equi infection in a patient with Crohn's disease treated with infliximab...J Infect. 2013 Dec;67(6):524-8. *J. Infect.* 70, 689–690 2p. doi:10.1016/j.jinf.2014.12.008.

Gullberg, R. (1978). Possible Role of Alterations of the Intestinal Flora in Rheumatoid Arthritis. *Rheumatology (Oxford)*. XVII, 5–10.

Gupta, A., Kang, S., Wagner, J., Kirkwood, C. D., Morrison, M., McSweeney, C., et al. (2014). Analysis of mucosal microbiota in inflammatory bowel disease using a custom phylogenetic microarray. *Austin J. Gastroenterol.* 1, 6.

Hafer, A., Krämer, S., Duncker, S., Krüger, M., Manns, M. P., and Bischoff, S. C. (2007). Effect of oral lactulose on clinical and immunohistochemical parameters in patients with inflammatory bowel disease: a pilot study. *BMC Gastroenterol.* 7, 1–11. doi:10.1186/1471-230X-7-36.

Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J., and Knight, R. (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* 5, 235–237. doi:10.1038/nmeth.1184.

Hamburg, J. van, Paulissen, S., Davelaar, N., Hazes, J., and Lubberts, E. (2016). IL-17A-low CCR6+ TH cell populations of patients with rheumatoid arthritis are pathogenic, multidrug resistant and associated with dmard and glucocorticoid treatment response. *Ann. Rheum. Dis.* 75, A22–A23. doi:doi:10.1136/annrheumdis-2016-209124.54.

Hao, W., and Lee, Y. (2004). Microflora of the Gastrointestinal Tract A Review. *Methods Mol Biol* 268, 491–502.

Harbord, M., Annese, V., Vavricka, S. R., Allez, M., Barreiro-De Acosta, M., Boberg, K. M., et al. (2016). The First European Evidence-based Consensus on Extra-intestinal Manifestations in Inflammatory Bowel Disease for the European Crohn's and Colitis Organisation [ECCO]. *J. Crohn's Colitis*, 1–16. doi:10.1093/ecco-jcc/jjv213.

Hayashi, H., Sakamoto, M., and Benno, Y. (2004). Evaluation of three different forward primers by terminal restriction fragment length polymorphism analysis for determination of fecal bifidobacterium spp. in healthy subjects. *Microbiol. Immunol.* 48, 1–6. doi:10.1111/j.1348-0421.2004.tb03481.x.

Hedstrom, A. K., Alfredsson, L., and Olsson, T. (2016). Environmental factors and their interactions with risk genotypes in MS susceptibility. *Curr. Opin. Neurol.* 29, 293–298. doi:10.1097/WCO.0000000000000329.

Helander, H. F., and Fandriks, L. (2014). Surface area of the digestive tract - revisited. *Scand J Gastroenterol* 49, 681–689. doi:10.3109/00365521.2014.898326.

Herrmann, I., Kellert, M., Schmidt, H., Mildner, A., Hanisch, U. K., Brück, W., et al. (2006). Streptococcus pneumoniae infection aggravates experimental autoimmune encephalomyelitis via toll-like receptor 2. *Infect. Immun.* 74, 4841–4848. doi:10.1128/IAI.00026-06.

Ho, J., Reimer, R. A., Doulla, M., and Huang, C. (2016). Effect of prebiotic intake on gut microbiota , intestinal permeability and glycemic control in children with type 1 diabetes : study protocol for a randomized controlled trial. *Trials* 17, 1–8. doi:10.1186/s13063-016-1486-y.

Hoarau, G., Mukherjee, P. K., Gower-Rousseau, C., Hager, C., Chandra, J., Retuerto, M. A., et al. (2016). Bacteriome and Mycobiome Interactions Underscore Microbial Dysbiosis in Familial Crohn's Disease. *MBio* 7, e01250-16. doi:10.1128/MBIO.01250-16.

Höfer, D., Jöns, T., Kraemer, J., and Drenckhahn, D. (1998). From Cytoskeleton to Polarity and Chemoreception in the Gut Epithelium. *Ann. N. Y. Acad. Sci.* 859.

Hoffmann, C., Dollive, S., Grunberg, S., Chen, J., Li, H., Wu, G. D., et al. (2013). Archaea and Fungi of the Human Gut Microbiome: Correlations with Diet and Bacterial Residents. *PLoS One* 8, e66019. doi:10.1371/journal.pone.0066019.

Hollon, J., Puppa, E. L., Greenwald, B., Goldberg, E., Guerrerio, A., and Fasano, A. (2015). Effect of Gliadin on Permeability of Intestinal Biopsy Explants from Celiac Disease Patients and Patients with Non-Celiac Gluten Sensitivity. *Nutrients* 7, 1565–1576. doi:10.3390/nu7031565.

Holmes, E. A., Xiang, F., and Lucas, R. M. (2015). Variation in Incidence of Pediatric Crohn's Disease in Relation to Latitude and Ambient Ultraviolet Radiation: A Systematic Review and Analysis. *Inflamm. Bowel Dis.* 21, 809–817. doi:10.1097/MIB.0000000000000320.

Holmøy, T., and Torkildsen, Ø. (2016). Can vitamin D reduce inflammation in relapsing-remitting multiple sclerosis? *Expert Rev. Neurother.* 16, 233–235. doi:10.1586/14737175.2016.1146134.

Howard, J., Trevick, S., and Younger, D. S. (2016). Epidemiology of Multiple Sclerosis. *Neurol. Clin.* 34, 919–939. doi:10.1016/j.ncl.2016.06.016.

Hughes-Austin, J. M., Deane, K. D., Derber, L. a, Kolfenbach, J. R., Zerbe, G. O., Sokolove, J., et al. (2013). Multiple cytokines and chemokines are associated with rheumatoid arthritis-related autoimmunity in first-degree relatives without rheumatoid arthritis: Studies of the Aetiology of Rheumatoid Arthritis (SERA). *Ann. Rheum. Dis. Dis.* 72, 901–7. doi:10.1136/annrheumdis-2012-201505.

Huh, H. J., Lee, S. T., Lee, J. H., Ki, C. S., and Lee, N. Y. (2010). *Clostridium symbiosum* Isolated from Blood. *Korean J. Clin. Microbiol.* 13, 90. doi:10.5145/KJCM.2010.13.2.90.

Huson, D., Auch, A., Qi, J., and Schuster, S. (2007). MEGAN analysis of metagenome data. *Genome Res.* 17, 377–386. doi:10.1101/gr.5969107.

Huttenhower, C., Gevers, D., Knight, R., Abubucker, S., Badger, J. H., Chinwalla, A. T., et al. (2012). Structure, function and diversity of the healthy human microbiome. *Nature* 486, 207–214. doi:10.1038/nature11234.

Hvatum, M., Kanerud, L., Hällgren, R., and Brandtzaeg, P. (2006). The gut–joint axis: cross reactive food antibodies in rheumatoid arthritis. *Gut Immunol.* 55, 1240–1248. doi:10.1136/gut.2005.076901.

Jaccard, P. (1912). The distribution of the flora in the alpine zone. *New Phytol.* 11, 37–50.

Jafari, N. V, Kuehne, S. A., Minton, N. P., Allan, E., and Bajaj-elliott, M. (2016). Anaerobe Clostridium difficile -mediated effects on human intestinal epithelia : Modelling host-pathogen interactions in a vertical diffusion chamber. *Anaerobe* 37, 96–102. doi:10.1016/j.anaerobe.2015.12.007.

Jangi, S., Gandhi, R., Cox, L. M., Li, N., von Glehn, F., Yan, R., et al. (2016). Alterations of the human gut microbiome in multiple sclerosis. *Nat. Commun.* 7, 12015. doi:10.1038/ncomms12015.

Jarrett, F., Ducasa, G. M., Buller, D. B., and Berwick, M. (2014). The Effect of Oral Supplementation of Vitamin D3 on Serum Levels of Vitamin D: A Review. *Epidemiol. Open Access* 4, 2–6. doi:10.4172/2161-1165.1000148.

Jenks, K., Stebbings, S., Burton, J., Schultz, M., Herbison, P., and Highton, J. (2010). Probiotic therapy for the treatment of spondyloarthritis: A randomized controlled trial. *J. Rheumatol.* 37, 2118–2125. doi:10.3899/jrheum.100193.

Jensen, S. R., Mirsepasi-Lauridsen, H. C., Thysen, A. H., Brynskov, J., Krogfelt, K. A., Petersen, A. M., et al. (2015). Distinct inflammatory and cytopathic characteristics of Escherichia coli isolates from inflammatory bowel disease patients. *Int. J. Med. Microbiol.* 305, 925–936. doi:10.1016/j.ijmm.2015.10.002.

Jhangi, S., Gandhi, R., Glanz, B., Cook, S., Nejad, P., Ward, D., et al. (2014a). Increased Archaea species and changes with therapy in gut microbiome of multiple sclerosis subjects. *Neurology* 82, S24.001.

Jhangi, S., Gandhi, R., Glanz, B., Cook, S., Nejad, P., Ward, D., et al. (2014b). Increased Archaea species and changes with therapy in gut microbiome of multiple sclerosis subjects (S24. 001). *Neurology* 82, S24.001.

Jin, D., Wu, S., Zhang, Y. G., Lu, R., Xia, Y., Dong, H., et al. (2015). Lack of Vitamin D Receptor Causes Dysbiosis and Changes the Functions of the Murine Intestinal Microbiome. *Clin. Ther.* 37, 996–1009.e7. doi:10.1016/j.clinthera.2015.04.004.

Johnson, C. C., and Ownby, D. R. (2016). Allergies and Asthma: Do Atopic Disorders Result from Inadequate Immune Homeostasis arising from Infant Gut Dysbiosis? *Expert Rev. Clin. Immunol.* 12, 379–388. doi:10.1586/1744666X.2016.1139452.

Johnstone, C., Hendry, C., Farley, A., and McLafferty, E. (2016). The digestive system: part 1. *Nurs. Stand.* 28, 37–45.

Joller, N., Lozano, E., Burkett, P. R., Patel, B., Xiao, S., Zhu, C., et al. (2014). Treg cells expressing the coinhibitory molecule TIGIT selectively inhibit proinflammatory Th1 and Th17 cell responses. *Immunity* 40, 569–581. doi:10.1016/j.immuni.2014.02.012.

Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., et al. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut* 60, 631–7. doi:10.1136/gut.2010.223263.

Joshua, V., Reynisdottir, G., Ytterberg, J., and Al., E. (2014). Characterisation of lung inflammation and identification of shared citrullinated targets in the lungs and joints of

early rheumatoid arthritis. *Ann. Rheum. Dis.* 73, A4-5.
Kabeerdoss, J., Sankaran, V., Pugazhendhi, S., and Ramakrishna, B. S. (2013). Clostridium leptum group bacteria abundance and diversity in the fecal microbiota of patients with inflammatory bowel disease: a case-control study in India. *BMC Gastroenterol.* 13, 20. doi:10.1186/1471-230X-13-20.
Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277-80. doi:10.1093/nar/gkh063.
Kang, S., Denman, S. E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., et al. (2010). Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm. Bowel Dis.* 16, 2034–2042. doi:10.1002/ibd.21319.
Kaplan, M. H. (2013). Th9 cells: Differentiation and disease. *Immunol. Rev.* 252, 104–115. doi:10.1111/imr.12028.
Katyal, I., Chaban, B., and Hill, J. E. (2016). Comparative Genomics of cpn60-Defined Enterococcus hirae Ecotypes and Relationship of Gene Content Differences to Competitive Fitness. *Microb. Ecol.* 72, 917–930. doi:10.1007/s00248-015-0708-2.
Keely, S., Walker, M. M., Marks, E., and Talley, N. J. (2015). Immune dysregulation in the functional GI disorders. *Eur. J. Clin. Invest.* 45, 1350–1359. doi:10.1111/eci.12548.
Kelly, C. J., Zheng, L., Campbell, E. L., Saeedi, B., Scholz, C. C., Bayless, a J., et al. (2015). Crosstalk between Microbiota-Derived Short-Chain Fatty Acids and Intestinal Epithelial HIF Augments Tissue Barrier Function. *Cell Host Microbe* 17, 662–671. doi:10.1016/j.chom.2015.03.005.
Kembel, S. W., Wu, M., Eisen, J. A., and Green, J. L. (2012). Incorporating 16S Gene Copy Number Information Improves Estimates of Microbial Diversity and Abundance. *PLoS Comput. Biol.* 8, 16–18. doi:10.1371/journal.pcbi.1002743.
Kernbauer, E., Ding, Y., and Cadwell, K. (2014). An enteric virus can replace the beneficial function of commensal bacteria. *Nature* 516, 94–98. doi:10.1038/nature13960.
Khan, K. J., Ullman, T. a, Ford, A. C., Abreu, M. T., Abadir, A., Marshall, J. K., et al. (2011). Antibiotic therapy in inflammatory bowel disease: a systematic review and meta-analysis. *Am. J. Gastroenterol.* 106, 661–673. doi:10.1038/ajg.2011.72.
Kleyer, A., Finzel, S., Rech, J., Manger, B., Krieter, M., Faustini, F., et al. (2014). Bone loss before the clinical onset of rheumatoid arthritis in subjects with anticitrullinated protein antibodies. *Ann. Rheum. Dis.* 73, 854–60. doi:10.1136/annrheumdis-2012-202958.
Kloppenburg, M., Breedveld, F. C., Terwiel, J. P., Mallee, C., and Dijkmans, B. A. (1994). Minocycline in active rheumatoid arthritis. *Arthritis Rheumatol.* 37, 629–636. doi:10.1002/art.1780370505.
Knösel, T., Schewe, C., Petersen, N., Dietel, M., and Petersen, I. (2009). Prevalence of infectious pathogens in Crohn's disease. *Pathol. Res. Pract.* 205, 223–230. doi:10.1016/j.prp.2008.04.018.
Knudsen, B. E., Bergmark, L., Munk, P., Lukjancenko, O., Prieme, A., Aarestrup, F. M., et al. (2016). Impact of Sample Type and DNA Isolation Procedure on Genomic Inference of Microbiome Composition. *mSystems* 1, 64394. doi:10.1101/064394.
Kobayashi, T., Kato, I., Nanno, M., Shida, K., Shibuya, K., Matsuoka, Y., et al. (2010). Oral administration of probiotic bacteria, Lactobacillus casei and Bifidobacterium breve, does not exacerbate neurological symptoms in experimental autoimmune encephalomyelitis. *Immunopharmacol. Immunotoxicol.* 32, 116–24. doi:10.3109/08923970903200716.
Koch-Henriksen, N., and Sorensen, P. S. (2011). Why does the north-south gradient of incidence

of multiple sclerosis seem to have disappeared on the Northern hemisphere? *J. Neurol. Sci.* 311, 58–63. doi:10.1016/j.jns.2011.09.003.
Koch-Henriksen, N., and Sørensen, P. S. (2010). The changing demographic pattern of multiple sclerosis epidemiology. *Lancet Neurol.* 9, 520–532. doi:10.1016/S1474-4422(10)70064-8.
Kohashi, O., Kohashi, Y., Takahashi, H., Ozawa, A., and Shigematsu, N. (1985). Reverse effect of gram-positive bacteria vs. gram-negative bacteria on adjuvant-induced arthritis in germfree rats. *Microbiol Immunol* 29, 487–497.
Kohashi, O., Kohashi, Y., Takahashi, H., Ozawa, A., and Shigematsu, N. (1986). Suppressive effect of Escherichia coli on adjuvant-induced arthritis in germ-free ratse. *Arthritis Rheum.* 29, 547–553.
Kohashi, O., Kuwata, J., Umehara, K., Uemura, F., Takahashi, T., and Ozawa, A. (1979). Susceptibility to adjuvant-induced arthritis among germfree, specific-pathogen-free, and conventional rats. *Infect. Immun.* 26, 791–794.
Kolho, K.-L., Korpela, K., Jaakkola, T., Pichai, M. V. a, Zoetendal, E. G., Salonen, A., et al. (2015). Fecal Microbiota in Pediatric Inflammatory Bowel Disease and Its Relation to Inflammation. *Am. J. Gastroenterol.* 110, 921–930. doi:10.1038/ajg.2015.149.
Kostic, A. D., Xavier, R. J., and Gevers, D. (2014). The Microbiome in Inflammatory Bowel Disease: Current Status and the Future Ahead. *Gastroenterology* 146, 1489–1499. doi:10.1053/j.gastro.2014.02.009.
Kotzamani, D., Panou, T., Mastorodemos, V., Tzagournissakis, M., Nikolakaki, H., Spanaki, C., et al. (2012). Rising incidence of multiple sclerosis in females associated with urbanization. *Neurology* 78, 1728–1735.
Krause, D. O., Little, A. C., Dowd, S. E., and Bernstein, C. N. (2011). Complete genome sequence of adherent invasive escherichia coli UM146 isolated from ileal crohn's disease biopsy tissue. *J. Bacteriol.* 193, 583. doi:10.1128/JB.01290-10.
Kuhn, K. A., Kulik, L., Tomooka, B., Braschler, K. J., Arend, W. P., Robinson, W. H., et al. (2006). Antibodies against citrullinated proteins enhance tissue injury in experimental autoimmune arthritis. *J. Clin. Invest.* 116, 961–973. doi:10.1172/JCI25422.
Kurtzke, J. F. (1995). MS epidemiology world wide. One view of current status. *Acta Neurol. Scand.*, 23–33. doi:10.1111/j.1600-0404.1995.tb05853.x.
Kwon, H. K., Kim, G. C., Kim, Y., Hwang, W., Jash, A., Sahoo, A., et al. (2013). Amelioration of experimental autoimmune encephalomyelitis by probiotic mixture is mediated by a shift in T helper cell immune response. *Clin. Immunol.* 146, 217–227. doi:10.1016/j.clim.2013.01.001.
Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. a, et al. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat. Biotechnol.* 31, 814–21. doi:10.1038/nbt.2676.
Larcombe, S., Hutton, M. L., and Lyras, D. (2016). Involvement of Bacteria Other Than Clostridium difficile in Antibiotic-Associated Diarrhoea. *Trends Microbiol.* xx, 1–14. doi:10.1016/j.tim.2016.02.001.
Laurin, M., Everett, M. L. O. U., and Parker, W. (2011). The Cecal Appendix : One More Immune Component With a Function Disturbed By Post-Industrial Culture. *Anat. Rec.* 294, 567–579. doi:10.1002/ar.21357.
Lavasani, S., Dzhambazov, B., Nouri, M., Fåk, F., Buske, S., Molin, G., et al. (2010). A novel probiotic mixture exerts a therapeutic effect on experimental autoimmune encephalomyelitis mediated by IL-10 producing regulatory T cells. *PLoS One* 5.

