

DYNAMICS OF THE GUT VIROME AND BACTERIAL MICROBIOME IN
PEDIATRIC CROHN'S DISEASE PATIENTS

By

Casey Michael Arthur Jones

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
December 2018

© Copyright by Casey Michael Arthur Jones, 2018

Dedication

To my family, friends, coaches, teammates, and mentors,
who have been there through it all.

TABLE OF CONTENTS

LIST OF TABLES	vi
LIST OF FIGURES	vii
ABSTRACT	ix
LIST OF ABBREVIATIONS USED	x
ACKNOWLEDGEMENTS	xii
CHAPTER 1: INTRODUCTION	1
1.1 CROHN'S DISEASE	1
1.1.1 IBD EPIDEMIOLOGY	1
1.1.2 DIAGNOSIS OF IBD	2
1.1.3 PATHOGENESIS OF CROHN'S DISEASE	3
1.1.3.1 GENETICS OF IBD	3
1.1.3.2 IMMUNOLOGY	5
1.1.3.3 ENVIRONMENT	6
1.1.4 TREATMENTS FOR CROHN'S DISEASE	7
1.1.4.1 PHARMACOLOGICAL THERAPIES	8
1.1.4.2 EXCLUSIVE ENTERAL NUTRITION (EEN) IN TREATMENT FOR IBD	11
1.2 THE GUT MICROBIOME	13
1.3 THE GUT VIROME	17
1.4 ANALYSING THE MICROBIOME	19
1.4.1 DNA SEQUENCING	19
1.4.2 BIOINFORMATIC APPROACHES	20
1.5 THE MICROBIOME IN IBD	22

1.6	OBJECTIVES	25
CHAPTER 2: MATERIALS AND METHODS		26
2.1	MAREEN STUDY RECRUITMENT	26
2.2	VIRUS LIKE PARTICLE ENRICHMENT PROTOCOL	26
2.3	LIBRARY PREPARATION AND SEQUENCING	27
2.4	VIRAL SEQUENCE PROCESSING AND ANALYSIS	27
2.5	16S WETLAB PROTOCOL	29
2.6	16S RIBOSOMAL RNA SEQUENCING ANNOTATION	29
2.7	STATISTICAL METHODS	30
2.8	MACHINE LEARNING	31
CHAPTER 3: THE GUT VIROME		33
3.1	VALIDATION OF VLP PREPARATION	34
3.2	COMPARISON OF BIOINFORMATIC METHODS FOR VIRAL ANNOTATION	37
3.3	REFERENCE-INDEPENDENT APPROACHES ARE UNINFORMATIVE FOR CLASSIFYING THE GUT VIROME	39
3.4	REFERENCE SEARCHES PRODUCE VALUABLE INSIGHT ON THE PEDIATRIC CD GUT VIROME	42
3.5	MACHINE LEARNING IDENTIFIES SUBTLE CHANGES IN THE GUT VIROME RELATED TO CLINICAL OUTCOMES	50
3.6	SUMMARY	51
CHAPTER 4: THE BACTERIOME OF PEDIATRIC CD		52
4.1	ALPHA DIVERSITY OF THE BACTERIOME DIFFERS WITH FECAL CALPROTECTIN LEVELS	52
4.2	BACTERIAL TAXA DIFFER WITH FECAL CALPROTECTIN LEVEL AND REMISSION STATUS	54

4.3	BACTERIAL TAXA SIGNIFICANTLY MODEL FCP LEVELS	56
4.4	SUMMARY	58
CHAPTER 5: VIROME-BACTERIOME INTERACTIONS		59
5.1	HIGH LEVEL BACTERIA – VIROME ASSOCIATIONS	59
5.2	KNOWN PHAGE-BACTERIA INTERACTIONS DO NOT CORRELATE WELL IN PEDIATRIC CD PATIENTS.....	62
5.3	ENTEROCOCCUS PHAGE IS ASSOCIATED WITH ITS HOST AND DISEASE ACTIVITY IN ONE PATIENT	62
5.4	SUMMARY	64
CHAPTER 6: DISCUSSION.....		65
6.1	VIRUS LIKE PARTICLE PREPARATIONS	65
6.2	REFERENCE BASED BIOINFORMATICS ARE OPTIMAL FOR ANNOTATION OF THE GUT VIROME	68
6.3	VIRAL DIVERSITY IS ASSOCIATED WITH SEVERAL CLINICAL CHARACTERISTICS	69
6.4	BACTERIOME OF CD	72
6.5	THE GUT VIROME AND BACTERIOME DISPLAY MODEST INTERACTIONS IN THE PEDIATRIC CD GUT.....	74
6.6	LIMITATIONS AND FUTURE DIRECTIONS	76
6.7	CONCLUSIONS.....	78
REFERENCES		79