doi:10.1371/journal.pone.0009009.
Lee, S., Hyun, Y., Min, J., Ku, K., Ho, S., and Jae, H. (2016). VvpE mediates the intestinal colonization of Vibrio vulnificus by the disruption of tight junctions. *Int. J. Med. Microbiol.* 306, 10–19. doi:10.1016/j.ijmm.2015.10.006.
Lee, Y. K., Menezes, J. S., Umesaki, Y., and Mazmanian, S. K. (2011). Proinflammatory T-cell responses to gut microbiota promote experimental autoimmune encephalomyelitis. *Proc. Natl. Acad. Sci. U. S. A.* 108 Suppl, 4615–22. doi:10.1073/pnas.1000082107.
Lepage, P., Colombet, J., Marteau, P., Sime-Ngando, T., Dore, J., and Leclerc, M. (2008). Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut* 57, 424–425. doi:10.1136/gut.2007.134668.
Lepage, P., Hösler, R., Spehlmann, M. E., Rehman, A., Zvirbliene, A., Begun, A., et al. (2011). Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology* 141, 227–236. doi:10.1053/j.gastro.2011.04.011.
Lepage, P., Seksik, P., Sutren, M., de la Cochetière, M.-F., Jian, R., Marteau, P., et al. (2005). Biodiversity of the mucosa-associated microbiota is stable along the distal digestive tract in healthy individuals and patients with IBD. *Inflamm. Bowel Dis.* 11, 473–480. doi:10.1097/01.mib.0000159662.62651.06.
Lerner, A., Neidhöfer, S., and Matthias, T. (2016). Beyond the Joint: What's Happening in the Gut. *Int. J. Celiac Dis.* 4, xx. doi:10.12691/ijcd-4-4-1.
Li, G., Yang, M., Zhou, K., Zhang, L., Tian, L., Lv, S., et al. (2015). Diversity of duodenal and rectal microbiota in biopsy tissues and luminal contents in healthy volunteers. *J. Microbiol. Biotechnol.* 25, 1136–1145. doi:10.4014/jmb.1412.12047.
Li, Y., Ge, Q., Cao, J., Zhou, Y., Du, Y., Shen, B., et al. (2016). Association of Fusobacterium nucleatum infection with colorectal cancer in Chinese patients. *World J. Gastroenterol.* 22, 3227–33. doi:10.3748/wjg.v22.i11.3227.
Lindsay, J. O., Whelan, K., Stagg, A. J., Gobin, P., Al-Hassi, H. O., Rayment, N., et al. (2006). Clinical, microbiological, and immunological effects of fructo-oligosaccharide in patients with Crohn's disease. *Gut* 55, 348–355. doi:10.1136/gut.2005.074971.
Liu, X., Zou, Q., Zeng, B., Fang, Y., and Wei, H. (2013). Analysis of fecal lactobacillus community structure in patients with early rheumatoid arthritis. *Curr. Microbiol.* 67, 170–176. doi:10.1007/s00284-013-0338-1.
Llopis, M., Antolin, M., Carol, M., Borruel, N., Casellas, F., Martinez, C., et al. (2009). Lactobacillus casei downregulates commensals' inflammatory signals in Crohn's disease mucosa. *Inflamm. Bowel Dis.* 15, 275–283. doi:10.1002/ibd.20736.
Loftus, C. G., Loftus Jr., E. V., Harmsen, W. S., Zinsmeister, A. R., Tremaine, W. J., Melton III, L. J., et al. (2007). Update on the Incidence and Prevalence of Crohn' s Disease. *Inflamm. Bowel Dis.* 13, 254–261. doi:10.1002/ibd.20029.
Loftus Jr, E., Silverstein, M. D., Sandborn, W. J., Tremaine, W. J., Harmsen, W. S., and Zinsmeister, A. R. (2000). Ulcerative colitis in Olmsted County, Minnesota, 1940–1993: incidence, prevalence, and survival. *Gut* 46, 336–343.
Loftus Jr, E. V, Silverstein, M. D., Sandborn, W. J., Tremaine, W. J., Harmsen, W. S., and Zinsmeister, A. R. (1998). Crohn's Disease in Olmsted County, Minnesota, 1940–1993: Incidence, Prevalence, and Survival. *Gastroenterology* 114, 1161–1168.
Lopez-Siles, M., Martinez-Medina, M., Abella, C., Busquets, D., Sabat-Mir, M., Duncan, S. H., et al. (2015). Mucosa-associated Faecalibacterium prausnitzii phylotype richness is reduced in patients with inflammatory bowel disease. *Appl. Environ. Microbiol.* 81, 7582–7592.

doi:10.1128/AEM.02006-15.
Lozupone, C., and Knight, R. (2005). UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi:10.1128/AEM.71.12.8228.
Lutgens, M., Vermeire, S., Van Oijen, M., Vleggaar, F., Siersema, P., van Assche, G., et al. (2015). A rule for determining risk of colorectal cancer in patients with inflammatory bowel disease. *Clin Gastroenterol Hepatol* 13, 148–154 e1. doi:10.1016/j.cgh.2014.06.032.
Lyte, M., Vulchanova, L., and Brown, D. R. (2011). Stress at the intestinal surface: Catecholamines and mucosa-bacteria interactions. *Cell Tissue Res.* 343, 23–32. doi:10.1007/s00441-010-1050-0.
Machiels, K., Joossens, M., Sabino, J., De Preter, V., Arijs, I., Eeckhaut, V., et al. (2013). A decrease of the butyrate-producing species Roseburia hominis and Faecalibacterium prausnitzii defines dysbiosis in patients with ulcerative colitis. *Gut* 0, 1–9. doi:10.1136/gutjnl-2013-304833.
Macpherson, A. J., McCoy, K. D., Johansen, F.-E., and Brandtzaeg, P. (2008). The immune geography of IgA induction and function. *Mucosal Immunol.* 1, 11–22. doi:10.1038/mi.2007.6.
Magoč, T., and Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi:10.1093/bioinformatics/btr507.
Makrygiannakis, D., af Klint, E., Lundberg, I. E., Löfberg, R., Ulfgren, A.-K., Klareskog, L., et al. (2006). Citrullination is an inflammation-dependent process. *Ann. Rheum. Dis.* 65, 1219–1222. doi:10.1136/ard.2005.049403.
Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., et al. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 55, 205–11. doi:10.1136/gut.2005.073817.
Mansson, I., and Colldahl, H. (1965). The intestinal flora in patients with bronchial asthma and rheumatoid arthritis. *Allergy* 20, 94–104. doi:10.1111/j.1398-9995.1965.tb03360.x.
Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. a, et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–80. doi:10.1038/nature03959.
Marin-Jiminez, I., García Sánchez, V., Gisbert, J., Lázaro Pérez Calle, J., Luján, M., Gordillo Ábalos, J., et al. (2014). Prevalence of different immune mediated inflammatory diseases in patients with inflammatory bowel disease. AQUILES study. *Gastroenterol Hepatol* 37, 495–502. doi:10.1016/S1873-9946(13)60647-9.
Martinez-Medina, M., and Garcia-Gil, L. (2014). Escherichia coli in chronic inflammatory bowel diseases: an update on adherent invasive Escherichia coli pathogenicity. *World J Gastrointest Pathophysiol* 5, 213–2227.
Mason, D., Pearson, J., Richardson, A., and Alla, S. (2016). New Zealand Nationwide Population- Based Multiple Sclerosis Incidence Study Shows Increasing NorSouth Latitude Gradient. *Neurology* 86, Supplement P1.399.
Maukonen, J., Satokari, R., Mättö, J., Söderlund, H., Mattila-Sandholm, T., and Saarela, M. (2006). Prevalence and temporal stability of selected clostridial groups in irritable bowel syndrome in relation to predominant faecal bacteria. *J. Med. Microbiol.* 55, 625–633. doi:10.1099/jmm.0.46134-0.
Mayer, E. A., Savidge, T., and Shulman, R. J. (2014). Brain-gut microbiome interactions and functional bowel disorders. *Gastroenterology* 146, 1500–1512.

doi:10.1053/j.gastro.2014.02.037.
McInnes, I. B., and Schett, G. (2011). The pathogenesis of rheumatoid arthritis. *N.Engl.J.Med.* 365, 2205–2219.
McMurdie, P. J., and Holmes, S. (2013). Phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS One* 8, e61217. doi:10.1371/journal.pone.0061217.
Meij, T. De, de Groot, E., Benninga, M., Budding, D., de Boer, N., van Bodegraven, A., et al. (2015). Microbiota dynamics in paediatric Crohn's disease from active disease upon achieving clinical remission. *Eur. Crohn's Colitis Organ.*, S437–S438. doi:http://dx.doi.org.uml.idm.oclc.org/10.1093/ecco-jcc/jju027.830.
Metz, L. M., Li, D., Traboulsee, A., Myles, M. L., Duquette, P., Godin, J., et al. (2009). Glatiramer acetate in combination with minocycline in patients with relapsing--remitting multiple sclerosis: results of a Canadian, multicenter, double-blind, placebo-controlled trial. *Mult. Scler.* 15, 1183–94. doi:10.1177/1352458509106779.
Michail, S., Durbin, M., Turner, D., Griffiths, A. M., Mack, D. R., Hyams, J., et al. (2012). Alterations in the gut microbiome of children with severe ulcerative colitis. *Inflamm. Bowel Dis.* 18, 1799–1808. doi:10.1002/ibd.22860.
Mima, K., Nishihara, R., Qian, Z. R., Cao, Y., Sukawa, Y., Nowak, J. A., et al. (2015). Fusobacterium nucleatum in colorectal carcinoma tissue and patient prognosis. *Gut* 0, 1–8. doi:10.1136/gutjnl-2015-310101.
Minichiello, E., Semerano, L., and Boissier, M.-C. (2016). Time trends in the incidence, prevalence, and severity of rheumatoid arthritis: A systematic literature review. *Jt. Bone Spine*, in press. doi:10.1016/j.jbspin.2016.07.007.
Miyake, S., Kim, S., Suda, W., Oshima, K., Nakamura, M., Matsuoka, T., et al. (2015). Dysbiosis in the gut microbiota of patients with multiple sclerosis, with a striking depletion of species belonging to clostridia XIVa and IV clusters. *PLoS One* 10, 1–16. doi:10.1371/journal.pone.0137429.
Moayyedi, P., Surette, M. G., Kim, P. T., Libertucci, J., Wolfe, M., Onischi, C., et al. (2015). Fecal Microbiota Transplantation Induces Remission in Patients With Active Ulcerative Colitis in a Randomized Controlled Trial. *Gastroenterology* 149, 102–109. doi:10.1053/j.gastro.2015.04.001.
Mokili, J. L., Rohwer, F., and Dutilh, B. E. (2012). Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* 2, 63–77. doi:10.1016/j.coviro.2011.12.004.
Molodecky, N. a., Soon, I. S., Rabi, D. M., Ghali, W. a., Ferris, M., Chernoff, G., et al. (2012). Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* 142, 46–54.e42. doi:10.1053/j.gastro.2011.10.001.
Mondot, S., Lepage, P., Seksik, P., Allez, M., Tréton, X., Bouhnik, Y., et al. (2016). Structural robustness of the gut mucosal microbiota is associated with Crohn's disease remission after surgery. *Gut* 65, 954–62. doi:10.1136/gutjnl-2015-309184.
Morampudi, V., Graef, F. A., Stahl, M., Dalwadi, U., Conlin, V. S., Huang, T., et al. (2016). Tricellular Tight Junction Protein Tricellulin is targeted by the Enteropathogenic E. coli Effector EspG1 Leading to Epithelial Barrier Disruption. *Infect. Immun.* doi:10.1128/IAI.00700-16.
Morgan, M. (1991). The treatment of chronic hepatic encephalopathy. *Hepatogastroenterology* 38, 377–387.
Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V, et al. (2012a).

Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13, R79. doi:10.1186/gb-2012-13-9-r79.

Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V, et al. (2012b). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* 13, R79. doi:10.1186/gb-2012-13-9-r79.

Morisita, M. (1959). Measuring of the dispersion and analysis of distribution patterns. *Mem. Fac. Sci. Kyushu Univ. Ser. E. Biol.* 2, 215–235.

Mowry, E., Waubant, E., Chehoud, C., DeSantis, T., Kuczynski, J., and Warrington, J. (2012). Gut bacterial populations in multiple sclerosis and in health (P05.106). *Neurology* 78, P05.106.

Mukhopadhya, I., Hansen, R., Meharg, C., Thomson, J. M., Russell, R. K., Berry, S. H., et al. (2014). The fungal microbiota of de-novo paediatric inflammatory bowel disease. *Microbes Infect.* 17, 304–310. doi:10.1016/j.micinf.2014.12.001.

Mylonaki, M., Rayment, N. B., Rampton, D. S., Hudspith, B. N., and Brostoff, J. (2005). Molecular characterization of rectal mucosa-associated bacterial flora in inflammatory bowel disease. *Inflamm. Bowel Dis.* 11, 481–487. doi:http://dx.doi.org/10.1097/01.MIB.0000159663.62651.4f.

Naftali, T., Reshef, L., Kovacs, A., Porat, R., Amir, I., Konikoff, F. M., et al. (2016). Distinct Microbiotas are Associated with Ileum-Restricted and Colon-Involving Crohn's Disease. *Inflamm. Bowel Dis.* 22, 293–302. doi:10.1097/MIB.0000000000000662.

Nagao-Kitamoto, H., Shreiner, A. B., Gillilland, M. G., Kitamoto, S., Ishii, C., Hirayama, A., et al. (2016). Functional Characterization of Inflammatory Bowel Disease–Associated Gut Dysbiosis in Gnotobiotic Mice. *C. Cell. Mol. Gastroenterol. Hepatol.*, 1–14. doi:10.1016/j.jcmgh.2016.02.003.

Nguyen, G. C., Chong, C. A., and Chong, R. Y. (2014). National estimates of the burden of inflammatory bowel disease among racial and ethnic groups in the United States. *J. Crohn's Colitis* 8, 288–295. doi:10.1016/j.crohns.2013.09.001.

Nichols, F. C., Housley, W. J., O'Conor, C. a, Manning, T., Wu, S., and Clark, R. B. (2009). Unique lipids from a common human bacterium represent a new class of Toll-like receptor 2 ligands capable of enhancing autoimmunity. *Am. J. Pathol.* 175, 2430–8. doi:10.2353/ajpath.2009.090544.

Nishimura, K., Sugiyama, D., Kogata, Y., Tsuji, G., and Nakazawa, T. (2007). Annals of Internal Medicine Review Meta-analysis: Diagnostic Accuracy of Anti-Cyclic Citrullinated Peptide Antibody and Rheumatoid Factor for Rheumatoid Arthritis. *Ann. Intern. Med.* 146, 797–808. doi:10.1016/S0084-3873(08)79116-8.

van Nood, E., Vrieze, A., Nieuwdorp, M., Fuentes, S., Zoetendal, E. G., de Vos, W. M., et al. (2013). Duodenal infusion of donor feces for recurrent Clostridium difficile. *N. Engl. J. Med.* 368, 407–15. doi:10.1056/NEJMoa1205037.

Norman, J. M., Handley, S. a, Parkes, M., Virgin, H. W., Norman, J. M., Handley, S. a, et al. (2015). Disease-Specific Alterations in the Enteric Virome in Article Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell* 160, 447–460. doi:10.1016/j.cell.2015.01.002.

O'Dell, J. R., Blakely, K. W., Mallek, J. A., Eckhoff, P. J., Leff, R. D., Wees, S. J., et al. (2001). Treatment of early seropositive rheumatoid arthritis: A two year, double blind comparison of minocycline and hydroxychloroquine. *Arthritis Rheum.* 44, 2235–2241. Available at: http://onlinelibrary.wiley.com/doi/10.1002/1529-0131(200110)44:10%3C2235::AID-

ART385%3E3.0.CO;2-A/full%5Cnhttp://onlinelibrary.wiley.com/doi/10.1002/1529-0131(200110)44:10%3C2235::AID-ART385%3E3.0.CO;2-A/full.
O'Dell, J. R., Haire, C. E., Palmer, W., Drymalski, W., Wees, S., Blakely, K., et al. (1997). Treatment of early rheumatoid arthritis with minocycline or placebo: results of a randomized, double-blind, placebo-controlled trial. *Arthritis Rheum.* 40, 842–848.
Ochoa-Repáraz, J., Mielcarz, D. W., Ditrio, L. E., Burroughs, A. R., Foureau, D. M., Haque-Begum, S., et al. (2009). Role of Gut Commensal Microflora in the Development of Experimental Autoimmune Encephalomyelitis. *J. Immunol.* 183, 6041–6050. doi:10.4049/jimmunol.0900747.
Ochoa-Repáraz, J., Mielcarz, D. W., Haque-Begum, S., and Kasper, L. H. (2010a). Induction of a regulatory B cell population in experimental allergic encephalomyelitis by alteration of the gut commensal microflora. *Gut Microbes* 1, 103–108. doi:10.4161/gmic.1.2.11515.
Ochoa-Repáraz, J., Mielcarz, D. W., Wang, Y., Begum-Haque, S., Dasgupta, S., Kasper, D. L., et al. (2010b). A polysaccharide from the human commensal Bacteroides fragilis protects against CNS demyelinating disease. *Mucosal Immunol.* 3, 487–495. doi:10.1038/mi.2010.29.
Ogrendik, M. (2007a). Effects of clarithromycin in patients with active rheumatoid arthritis. *Curr. Med. Res. Opin.* 23, 515–522. doi:10.1185/030079906X167642.
Ogrendik, M. (2007b). Levofloxacin treatment in patients with rheumatoid arthritis receiving methotrexate. *South. Med. J.* 100, 135–139.
Ogrendik, M. (2009). Efficacy of roxithromycin in adult patients with rheumatoid arthritis who had not received disease-modifying antirheumatic drugs: A 3-month, randomized, double-blind, placebo-controlled trial. *Clin. Ther.* 31, 1754–1764. doi:10.1016/j.clinthera.2009.08.014.
Ogrendik, M., and Karagoz, N. (2011). Treatment of rheumatoid arthritis with roxithromycin: A randomized trial. *Postgrad. Med.* 123, 220–227. doi:10.3810/pgm.2011.09.2478.
Ohkusa, T., Sato, N., Ogihara, T., Morita, K., Ogawa, M., and Okayasu, I. (2002). Fusobacterium varium localized in the colonic mucosa of patients with ulcerative colitis stimulates species-specific antibody. *J. Gastroenterol. Hepatol.* 17, 849–853. doi:10.1046/j.1440-1746.2002.02834.x.
Ott, S. J., Kuhbacher, T., Musfeldt, M., Rosenstiel, P., Hellmig, S., Rehman, A., et al. (2008). Fungi and inflammatory bowel diseases: Alterations of composition and diversity. *Scand J Gastroenterol* 43, 831–841. doi:10.1080/00365520801935434.
Palacios, N., Alonso, A., BrØnnum-Hansen, H., and Ascherio, A. (2011). Smoking and Increased Risk of Multiple Sclerosis: Parallel Trends in the Sex Ratio Reinforce the Evidence. *Ann. Epidemiol.* 21, 536–542. doi:10.1016/j.annepidem.2011.03.001.
Parkes, M., Cortes, A., van Heel, D. A., and Brown, M. A. (2013). Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet. M. al., 2013. Genet. insights into common pathways complex relationships among immune-mediated Dis. Nat. Rev. Genet. 14(9), pp.661–73. Available http//www.nature.com/nrg/journal/v14/n9/pdf/nrg3502* 14, 661–73. doi:10.1038/nrg3502.
Paula, A., Moreira, B., Fiche, T., Texeira, S., Ferreira, A. B., and Gerais, M. (2012). Influence of a high-fat diet on gut microbiota , intestinal permeability and metabolic endotoxaemia. *Br. J. Nutr.* 108, 801–809. doi:10.1017/S0007114512001213.
Pearson, C., Uhlig, H. H., and Powrie, F. (2012). Lymphoid microenvironments and innate lymphoid cells in the gut. *Trends Immunol.* 33, 289–296. doi:10.1016/j.it.2012.04.004.

Pei, A., Li, H., Oberdorf, W. E., Alekseyenko, A. V., Parsons, T., Yang, L., et al. (2012). Diversity of 5S rRNA genes within individual prokaryotic genomes. *FEMS Microbiol. Lett.* 335, 11–18. doi:10.1111/j.1574-6968.2012.02632.x.

Pérez-Cobas, A. E., Artacho, A., Ott, S. J., Moya, A., Gosalbes, M. J., and Latorre, A. (2014). Structural and functional changes in the gut microbiota associated to Clostridium difficile infection. *Front. Microbiol.* 5, 1–15. doi:10.3389/fmicb.2014.00335.

Petersen, A. M., Halkjær, S. I., and Gluud, L. L. (2015). Intestinal colonization with phylogenetic group B2 Escherichia coli related to inflammatory bowel disease: a systematic review and meta-analysis. *Scand. J. Gastroenterol.* 50, 1199–1207. doi:10.3109/00365521.2015.1028993.

Pinto, R. Del, Pietropaoli, D., Chandar, A. K., Ferri, C., and Cominelli, F. (2015). Association Between Inflammatory Bowel Disease and Vitamin D Deficiency: A Systematic Review and Meta-analysis. *Inflamm. Bowel Dis.* 21, 2708–2717. doi:10.1097/MIB.0000000000000546.

Pruesse, E., Quast, C., Knittel, K., Fuchs, B. M., Ludwig, W., Peplies, J., et al. (2007). SILVA: A comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35, 7188–7196. doi:10.1093/nar/gkm864.

Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing: Commentary. *Inflamm. Bowel Dis. Monit.* 11, 28. doi:10.1038/nature08821.

Qin, J., Li, Y., Cai, Z., Li, S. S., Zhu, J., Zhang, F., et al. (2012). A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 490, 55–60. doi:10.1038/nature11450.

Quera, R., Espinoza, R., Estay, C., and Rivera, D. (2014). Bacteremia as an adverse event of fecal microbiota transplantation in a patient with Crohn's disease and recurrent Clostridium difficile infection. *J. Crohns. Colitis* 8, 252–3. doi:10.1016/j.crohns.2013.10.002.

Rajilić-Stojanović, M., Biagi, E., Heilig, H. G. H. J., Kajander, K., Kekkonen, R. A., Tims, S., et al. (2011). Global and deep molecular analysis of microbiota signatures in fecal samples from patients with irritable bowel syndrome. *Gastroenterology* 141, 1792–1801. doi:10.1053/j.gastro.2011.07.043.

Rath, H. C., Herfarth, H. H., Ikeda, J. S., Grenther, W. B., Hamm, T. E., Balish, E., et al. (1996). Normal luminal bacteria, especially bacteroides species, mediate chronic colitis, gastritis, and arthritis in HLA-B27/human B2 microglobulin transgenic rats. *J. Clin. Invest.* 98, 945–953. doi:10.1172/JCI118878.

Reunanen, J., Kainulainen, V., Huuskonen, L., Ottman, N., Belzer, C., Huhtinen, H., et al. (2015). Akkermansia muciniphila Adheres to Enterocytes and Strengthens the Integrity of the Epithelial Cell Layer. *Appl. Environ. Microbiol.* 81, 3655–62. doi:10.1128/AEM.04050-14.

Rezende, R. M., Oliveira, R. P., Medeiros, S. R., Gomes-Santos, A. C., Alves, A. C., Loli, F. G., et al. (2013). Hsp65-producing Lactococcus lactis prevents experimental autoimmune encephalomyelitis in mice by inducing CD4+LAP+ regulatory T cells. *J. Autoimmun.* 40, 45–57. doi:10.1016/j.jaut.2012.07.012.

Rhee, S. H., Pothoulakis, C., and Mayer, E. A. (2009). Principles and clinical implications of the brain-gut-enteric microbiota axis. *Nat. Rev. Gastroenterol. Hepatol.* 6, 306–14. doi:10.1038/nrgastro.2009.35.

Rivière, A., Selak, M., Lantin, D., Leroy, F., and De Vuyst, L. (2016). Bifidobacteria and

butyrate-producing colon bacteria: Importance and strategies for their stimulation in the human gut. *Front. Microbiol.* 7. doi:10.3389/fmicb.2016.00979.
Rodriguez, M., Wootla, B., and Anderson, G. (2016). Multiple Sclerosis, Gut Microbiota and Permeability: Role of Tryptophan Catabo- lites , Depression and the Driving Down of Local Melatonin. *Curr. Pharm. Des.* 22, 1–8. doi:10.2174/1381612822666160915.
Roduit, C., Frei, R., and Mutius, E. Von (2016). The Hygiene Hypothesis. *Environ. Influ. Immune Syst.*, 77–96. doi:10.1007/978-3-7091-1890-0.
Rooks, M. G., Veiga, P., Wardwell-scott, L. H., Tickle, T., Segata, N., Michaud, M., et al. (2014). Gut microbiome composition and function in experimental colitis during active disease and treatment-induced remission. *ISME J.* 8, 1403–17. doi:10.1038/ismej.2014.3.
Rosen, G., Garbarine, E., Caseiro, D., Polikar, R., and Sokhansanj, B. (2008). Metagenome fragment classification using N-mer frequency profiles. *Adv. Bioinformatics*, 205969. doi:10.1155/2008/205969.
Rossen, N. G., MacDonald, J. K., De Vries, E. M., D'Haens, G. R., De Vos, W. M., Zoetendal, E. G., et al. (2015). Fecal microbiota transplantation as novel therapy in gastroenterology: A systematic review. *World J. Gastroenterol.* 21, 5359–5371. doi:10.3748/wjg.v21.i17.5359.
Roth, E. B., Stenberg, P., Book, C., and Sjöberg, K. (2006). Antibodies against transglutaminases, peptidylarginine deiminase and citrulline in rheumatoid arthritis - New pathways to epitope spreading. *Clin. Exp. Rheumatol.* 24, 12–18.
Rothberg, J. M., Hinz, W., Rearick, T. M., Schultz, J., Mileski, W., Davey, M., et al. (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475, 348–352. doi:10.1038/nature10242.
Russell, R. K., Ip, B., Aldhous, M. C., MacDougall, M., Drummond, H. E., Arnott, I. D. R., et al. (2009). Anti-Saccharomyces cerevisiae antibodies status is associated with oral involvement and disease severity in Crohn disease. *J. Pediatr. Gastroenterol. Nutr.* 48, 161–7. doi:10.1097/MPG.0b013e318183e112.
Salameh, A., Klotz, S. A., and Zangeneh, T. T. (2012). Disseminated Infection Caused by Eggerthella lenta in a Previously Healthy Young Man: A Case Report. *Case Rep. Infect. Dis.* 2012, 517637. doi:10.1155/2012/517637.
Sanger, F., Nicklen, S., and Coulson, A. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5463–7. doi:10.1073/pnas.74.12.5463.
Sasdelli, A. S., Barbanti, F. A., and Marchesini, G. (2016). *Gut–Brain Axis: A New Revolution to Understand the Pathogenesis of Autism and Other Severe Neurological Diseases*. doi:10.1007/978-3-319-30361-1_8.
Saviola, G., Abdi-Ali, L., Campostrini, L., Sacco, S., Baiardi, P., Manfredi, M., et al. (2013). Clarithromycin in rheumatoid arthritis: The addition to methotrexate and low-dose methylprednisolone induces a significant additive value - A 24-month single-blind pilot study. *Rheumatol. Int.* 33, 2833–2838. doi:10.1007/s00296-013-2822-0.
Saviola, G., Ali, L. A., Rossini, P., Campostrini, L., Coppini, A., Gori, M., et al. (2002). Clarithromycin in rheumatoid arthritis patients not responsive to disease-modifying antirheumatic drugs : An open, uncontrolled pilot study. *Clin Exp Rheumatol* 20, 373–378.
Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., et al. (2009). Introducing mothur: Open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537–7541. doi:10.1128/AEM.01541-09.
Schoepfer, A. M., Schaffer, T., Seibold-Schmid, B., Müller, S., and Seibold, F. (2008).

Antibodies to flagellin indicate reactivity to bacterial antigens in IBS patients. *Neurogastroenterol. Motil.* 20, 1110–1118. doi:10.1111/j.1365-2982.2008.01166.x.
Seeley, R., Stephens, T., and Tate, P. (2007). *Essentials of Anatomy and Physiology*. Sixth. Boston MA: McGraw Hill.
Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12, R60. doi:10.1186/gb-2011-12-6-r60.
Sekirov, I., Russell, S., and Antunes, L. (2010). Gut microbiota in health and disease. *Physiol. Rev.* 90, 859–904. doi:10.1152/physrev.00045.2009.
Seksik, P., Lepage, P., De La Cochetière, M. F., Bourreille, A., Sutren, M., Galmiche, J. P., et al. (2005). Search for localized dysbiosis in Crohn's disease ulcerations by temporal temperature gradient gel electrophoresis of 16S rRNA. *J. Clin. Microbiol.* 43, 4654–4658. doi:10.1128/JCM.43.9.4654-4658.2005.
Selma, M. V., Tomás-Barberán, F. A., Beltrán, D., García-Villalba, R., and Espín, J. C. (2014). Gordonibacter urolithinfaciens sp. nov., a urolithin-producing bacterium isolated from the human gut. *Int. J. Syst. Evol. Microbiol.* 64, 2346–2352. doi:10.1099/ijs.0.055095-0.
Sepehri, S., Kotlowski, R., Bernstein, C. N., and Krause, D. O. (2007). Microbial diversity of inflamed and noninflamed gut biopsy tissues in inflammatory bowel disease. *Inflamm. Bowel Dis.* 13, 675–683. doi:10.1002/ibd.20101.
Shahir, N., Carroll, I., Frank, D., Newberry, R., Sartor, R., Herfarth, H., et al. (2016). A Distinct Microbiome Signature Characterizes Patients with Penetrating and Fistulizing Crohn's Disease. *Inflamm. Bowel Dis.*, P193.
Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell Syst. Tech. J.* 27, 623–656. doi:10.1002/j.1538-7305.1948.tb01338.x.
Sharpton, T. J. (2014). An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* 5. doi:10.3389/fpls.2014.00209.
Shen, J., Zuo, Z. X., and Mao, A. P. (2014). Effect of probiotics on inducing remission and maintaining therapy in ulcerative colitis, Crohn's disease, and pouchitis: meta-analysis of randomized controlled trials. *Inflamm Bowel Dis* 20, 21–35. doi:10.1097/01.MIB.0000437495.30052.be.
Sher, J., Sczesnak, A., Longman, R., Segata, N., Ubeda, C., Bielski, C., et al. (2013). Expansion of intestinal Prevotella copri correlates with enhanced susceptibility to arthritis. *Elife* 2, e01202. doi:http://dx.doi.org/10.7554/eLife.01202.
Shi, J., van Veelen, P. A., Mahler, M., Janssen, G. M. C., Drijfhout, J. W., Huizinga, T. W. J., et al. (2014). Carbamylation and antibodies against carbamylated proteins in autoimmunity and other pathologies. *Autoimmun. Rev.* 13, 225–230. doi:10.1016/j.autrev.2013.10.008.
Shreiner, A. B., Kao, J. Y., and Young, V. B. (2015). The gut microbiome in health and disease. *Curr. Opin. Gastroenterol.* 31, 69–75. doi:10.1053/j.gastro.2014.03.032.
Simpson, E. (1949). Measurement of diversity. *Nature* 163, 688.
Smith, P. M., Howitt, M. R., Panikov, N., Michaud, M., Gallini, C. A., Bohlooly-y, M., et al. (2013). The Microbial Metabolites, Short-Chain Fatty Acids, Regulate Colonic T Cell Homeostasis. *Science (80-.).* 341, 569–573.
Sokol, H., Leducq, V., Aschard, H., Pham, H., Jegou, S., Landman, C., et al. (2016). Fungal microbiota dysbiosis in IBD. *Gut* 0, 1–10. doi:10.1136/gutjnl-2015-310746.
Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermúdez-Humarán, L. G., Gratadoux, J.-J., et al. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium

identified by gut microbiota analysis of Crohn disease patients. *Proc. Natl. Acad. Sci. U. S. A.* 105, 16731–16736. doi:10.1073/pnas.0804812105.

Sokolove, J., Zhao, X., Chandra, P. E., and Robinson, W. H. (2011). Immune complexes containing citrullinated fibrinogen costimulate macrophages via toll-like receptor 4 and Fc?? receptor. *Arthritis Rheum.* 63, 53–62. doi:10.1002/art.30081.

Somers, E. C., Thomas, S. L., Smeeth, L., and Hall, A. J. (2006). Autoimmune Diseases Co-occurring Within Individuals and Within Families. *Epidemiology* 17, 202–217. doi:10.1097/01.ede.0000193605.93416.df.

Sonnenberg, A., Turner, K. O., and Genta, R. M. (2016). Differences in the socio-economic distribution of inflammatory bowel disease and microscopic colitis. *Color. Dis.*, Accepted article. doi:10.1111/codi.13378.

Soon, I. S., Molodecky, N. A., Rabi, D. M., Ghali, W. A., Barkema, H. W., and Kaplan, G. G. (2012). The relationship between urban environment and the inflammatory bowel diseases: a systematic review and meta-analysis. *BMC Gastroenterol.* 12.

Sørensen, T. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. *Biol. Skr* 5, 1–34.

Sorrentino, R. (2014). Genetics of autoimmunity: An update. *Immunol. Lett.* 158, 116–119. doi:10.1016/j.imlet.2013.12.005.

Stackebrandt, E., and Goebel, B. (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *Int. J. Syst. Bacteriol.* 44, 846–849. doi:10.1099/00207713-44-4-846.

Strati, F., Paola, M. Di, Stefanini, I., Albanese, D., Rizzetto, L., Lionetti, P., et al. (2016). Age and Gender Affect the Composition of Fungal Population of the Human Gastrointestinal Tract. *Front. Microbiol.* 7, 1227. doi:10.3389/fmicb.2016.01227.

Strauss, J., Kaplan, G. G., Beck, P. L., Rioux, K., Panaccione, R., Devinney, R., et al. (2011). Invasive potential of gut mucosa-derived fusobacterium nucleatum positively correlates with IBD status of the host. *Inflamm. Bowel Dis.* 17, 1971–1978. doi:10.1002/ibd.21606.

Su, H. Y., Gupta, V., Day, A. S., and Gearry, R. B. (2016). Rising Incidence of Inflammatory Bowel Disease in Canterbury, New Zealand. *Inflamm. Bowel Dis.* 22, 2238–2244. doi:10.1097/MIB.0000000000000829.

Svartz, N. (1948). The treatment of rheumatic polyarthritis with acid azo compounds. *Rheumatism* 4, 180–185.

Szekanecz, Z., Soós, L., Szabó, Z., Fekete, A., Kapitány, A., Végvári, A., et al. (2008). Anti-citrullinated protein antibodies in rheumatoid arthritis: As good as it gets? *Clin. Rev. Allergy Immunol.* 34, 26–31. doi:10.1007/s12016-007-8022-5.

Szilagyi, A., and Xue, X. (2016). Geographic associations between lactase phenotype , multiple sclerosis , and inflammatory bowel diseases ; Does obesity trump geography ? *Med. Hypotheses* 96, 68–72. doi:10.1016/j.mehy.2016.10.001.

Tan, I. J., Peeva, E., and Zandman-Goddard, G. (2014). Hormonal modulation of the immune system - A spotlight on the role of progestogens. *Autoimmun. Rev.* 14, 536–542. doi:10.1016/j.autrev.2015.02.004.

Tang, M. S., Poles, J., Leung, J. M., Wolff, M. J., Davenport, M., Lee, S. C., et al. (2015). Inferred metagenomic comparison of mucosal and fecal microbiota from individuals undergoing routine screening colonoscopy reveals similar differences observed during active inflammation. *Gut Microbes* 6, 48–56.