LIST OF TABLES

Table 1 MAREEN Study cohort information	33
Table 2 Shannon diversity is correlated with selected clinical phenotypes	45
Table 3 Top 20 viral species important for regression of FCP levels.....	50
Table 4 Differentially abundant bacterial taxa between samples with high and low fecal calprotectin samples.....	55
Table 5 Top 20 bacterial taxa important for regression of FCP levels	57

LIST OF FIGURES

Figure 1 Virally-enriched samples show a greater percent read mapping to viral protein database	35
Figure 2 Stacked bar chart depicting relative abundance of top 10 most abundant phages in a blank run sample	36
Figure 3 Stacked bar chart depicting relative abundance of 10 most abundant viruses in Modulen EEN formula	37
Figure 4 Performance of selected bioinformatic tools of annotating viral metagenomic sequencing reads in virally enriched samples.....	38
Figure 5. Shannon Diversity of each MAREEN patient across all timepoints.....	39
Figure 6 Viral genome bin diversity does not differ between samples with low FCP and high FCP levels	40
Figure 7 Viral genome bin dissimilarity in all MAREEN patients	41
Figure 8 Boxplot of most abundant viral families observed in MAREEN patients..	43
Figure 9 Viral species diversity is higher in samples with low disease activity compared to samples with high disease activity.....	44
Figure 10 Stacked bar charts of viral families grouped by inflammation level and time sampled.....	46
Figure 11 Patients that undergo sustained remission after EEN administration do not have distinct gut virome compositions.....	47
Figure 12 The gut virome significantly varies by patient.....	48
Figure 13 Selected phage trend toward significant difference between high and low FCP levels	49
Figure 14 Alpha diversity of the bacteriome is significantly higher in samples with low FCP compared to high FCP	52
Figure 15 Stacked bar charts of baseline gut bacteriome in pediatric CD patients	53
Figure 16 Weighted UniFrac principal coordinates axis plot of gut bacteriome	54

Figure 17 Patients that undergo sustained remission after EEN administration have distinct bacterial signatures 56

Figure 18 The abundance of Enterobacteria phage mEpx1 is significantly correlated with its bacterial host, *Escherichia-Shigella sp.*..... 60

Figure 19 Spearman correlation plots of the relative abundance of the 50 most abundant viral taxa and 20 most informative ASVs in the machine learning model of FCP regression 61

Figure 20 Abundance of Enterococcus phage EF62phi is significantly associated with abundance of *Enterococcus sp.* and trends with clinical diagnostic values in one patient..... 63

ABSTRACT

The human gut microbiome is a dynamic community of bacteria, viruses, and eukaryotic microbes that occupy the gastrointestinal tract. The bacterial portion of the microbiome in inflammatory bowel disease (IBD) patients has been well studied, but less is known about the gut virome, the viral portion of the gut microbiome. An expansive community of viruses infecting bacteria (bacteriophage) exists in the human gut and is thought to be associated with disease. Exclusive enteral nutrition (EEN) is an evolving first-line treatment for pediatric Crohn's Disease (CD) that has a modulatory effect on the gut microbiome, whose effect has yet to be characterized on the gut virome. We sought out to characterize the gut virome in 22 patients undergoing treatment for pediatric CD over a 96-week sampling period. Metagenomic sequencing was carried out on stool samples enriched for viral particles, and 16S rRNA sequencing was done on stool to identify bacterial taxonomy. We found that viral species diversity was enriched in samples with a lower disease activity score, and that diversity was associated with several clinical markers. Bacterial diversity was also enriched in samples with low inflammation. We used machine learning models to detect viral and bacterial taxa linked to gut inflammation and used these features to uncover phage-bacteria dynamics previously uncharacterized in the human gut. This longitudinal analysis of gut phage and bacteria was shown for the first time here in a cohort of patients undergoing treatment for pediatric CD. This work suggests that the gut virome may contribute to intestinal inflammation and play a role in CD.

LIST OF ABBREVIATIONS USED

5-ASA	5-aminosalicylic acid
6-MP	6-mercaptopurine
6-TG	6-thioguanine
ADA	Adalimumab
AIEC	Adherent invasive Escherichia coli
ASCA	anti-Saccharomyces cerevisiae antibody
ASV	Amplicon sequence variants
AZA	Azathioprine
CD	Crohn's Disease
cGCR	Cytosolic glucocorticoid receptor alpha
CRP	C-reactive protein
dsDNA	Double-stranded DNA
EEN	Exclusive enteral nutrition
EN	Enteral nutrition
ESR	Erythrocyte sedimentation rate
FCP	Fecal calprotectin
FDR	False discovery rate
GI	Gastrointestinal
GWAS	Genome wide association study
IBD	Inflammatory Bowel Disease
IFX	Infliximab
IL	Interleukin
IMG/VR	Integrated microbial genome/virus
IMID	Immune-mediated inflammatory disorders
KTW	Kill the winner
LPS	Lipopolysaccharide
MAREEN	A metagenomic approach to diagnosis, induction and maintenance of deep remission following exclusive enteral nutrition in pediatric Crohn's Disease
MARVEL	Metagenomic Analysis and Retrieval of Viral Elements