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.-P., et al. (2009). Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* 11, 2574–2584.
Taurog, J. D., Richardson, J. A., Croft, J. T., Simmons, W. A., Zhou, M., Fernández-Sueiro, J. L., et al. (1994). The germfree state prevents development of gut and joint inflammatory disease in HLA-B27 transgenic rats. *J. Exp. Med.* 180, 2359–64. doi:10.1084/jem.180.6.2359.
Taylor, D. E., Ge, Z., and Purych, D. (1997). Cloning and sequence analysis of the two copies of 23S rRNA genes from Helicobacter pylori and association of clarithromycin resistance with 23 rRNA mutations. *Antimicrob Agents Chemother* 41, 2621–2628.
Terato, K., Do, C. T., and Shionoya, H. (2015). Slipping through the Cracks : Linking Low Immune Function and Intestinal Bacterial Imbalance to the Etiology of Rheumatoid Arthritis. *Autoimmune Dis.* 2015. doi:http://dx.doi.org/10.1155/2015/636207.
Thota, V. R., Dacha, S., Natarajan, A., and Nerad, J. (2011). Eggerthella lenta bacteremia in a Crohn's disease patient after ileocecal resection. *Future Microbiol.* 6, 595–597. doi:10.2217/fmb.11.31.
Tilley, B. C., Alarcon, G. S., Heyse, S. P., Trentham, D. E., Neuner, R., Kaplan, D. A., et al. (1995). Minocycline in Rheumatoid Arthritis: A 48-Week, Double-Blind, Placebo-Controlled Trial. *Ann. Intern. Med.* 122, 81–89. doi:doi:10.7326/0003-4819-122-2-199501150-00001.
Tobon, G. J., Youinou, P., and Saraux, A. (2010). The environment, geo-epidemiology, and autoimmune disease: Rheumatoid arthritis. *J. Autoimmun.* 35, 10–14. doi:10.1016/j.jaut.2009.12.009.
Torchinsky, M. B., and Blander, J. M. (2010). T helper 17 cells: Discovery, function, and physiological trigger. *Cell. Mol. Life Sci.* 67, 1407–1421. doi:10.1007/s00018-009-0248-3.
Tortora, G., and Derrickson, B. (2009). *Principles of Anatomy and Physiology. Volume 2. Maintenance and Continuity of the Human Body*. 12th ed. Hoboken NJ: John Wiley and Sons.
Tortora, G., and Derrickson, B. (2013). *Essentials of Anatomy and Physiology*. Ninth. Hoboken NJ: International Student Version. John Wiley and Sons.
Tremlett, H., Fadrosh, D., Farqui, A., Hart, J., Roalstad, S., Graves, J., et al. (2016a). Associations Between Immune Markers and Gut Microbiota in Pediatric Multiple Sclerosis and Controls (S29.008). *Neurology* 86, S29.008.
Tremlett, H., Fadrosh, D., Lynch, S., Hart, K., Graves, J., Lulu, S., et al. (2015). Gut microbiome in early pediatric multiple sclerosis: a case-control study (P4.027). *Neurology* 84. Available at: http://www.neurology.org/content/84/14_Supplement/P4.027.abstract.
Tremlett, H., Fadrosh, D. W., Faruqi, A. A., Hart, J., Roalstad, S., Graves, J., et al. (2016b). Gut microbiota composition and relapse risk in pediatric MS: A pilot study. *J. Neurol. Sci.* 363, 153–157. doi:10.1016/j.jns.2016.02.042.
Tremlett, H., Fadrosh, D. W., Faruqi, A. A., Zhu, F., Hart, J., Roalstad, S., et al. (2016c). Gut microbiota in early pediatric multiple sclerosis: a case control study. *Eur. J. Neurol.* 23, 1308–1321. doi:10.1111/ene.13026.
Trouw, L. A., Haisma, E. M., Levarht, E. W. N., Van Der Woude, D., Ioan-Facsinay, A., Daha, M. R., et al. (2009). Anti-cyclic citrullinated peptide antibodies from rheumatoid arthritis patients activate complement via both the classical and alternative pathways. *Arthritis Rheum.* 60, 1923–1931. doi:10.1002/art.24622.
Tsigos, C., and Chrousos, G. P. (2002). Hypothalamic-pituitary-adrenal axis, neuroendocrine

factors and stress. *J. Psychosom. Res.* 53, 865–871. doi:10.1016/S0022-3999(02)00429-4.
Vaahtovuo, J., Munukka, E., Korkeamaki, M., Luukkainen, R., and Toivanen, P. (2008). Fecal microbiota in early rheumatoid arthritis. *J. Rheumatol.* 35, 1500–1505.
Vaghef-Mehrabany, E., Alipour, B., Homayouni-Rad, A., Sharif, S.-K. K., Asghari-Jafarabadi, M., and Zavvari, S. (2014). Probiotic supplementation improves inflammatory status in patients with rheumatoid arthritis. *Nutrition* 30, 430–435. doi:10.1016/j.nut.2013.09.007.
Vanaclocha, F., Crespo-Erchiga, V., Jiménez-Puya, R., Puig, L., Sánchez-Carazo, J. L., Ferrán, M., et al. (2015). Immune-mediated inflammatory diseases and other comorbidities in patients with psoriasis: baseline characteristics of patients in the AQUILES study. *Actas Dermosifiliogr.* 106, 35–43. doi:10.1016/j.ad.2014.06.003.
Vanuytsel, T., Wanrooy, S. Van, Vanheel, H., Vanormelingen, C., Houben, E., Rasoel, S. S., et al. (2014). Psychological stress and corticotropin-releasing hormone increase intestinal permeability in humans by a mast cell-dependent mechanism. *Neurogastroenterology* 63, 1293–1299. doi:10.1136/gutjnl-2013-305690.
Vasquez, N., Mangin, I., Lepage, P., Seksik, P., Duong, J. P., Blum, S., et al. (2007). Patchy distribution of mucosal lesions in ileal Crohn's disease is not linked to differences in the dominant mucosa-associated bacteria: A study using fluorescence in situ hybridization and temporal temperature gradient gel electrophoresis. *Inflamm. Bowel Dis.* 13, 684–692. doi:10.1002/ibd.20084.
Vavassori, P., Mencarelli, a., Renga, B., Distrutti, E., and Fiorucci, S. (2009). The Bile Acid Receptor FXR Is a Modulator of Intestinal Innate Immunity. *J. Immunol.* 183, 6251–6261. doi:10.4049/jimmunol.0803978.
Vital, M., Penton, C. R., Wang, Q., Young, V. B., Antonopoulos, D. a, Sogin, M. L., et al. (2013). A gene-targeted approach to investigate the intestinal butyrate-producing bacterial community. *Microbiome* 1, 8. doi:10.1186/2049-2618-1-8.
Volokhov, D. V, Simonyan, V., Davidson, M. K., and Chizhikov, V. E. (2012). RNA polymerase beta subunit (rpoB) gene and the 16S-23S rRNA intergenic transcribed spacer region (ITS) as complementary molecular markers in addition to the 16S rRNA gene for phylogenetic analysis and identification of the species of the family Mycoplasm. *Mol. Phylogenet. Evol.* 62, 515–28. doi:10.1016/j.ympev.2011.11.002.
Wagner, J., Maksimovic, J., Farries, G., Sim, W. H., Bishop, R. F., Cameron, D. J., et al. (2013). Bacteriophages in Gut Samples From Pediatric Crohn's Disease Patients. *Inflamm. Bowel Dis.* 19, 1598–1608. doi:10.1097/MIB.0b013e318292477c.
Wallin, M., Culpepper, W., and Kurtzke, J. (2015). Risk Factors for Disease Progression in a Diverse US Population-based Multiple Sclerosis Cohort. *Neurology* 84, S45.003.
Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi:10.1128/AEM.00062-07.
Wang, Y., and Kasper, L. H. (2014). The role of microbiome in central nervous system disorders. *Brain. Behav. Immun.* 38, 1–12. doi:10.1016/j.bbi.2013.12.015.
Waugh, A., and Grant, A. (2010). *Ross and Wilson Anatomy and Physiology in Health and Ilness*. 11th ed. Edinburgh: Churchill Livingstone Elsevier.
Weinstock, G. M. (2012). Genomic approaches to studying the human microbiota. *Nature* 489, 250–256. doi:10.1038/nature11553.
Weng, X., Liu, L., Barcellos, L. F., Allison, J. E., and Herrinton, L. J. (2007). Clustering of inflammatory bowel disease with immune mediated diseases among members of a northern

California-managed care organization. *Am. J. Gastroenterol.* 102, 1429–1435. doi:10.1111/j.1572-0241.2007.01215.x.
Wexler, H. M. (2007). Bacteroides: The good, the bad, and the nitty-gritty. *Clin. Microbiol. Rev.* 20, 593–621. doi:10.1128/CMR.00008-07.
Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* 30, 280–338.
Wilde, K. De, Debusschere, K., Beeckman, S., Jacques, P., and Elewaut, D. (2015). Integrating the pathogenesis of spondyloarthritis : gut and joint united ? *Curr. Opin. Rheumatol.* 27, 189–196. doi:10.1097/BOR.0000000000000144.
Woese, C. R., and Fox, G. E. (1977). Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc. Natl. Acad. Sci. U. S. A.* 74, 5088–5090. doi:10.1073/pnas.74.11.5088.
Wu, H. J., Ivanov, I. I., Darce, J., Hattori, K., Shima, T., Umesaki, Y., et al. (2010). Gut-residing segmented filamentous bacteria drive autoimmune arthritis via T helper 17 cells. *Immunity* 32, 815–827. doi:10.1016/j.immuni.2010.06.001.
Wu, W., Chen, F., Liu, Z., and Cong, Y. (2016). Microbiota-specific Th17 Cells. *Inflamm. Bowel Dis.* 22, 1473–1482. doi:10.1097/MIB.0000000000000775.
Wu, X., Chen, H., and Xu, H. (2015). The genomic landscape of human immune-mediated diseases. *J. Hum. Genet.* 60, 675–681. doi:10.1038/jhg.2015.99.
Zeng, Q. Y., Chen, R., Darmawan, J., Xiao, Z. Y., Chen, S. B., Wigley, R., et al. (2008). Rheumatic diseases in China. *Arthritis Res. Ther.* 10, R17. doi:10.1186/ar2368.
Zhang, B. J., Markovic-plese, S., Lacet, B., Raus, J., Weiner, H. L., and Hailer, D. a (1994). Increased Frequency of Interleukin 2-responsive T Cells Specific for Myelin Basic Protein and Proteolipid Protein in Peripheral Blood and Cerebrospinal Fluid of Patients with Multiple Sclerosis. *J. Exp. Med.* 179, 973–984. doi:doi: 10.1084/jem.179.3.973.
Zhang, X., Zhang, D., Jia, H., Feng, Q., Wang, D., Liang, D., et al. (2015). The oral and gut microbiomes are perturbed in rheumatoid arthritis and partly normalized after treatment. *Nat. Med.* 21, 895–905. doi:10.1038/nm.3914.
Zhu, J., and Paul, W. E. (2008). CD4 T cells: fates, functions, and faults. *Blood* 112, 1557–1569. doi:doi:10.1182/blood-2008-05-078154.
Zitomersky, N. L., Atkinson, B. J., Franklin, S. W., Mitchell, P. D., Snapper, S. B., Comstock, L. E., et al. (2013). Characterization of Adherent Bacteroidales from Intestinal Biopsies of Children and Young Adults with Inflammatory Bowel Disease. *PLoS One* 8, e63686. doi:10.1371/journal.pone.0063686.

APPENDICES

APPENDIX I

Rheumatoid arthritis patient meta-data

disease	patient	sex	age	hands	feet	shoulders	elbows	wrists	knees	hips	misc	eyes	lungs	cancer	antibiotics
RA	1003	F	61	yes	no	yes	yes	no	yes	no	no	no	no	no	no
RA	1021	F	59	yes	yes	yes	yes	no	yes	yes	yes	yes	no	no	yes
RA	1024	F	60	yes	yes	yes	no	yes	yes	no	yes	no	yes	no	no
RA	1025	F	55	yes	yes	yes	yes	yes	yes	no	yes	yes	no	no	no
RA	1027	F	49	yes	yes	yes	yes	yes	yes	no	yes	no	no	no	no
RA	1028	F	58	yes	yes	no	no	yes	yes	yes	yes	no	no	no	yes
RA	1071	M	71	yes	yes	yes	no	no	no	yes	no	no	no	no	no
RA	1075	F	76	yes	yes	yes	no	no	yes	yes	yes	no	no	no	no
RA	1091	M	61	yes	yes	yes	yes	yes	yes	no	yes	no	no	no	no
RA	1113	F	65	yes	yes	no	no	no	yes	no	no	no	no	no	no
RA	1128	F	69	yes	yes	yes	no	no	yes	yes	no	no	no	no	yes
RA	1138	F	73	yes	yes	no	no	no	no	no	no	no	no	no	no
RA	1140	M	63	yes	yes	no	no	no	yes	yes	yes	yes	no	no	no
RA	1148	M	69	yes	yes	yes	no	yes	yes	no	yes	no	no	no	no
RA	1160	F	38	yes	yes	yes	yes	yes	yes	yes	yes	no	no	no	no
RA	1173	F	57	yes	yes	no	yes	yes	yes	no	yes	no	no	no	no
RA	1198	M	77	yes	yes	no	no	yes	no	no	no	no	no	no	yes
RA	1204	F	69	yes	yes	yes	yes	yes	yes	yes	yes	yes	no	no	no
RA	1235	M	58	yes	yes	yes	yes	no	yes	no	no	yes	yes	no	no
RA	1257	M	63	yes	no	yes	no	yes	yes	yes	yes	no	no	yes	no
RA	1331	F	57	yes	no	no	no	yes	no	yes	no	no	no	no	no

APPENDIX II

Stool microbiota species count data

Table includes only species with an average abundance > 0.10.

Species	CD	UC	MS	RA	HC
Blautia_luti	0.483	0.648	0.675	0.726	1.029
Eubacterium_hallii	0.445	0.612	0.693	0.830	0.844
Faecalibacterium_prausnitzii	0.332	0.767	0.427	0.534	1.133
Collinsella_aerofaciens	0.428	0.474	0.596	0.593	0.853
Anaerostipes_hadrus	0.510	0.573	0.544	0.737	0.505
Ruminococcus_obeum	0.389	0.543	0.557	0.479	0.748
Blautia_faecis	0.285	0.527	0.314	0.422	0.507
Eubacterium_rectale	0.487	0.474	0.239	0.308	0.401
Ruminococcus_torques	0.263	0.319	0.495	0.298	0.428
Clostridium_ruminantium	0.103	0.379	0.253	0.245	0.505
Blautia_glucerasea	0.379	0.130	0.338	0.283	0.166
Dorea_longicatena	0.208	0.257	0.202	0.245	0.340
Clostridium_hylemonae	0.355	0.158	0.256	0.244	0.185
Ruminococcus_bromii	0.083	0.175	0.319	0.244	0.308
Gemmiger_formicilis	0.116	0.138	0.201	0.185	0.404
Blautia_schinkii	0.122	0.148	0.272	0.261	0.238
Akkermansia_muciniphila	0.169	0.122	0.195	0.235	0.206
Clostridium_sporosphaeroides	0.100	0.190	0.196	0.204	0.208
Coprococcus_comes	0.118	0.191	0.187	0.185	0.208
Clostridium_saccharogumia	0.108	0.189	0.190	0.128	0.238
Ruminococcus_lactaris	0.110	0.103	0.136	0.161	0.190
Clostridium_caenicola	0.042	0.077	0.191	0.198	0.159
Dorea_formicigenerans	0.077	0.140	0.119	0.138	0.184
Clostridium_disporicum	0.057	0.119	0.190	0.145	0.138
Oscillibacter_valericigenes	0.057	0.092	0.108	0.155	0.205
Clostridium_bartlettii	0.139	0.134	0.113	0.100	0.113
Eubacterium_coprostanoligenes	0.070	0.104	0.099	0.092	0.175
Clostridium_methylpentosum	0.050	0.088	0.104	0.124	0.154
Eubacterium_ramulus	0.060	0.102	0.105	0.107	0.135
Denitrobacterium_detoxificans	0.041	0.079	0.105	0.110	0.157
Ruminococcus_gnavus	0.217	0.055	0.057	0.088	0.052
Catenibacterium_mitsuokai	0.037	0.055	0.041	0.099	0.228
Turicibacter_sanguinis	0.032	0.110	0.118	0.094	0.096
Clostridium_leptum	0.057	0.093	0.086	0.094	0.116
Eggerthella_lenta	0.075	0.080	0.103	0.139	0.048
Subdoligranulum_variabile	0.064	0.100	0.069	0.088	0.123
Intestinimonas_butyriciproducens	0.057	0.057	0.087	0.108	0.129
Eubacterium_biforme	0.043	0.074	0.084	0.085	0.145
Eubacterium_desmolans	0.066	0.091	0.052	0.081	0.137
Clostridium_glycolicum	0.065	0.128	0.065	0.064	0.071

Roseburia_faecis	0.049	0.071	0.061	0.073	0.128
Papillibacter_cinnamivorans	0.025	0.057	0.072	0.096	0.125
Clostridium_thermocellum	0.038	0.089	0.065	0.076	0.104
Clostridium_glycyrrhizinilyticum	0.052	0.073	0.071	0.080	0.075
Gordonibacter_pamelaeae	0.031	0.059	0.069	0.087	0.103
Coprococcus_catus	0.038	0.061	0.083	0.078	0.087
Bacteroides_vulgatus	0.042	0.054	0.032	0.135	0.074
Bifidobacterium_bifidum	0.026	0.112	0.065	0.065	0.065
Olsenella_profusa	0.034	0.030	0.061	0.097	0.086
Sporobacter_termitidis	0.021	0.041	0.046	0.057	0.130
Coprococcus_eutactus	0.048	0.045	0.038	0.038	0.124
Lactonifactor_longoviformis	0.069	0.042	0.068	0.058	0.033
Clostridium_aminobutyricum	0.031	0.042	0.058	0.058	0.079
Flavonifractor_plautii	0.032	0.046	0.053	0.072	0.064
Coprobacillus_cateniformis	0.029	0.068	0.061	0.064	0.040
Eubacterium_eligens	0.023	0.044	0.040	0.070	0.084
Clostridium_ramosum	0.068	0.036	0.058	0.070	0.026
Eubacterium_cylindroides	0.025	0.045	0.069	0.089	0.028
Anaerosporobacter_mobilis	0.028	0.046	0.053	0.060	0.058
Bacteroides_uniformis	0.036	0.038	0.032	0.032	0.096
Clostridium_citroniae	0.049	0.034	0.037	0.089	0.023
Collinsella_tanakaei	0.026	0.057	0.046	0.061	0.038
Blautia_stercoris	0.040	0.047	0.034	0.040	0.066
Clostridium_innocuum	0.045	0.051	0.049	0.052	0.029
Clostridium_hathewayi	0.032	0.026	0.045	0.044	0.071
Eubacterium_ventriosum	0.030	0.037	0.044	0.050	0.057
Dehalobacterium_formicoaceticum	0.015	0.021	0.054	0.064	0.057
Clostridium_spiroforme	0.029	0.038	0.046	0.053	0.039
Slackia_isoflavoniconvertens	0.039	0.026	0.053	0.024	0.055
Butyrivibrio_crossotus	0.025	0.032	0.034	0.046	0.058
Ruminococcus_albus	0.018	0.034	0.037	0.045	0.060
Actinomyces_odontolyticus	0.036	0.036	0.039	0.048	0.034
Cohnella_laeviribosi	0.017	0.040	0.040	0.029	0.066
Lachnobacterium_bovis	0.021	0.034	0.038	0.037	0.059
Blautia_hydrogenotrophica	0.026	0.038	0.055	0.041	0.027
Bacteroides_cellulosolvens	0.015	0.035	0.037	0.037	0.061
Eubacterium_siraeum	0.030	0.022	0.042	0.032	0.057
Peptococcus_niger	0.026	0.023	0.065	0.030	0.034
Roseburia_intestinalis	0.032	0.036	0.030	0.030	0.047
Dialister_invisus	0.037	0.071	0.013	0.019	0.036
Clostridium_oroticum	0.031	0.033	0.044	0.026	0.038