MCT	Medium chain triglycerides
MDP	Muramyl dipeptide
MTX	Methotrexate
NF- κ B	Nuclear Factor- κ B
NGS	Next generation sequencing
NOD2	Nucleotide-binding oligomerization domain containing 2
NR	Non-sustained remission
NSAID	Non-steroidal anti-inflammatory drugs
PAMPs	Pathogen associated molecular patterns
pANCA	Atypical perinuclear anti-neutrophil cytoplasmic antibody
PCoA	Principle coordinate analysis
PIBD	Pediatric inflammatory bowel disease
PRR	Pattern recognition receptor
PTW	Piggyback the winner
RF	Random forests
rRNA	Ribosomal ribonucleic acid
SBS	Sequencing by synthesis
SCFA	Short-chain fatty acid
SR	Sustained remission
ssDNA	Single-stranded DNA
SV40	Simian virus 40
TGF- β 2	Transforming growth factor-beta 2
Th	T helper cell
TLR	Toll like receptor
TMPT	Thiopurine S-methyltransferase
TNF- α	Tumor necrosis factor-alpha
Treg	T regulatory
UC	Ulcerative Colitis
VLP	Virus like particle
wPCDAI	Weighted pediatric Crohn's Disease activity index

ACKNOWLEDGEMENTS

It has been a privilege to work in the Department of Pharmacology for three years during my Honours work in Dr. Langille's lab and more recently as a graduate student. I will always be grateful for his support in my academic endeavours and beyond. Morgan was continually encouraging of me presenting, publishing, and promoting my work through scientific communication across the globe and here in Halifax.

I am also grateful for Dr. Johan Van Limbergen's guidance and support throughout my degrees, for stimulating my interest in inflammatory bowel disease and the medical profession. His receptivity of me observing his clinical care first hand at the IWK opened my eyes to translational science and provided rationale for this work.

An immense thanks goes out to members of the Langille Lab who have broadened my knowledge and ability in ways the classroom never has. Thank you to Dr. Andre Comeau and Gavin Douglas for being open and encouraging of every question I've asked them. Gavin has provided me with significant bioinformatic assistance on several aspects of my work, which I am very grateful for. Thanks goes out to Andre and Dr. Akhilesh Dhanani for their persistent work that makes studies like these possible. To all my past and present lab mates, Emily Lamoureux, Jacob Nearing, Montse Palau de Miguel, thank you for your unwavering support and friendship.

A special thanks goes out to Dr. Jessica Connors from the IWK who collected all the metadata for this study, recruited patients, and managed the MAREEN study. The work of Scott Whitehouse in preparation of the virally enriched samples is highly appreciated, in addition to him training me these protocols. I would like to thank the patients and families of the IWK who participated in the MAREEN study. Your commitment to advancing research is inspiring.

I would like to thank all members of the Department of Pharmacology for their guidance, feedback, and hard work to promote this great Department. A special thank you goes out to Luisa Vaughan and Sandi Leaf for their continual administrative support in keeping our department running.

Finally, an immense thank you goes to my friends, family, and football teammates whose continual support and understanding has helped me through my education.

CHAPTER 1: INTRODUCTION

1.1 CROHN'S DISEASE

Inflammatory bowel diseases (IBDs) are a set of increasingly common disorders characterized by bouts of chronic intestinal inflammation. The two major types of IBD are ulcerative colitis (UC) and Crohn's disease (CD), which have unique clinical presentations in the intestinal tract, yet share symptoms like abdominal pain, diarrhea, and malnutrition (Cho, 2008). Ulcerative colitis and Crohn's disease differ in disease location and pathology, where CD affects any part of the gastrointestinal tract from mouth to anus, often discontinuously, while UC is limited to the colon (Cho, 2008). CD is characterized by granulomatous inflammation that spans transmurally, while the inflammation associated with UC is continuous and is confined to mucosal and submucosal surfaces (Wilson & Russell, 2017). Other histological features more common in pediatric CD versus UC include the presence of patchy ileitis or colitis, fissuring ulceration, stricture and fistula formation (Bousvaros et al., 2007).

Most IBD patients are diagnosed with either CD or UC on the basis of clinical, endoscopic, laboratory, and radiologic guidelines. In 5-23% of pediatric IBD patients, a determinate diagnosis of CD or UC cannot be reached due to overlapping histological features of UC and CD, or initial macroscopic appearance of the disease resembling both subtypes. Over time, most of these patients evolve into a UC or CD consistent diagnosis, but 20-60% of these patients retain a diagnosis of indeterminate colitis (IC) or IBD-unclassified (IBD-U) for up to 10 years post diagnosis (Kirschner, 2016).

1.1.1 IBD EPIDEMIOLOGY

IBD can develop at any age, yet 10-20% of patients are diagnosed in childhood (Wilson & Russell, 2017). UC is the more common subtype of IBD in adults, while CD is more common in pediatric IBD (PIBD) in most regions of the world (Benchimol et al.,

2011). In comparison to adult-onset IBD, PIBD is characterized by a more extensive anatomical involvement and rapid early progression of disease after diagnosis (Van Limbergen et al., 2008). The incidence of PIBD is increasing worldwide and Canada has amongst the highest incidence in the world, especially in the province of Nova Scotia. (Benchimol et al., 2017). A recent analysis of five Canadian provinces by Benchimol et al. (2017) showed that the incidence of PIBD in Canada from 1999-2010 was 9.68 per 100,000 children across the five provinces studied (Alberta, Manitoba, Nova Scotia, Ontario, and Quebec), but reached 15.18 in Nova Scotia. Regions in the northern hemisphere show a similar trend of increased incidence, as shown in countries like Scotland (Henderson et al., 2012) and Scandinavia (Malmborg, Grahnquist, Lindholm, Montgomery, & Hildebrand, 2013; Perminow et al., 2009). Conversely, incidence rates from other regions of Europe and the southern United States are significantly lower (Benchimol et al., 2011). The explanation for higher rates of IBD in northern regions is unknown, but it has been hypothesized that lack of sunlight exposure, vitamin D deficiency, and diet may contribute (Aujnarain, Mack, & Benchimol, 2013). The Nurses' Health Study found that women with high levels of vitamin D plasma levels had a significantly lower risk of developing CD compared with those in the lowest quartile of vitamin D levels (Ananthakrishnan, Khalili, et al., 2012). It is hypothesized that sunlight exposure leading to subsequent increased vitamin D levels may protect against CD by downregulating immune responses (Peyrin-Biroulet, Oussalah, & Bigard, 2009).

1.1.2 DIAGNOSIS OF IBD

There is a vast clinical diversity and indeterminate set of symptoms in children presenting with IBD, although many symptomatic features are common. Abdominal pain, rectal bleeding, diarrhea, and anemia are observed in both CD and UC (Yu & Rodriguez, 2017). Weight loss and poor growth development are common in CD but infrequent in UC. Endoscopic features common to CD include patchy inflammation, active ileitis, fissures, strictures, and cobblestoning (Yu & Rodriguez, 2017). Although endoscopy is the gold standard for IBD diagnosis, less invasive laboratory tests can offer insight into IBD disease

activity. Tests useful for IBD diagnosis include serum iron, C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), and immune markers such as atypical perinuclear anti-neutrophil cytoplasmic antibody (pANCA) and anti-Saccharomyces cerevisiae antibody (ASCA) (Yu & Rodriguez, 2017). A less invasively retrieved biomarker for IBD is fecal calprotectin (FCP), a protein mainly derived from neutrophils. FCP levels are increased with inflammation of the colon (Fagerberg, Lööf, Myrdal, Hansson, & Finkel, 2005), and showed a diagnostic specificity of 94% and specificity of 64% in detecting active mucosal inflammation in 58 PIBD patients (Berni Canani et al., 2008).

1.1.3 PATHOGENESIS OF CROHN'S DISEASE

Although UC and CD present with different clinical manifestations, they share four main underpinnings of disease. These include: Human genetics, immune response, environmental exposures, and the gut microbiota (Wilson & Russell, 2017). The prevailing theory is that IBD is thought to manifest in genetically susceptible individuals that exhibit an exaggerated immune response to intestinal microbes when exposed to environmental triggers (Sartor & Wu, 2017).

1.1.3.1 GENETICS OF IBD

The genetic underpinnings of IBD are complex, however recent findings have uncovered over 230 loci associated with IBD risk (De Lange et al., 2017; Jostins et al., 2012; Liu et al., 2015). A 2012 GWAS in a European cohort of over 75,000 IBD patients and controls found 163 loci that contribute to IBD risk (Jostins et al., 2012), detecting considerable overlap with loci that confer risk for other immune-mediated diseases. Later findings by Liu et al. (2015) revealed an additional 38 loci that confer IBD risk across individuals of varying descent. Most recently, a GWAS of 25,305 individuals identified an

additional 25 susceptibility loci for IBD in genes related to immunity and inflammation regulation (De Lange et al., 2017).