Streptococcus_mutans	0.028	0.027	0.031	0.052	0.029
Parvimonas_micra	0.030	0.033	0.036	0.021	0.042
Eubacterium_contortum	0.015	0.029	0.044	0.044	0.030
Rothia_mucilaginosa	0.027	0.028	0.031	0.044	0.030
Oscillospira_guilliermondii	0.009	0.032	0.028	0.024	0.065
Bacteroides_caccae	0.011	0.025	0.064	0.019	0.035
Lutispora_thermophila	0.013	0.011	0.039	0.051	0.040
Clostridium_lactatifermentans	0.014	0.022	0.029	0.036	0.049
Streptococcus_thermophilus	0.028	0.030	0.024	0.050	0.014
Roseburia_inulinivorans	0.019	0.037	0.023	0.020	0.046
Clostridium_cellulosi	0.009	0.029	0.026	0.033	0.045
Roseburia_hominis	0.017	0.030	0.029	0.025	0.041
Anaerofustis_stercorihominis	0.017	0.024	0.030	0.036	0.033
Clostridium_scindens	0.034	0.023	0.029	0.033	0.020
Bacteroides_xylanolyticus	0.019	0.023	0.029	0.022	0.041
Actinomyces_lingnae	0.029	0.024	0.024	0.031	0.026
Clostridium_viride	0.017	0.032	0.018	0.029	0.037
Anaerotruncus_colihominis	0.013	0.028	0.028	0.033	0.030
Clostridium_nexile	0.047	0.011	0.037	0.022	0.011
Prevotella_copri	0.016	0.015	0.032	0.014	0.050
Lactobacillus_ruminis	0.007	0.046	0.007	0.007	0.060
Bifidobacterium_dentium	0.019	0.030	0.018	0.042	0.016
Bacteroides_ovatus	0.028	0.018	0.018	0.019	0.040
Barnesiella_intestinihominis	0.015	0.012	0.022	0.028	0.045
Brassicibacter_mesophilus	0.014	0.018	0.021	0.024	0.044
Escherichia/Shigella_coli	0.050	0.014	0.010	0.026	0.022
Ruminococcus_flavefaciens	0.007	0.017	0.021	0.024	0.051
Streptococcus_sanguinis	0.030	0.019	0.027	0.022	0.019
Oxobacter_pfennigii	0.007	0.022	0.024	0.033	0.033
Alkalibacter_saccharofermentans	0.006	0.014	0.040	0.029	0.028
Clostridium_populeti	0.013	0.015	0.017	0.030	0.041
Slackia_piriformis	0.016	0.017	0.015	0.025	0.042
Parabacteroides_distasonis	0.014	0.017	0.022	0.025	0.037
Defluviitalea_saccharophila	0.013	0.015	0.025	0.027	0.031
Clostridium_purinilyticum	0.015	0.021	0.024	0.029	0.022
Alistipes_putredinis	0.014	0.018	0.017	0.024	0.038
Ruminococcus_champanellensis	0.010	0.010	0.031	0.027	0.030
Caloramator_fervidus	0.005	0.019	0.028	0.025	0.030
Holdemania_filiformis	0.018	0.026	0.015	0.015	0.029
Pelotomaculum_terephthalicicum	0.006	0.011	0.037	0.031	0.018
Thermotalea_metallivorans	0.011	0.020	0.020	0.028	0.024

Melghirimyces_thermohalophilus	0.010	0.013	0.032	0.029	0.018
Solobacterium_moorei	0.018	0.022	0.018	0.022	0.021
Bacteroides_dorei	0.023	0.013	0.017	0.022	0.026
Christensenella_minuta	0.005	0.014	0.026	0.025	0.032

APPENDIX III

KEGG pathways identified in immune-mediated inflammatory diseases

KEGG Pathway	CD[a]	UC	MS	RA	HC	*P*-value[b]	*P*-corr[c]
1,1,1 Trichloro 2 2 bis 4 chlorophenyl ethane DDT degradation	0.0E+00	4.6E-11	0.0E+00	0.0E+00	3.1E-11	9.9E-02	3.2E-01
ABC transporters	2.0E-04	2.0E-04	1.6E-04	1.7E-04	1.8E-04	5.3E-02	2.6E-01
Adipocytokine signaling pathway	1.6E-06	2.2E-06	1.8E-06	1.9E-06	2.0E-06	1.9E-02	1.5E-01
African trypanosomiasis	1.0E-07	1.9E-07	1.8E-07	1.7E-07	1.7E-07	2.0E-02	1.5E-01
Alanine aspartate and glutamate metabolism	4.9E-05	5.3E-05	4.3E-05	4.3E-05	4.7E-05	2.3E-01	4.3E-01
Aldosterone regulated sodium reabsorption	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	6.8E-02	2.8E-01
alpha Linolenic acid metabolism	9.0E-09	5.0E-09	5.3E-09	8.3E-09	6.4E-09	1.5E-01	3.4E-01
Alzheimer s disease	2.2E-06	2.3E-06	1.9E-06	2.1E-06	2.0E-06	3.3E-01	5.0E-01
Amino acid metabolism	1.4E-05	1.3E-05	1.0E-05	1.2E-05	1.2E-05	2.2E-01	4.2E-01
Amino acid related enzymes	7.2E-05	7.5E-05	6.1E-05	6.1E-05	6.7E-05	2.3E-01	4.3E-01
Amino sugar and nucleotide sugar metabolism	7.5E-05	7.6E-05	6.2E-05	6.4E-05	6.6E-05	2.4E-01	4.3E-01
Aminoacyl tRNA biosynthesis	5.9E-05	6.2E-05	5.1E-05	5.2E-05	5.5E-05	4.5E-01	5.5E-01
Aminobenzoate degradation	4.3E-06	3.8E-06	3.5E-06	4.2E-06	3.8E-06	3.7E-01	5.2E-01
Amoebiasis	1.1E-06	1.0E-06	9.4E-07	1.0E-06	1.0E-06	9.5E-01	9.5E-01
Amyotrophic lateral sclerosis ALS	2.3E-07	2.1E-07	2.0E-07	2.3E-07	2.8E-07	3.0E-01	4.7E-01
Antigen processing and presentation	1.9E-06	2.0E-06	1.7E-06	1.7E-06	1.8E-06	4.1E-01	5.4E-01
Apoptosis	2.3E-08	2.0E-08	2.0E-08	4.9E-08	1.3E-08	1.6E-03	3.1E-02
Arachidonic acid metabolism	2.0E-07	2.2E-07	2.6E-07	2.8E-07	2.2E-07	1.1E-01	3.3E-01
Arginine and proline metabolism	6.5E-05	6.7E-05	5.4E-05	5.5E-05	6.0E-05	2.2E-01	4.2E-01
Ascorbate and aldarate metabolism	6.0E-06	5.1E-06	4.8E-06	4.8E-06	4.9E-06	1.3E-01	3.4E-01
Atrazine degradation	2.0E-06	2.0E-06	1.7E-06	1.9E-06	1.9E-06	4.9E-01	5.7E-01
Bacterial chemotaxis	1.7E-05	1.7E-05	1.4E-05	1.5E-05	1.6E-05	5.3E-02	2.6E-01
Bacterial invasion of epithelial cells	3.5E-08	2.5E-08	2.6E-08	4.8E-08	1.7E-08	2.2E-03	3.6E-02
Bacterial motility proteins	1.9E-05	2.2E-05	1.7E-05	1.7E-05	2.1E-05	1.1E-01	3.3E-01
Bacterial secretion system	2.4E-05	2.5E-05	2.2E-05	2.4E-05	2.3E-05	3.6E-01	5.2E-01
Bacterial toxins	6.5E-06	6.7E-06	5.9E-06	6.3E-06	6.0E-06	6.9E-01	7.3E-01
Basal transcription factors	6.0E-10	3.9E-08	1.1E-08	1.2E-08	4.3E-08	2.9E-11	8.0E-09
Base excision repair	2.2E-05	2.2E-05	1.9E-05	1.9E-05	2.1E-05	2.7E-01	4.5E-01

Benzoate degradation	7.9E-06	8.1E-06	6.9E-06	8.1E-06	7.6E-06	5.6E-01	6.3E-01
beta Alanine metabolism	9.7E-06	9.0E-06	7.4E-06	8.0E-06	8.4E-06	4.7E-02	2.5E-01
beta Lactam resistance	2.4E-06	2.3E-06	1.8E-06	2.1E-06	2.3E-06	2.7E-01	4.5E-01
Betalain biosynthesis	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	1.6E-01	3.5E-01
Bile secretion	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	7.8E-01	8.0E-01
Biosynthesis and biodegradation of secondary metabolites	3.6E-06	3.6E-06	3.0E-06	3.1E-06	3.3E-06	1.2E-01	3.3E-01
Biosynthesis of 12 14 and 16 membered macrolides	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	4.7E-01	5.5E-01
Biosynthesis of ansamycins	5.5E-06	5.3E-06	4.8E-06	5.1E-06	5.0E-06	1.2E-01	3.3E-01
Biosynthesis of siderophore group nonribosomal peptides	2.3E-07	1.8E-07	2.2E-07	3.2E-07	2.0E-07	8.6E-03	9.8E-02
Biosynthesis of type II polyketide backbone	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	4.7E-01	5.5E-01
Biosynthesis of type II polyketide products	0.0E+00	0.0E+00	3.1E-11	3.1E-11	0.0E+00	5.3E-04	1.6E-02
Biosynthesis of unsaturated fatty acids	5.5E-06	5.6E-06	4.8E-06	5.5E-06	5.0E-06	2.0E-01	4.1E-01
Biosynthesis of vancomycin group antibiotics	3.4E-06	3.5E-06	2.8E-06	3.3E-06	3.1E-06	3.5E-01	5.1E-01
Biotin metabolism	8.8E-06	8.3E-06	6.7E-06	7.3E-06	7.8E-06	8.2E-02	3.1E-01
Bisphenol degradation	4.5E-06	4.1E-06	3.7E-06	3.9E-06	3.8E-06	4.1E-01	5.4E-01
Bladder cancer	2.8E-09	1.2E-09	1.5E-09	2.5E-09	9.9E-10	1.1E-04	4.4E-03
Butanoate metabolism	2.7E-05	2.6E-05	2.2E-05	2.3E-05	2.5E-05	2.9E-01	4.6E-01
Butirosin and neomycin biosynthesis	2.7E-06	2.8E-06	2.5E-06	2.4E-06	2.8E-06	5.8E-01	6.5E-01
C5 Branched dibasic acid metabolism	2.0E-05	1.9E-05	1.5E-05	1.6E-05	1.7E-05	8.5E-02	3.1E-01
Caffeine metabolism	4.2E-10	2.8E-10	1.4E-08	5.5E-09	2.8E-09	6.6E-03	8.6E-02
Caprolactam degradation	5.5E-07	3.9E-07	5.0E-07	5.4E-07	4.4E-07	2.0E-02	1.5E-01
Carbohydrate digestion and absorption	3.1E-07	3.1E-07	3.1E-07	3.7E-07	2.8E-07	9.2E-01	9.2E-01
Carbohydrate metabolism	1.1E-05	9.9E-06	8.3E-06	9.3E-06	9.0E-06	1.1E-02	1.1E-01
Carbon fixation in photosynthetic organisms	2.9E-05	3.0E-05	2.6E-05	2.6E-05	2.7E-05	2.5E-01	4.3E-01
Carbon fixation pathways in prokaryotes	4.2E-05	4.3E-05	3.7E-05	3.7E-05	4.0E-05	2.9E-01	4.6E-01
Cardiac muscle contraction	2.0E-10	1.9E-10	2.2E-10	1.9E-10	9.3E-11	9.4E-02	3.1E-01
Carotenoid biosynthesis	3.0E-08	2.7E-08	1.1E-07	7.7E-08	8.2E-08	5.0E-04	1.6E-02

Cell cycle Caulobacter	2.4E-05	2.6E-05	2.0E-05	2.0E-05	2.3E-05	2.0E-01	4.1E-01
Cell division	5.2E-06	4.9E-06	3.7E-06	4.2E-06	4.7E-06	1.4E-01	3.4E-01
Cell motility and secretion	5.7E-06	5.3E-06	4.8E-06	5.2E-06	5.1E-06	1.1E-01	3.3E-01
Cellular antigens	1.3E-07	2.1E-07	1.7E-07	1.7E-07	1.8E-07	7.1E-02	2.9E-01
Chagas disease American trypanosomiasis	8.1E-08	1.2E-07	8.3E-08	8.4E-08	8.1E-08	4.2E-01	5.4E-01
Chaperones and folding catalysts	4.7E-05	4.7E-05	3.9E-05	3.9E-05	4.3E-05	1.3E-01	3.4E-01
Chloroalkane and chloroalkene degradation	1.2E-05	1.2E-05	1.0E-05	1.1E-05	1.1E-05	3.5E-01	5.1E-01
Chlorocyclohexane and chlorobenzene degradation	1.2E-06	1.2E-06	9.1E-07	1.2E-06	1.1E-06	3.7E-01	5.2E-01
Chromosome	7.4E-05	7.4E-05	6.3E-05	6.2E-05	6.8E-05	1.4E-01	3.4E-01
Circadian rhythm plant	4.2E-10	2.9E-10	1.4E-08	5.5E-09	2.8E-09	1.2E-02	1.1E-01
Citrate cycle TCA cycle	2.1E-05	2.0E-05	1.8E-05	2.0E-05	2.0E-05	4.8E-01	5.6E-01
Clavulanic acid biosynthesis	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	9.4E-02	3.1E-01
Colorectal cancer	6.2E-11	6.2E-11	7.7E-11	6.2E-11	3.1E-11	4.5E-01	5.5E-01
Cyanoamino acid metabolism	1.5E-05	1.3E-05	1.0E-05	1.2E-05	1.2E-05	2.6E-02	1.9E-01
Cysteine and methionine metabolism	4.4E-05	4.6E-05	3.8E-05	4.1E-05	4.0E-05	1.4E-01	3.4E-01
Cytochrome P450	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	1.5E-01	3.4E-01
Cytoskeleton proteins	2.0E-05	2.1E-05	1.7E-05	1.7E-05	2.0E-05	2.1E-01	4.2E-01
D Alanine metabolism	4.8E-06	4.9E-06	4.1E-06	4.4E-06	4.5E-06	3.3E-01	5.0E-01
D Arginine and D ornithine metabolism	2.7E-08	4.2E-08	2.1E-08	2.3E-08	2.6E-08	2.1E-01	4.1E-01
D Glutamine and D glutamate metabolism	6.5E-06	6.8E-06	5.6E-06	5.9E-06	6.1E-06	2.7E-01	4.5E-01
Dioxin degradation	3.6E-06	3.8E-06	3.2E-06	3.3E-06	3.4E-06	5.6E-01	6.3E-01
DNA repair and recombination proteins	1.3E-04	1.4E-04	1.1E-04	1.2E-04	1.2E-04	2.4E-01	4.3E-01
DNA replication	3.0E-05	3.2E-05	2.6E-05	2.8E-05	2.8E-05	2.5E-01	4.3E-01
DNA replication proteins	5.6E-05	5.9E-05	4.8E-05	4.8E-05	5.3E-05	1.7E-01	3.8E-01
Drug metabolism cytochrome P450	6.8E-07	7.0E-07	7.0E-07	7.3E-07	6.9E-07	6.9E-01	7.3E-01
Drug metabolism other enzymes	1.3E-05	1.3E-05	1.1E-05	1.1E-05	1.2E-05	9.4E-02	3.1E-01
Electron transfer carriers	9.7E-07	1.1E-06	1.6E-06	1.6E-06	1.6E-06	3.1E-03	4.7E-02
Endocytosis	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	8.7E-01	8.9E-01
Energy metabolism	3.9E-05	4.0E-05	3.3E-05	3.3E-05	3.7E-05	3.7E-01	5.2E-01

Epithelial cell signaling in Helicobacter pylori infection	5.9E-06	5.6E-06	4.5E-06	4.8E-06	5.3E-06	2.1E-01	4.2E-01
Ether lipid metabolism	1.4E-09	2.0E-08	1.9E-08	1.6E-08	2.8E-08	7.7E-05	3.5E-03
Ethylbenzene degradation	3.9E-07	6.2E-07	5.3E-07	6.4E-07	6.4E-07	1.2E-03	2.6E-02
Fatty acid biosynthesis	2.3E-05	2.3E-05	2.0E-05	1.9E-05	2.2E-05	2.8E-01	4.6E-01
Fatty acid elongation in mitochondria	4.0E-10	2.3E-10	1.4E-08	5.5E-09	2.8E-09	1.2E-02	1.1E-01
Fatty acid metabolism	1.0E-05	1.0E-05	9.0E-06	1.1E-05	9.8E-06	6.6E-01	7.1E-01
Fc gamma R mediated phagocytosis	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	8.7E-01	8.9E-01
Flagellar assembly	4.5E-06	6.5E-06	4.3E-06	3.6E-06	6.2E-06	1.5E-02	1.3E-01
Flavone and flavonol biosynthesis	1.1E-07	1.9E-07	8.9E-08	6.8E-08	1.7E-07	5.2E-03	7.1E-02
Flavonoid biosynthesis	3.9E-08	5.5E-08	1.3E-07	8.0E-08	1.4E-07	6.0E-05	3.5E-03
Fluorobenzoate degradation	3.5E-08	4.5E-09	2.9E-08	3.8E-08	7.9E-09	8.5E-04	2.2E-02
Folate biosynthesis	1.8E-05	1.8E-05	1.5E-05	1.7E-05	1.6E-05	7.8E-02	3.1E-01
Fructose and mannose metabolism	5.8E-05	5.8E-05	5.0E-05	5.0E-05	5.2E-05	2.3E-01	4.3E-01
Function unknown	6.0E-05	6.0E-05	4.9E-05	5.5E-05	5.5E-05	1.1E-01	3.3E-01
G protein coupled receptors	0.0E+00	1.1E-09	1.6E-09	1.9E-10	6.2E-10	3.8E-09	5.2E-07
Galactose metabolism	4.7E-05	4.7E-05	3.9E-05	4.0E-05	4.0E-05	1.4E-01	3.4E-01
General function prediction only	1.8E-04	1.8E-04	1.4E-04	1.5E-04	1.6E-04	1.4E-01	3.4E-01
Geraniol degradation	3.7E-07	2.3E-07	4.0E-07	3.6E-07	2.8E-07	2.1E-03	3.5E-02
Germination	3.2E-06	3.0E-06	2.4E-06	2.4E-06	2.8E-06	6.1E-02	2.8E-01
Glutamatergic synapse	5.3E-06	5.4E-06	4.3E-06	4.5E-06	5.0E-06	3.1E-01	4.8E-01
Glutathione metabolism	7.0E-06	6.7E-06	5.6E-06	6.8E-06	6.0E-06	4.4E-02	2.4E-01
Glycan biosynthesis and metabolism	3.0E-07	3.3E-07	2.8E-07	3.1E-07	3.7E-07	8.5E-02	3.1E-01
Glycerolipid metabolism	2.3E-05	2.2E-05	1.9E-05	1.8E-05	2.0E-05	1.3E-01	3.4E-01
Glycerophospholipid metabolism	2.8E-05	2.8E-05	2.4E-05	2.3E-05	2.6E-05	8.9E-02	3.1E-01
Glycine serine and threonine metabolism	4.0E-05	4.0E-05	3.3E-05	3.7E-05	3.6E-05	1.3E-01	3.4E-01
Glycolysis Gluconeogenesis	5.4E-05	5.5E-05	4.8E-05	5.1E-05	4.9E-05	3.0E-01	4.7E-01
Glycosaminoglycan degradation	6.5E-07	7.1E-07	6.7E-07	6.6E-07	6.2E-07	7.4E-01	7.7E-01
Glycosphingolipid biosynthesis ganglio series	2.1E-07	1.8E-07	2.6E-07	2.5E-07	2.1E-07	1.2E-01	3.3E-01
Glycosphingolipid biosynthesis globo series	4.3E-06	3.8E-06	3.3E-06	3.9E-06	3.5E-06	6.2E-02	2.8E-01