A prevailing paradigm among many researchers is that IBD involves an uncontrolled immune response in those that carry unfavourable alleles at IBD risk loci in interactions with the gut microbiota (Sartor & Wu, 2017). This is substantiated by polymorphisms in genes coding for pattern recognition receptors (PRRs) and cytokines explaining the greatest variance in IBD amongst the 201 identified loci (McGovern, Kugathasan, & Cho, 2015). PRRs play an integral role in the innate immune system, as they generate immune responses against molecular patterns unique to bacteria called pathogen-associated molecular patterns (PAMPs) (Owen, Punt, & Stranford, 2009). It is worth noting that PAMPs can be expressed by microbes whether or not the microbe is pathogenic. PAMPs that induce an inflammatory response include bacterial cell wall components, lipopolysaccharides (LPS), and peptidoglycans. PRRs also recognize ligands from viruses, such as double stranded RNA (Owen, Punt, & Stranford, 2009). The nucleotide-binding oligomerization domain containing 2 (*NOD2*) gene codes for an intracellular PRR, and loss of function mutations in the gene are considered to lead to increased inflammation due to impaired clearance of intestinal bacteria that are harmful to the human host (de Souza & Fiocchi, 2016). The *NOD2* protein plays a role in the immune response by recognizing bacteria that possess muramyl dipeptide (MDP), a constituent of bacterial peptidoglycan of both gram-positive and gram-negative bacteria (Al Nabhani, Dietrich, Hugot, & Barreau, 2017). When stimulated by MDP, *NOD2* activates intracellular pathways that promote inflammation. Other markers identified by GWAS are found in genes related to autophagy and inflammatory signalling pathways such as IL-23-mediated Th cell responses (Uhlig & Muike, 2017). Comparison of IBD risk loci with other inflammatory disorders such as ankylosing spondylitis, psoriasis, and Celiac disease shows considerable overlap amongst pathways and genes (Ellinghaus et al., 2016; Y. R. Li et al., 2015), further pointing to the immune involvement in IBD.

Despite the high number of IBD risk loci found to date, the 163 genetic loci found by Jostins et al. (2012) account for under 14% in disease variance. Furthermore, the concordance rate of IBD between monozygotic twins, which is a representation of the

probability that both twins in a pair will have the disease, is well under 50%, with the least concordance in CD (Halme et al., 2006).

1.1.3.2 IMMUNOLOGY

There is an extensive involvement of the immune system in IBD, as the disease is now considered as one of the immune-mediated inflammatory disorders (IMID) (Wilson & Russell, 2017). In healthy gut mucosa, there is a baseline level of physiologic inflammation driven by an interplay between appropriate responses to immune stimulation and down-regulatory forces in the gut tissue. In IBD, defective barrier function can allow increased permeability of the gut to bacterial or other antigens, leading to an uncontrolled stimulation of intestinal immune cells. In CD, the down-regulation of immune responses is defective, and can result in inflammation and the tissue injury characteristic of CD (Wilson & Russell, 2017). This defective immunity in CD may be due to genetic loci previously discussed, in addition to environmental exposures such as vitamin D deficiency.

There are changes in both innate and adaptive immunity in CD. The mucosal adaptive immune response in the gut involves T helper cell (Th), Th2, Th17, and T regulatory (Treg) pathways. Th1 cytokines like TNF- α promote inflammation and the induction of CD4⁺ T cells to the Th1 phenotype. The success of biologics targeted against TNF- α , like adalimumab and infliximab, highlight the importance of this cytokine to CD pathogenesis. Other immune regulators like the $\alpha 4\beta 7$ integrins coordinate migration of lymphocytes to the gut, and an antibody therapy targeting them has shown proven success in CD and UC (Wilson & Russell, 2017). A distinct development of antibodies against microbial antigens in CD patients has been shown. Increased levels of serum antibodies against *Saccharomyces cerevisiae*, flagellin protein (CBir1), and *E. coli* outer membrane protein C have been demonstrated in CD patients (Lodes et al., 2004; Mow et al., 2004). Studies of the humoral immune response in CD supports the thought that microbial antigens, and the human gut microbiome, are involved in IBD pathogenesis (de Souza & Fiocchi, 2016).

Evidence from epidemiological work indicates a correlation between a decrease in infectious disease, antibiotic use, vaccination, and general improvement in sanitary conditions with an increase in the incidence of autoimmune and chronic inflammatory disease. These findings taken together constitute the basis of the “hygiene hypothesis”, that suggests overly sterile environments may have led to a decrease in the efficiency of immunoregulation, suggesting a link between the human immune system and environment (Rook, 2012). The immunopathogenesis of IBD was recently reviewed in detail by de Souza and Fiocchi (2016).

1.1.3.3 ENVIRONMENT

Environmental factors that confer risk for PIBD are numerous, and include diet, vitamin D exposure, antibiotic use, and lifestyle. Environmental factors that confer risk for IBD include smoking (Mahid, Minor, Soto, Hornung, & Galandiuk, 2006), NSAID use (Ananthakrishnan, Higuchi, et al., 2012), and intake of dietary fibre (Ananthakrishnan et al., 2013). Although epidemiological studies focused on IBD risk and diet have been heterogeneous in design and riddled with limitations (Ananthakrishnan, 2015), an inverse association with a lower intake of fibre and increased IBD risk has been consistent. In 130 newly diagnosed pediatric CD patients, it was shown that increased intake of vegetables, fruits, fish, and dietary fibre protected from CD risk when comparing to 202 controls (Amre et al., 2007). A proposed mechanism for protection from CD by dietary fibre intake may be that the short-chain fatty acid (SCFA) metabolites of soluble fibre from fruits and vegetables inhibit transcription of proinflammatory molecules (Galvez, Rodríguez-Cabezas, & Zarzuelo, 2005). Additionally, soluble plant fibre helps maintain structure of the epithelial barrier and reduces translocation of *E. coli* across intestinal microfold cells of gut-associated lymphoid tissue (Roberts et al., 2010). The same *in vitro* study also found that the food emulsifier polysorbate 80 had the opposite effect on gut permeability. Dietary fat, especially saturated fats, may also play a role in IBD pathogenesis (Ananthakrishnan, 2015). As previously mentioned, vitamin D levels may also play a role in IBD risk. Deficiency of the active metabolite 1,25-dihydroxy vitamin D3 (1,25(OH)₂D₃) or knockout

of the vitamin D receptor is associated with an increased risk of colitis, but administration of 1,25(OH)₂D₃ dampens inflammation and suppresses proinflammatory gene expression (Cantorna, Munsick, Bemiss, & Mahon, 2000; Froicu & Cantorna, 2007; Froicu, Zhu, & Cantorna, 2006).

The exposome, first defined in 2005 (Wild, 2005), refers to the entirety of all environmental exposures of an individual during their lifetime. The exposome factors relevant to IBD include urbanization, exposure to pollution, hygiene, antibiotic exposure, and hypoxia. Epidemiological studies consistently show that IBD is more common in urban centers and in newly industrialized regions including Asia, the Middle East, and South America (Zuo, Kamm, Colombel, & Ng, 2018). A meta-analysis of 40 studies investigating urban habitation and IBD found that the pooled incidence rate ratio for urban compared with rural environments was 1.42 for CD (Soon et al., 2012), suggesting a link between urbanization and acquiring CD. The adoption of a Western diet laden in saturated fats and refined carbohydrates, a less active lifestyle, improved hygiene and increased levels of pollution often accompany the transition from a rural to urban environment (Zuo, Kamm, et al., 2018).

Taken together, many of these environmental factors and genetic predispositions can affect a key aspect of human health: the gut microbiome. The intestinal immune system must balance the response to harmful pathogens while co-existing with commensal bacteria and food antigens that enter the gut (Xavier & Podolsky, 2007). IBD patients have a heightened sensitivity to antigens that reach the gastrointestinal tract, resulting in the ensuing bouts of inflammation that patients suffer from. It is thought that genetic predispositions cause this altered immune response to the gut microbiota (McGovern et al., 2015).

1.1.4 TREATMENTS FOR CROHN'S DISEASE

At present, there is no cure for IBD, but in the last decade more treatment options have arisen. Drugs used for the treatment of IBD include 5-aminosalicylate, antibiotics,

corticosteroids, immunomodulators, biologics, and nutritional therapies (Lahad & Weiss, 2015). In pediatric CD patients, the main objective of treatment is to induce and maintain clinical remission of disease, while optimizing the patient's quality of life and growth with adequate nutrition. This involves the achievement of optimal clinical, laboratory, and histological control of inflammatory disease while minimizing the adverse effects of treatment.

1.1.4.1 PHARMACOLOGICAL THERAPIES

5-aminosalicylic acid (5-ASA) agents like sulfasalazine and mesalamine are considered by many gastroenterologists as a first-line treatment for mild to moderate CD, although meta-analyses in adults show no benefit of the drug over placebo in CD (Akobeng & Gardener, 2005). Only two clinical trials on 5-ASA efficacy in pediatric CD have been carried out and show marginal benefits in improving disease activity (Cezard et al., 2009; Griffiths, Koletzko, Sylvester, Marcon, & Sherman, 1993). 5-ASA agents are rapidly absorbed in the GI tract and are believed to exhibit anti-inflammatory effects. Slow-release formulae allow 5-ASA to reach the terminal ileum and colon in the GI tract where they can exert their anti-inflammatory effects. The exact mechanism of action of 5-ASA is unclear, but it has been shown to inhibit many immune processes *in vitro*, such as inhibition of T cell and natural killer cell cytotoxic cells (Aparicio-Pagés et al., 1990; MacDermott, Kane, Steele, & Stenson, 1986). 5-ASA has also shown to exert inhibitory effects on the function of macrophages and neutrophils (Greenfield et al., 1993; Neal, Winterbourn, & Vissers, 1987). Many of the effects of 5-ASA can be explained by the inhibition of the transcription factor nuclear Factor- κ B (NF- κ B). Sulfasalazine, a 5-ASA agent, was shown to block translocation of NF- κ B into the nucleus, therefore suppressing its transcription-mediating effects (Wahl, Liptay, Adler, & Schmid, 1998). This result was also shown *in vivo* in inflamed mucosa of UC patients (Bantel et al., 2000).