Glycosphingolipid biosynthesis lacto and neolacto series	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	7.2E-01	7.6E-01
Glycosylphosphatidylinositol GPI anchor biosynthesis	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	4.7E-01	5.5E-01
Glycosyltransferases	1.4E-05	1.3E-05	1.1E-05	1.3E-05	1.2E-05	2.0E-02	1.5E-01
Glyoxylate and dicarboxylate metabolism	2.5E-05	2.6E-05	2.2E-05	2.3E-05	2.5E-05	1.3E-01	3.4E-01
GnRH signaling pathway	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	8.7E-01	8.9E-01
Hematopoietic cell lineage	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	5.9E-01	6.5E-01
Histidine metabolism	2.9E-05	3.1E-05	2.6E-05	2.5E-05	2.8E-05	3.3E-01	5.0E-01
Homologous recombination	4.2E-05	4.4E-05	3.7E-05	3.8E-05	3.9E-05	3.0E-01	4.7E-01
Huntington s disease	3.8E-07	3.5E-07	3.6E-07	4.1E-07	3.6E-07	3.7E-01	5.2E-01
Hypertrophic cardiomyopathy HCM	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	6.5E-01	7.1E-01
Indole alkaloid biosynthesis	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	7.4E-01	7.7E-01
Influenza A	6.2E-11	6.2E-11	7.7E-11	6.2E-11	3.1E-11	4.5E-01	5.5E-01
Inorganic ion transport and metabolism	8.9E-06	8.8E-06	7.5E-06	8.5E-06	7.9E-06	1.8E-01	3.9E-01
Inositol phosphate metabolism	4.1E-06	4.0E-06	3.5E-06	4.4E-06	3.5E-06	1.2E-01	3.3E-01
Insulin signaling pathway	4.0E-06	4.2E-06	3.7E-06	3.7E-06	3.9E-06	3.9E-01	5.4E-01
Ion channels	6.4E-07	6.7E-07	6.0E-07	7.3E-07	5.8E-07	4.1E-01	5.4E-01
Isoflavonoid biosynthesis	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	9.2E-01	9.2E-01
Isoquinoline alkaloid biosynthesis	2.1E-06	2.1E-06	1.7E-06	2.0E-06	1.9E-06	6.0E-02	2.8E-01
Limonene and pinene degradation	2.9E-06	2.8E-06	2.8E-06	2.9E-06	2.7E-06	6.6E-01	7.1E-01
Linoleic acid metabolism	3.7E-06	3.7E-06	3.1E-06	3.6E-06	3.5E-06	4.3E-01	5.5E-01
Lipid biosynthesis proteins	2.5E-05	2.5E-05	2.1E-05	2.2E-05	2.2E-05	3.5E-01	5.1E-01
Lipid metabolism	8.1E-06	7.7E-06	6.7E-06	6.7E-06	7.1E-06	1.4E-01	3.4E-01
Lipoic acid metabolism	3.4E-07	4.0E-07	4.1E-07	4.4E-07	3.8E-07	1.5E-01	3.4E-01
Lipopolysaccharide biosynthesis	8.5E-07	5.2E-07	5.5E-07	6.4E-07	7.6E-07	4.5E-01	5.5E-01
Lipopolysaccharide biosynthesis proteins	5.1E-06	4.3E-06	4.1E-06	4.7E-06	4.3E-06	1.0E-02	1.1E-01
Lysine biosynthesis	4.7E-05	4.8E-05	3.8E-05	3.8E-05	4.4E-05	1.9E-01	4.0E-01
Lysine degradation	4.4E-06	4.0E-06	3.7E-06	4.1E-06	4.2E-06	2.5E-01	4.3E-01
Lysosome	2.6E-06	2.2E-06	2.0E-06	2.4E-06	2.0E-06	2.1E-01	4.1E-01

MAPK signaling pathway yeast	2.1E-06	1.9E-06	1.6E-06	1.8E-06	1.8E-06	1.1E-01	3.3E-01
Meiosis yeast	4.9E-09	3.9E-09	3.5E-08	2.0E-08	1.1E-08	1.8E-03	3.2E-02
Melanogenesis	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	1.5E-01	3.4E-01
Membrane and intracellular structural molecules	1.7E-05	1.6E-05	1.3E-05	1.6E-05	1.6E-05	3.1E-02	2.0E-01
Metabolism of cofactors and vitamins	7.0E-06	7.0E-06	5.5E-06	6.2E-06	6.5E-06	4.6E-01	5.5E-01
Metabolism of xenobiotics by cytochrome P450	6.7E-07	6.9E-07	6.4E-07	7.2E-07	6.8E-07	6.7E-01	7.2E-01
Methane metabolism	7.3E-05	7.2E-05	5.9E-05	5.9E-05	6.6E-05	1.2E-01	3.4E-01
Mineral absorption	7.1E-08	1.2E-07	9.1E-08	1.1E-07	1.4E-07	6.2E-02	2.8E-01
Mismatch repair	3.8E-05	3.9E-05	3.3E-05	3.2E-05	3.6E-05	1.5E-01	3.4E-01
mRNA surveillance pathway	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	4.7E-01	5.5E-01
N Glycan biosynthesis	9.9E-07	1.0E-06	8.4E-07	9.5E-07	9.9E-07	6.6E-01	7.1E-01
Naphthalene degradation	5.4E-06	5.5E-06	4.9E-06	5.5E-06	5.0E-06	8.8E-01	8.9E-01
Nicotinate and nicotinamide metabolism	2.0E-05	2.1E-05	1.7E-05	1.9E-05	1.8E-05	2.5E-01	4.3E-01
Nitrogen metabolism	3.4E-05	3.3E-05	2.7E-05	2.8E-05	3.0E-05	8.0E-02	3.1E-01
Nitrotoluene degradation	7.3E-06	6.4E-06	5.4E-06	5.7E-06	6.4E-06	4.9E-02	2.6E-01
NOD like receptor signaling pathway	2.0E-06	2.0E-06	1.8E-06	1.7E-06	1.9E-06	4.0E-01	5.4E-01
Nonhomologous endjoining	1.1E-07	1.4E-07	3.5E-07	2.1E-07	2.6E-07	1.6E-05	1.5E-03
Novobiocin biosynthesis	6.4E-06	6.7E-06	5.3E-06	5.6E-06	5.9E-06	1.5E-01	3.4E-01
Nucleotide excision repair	1.8E-05	2.0E-05	1.6E-05	1.8E-05	1.7E-05	3.9E-01	5.4E-01
Nucleotide metabolism	1.8E-06	1.4E-06	1.2E-06	1.5E-06	1.4E-06	7.4E-03	9.1E-02
One carbon pool by folate	2.6E-05	2.8E-05	2.3E-05	2.4E-05	2.4E-05	2.7E-01	4.5E-01
Other glycan degradation	1.2E-05	1.1E-05	9.0E-06	1.0E-05	1.0E-05	5.8E-02	2.8E-01
Other ion coupled transporters	5.7E-05	5.5E-05	4.4E-05	4.7E-05	5.2E-05	6.6E-02	2.8E-01
Other transporters	1.2E-05	1.2E-05	1.1E-05	1.1E-05	1.2E-05	3.3E-01	5.0E-01
Others	4.6E-05	4.4E-05	3.6E-05	4.3E-05	4.0E-05	5.7E-02	2.8E-01
Oxidative phosphorylation	5.9E-05	5.5E-05	4.5E-05	4.8E-05	5.0E-05	9.4E-02	3.1E-01
p53 signaling pathway	6.2E-11	6.2E-11	7.7E-11	6.2E-11	3.1E-11	4.5E-01	5.5E-01
Pantothenate and CoA biosynthesis	3.4E-05	3.3E-05	2.7E-05	2.8E-05	3.0E-05	1.0E-01	3.2E-01
Parkinson s disease	3.1E-10	2.5E-10	2.9E-10	2.2E-10	1.5E-10	1.3E-01	3.4E-01
Pathways in cancer	2.0E-06	2.0E-06	1.8E-06	1.8E-06	1.9E-06	3.4E-01	5.1E-01

Penicillin and cephalosporin biosynthesis	1.6E-06	1.5E-06	1.3E-06	1.4E-06	1.4E-06	3.6E-01	5.2E-01
Pentose and glucuronate interconversions	3.0E-05	3.0E-05	2.6E-05	2.7E-05	2.7E-05	8.1E-02	3.1E-01
Pentose phosphate pathway	4.5E-05	4.6E-05	3.9E-05	3.8E-05	4.1E-05	1.9E-01	4.0E-01
Peptidases	9.1E-05	9.2E-05	7.5E-05	7.7E-05	8.4E-05	1.3E-01	3.4E-01
Peptidoglycan biosynthesis	4.2E-05	4.3E-05	3.6E-05	3.6E-05	3.9E-05	2.5E-01	4.3E-01
Peroxisome	6.2E-06	6.3E-06	5.0E-06	6.2E-06	5.7E-06	2.8E-01	4.5E-01
Pertussis	3.3E-08	1.2E-08	6.6E-09	8.3E-09	7.6E-09	2.7E-02	1.9E-01
Phenylalanine tyrosine and tryptophan biosynthesis	4.4E-05	4.5E-05	3.8E-05	3.8E-05	4.2E-05	9.9E-02	3.2E-01
Phenylalanine metabolism	7.2E-06	7.3E-06	6.5E-06	6.6E-06	7.0E-06	3.5E-01	5.1E-01
Phenylpropanoid biosynthesis	8.2E-06	7.7E-06	6.2E-06	6.9E-06	7.2E-06	4.3E-02	2.4E-01
Phosphatidylinositol signaling system	4.1E-06	4.2E-06	3.4E-06	3.9E-06	3.7E-06	2.3E-01	4.3E-01
Phosphonate and phosphinate metabolism	2.5E-06	2.4E-06	2.1E-06	2.3E-06	2.1E-06	4.5E-01	5.5E-01
Phosphotransferase system PTS	3.4E-05	3.0E-05	2.9E-05	3.2E-05	2.9E-05	4.0E-01	5.4E-01
Photosynthesis	2.6E-05	2.5E-05	2.0E-05	2.1E-05	2.2E-05	1.9E-01	4.0E-01
Photosynthesis antenna proteins	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	3.9E-01	5.4E-01
Photosynthesis proteins	2.6E-05	2.5E-05	2.0E-05	2.1E-05	2.2E-05	1.9E-01	4.0E-01
Plant pathogen interaction	7.0E-06	7.0E-06	5.9E-06	5.8E-06	6.3E-06	1.1E-01	3.3E-01
Polycyclic aromatic hydrocarbon degradation	5.8E-06	5.9E-06	4.9E-06	4.9E-06	5.3E-06	4.7E-01	5.5E-01
Polyketide sugar unit biosynthesis	1.2E-05	1.2E-05	9.5E-06	1.0E-05	1.1E-05	2.7E-01	4.5E-01
Pores ion channels	7.7E-06	6.4E-06	5.9E-06	7.0E-06	5.9E-06	1.6E-02	1.3E-01
Porphyrin and chlorophyll metabolism	5.6E-05	5.3E-05	4.3E-05	4.3E-05	4.9E-05	3.9E-02	2.2E-01
PPAR signaling pathway	4.1E-06	4.3E-06	3.7E-06	3.9E-06	4.1E-06	7.9E-01	8.2E-01
Prenyltransferases	1.3E-05	1.3E-05	1.1E-05	1.2E-05	1.2E-05	4.0E-01	5.4E-01
Primary bile acid biosynthesis	2.1E-06	2.0E-06	1.7E-06	1.8E-06	1.8E-06	1.8E-01	3.9E-01
Primary immunodeficiency	2.2E-06	2.0E-06	1.6E-06	1.7E-06	1.7E-06	8.4E-03	9.8E-02
Prion diseases	1.1E-07	1.4E-07	1.1E-07	1.5E-07	1.7E-07	2.2E-01	4.3E-01
Progesterone mediated oocyte maturation	1.9E-06	2.0E-06	1.7E-06	1.7E-06	1.8E-06	4.1E-01	5.4E-01
Propanoate metabolism	2.4E-05	2.5E-05	2.1E-05	2.1E-05	2.3E-05	2.7E-01	4.5E-01
Prostate cancer	1.9E-06	2.0E-06	1.7E-06	1.7E-06	1.8E-06	4.1E-01	5.4E-01

Proteasome	2.1E-06	2.3E-06	1.8E-06	2.1E-06	2.1E-06	5.9E-01	6.5E-01
Protein digestion and absorption	1.8E-08	1.1E-07	5.3E-08	5.0E-08	7.8E-08	1.2E-03	2.6E-02
Protein export	2.7E-05	2.8E-05	2.4E-05	2.3E-05	2.5E-05	2.8E-01	4.6E-01
Protein folding and associated processing	3.4E-05	3.4E-05	2.8E-05	3.0E-05	3.0E-05	3.0E-01	4.7E-01
Protein kinases	1.8E-05	1.7E-05	1.4E-05	1.5E-05	1.5E-05	3.6E-02	2.2E-01
Protein processing in endoplasmic reticulum	2.1E-06	2.1E-06	2.1E-06	2.0E-06	2.1E-06	8.8E-01	8.9E-01
Proximal tubule bicarbonate reclamation	1.7E-07	1.8E-07	2.1E-07	2.0E-07	2.5E-07	5.9E-02	2.8E-01
Purine metabolism	1.1E-04	1.1E-04	9.1E-05	9.6E-05	1.0E-04	1.9E-01	4.0E-01
Pyrimidine metabolism	8.3E-05	8.8E-05	7.2E-05	7.4E-05	7.8E-05	2.4E-01	4.3E-01
Pyruvate metabolism	5.2E-05	5.3E-05	4.6E-05	4.5E-05	4.8E-05	2.4E-01	4.3E-01
Renal cell carcinoma	5.5E-08	5.6E-08	4.2E-08	5.3E-08	5.9E-08	6.3E-01	6.8E-01
Renin angiotensin system	3.1E-11	0.0E+00	1.5E-11	3.1E-11	0.0E+00	5.3E-01	6.1E-01
Replication recombination and repair proteins	4.2E-05	4.3E-05	3.4E-05	3.8E-05	3.8E-05	1.3E-01	3.4E-01
Restriction enzyme	6.9E-06	6.6E-06	5.9E-06	5.9E-06	6.3E-06	4.1E-01	5.4E-01
Retinol metabolism	6.4E-07	6.9E-07	6.4E-07	6.8E-07	6.4E-07	6.2E-01	6.8E-01
Riboflavin metabolism	1.0E-05	9.9E-06	7.9E-06	8.7E-06	8.9E-06	3.2E-02	2.0E-01
Ribosome	1.1E-04	1.2E-04	9.4E-05	1.0E-04	1.0E-04	4.2E-01	5.4E-01
Ribosome Biogenesis	7.0E-05	6.9E-05	5.8E-05	5.7E-05	6.3E-05	1.5E-01	3.4E-01
Ribosome biogenesis in eukaryotes	2.3E-06	2.4E-06	1.9E-06	2.2E-06	2.1E-06	3.8E-01	5.2E-01
RIG I like receptor signaling pathway	4.8E-09	2.8E-08	2.2E-08	2.0E-08	3.2E-08	6.8E-05	3.5E-03
RNA degradation	2.1E-05	2.2E-05	1.8E-05	1.8E-05	2.0E-05	2.7E-01	4.5E-01
RNA polymerase	7.9E-06	8.3E-06	6.9E-06	7.6E-06	7.5E-06	4.5E-01	5.5E-01
RNA transport	8.4E-06	7.5E-06	6.3E-06	6.8E-06	7.2E-06	3.3E-02	2.0E-01
Secondary bile acid biosynthesis	2.1E-06	2.0E-06	1.7E-06	1.8E-06	1.8E-06	2.0E-01	4.0E-01
Secretion system	5.1E-05	4.9E-05	4.2E-05	4.8E-05	4.7E-05	1.2E-01	3.4E-01
Selenocompound metabolism	1.8E-05	1.9E-05	1.6E-05	1.8E-05	1.7E-05	3.6E-01	5.2E-01
Shigellosis	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	6.8E-02	2.8E-01
Signal transduction mechanisms	2.6E-05	2.6E-05	2.1E-05	2.2E-05	2.4E-05	6.4E-02	2.8E-01
Small cell lung cancer	6.2E-11	6.2E-11	7.7E-11	6.2E-11	3.1E-11	4.5E-01	5.5E-01
Sphingolipid metabolism	1.0E-05	9.6E-06	7.7E-06	9.6E-06	8.7E-06	5.0E-02	2.6E-01

Sporulation	5.2E-05	4.8E-05	4.1E-05	4.2E-05	4.6E-05	9.0E-02	3.1E-01
Staphylococcus aureus infection	3.5E-07	2.8E-07	3.3E-07	5.7E-07	2.1E-07	8.7E-04	2.2E-02
Starch and sucrose metabolism	6.1E-05	5.9E-05	4.9E-05	5.2E-05	5.1E-05	8.8E-02	3.1E-01
Steroid biosynthesis	1.2E-09	7.3E-10	4.2E-08	1.7E-08	8.4E-09	1.1E-02	1.1E-01
Steroid hormone biosynthesis	3.2E-08	1.6E-08	8.8E-08	7.1E-08	3.8E-08	1.1E-02	1.1E-01
Stilbenoid diarylheptanoid and gingerol biosynthesis	1.3E-08	2.5E-08	2.7E-08	2.4E-08	3.5E-08	4.7E-03	6.7E-02
Streptomycin biosynthesis	1.7E-05	1.7E-05	1.4E-05	1.5E-05	1.5E-05	1.9E-01	4.0E-01
Styrene degradation	1.4E-06	1.2E-06	1.1E-06	1.3E-06	1.2E-06	3.3E-01	5.0E-01
Sulfur metabolism	1.2E-05	1.2E-05	9.6E-06	1.1E-05	1.1E-05	1.1E-01	3.3E-01
Sulfur relay system	1.5E-05	1.4E-05	1.2E-05	1.3E-05	1.3E-05	4.7E-02	2.5E-01
Synthesis and degradation of ketone bodies	4.2E-07	4.1E-07	3.9E-07	4.9E-07	4.9E-07	6.6E-01	7.1E-01
Systemic lupus erythematosus	7.7E-10	3.6E-10	1.4E-08	5.5E-09	2.8E-09	1.7E-02	1.4E-01
Taurine and hypotaurine metabolism	4.6E-06	4.8E-06	4.1E-06	4.4E-06	4.2E-06	2.2E-01	4.2E-01
Terpenoid backbone biosynthesis	2.5E-05	2.7E-05	2.2E-05	2.2E-05	2.4E-05	3.7E-01	5.2E-01
Tetracycline biosynthesis	1.0E-05	9.3E-06	8.2E-06	7.9E-06	8.9E-06	4.8E-01	5.6E-01
Thiamine metabolism	2.6E-05	2.7E-05	2.1E-05	2.2E-05	2.3E-05	8.6E-02	3.1E-01
Toluene degradation	3.5E-06	3.3E-06	3.2E-06	3.4E-06	3.3E-06	7.7E-01	8.0E-01
Toxoplasmosis	6.2E-11	6.2E-11	7.7E-11	6.2E-11	3.1E-11	4.5E-01	5.5E-01
Transcription factors	1.0E-04	1.0E-04	8.4E-05	8.5E-05	9.1E-05	8.5E-02	3.1E-01
Transcription machinery	4.6E-05	4.6E-05	3.8E-05	3.9E-05	4.4E-05	2.3E-01	4.3E-01
Transcription related proteins	2.0E-07	1.8E-07	2.0E-07	2.4E-07	1.7E-07	3.7E-02	2.2E-01
Translation factors	2.4E-05	2.6E-05	2.1E-05	2.1E-05	2.3E-05	3.3E-01	5.0E-01
Translation proteins	4.1E-05	4.2E-05	3.6E-05	3.7E-05	3.7E-05	2.0E-01	4.1E-01
Transporters	4.1E-04	4.1E-04	3.3E-04	3.5E-04	3.8E-04	7.6E-02	3.1E-01
Tropane piperidine and pyridine alkaloid biosynthesis	4.6E-06	4.8E-06	4.1E-06	4.4E-06	4.4E-06	4.2E-01	5.4E-01
Tryptophan metabolism	5.7E-06	5.5E-06	4.9E-06	5.3E-06	5.6E-06	3.1E-01	4.9E-01
Tuberculosis	6.8E-06	7.1E-06	6.2E-06	6.2E-06	6.6E-06	5.6E-01	6.3E-01
Two component system	7.0E-05	6.2E-05	5.0E-05	6.0E-05	6.0E-05	3.2E-02	2.0E-01