Gut bacteria play a major role in the pathogenesis of IBD (Friswell, Campbell, & Rhodes, 2010), so 5-ASA treatment is often complimented with antibiotic therapy. The

two main antibiotics used for acute CD and flares are Metronidazole (Flagyl ®, 10-20 mg/kg/day) or Ciprofloxacin (20 mg/kg/day), either individually or in combination (Lahad & Weiss, 2015). Antibiotics are frequently indicated when a CD patient presents with perianal disease such as fistulae and abscess. Efficacy of antibiotics appears to be inadequate in the long term, and especially for moderate to severe CD.

If treatment from 5-ASA and/or antibiotics is insufficient, or CD state is more severe, systemic corticosteroids or budesonide may be indicated. Corticosteroids exert a general potent anti-inflammatory effect that is typically effective for inducing remission in CD. Commonly used in pediatric CD are oral prednisolone and its prodrug, prednisone. Prednisone is a synthetic glucocorticoid, which is rapidly hydrolyzed in the liver to prednisolone. The lipophilic molecule prednisolone exerts its effect by passing through the cell lipid membrane, where it then binds to the cytosolic glucocorticoid receptor alpha (cGCR). The ligand-activated cGCR then migrates to the nucleus, where it exerts the majority of its anti-inflammatory effects by inducing synthesis of anti-inflammatory proteins or suppressing transcription of inflammatory genes (Stahn, Löwenberg, Hommes, & Buttgerit, 2007). In CD patients, systemic corticoids are generally effective in treating active disease and achieving clinical remission when first-line therapies are unsuccessful. Minimizing the use of corticosteroids is desirable however, due to their adverse effects including hypertension, glucose intolerance, decreased linear growth, and increased infection risk (Lahad & Weiss, 2015). Upon clinical improvement, the daily corticosteroid dose is reduced slowly until discontinued completely. The induction and tapering time of steroids depends on patient response and physician's discretion.

Following induction of remission with corticosteroids or nutritional therapies, maintenance of remission is frequently carried out with immunomodulatory drugs. Immunomodulators used in pediatric CD include methotrexate, 6-mercaptopurine, and azathioprine. Methotrexate (MTX) is a competitive inhibitor of dihydrofolate reductase, the rate-limiting enzyme in the production of tetrahydrofolate. This decreases *de novo* production of nucleotide bases and interferes with DNA synthesis (Chan & Cronstein, 2013). The anti-inflammatory effects of MTX also include decreased pro-inflammatory cytokine production and lymphocyte apoptosis. Although MTX is highly effective at

disrupting inflammatory cell proliferation in target tissues, it has off-target cell cycle inhibition that can cause harmful adverse effects like bone marrow suppression and teratogenicity. Supplementing with folate can diminish these adverse effects (Turner et al., 2007). Methotrexate is most often administered intramuscularly or subcutaneously, with the effective dose of MTX being 15 mg/m² weekly with a maximal dose of 25 mg. Clinical response to MTX typically occurs within 3-4 weeks and continues for at least 4-6 months. Responders to treatment may alternate to oral dosing with the lowest effective dose being maintained.

Other common immunomodulators used for pediatric CD include the closely related compounds azathioprine (AZA) and 6-mercaptopurine (6-MP) which are classified as thiopurine anti-metabolites. AZA is a prodrug that is rapidly converted to 6-MP following oral administration, mediated by non-enzymatic reductive cleavage by glutathione and related compounds in the intestinal wall, liver, and bloodstream (Atreya & Neurath, 2008). 6-MP is then converted to its active metabolite 6-thioguanine (6-TG) that acts as a purine analog and results in general inhibition of nucleic acid synthesis (Maltzman & Koretzky, 2003). AZA and 6-MP are also potent inducers of apoptosis in stimulated CD4⁺ T cells (Tiede et al., 2003) and can alter lymphocyte function (Nielsen, Vainer, & Rask-Madsen, 2001). Clinical efficacy is expected 8 to 12 weeks after induction of treatment in pediatric CD since AZA and 6-MP are slow acting. A Cochrane review assessing 7 trials of AZA in adults shows that its use is associated with decreased hospitalization, surgery, and decreased use of steroids (Prefontaine, Sutherland, Macdonald, & Cepoiu, 2009). Due to their purine antagonism, severe adverse effects of AZA and 6-MP include bone marrow suppression, pancreatitis, and hepatitis (Lahad & Weiss, 2015). These effects can be worsened if a patient has genetic polymorphisms of thiopurine S-methyltransferase (TPMT). Polymorphisms of TPMT lead to decreased methylation and decreased inactivation of the active 6-MP, which can result in bone marrow suppression. Before starting AZA or 6-MP treatment, TPMT blood test or gene tests are frequently indicated.