Type I diabetes mellitus	2.2E-06	2.3E-06	1.9E-06	2.0E-06	2.1E-06	5.7E-01	6.4E-01
Type II diabetes mellitus	2.2E-06	2.2E-06	1.9E-06	2.0E-06	1.9E-06	4.9E-01	5.7E-01
Tyrosine metabolism	1.5E-05	1.6E-05	1.3E-05	1.5E-05	1.4E-05	4.3E-01	5.5E-01
Ubiquinone and other terpenoid quinone biosynthesis	2.5E-06	2.1E-06	2.3E-06	2.3E-06	2.3E-06	4.7E-01	5.5E-01
Ubiquitin system	8.7E-08	1.2E-07	8.3E-08	1.1E-07	9.4E-08	4.3E-01	5.5E-01
Valine leucine and isoleucine biosynthesis	4.3E-05	4.2E-05	3.5E-05	3.7E-05	3.8E-05	1.5E-01	3.4E-01
Valine leucine and isoleucine degradation	6.1E-06	6.1E-06	5.4E-06	5.9E-06	5.9E-06	3.6E-01	5.2E-01
Various types of N glycan biosynthesis	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	2.5E-01	4.3E-01
Vasopressin regulated water reabsorption	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	5.2E-01	5.9E-01
Vibrio cholerae infection	0.0E+00	0.0E+00	0.0E+00	0.0E+00	0.0E+00	6.0E-01	6.6E-01
Vibrio cholerae pathogenic cycle	3.0E-06	2.9E-06	2.4E-06	2.6E-06	2.6E-06	3.0E-02	2.0E-01
Viral myocarditis	6.2E-11	6.2E-11	7.7E-11	6.2E-11	3.1E-11	4.5E-01	5.5E-01
Vitamin B6 metabolism	7.7E-06	8.3E-06	7.1E-06	7.7E-06	7.6E-06	5.5E-01	6.2E-01
Xylene degradation	3.6E-06	3.8E-06	3.1E-06	3.2E-06	3.3E-06	6.8E-01	7.2E-01
Zeatin biosynthesis	2.1E-06	2.3E-06	1.9E-06	1.9E-06	2.0E-06	4.3E-01	5.5E-01

[a] median abundance
[b] *P*-values calculated via Kruskal-Wallis
[c] FDR corrected *P*-values

APPENDIX IV

Mucosal microbiota species abundance data

Table includes only species with an average abundance > 0.01.

	CD								UC								nonIBD			
	Inflamed				Noninflamed				Inflamed				Noninflamed							
Species	Cecum	Colon	Ileum	Rectum	Cecum	Colon	Ileum	Rectum	Cecum	Colon	Ileum	Rectum	Cecum	Colon	Ileum	Rectum	Cecum	Colon	Ileum	Rectum
Ruminococcus_torques	0.01	0.16	0.62	0.00	0.17	1.37	0.18	0.80	0.01	0.25	0.01	0.17	0.39	0.28	0.28	0.15	0.03	0.04	0.03	0.08
Faecalibacterium_prausnitzii	0.01	0.05	0.07	0.00	0.06	0.18	0.13	0.30	0.00	0.21	0.01	0.16	0.43	0.20	0.27	0.17	0.05	0.06	0.04	0.08
Ruminococcus_gnavus	0.01	0.01	0.26	0.00	0.03	0.53	0.02	0.34	0.00	0.10	0.00	0.12	0.23	0.13	0.14	0.05	0.01	0.02	0.01	0.01
Anaerostipes_hadrus	0.00	0.03	0.15	0.00	0.08	0.43	0.05	0.22	0.00	0.06	0.00	0.04	0.12	0.04	0.08	0.04	0.01	0.01	0.01	0.02
Clostridium_clostridioforme	0.00	0.02	0.18	0.00	0.05	0.45	0.02	0.24	0.00	0.05	0.00	0.03	0.08	0.06	0.06	0.02	0.02	0.02	0.02	0.02
Clostridium_symbiosum	0.03	0.02	0.10	0.00	0.08	0.38	0.01	0.17	0.00	0.10	0.00	0.08	0.13	0.09	0.07	0.03	0.01	0.00	0.01	0.00
Prevotella_ruminicola	0.02	0.01	0.06	0.01	0.01	0.09	0.01	0.06	0.01	0.11	0.02	0.18	0.22	0.10	0.24	0.11	0.00	0.01	0.01	0.02
Eubacterium_rectale	0.00	0.05	0.10	0.00	0.04	0.29	0.02	0.17	0.00	0.06	0.00	0.06	0.11	0.07	0.08	0.03	0.02	0.02	0.02	0.02
Clostridium_hathewayi	0.00	0.05	0.07	0.00	0.02	0.18	0.02	0.10	0.00	0.06	0.00	0.05	0.13	0.10	0.09	0.07	0.01	0.02	0.01	0.03
Flavonifractor_plautii	0.01	0.02	0.13	0.00	0.02	0.24	0.02	0.13	0.00	0.05	0.00	0.03	0.09	0.07	0.08	0.04	0.01	0.01	0.01	0.01
Bacteroides_vulgatus	0.00	0.03	0.08	0.00	0.06	0.29	0.03	0.15	0.00	0.03	0.00	0.03	0.05	0.05	0.05	0.02	0.01	0.02	0.01	0.03
Blautia_hansenii	0.00	0.01	0.16	0.00	0.02	0.26	0.01	0.22	0.00	0.04	0.00	0.01	0.07	0.04	0.03	0.02	0.01	0.02	0.01	0.02
Bacteroides_thetaiotaomicron	0.01	0.02	0.10	0.00	0.05	0.19	0.02	0.11	0.00	0.02	0.00	0.02	0.05	0.03	0.05	0.02	0.01	0.02	0.02	0.02
Roseburia_inulinivorans	0.01	0.05	0.03	0.00	0.01	0.09	0.03	0.08	0.00	0.07	0.00	0.06	0.12	0.07	0.07	0.03	0.00	0.01	0.01	0.02
Anaerosporobacter_mobilis	0.00	0.02	0.06	0.00	0.02	0.15	0.01	0.09	0.00	0.05	0.00	0.04	0.07	0.04	0.06	0.02	0.01	0.02	0.01	0.02
Clostridium_lactatifermentans	0.01	0.02	0.03	0.00	0.01	0.10	0.02	0.06	0.00	0.05	0.00	0.05	0.09	0.06	0.07	0.04	0.01	0.02	0.01	0.02
Eubacterium_hallii	0.00	0.02	0.05	0.00	0.03	0.15	0.02	0.07	0.00	0.04	0.01	0.03	0.06	0.05	0.05	0.02	0.01	0.01	0.01	0.02
Blautia_stercoris	0.00	0.03	0.04	0.00	0.01	0.10	0.03	0.08	0.00	0.04	0.00	0.03	0.08	0.04	0.06	0.03	0.01	0.01	0.01	0.03
Butyrivibrio_fibrisolvens	0.00	0.01	0.04	0.00	0.01	0.07	0.01	0.05	0.00	0.07	0.01	0.06	0.08	0.03	0.07	0.06	0.01	0.02	0.02	0.02
Clostridium_spiroforme	0.00	0.00	0.02	0.00	0.00	0.09	0.01	0.05	0.00	0.04	0.01	0.06	0.11	0.07	0.08	0.02	0.02	0.01	0.01	0.02
Alistipes_finegoldii	0.00	0.00	0.05	0.00	0.02	0.17	0.01	0.09	0.00	0.03	0.00	0.03	0.05	0.02	0.06	0.02	0.00	0.02	0.01	0.01
Clostridium_glycyrrhizinilyticum	0.00	0.03	0.04	0.00	0.02	0.16	0.01	0.07	0.00	0.04	0.00	0.02	0.06	0.02	0.06	0.03	0.00	0.02	0.01	0.01
Blautia_faecis	0.00	0.04	0.04	0.00	0.02	0.13	0.02	0.08	0.00	0.04	0.01	0.03	0.06	0.04	0.05	0.01	0.00	0.00	0.00	0.01
Eubacterium_ramulus	0.00	0.01	0.05	0.00	0.03	0.18	0.02	0.08	0.00	0.02	0.00	0.04	0.07	0.02	0.03	0.02	0.00	0.00	0.00	0.00
Bacteroides_fragilis	0.00	0.00	0.09	0.00	0.01	0.20	0.01	0.12	0.00	0.01	0.00	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.01
Clostridium_elmenteitii	0.00	0.01	0.08	0.00	0.01	0.19	0.01	0.11	0.00	0.02	0.00	0.02	0.04	0.02	0.02	0.01	0.01	0.01	0.01	0.00

Paenibacillus_mucilaginosus	0.02	0.02	0.03	0.00	0.01	0.09	0.03	0.05	0.00	0.05	0.00	0.03	0.07	0.04	0.05	0.04	0.00	0.01	0.00	0.02
Clostridium_citroniae	0.00	0.01	0.08	0.00	0.01	0.18	0.01	0.11	0.00	0.01	0.00	0.02	0.02	0.02	0.03	0.01	0.01	0.01	0.01	0.01
Bacteroides_dorei	0.01	0.01	0.08	0.00	0.01	0.18	0.01	0.09	0.00	0.02	0.00	0.02	0.03	0.03	0.03	0.01	0.01	0.01	0.01	0.01
Bacteroides_uniformis	0.00	0.02	0.07	0.00	0.02	0.15	0.01	0.08	0.00	0.02	0.00	0.02	0.03	0.02	0.04	0.01	0.01	0.01	0.01	0.01
Ruminococcus_obeum	0.01	0.02	0.03	0.00	0.03	0.09	0.02	0.07	0.00	0.04	0.00	0.03	0.06	0.03	0.05	0.03	0.01	0.01	0.01	0.01
Bacteroides_stercoris	0.00	0.01	0.05	0.00	0.03	0.07	0.01	0.04	0.00	0.01	0.00	0.02	0.02	0.04	0.02	0.02	0.04	0.03	0.04	0.04
Eubacterium_contortum	0.00	0.01	0.04	0.00	0.02	0.11	0.01	0.06	0.00	0.02	0.00	0.03	0.06	0.04	0.03	0.02	0.00	0.00	0.01	0.01
Eubacterium_eligens	0.01	0.02	0.05	0.00	0.03	0.11	0.01	0.07	0.00	0.04	0.00	0.02	0.04	0.02	0.04	0.02	0.01	0.00	0.00	0.01
Blautia_luti	0.00	0.01	0.07	0.00	0.03	0.13	0.01	0.08	0.00	0.02	0.00	0.02	0.03	0.02	0.02	0.01	0.00	0.00	0.01	0.01
Clostridium_innocuum	0.01	0.02	0.03	0.00	0.01	0.06	0.01	0.05	0.00	0.03	0.00	0.03	0.06	0.03	0.05	0.02	0.01	0.01	0.01	0.01
Parabacteroides_distasonis	0.01	0.01	0.04	0.00	0.01	0.08	0.01	0.08	0.00	0.02	0.00	0.02	0.04	0.02	0.03	0.01	0.01	0.01	0.02	0.01
Clostridium_aldenense	0.01	0.01	0.05	0.00	0.01	0.10	0.01	0.06	0.00	0.02	0.00	0.03	0.04	0.03	0.02	0.02	0.00	0.01	0.00	0.01
Odoribacter_splanchnicus	0.00	0.01	0.03	0.00	0.03	0.08	0.03	0.08	0.00	0.02	0.00	0.02	0.03	0.02	0.02	0.02	0.01	0.01	0.00	0.02
Intestinimonas_butyriciproduce ns	0.00	0.02	0.01	0.00	0.01	0.02	0.01	0.04	0.00	0.04	0.01	0.03	0.07	0.05	0.05	0.03	0.00	0.01	0.01	0.01
Clostridium_lavalense	0.00	0.02	0.01	0.00	0.00	0.05	0.00	0.03	0.00	0.02	0.00	0.03	0.07	0.05	0.04	0.02	0.00	0.01	0.00	0.00
Barnesiella_intestinihominis	0.00	0.01	0.02	0.00	0.00	0.05	0.01	0.03	0.00	0.03	0.00	0.05	0.05	0.01	0.03	0.02	0.01	0.02	0.00	0.02
Bacteroides_caccae	0.00	0.02	0.02	0.00	0.02	0.06	0.02	0.03	0.00	0.02	0.00	0.03	0.04	0.03	0.03	0.01	0.01	0.01	0.01	0.02
Blautia_hydrogenotrophica	0.00	0.01	0.02	0.00	0.01	0.05	0.01	0.03	0.00	0.03	0.00	0.03	0.05	0.02	0.04	0.02	0.01	0.01	0.01	0.01
Collinsella_aerofaciens	0.00	0.01	0.02	0.00	0.03	0.06	0.03	0.06	0.00	0.01	0.00	0.02	0.02	0.02	0.02	0.01	0.00	0.00	0.00	0.01
Akkermansia_muciniphila	0.00	0.01	0.03	0.00	0.01	0.07	0.01	0.02	0.00	0.02	0.00	0.03	0.04	0.03	0.04	0.02	0.01	0.00	0.00	0.01
Parabacteroides_merdae	0.00	0.01	0.02	0.00	0.02	0.06	0.02	0.05	0.00	0.02	0.00	0.03	0.03	0.02	0.02	0.01	0.01	0.00	0.00	0.01
Dolosicoccus_paucivorans	0.00	0.02	0.02	0.00	0.01	0.03	0.01	0.03	0.00	0.02	0.00	0.02	0.04	0.03	0.03	0.03	0.00	0.01	0.00	0.01
Blautia_glucerasea	0.00	0.01	0.06	0.00	0.02	0.10	0.00	0.06	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.01	0.00	0.00	0.00	0.00
Dorea_longicatena	0.00	0.02	0.03	0.00	0.01	0.06	0.02	0.06	0.00	0.02	0.00	0.02	0.03	0.01	0.03	0.01	0.00	0.00	0.00	0.01
Blautia_producta	0.00	0.01	0.04	0.00	0.02	0.05	0.00	0.03	0.00	0.02	0.00	0.03	0.03	0.02	0.03	0.01	0.00	0.01	0.00	0.01
Streptococcus_suis	0.01	0.01	0.02	0.00	0.01	0.05	0.01	0.03	0.00	0.02	0.00	0.03	0.05	0.02	0.03	0.02	0.00	0.01	0.00	0.01
Pseudoflavonifractor_capillosus	0.00	0.02	0.01	0.00	0.00	0.03	0.01	0.02	0.00	0.03	0.01	0.02	0.04	0.02	0.05	0.02	0.00	0.01	0.01	0.00
Holdemania_filiformis	0.00	0.01	0.02	0.00	0.00	0.06	0.01	0.04	0.00	0.02	0.00	0.03	0.04	0.02	0.04	0.02	0.00	0.00	0.00	0.00
Prevotella_nanceiensis	0.00	0.00	0.02	0.00	0.01	0.03	0.00	0.01	0.00	0.01	0.01	0.04	0.06	0.02	0.06	0.03	0.00	0.00	0.00	0.00

Ruminococcus_lactaris	0.00	0.02	0.00	0.00	0.01	0.03	0.01	0.03	0.00	0.05	0.00	0.02	0.03	0.03	0.03	0.01	0.01	0.01	0.01	0.01
Alistipes_putredinis	0.00	0.00	0.02	0.00	0.01	0.06	0.01	0.04	0.00	0.02	0.00	0.02	0.04	0.02	0.02	0.01	0.01	0.00	0.01	0.01
Clostridium_sphenoides	0.01	0.01	0.02	0.00	0.02	0.05	0.01	0.04	0.00	0.02	0.00	0.01	0.02	0.02	0.03	0.02	0.00	0.01	0.00	0.01
Streptococcus_hyointestinalis	0.00	0.00	0.05	0.00	0.00	0.08	0.01	0.05	0.00	0.02	0.00	0.01	0.02	0.01	0.02	0.00	0.00	0.01	0.00	0.00
Paenibacillus_assamensis	0.00	0.02	0.01	0.00	0.01	0.05	0.01	0.05	0.00	0.02	0.00	0.02	0.02	0.01	0.02	0.01	0.00	0.01	0.00	0.01
Ruminococcus_flavefaciens	0.00	0.02	0.01	0.00	0.01	0.02	0.01	0.03	0.00	0.02	0.00	0.02	0.04	0.02	0.02	0.02	0.00	0.01	0.00	0.01
Allobacillus_halotolerans	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.02	0.00	0.03	0.05	0.06	0.04	0.02	0.00	0.01	0.01	0.01
Coprobacter_fastidiosus	0.00	0.01	0.02	0.00	0.00	0.05	0.01	0.02	0.00	0.02	0.00	0.02	0.03	0.02	0.03	0.03	0.00	0.00	0.00	0.01
Bilophila_wadsworthia	0.00	0.00	0.02	0.00	0.02	0.05	0.01	0.04	0.00	0.02	0.00	0.02	0.03	0.02	0.02	0.01	0.00	0.00	0.00	0.01
Ruminococcus_bromii	0.00	0.01	0.01	0.00	0.01	0.03	0.01	0.04	0.00	0.02	0.00	0.02	0.03	0.01	0.02	0.02	0.00	0.01	0.00	0.01
Paenibacillus_mendelii	0.00	0.01	0.01	0.00	0.00	0.01	0.01	0.02	0.00	0.02	0.00	0.01	0.04	0.04	0.04	0.03	0.00	0.00	0.00	0.01
Anaerotruncus_colihominis	0.00	0.01	0.01	0.00	0.01	0.03	0.01	0.03	0.00	0.01	0.00	0.01	0.03	0.04	0.04	0.02	0.00	0.01	0.00	0.01
Roseburia_hominis	0.00	0.01	0.00	0.00	0.01	0.03	0.01	0.02	0.00	0.02	0.00	0.02	0.04	0.03	0.04	0.02	0.00	0.00	0.01	0.01
Bacteroides_ovatus	0.00	0.01	0.02	0.00	0.01	0.05	0.01	0.04	0.00	0.01	0.00	0.01	0.02	0.01	0.02	0.01	0.00	0.01	0.00	0.01
Robinsoniella_peoriensis	0.00	0.01	0.03	0.00	0.03	0.05	0.01	0.03	0.00	0.00	0.00	0.00	0.02	0.02	0.01	0.00	0.00	0.00	0.00	0.00
Natranaerovirga_pectinivora	0.00	0.00	0.03	0.00	0.01	0.05	0.00	0.04	0.00	0.02	0.00	0.01	0.02	0.02	0.03	0.01	0.00	0.00	0.00	0.00
Paenibacillus_barcinonensis	0.00	0.00	0.01	0.00	0.00	0.02	0.00	0.03	0.00	0.02	0.00	0.02	0.03	0.02	0.03	0.02	0.00	0.01	0.01	0.02
Clostridium_scindens	0.00	0.01	0.03	0.00	0.01	0.04	0.00	0.03	0.00	0.01	0.00	0.02	0.03	0.02	0.01	0.01	0.00	0.00	0.00	0.00
Catenibacterium_mitsuokai	0.00	0.01	0.01	0.00	0.01	0.03	0.01	0.02	0.00	0.01	0.00	0.02	0.03	0.02	0.03	0.01	0.01	0.00	0.00	0.01
Prevotella_aurantiaca	0.00	0.01	0.01	0.00	0.00	0.01	0.00	0.01	0.00	0.02	0.00	0.03	0.04	0.02	0.05	0.03	0.00	0.00	0.00	0.00
Oribacterium_sinus	0.01	0.00	0.02	0.00	0.01	0.04	0.01	0.04	0.00	0.01	0.00	0.01	0.03	0.02	0.02	0.01	0.00	0.00	0.00	0.00
Eggerthella_lenta	0.00	0.01	0.02	0.00	0.01	0.05	0.00	0.02	0.00	0.01	0.00	0.02	0.02	0.01	0.02	0.01	0.00	0.00	0.01	0.01
Clostridium_ramosum	0.00	0.01	0.02	0.00	0.00	0.05	0.00	0.02	0.00	0.01	0.00	0.01	0.03	0.02	0.02	0.01	0.00	0.00	0.01	0.00
Butyricimonas_virosa	0.00	0.00	0.02	0.00	0.01	0.05	0.01	0.03	0.00	0.02	0.00	0.01	0.02	0.01	0.02	0.01	0.00	0.01	0.00	0.01
Clostridium_lituseburense	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.07	0.04	0.06	0.01	0.00	0.00	0.00	0.00
Rhodovulum_sulfidophilum	0.00	0.02	0.03	0.00	0.01	0.06	0.01	0.03	0.00	0.01	0.00	0.00	0.01	0.01	0.01	0.02	0.00	0.00	0.00	0.00
Roseburia_intestinalis	0.00	0.01	0.01	0.00	0.00	0.04	0.00	0.02	0.00	0.03	0.00	0.02	0.03	0.01	0.03	0.02	0.00	0.00	0.00	0.00
Parasporobacterium_paucivorans	0.00	0.01	0.01	0.00	0.01	0.02	0.01	0.02	0.00	0.02	0.00	0.01	0.03	0.01	0.02	0.01	0.00	0.01	0.00	0.01
Peptoclostridium_difficile	0.00	0.00	0.03	0.00	0.00	0.06	0.01	0.06	0.00	0.01	0.00	0.00	0.01	0.02	0.01	0.00	0.00	0.00	0.00	0.00

Acidaminococcus_intestini	0.00	0.00	0.04	0.00	0.00	0.08	0.00	0.04	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.01	0.00
Marvinbryantia_formatexigens	0.00	0.00	0.02	0.00	0.02	0.06	0.00	0.03	0.00	0.02	0.00	0.02	0.02	0.00	0.02	0.01	0.00	0.00	0.00	0.00
Prevotella_baroniae	0.00	0.00	0.01	0.00	0.01	0.02	0.00	0.01	0.00	0.02	0.00	0.02	0.04	0.02	0.02	0.02	0.00	0.00	0.00	0.00
Tissierella_praeacuta	0.00	0.00	0.02	0.00	0.02	0.02	0.00	0.03	0.00	0.01	0.00	0.01	0.03	0.02	0.02	0.01	0.00	0.00	0.00	0.00
Clostridium_fimetarium	0.00	0.01	0.01	0.00	0.00	0.07	0.01	0.04	0.00	0.00	0.00	0.01	0.02	0.01	0.00	0.01	0.01	0.00	0.00	0.01

APPENDIX V

Sample copy of form used to obtain consent from study participants

Health Sciences Centre
Winnipeg

IBD Clinical & Research Centre

Phone 204-787-1643
Fax 204-787-1643

RESEARCH PARTICIPANT INFORMATION AND CONSENT FORM

Study Title: **Inflammation in Inflammatory Bowel Disease**

Principal Investigator: **Dr. Charles Bernstein**
Health Sciences Centre
804F – 715 McDermot Avenue
Winnipeg, MB R3E 3P4
(204) 789-3369 / 787-2060

INTRODUCTION

You are being asked to participate in a human research study. Please take your time to review this consent form and discuss any questions you may have with study staff. You may take your time to make your decision about participating in this research and you may discuss it with your regular doctor, family, and friends before making a decision. This consent form may contain words that you do not understand. Please ask the study doctor or study staff to explain any words or information that you do not clearly understand.

Once you understand the nature of the study and the study procedures, you will be asked to sign this form if you agree to participate. You will be given a signed and dated copy of this form to keep as a record.

Your participation in this study is entirely voluntary.

PURPOSE OF STUDY

You are being asked to participate in a study about the etiology and pathogenesis (cause) of inflammatory bowel disease (Crohn's disease and ulcerative colitis). We will be studying factors that have an important role in perpetuating the inflammatory response of inflammatory bowel disease.

Specifically, we will be studying inflammatory factors in the intestinal wall, in the serum, and on circulating lymphocytes (a type of white blood cell), and the stool microbes (bugs). We will be studying some of the important inflammatory proteins, as well as the genes or DNA in the cells that help produce these proteins. Using magnetic resonance spectroscopy as a method of studying inflammation, we will be assessing the biochemical profile (chemical makeup) of human tissue.

You are being asked to participate in this study either because you have an inflammatory condition of the bowel (Crohn's disease or ulcerative colitis), or you have Irritable Bowel Syndrome (IBS), or intestinal cancer, or a normal bowel. Your tissue biopsies and/or blood sample will be used as a control (normal) for the purpose of this study.

PART A: BLOOD SAMPLE STUDY

If you agree to participate in this study, blood from a vein in your arm will be obtained to study the inflammatory factors in the serum or on blood cells. After processing the blood sample and labelling it with a laboratory number, we will be freezing and storing the blood in

a secure freezer located at the John Buhler Research Building (part of the University of Manitoba) so that we may study it at a later time. The purpose of freezing the blood is to allow for the study of groups of blood samples from a number of patients at one time. The blood sample (serum and blood cells) will be stored confidentially (laboratory number only) and will be kept for up to 25 years for further research relating to inflammatory bowel disease. We will be extracting DNA from some of the blood samples to study the different aspects of inflammation. As well, blood samples may be shared confidentially with other researchers working in collaboration on studies in inflammatory bowel disease, either locally or internationally.

Study Procedures

If you take part in this study, you will have 20 ml (3 teaspoons) of blood drawn from a vein in your arm once.

Risks/Discomforts

The risks of having a venipuncture (blood drawn) are small and may include pain and/or a bruise at the needle site after the venipuncture. Very rarely, an infection at the venipuncture site may occur.

Benefits

By participating in this study, your blood sample may provide the study doctors with information about the inflammatory factors in serum or blood cells and/or DNA. There may or may not be direct benefit to you from participating in this study.

PART B: INTESTINAL BIOPSY/TISSUE STUDY

In this study, we will be taking intestinal biopsies from patients who may or may not have intestinal inflammation, but who are undergoing endoscopies and biopsies as indicated by their attending physician. During the course of this endoscopy, an additional 4-8 biopsies will be taken from the bowel. If there are areas of inflammation, the additional biopsies will then be taken from the rectum.

If you agree to participate in this study, tissue from your colon will be obtained to study the inflammatory factors in the tissue. After processing the tissue sample and labeling it with a laboratory number, we will be freezing and storing the tissue in a secure freezer located at the John Buhler Research Building (part of the University of Manitoba) so that we may study it at a later time. The purpose of freezing the tissue is to allow for the study of groups of tissue samples from a number of patients at one time. The tissue sample will be stored confidentially (laboratory number only) and will be kept for up to 25 years for further research relating to inflammatory bowel disease. We will be extracting DNA from some of the tissue samples to study the different aspects of inflammation. Other tissue samples will be studied by magnetic resonance spectroscopy at the National Research Council of Canada Institute for Biodiagnostics, 435 Ellice Avenue, Winnipeg, MB. As well, tissue samples may be shared confidentially with other researchers working in collaboration on studies in inflammatory bowel disease, either locally or internationally.

Study Procedures

If you take part in this study, during your endoscopy, you will have 4-8 tissue biopsies taken from the bowel. Additional biopsies will be taken from any inflamed sites or from the rectum if there is no inflammation.

Risks/Discomforts

The risk of intestinal biopsies is very minimal and is associated with excess bleeding from the biopsy site or intestinal perforation. This risk is minimal to begin with and the additional biopsies will only very slightly increase this risk. The biopsies are painless and the amount of tissue taken is very small. The additional biopsies will add approximately 2 minutes to the procedure.

You may not be included in this study if you are at increased risk of bleeding from intestinal biopsies. This includes patients with Hemophilia, other blood-clotting disorders, and patients chronically using Coumadin or Heparin (medications that specifically interfere with blood clotting).

Benefits

By participating in this study, your tissue samples may provide the study doctors with information about the inflammation and/or DNA in tissue. There may or may not be direct benefit to you from participating in this study.

Part C: STOOL STUDY

In this study, we will be assessing stool samples which will allow us to determine if there are microbes (bugs) that might be present in inflammatory bowel disease that are not present in healthy subjects. It is important to look in stool as well as in tissue for microbes because we have shown previously that sometimes the microbes in stool are different than the microbe profile in colon tissue.

Study Procedures

If you agree to participate in this study, you will collect two samples of your stool in the stool kits provided, two to three months apart, as well as complete a food frequency questionnaire. Both stool samples and the questionnaire can be returned to the study research area at GA1 Clinic C.

Risks/Discomforts

There are no risks to the stool sample collection.

Benefits

By participating in this study, your stool sample may provide the study doctors to determine if there are microbes (bugs) that might be present in inflammatory bowel disease that are not present in healthy subjects. There may or may not be direct benefit to you from participating in this study

COSTS/PAYMENT FOR PARTICIPATION

All clinic and professional fees, diagnostic and laboratory tests, which will be performed as part of this study are provided at no cost to you.

You will receive a $20 study honorarium for your initial stool sample and Food Frequency Questionnaire and an additional $20 study honorarium for your second stool sample.

CONFIDENTIALITY

Information gathered in this research study may be published or presented in public forums; however, your name will not be used or revealed. All tissue and blood samples will be identified with a laboratory number. Medical records that contain your identity will be treated as confidential in accordance with the Personal Health Information Act of Manitoba, and only be used for research purposes. Despite efforts to keep your personal information confidential, absolute confidentiality cannot be guaranteed. Your personal information may be disclosed if required by law. Organizations that may inspect and/or copy your research records for quality assurance and data analysis include such groups as University of Manitoba, Health Research Ethics Board.

VOLUNTARY PARTICIPATION/WITHDRAWAL FROM THE STUDY

Your decision to take part in this study is voluntary. You may refuse to participate or you may withdraw from the study at any time. Your decision to not participate or to withdraw from this study will not affect your other medical care at this site.

MEDICAL CARE FOR INJURY RELATED TO THE STUDY

In the case of injury or illness from this study, necessary medical treatment will be available at no additional cost to you. If you should become physically injured as a result of any research activity, the study doctor will provide any necessary treatment, at no charge, to help you promptly recover from the injury.

QUESTIONS

You are free to ask any questions that you may have about the study and your rights as a research participant. If any questions come up during or after the study, or if you have a research-related injury, contact the study doctor or study nurse at (204) 787-1643.

For questions about your rights as a research participant, you may contact the University of Manitoba, Health Research Ethics Board at (204) 789-3389.

Do not sign this consent form unless you have had a chance to ask questions and have received satisfactory answers to all of your questions.

STATEMENT OF CONSENT

I have read this consent form. I have had the opportunity to discuss this research study with Dr. Charles Bernstein and/or his study staff. I have had my questions answered by them in language I understand. The risks and benefits have been explained to me. I understand that I will be given a copy of this consent form after signing it. I understand that my participation in this study is voluntary, and that I may choose to withdraw at any time. I freely agree to participate in this research study.

I understand that information regarding my personal identity will be kept confidential but that confidentiality is not guaranteed. I authorize the inspection of my medical records by The University of Manitoba, Health Research Ethics Board.

By signing this consent form, I have not waived any legal rights that I have as a participant in a research study.

PART A	**Yes**	**No**
I agree to have my blood used in the inflammation studies.	☐	☐
I agree to have my blood stored in the freezer for 25 years for later use.	☐	☐
I agree to have my blood used in inflammatory studies either locally or internationally.	☐	☐
PART B		
I agree to have my tissue used in the inflammation studies.	☐	☐
I agree to have my tissue used in the magnetic resonance spectroscopy studies.	☐	☐
I agree to have my tissue stored in the freezer for 25 years for later use.	☐	☐
I agree to have my tissue used in inflammatory studies either locally or internationally.	☐	☐
PART C		
I agree to have my stool used in the inflammation studies.	☐	☐
I agree to have my stool stored in the freezer for 25 years for later use.	☐	☐
I agree to have my stool used in inflammatory studies either locally or internationally.	☐	☐

_____ _____ _____/_____/_____

Signature of participant Name (printed) day month year*

STATEMENT OF THE PERSON OBTAINING CONSENT

I, the undersigned, have fully explained the relevant details of this research study to the above-named participant, and believe that the participant has understood and has knowingly given their consent.

_____ _____ _____/_____/_____

Signature of person obtaining consent Name (printed) day month year*

Role in the study _____

APPENDIX VI

Copyright agreements

Copyright agreement for manuscript I

Frontiers Editorial Office
To: Jessica Forbes
Re: requesting permissions

Today at 6:55 AM

FE

Dear Jessica,

Thank you for your email.

Under the Frontiers Terms and Conditions, authors retain the copyright to their work. Furthermore, all Frontiers articles are Open Access and distributed under the terms of the Creative Commons Attribution License (CC-BY 3,0), which permits the use, distribution and reproduction of material from published articles, provided the original authors and source are credited, and subject to any copyright notices concerning any third-party content.

You can therefore reuse your paper, given that you make sure to properly acknowledge, in your manuscript, the original source . We recommend to use the format below: "As originally published in Metpally RPR, Nasser S, Malenica I, Courtright A, Carlson E, Ghaffari L, Villa S, Tembe W and Van Keuren-Jensen K (2013) . Front. Genet. 4:20. doi: 10.3389/fgene.2013.00020"

Please let me know if you have any other questions or concerns.

Kind Regards,

Damaris

Ethics & Integrity Manager: Gearóid Ó Faoleán, PhD

Frontiers

www.frontiersin.org
EPFL Innovation Square, Building I
Lausanne, Switzerland
Office T +44 79 34 46 47 49

For technical issues, please visit our Frontiers Help Center frontiers.zendesk.com

Copyright agreement for manuscript II

Copyright agreement for manuscript III

Thank you for your order!

Dear Jessica Forbes,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Order Date:	Oct 23, 2016
Order Number:	3975061118473
Publication:	Canadian Journal of Microbiology
Title:	Microbiome profiling of drinking water in relation to incidence of inflammatory bowel disease
Type of Use:	Thesis/Dissertation
Order Total:	0.00 USD

View or print complete details of your order and the publisher's terms and conditions.

Please note: You will be invoiced within 24 hours.

Sincerely,

Copyright Clearance Center

Copyright agreement for Figure 1

Peter Lamb 12:14 PM (7 hours ago)

to me

Hello Jessica,

Yes, you have my permission to reuse the illustration.

Thanks for asking.

All the best.

Peter

...

On 11/2/2016 5:23 PM, Jessica Forbes wrote:

> Hello,
> I was forwarded your contact information..
>
> > I was interesting in reusing figure 2 from the following article: *Nursing Standard*. 28, 24, 37-45. http://dx.doi.org.uml.idm.oclc.org/10.7748/ns2014.02.28.24.37.e7395 for use in a thesis and was wondering if I could get copyright permissions.
>
> Thanks,
> Jessica

Copyright agreement for Figure 3

Thank you for your order!

Dear Jessica Forbes,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Order Date:	Nov 5, 2016
Order Number:	3982680027678
Publication:	Nature
Title:	Genomic approaches to studying the human microbiota
Type of Use:	reuse in a dissertation / thesis
Order Total:	0.00 USD

View or print complete details of your order and the publisher's terms and conditions.

Please note: You will be invoiced within 24 hours.

Sincerely,

Copyright Clearance Center

Copyright agreement for Figure 4.

Thank you for your order!

Dear Jessica Forbes,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Order Date:	Oct 29, 2016
Order Number:	3978530246337
Publication:	Brain, Behavior, and Immunity
Title:	The role of microbiome in central nervous system disorders
Type of Use:	reuse in a thesis/dissertation
Order Total:	0.00 USD

View or print complete details of your order and the publisher's terms and conditions.

Please note: You will be invoiced within 24 hours.

Sincerely,

Copyright Clearance Center

Copyright agreement for Figure 5.

Thank you for your order!

Dear Jessica Forbes,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Order Date:	Oct 31, 2016
Order Number:	3979351404063
Publication:	Rheumatology
Title:	Gene, environment, microbiome and mucosal immune tolerance in rheumatoid arthritis:
Type of Use:	Thesis/Dissertation
Order Total:	0.00 USD

View or print complete details of your order and the publisher's terms and conditions.

Please note: You will be invoiced within 24 hours.

Sincerely,

Copyright Clearance Center

Copyright agreement for Figure 6

Thank you for your order!

Dear Jessica Forbes,

Thank you for placing your order through Copyright Clearance Center's RightsLink® service.

Order Summary

Order Date:	Oct 28, 2016
Order Number:	3977870669923
Publication:	American Journal of Epidemiology
Title:	A Population-based Ecologic Study of Inflammatory Bowel Disease: Searching for Etiologic Clues:
Type of Use:	Thesis/Dissertation
Order Total:	0.00 USD

View or print complete details of your order and the publisher's terms and conditions.

Please note: You will be invoiced within 24 hours.

Sincerely,

Copyright Clearance Center