In patients that have an inadequate response to the aforementioned treatments, biologic therapies are often indicated. Tumor necrosis factor-alpha (TNF- α) is a cytokine

with diverse pro-inflammatory effects in the GI tract, and has been implicated in the pathogenesis of CD (Lahad & Weiss, 2015). Antibodies against TNF- α neutralize its harmful effects and can interrupt the inflammatory cascade in the gut. Two anti-TNF antibody therapies have been approved for the induction of remission and maintenance of remission in CD: infliximab (IFX, trade name Remicade) and adalimumab (ADA, trade name Humira). Infliximab is a chimeric IgG1 monoclonal antibody raised against TNF- α consisting of human and murine sequences. It is indicated for CD patients with moderate to severe acute CD and patients with fistulizing CD that have had an inadequate response to conventional therapies. IFX is administered via intravenous infusion at week 0, 2, and 6, followed by maintenance infusions every 8 weeks. Adalimumab is a complete human IgG1 monoclonal antibody raised against TNF- α and is typically used to achieve and maintain remission in patients with moderate to severe acute luminal CD. ADA is administered by subcutaneous injection with maintenance injections given at 2-week intervals. Typically, patients respond to anti-TNF therapies within a few days to a month, and continue treatment in the event of sustained remission (Lahad & Weiss, 2015). Adverse effects of anti-TNF biologics are due to its immunosuppressive properties and include risk for serious infections or allergic reactions.

1.1.4.2 EXCLUSIVE ENTERAL NUTRITION (EEN) IN TREATMENT FOR IBD

A recent treatment to induce remission in CD patients is dietary therapy using exclusive enteral nutrition (EEN). As a primary therapy, enteral nutrition (EN) delivers a complete daily nutritional requirement through a liquid formula delivered orally or through nasogastric or gastric tubes over 6-12 weeks (Hansen & Duerksen, 2018). Enteral nutrition (EN) has shown to be efficacious in controlling disease activity, maintaining remission, and addressing malnutrition in PIBD patients. EN formulae vary in their contents, but can be classified as whole protein (polymeric), modified protein, and disease specific (Escuro & Hummell, 2016). Commonly used in pediatric patients are polymeric formulae such as Modulen IBD[®] (Nestle Health Sciences). Modulen consists of 44% carbohydrate, 14%

protein, and 42% fat, with the remaining 3% of contents consisting of various micronutrients (Nestlé Health, 2018). The primary carbohydrate source in Modulen are corn syrup and sucrose, and the major protein constituent is casein from milk. Fat sources include milk fat, medium chain triglycerides (MCT), and corn oil (Nestlé Health, 2018).

The casein protein in Modulen is specially processed to retain transforming growth factor-beta 2 (TGF- β 2), a cytokine with pleiotropic functions in the immune system. Anti-inflammatory effects are exerted by TGF- β 2 through its suppression of T-regulatory cell generation and repair of damaged tissues. The TGF- β cytokines also regulate growth of epithelial tissues and fluid transport across epithelial barriers (Clerici & Matthay, 2003; Fjellbirkeland et al., 2003). Additionally, TGF- β inhibits T-cell proliferation (Kunzmann et al., 2003), and inhibits differentiation of Th1 and Th2 cells (Gorelik, Fields, & Flavell, 2000; Gorelik, Constant, & Flavell, 2002).

Many practitioners prefer EEN as an induction therapy for pediatric CD as it results in similar remission rates as steroids without adverse effects. In a retrospective review of patients at the IWK Children's Hospital (Halifax, NS), it was found that 86.6% of EEN-treated patients achieved remission by week 12, compared with 58.1% of patients treated with corticosteroids (Connors et al., 2017). Furthermore, choice of EEN as induction therapy was associated with long-term avoidance of corticosteroid use over a 6-year follow up period. In 15 Australian children with newly diagnosed CD, inflammatory markers such as ESR, CRP, albumin, and platelet count were significantly reduced after 8 weeks of EEN therapy (Day et al., 2006).

The mechanism of action of EEN is not well characterized, but it is known that it involves the bacterial species that live in the GI tract, collectively called the gut microbiome. Therapy with EEN has shown to modulate the gut microbiome and improve disease activity. Furthermore, EEN may directly affect enterocytes, as it was shown *in vitro* that their inflammatory response was dampened to noxious stimuli if they were incubated with EEN formula (de Jong, Leach, & Day, 2007). Yamamoto et al. (2005) demonstrated that levels of pro-inflammatory cytokines such as IL-1B, IL-6, IL-8, and TNF-alpha are decreased in the ileum and large bowel of those treated with an elemental EEN formula. These findings were replicated in a second cohort using a polymeric EEN formula, in

addition to the finding that EEN leads to mucosal healing of the ileum and colon (Fell et al., 2000).

Treatment plans for pediatric CD are chosen based on disease activity, severity, location, patient growth, and psychosocial condition (Lahad & Weiss, 2015). Therapy for pediatric CD is commonly administered in a step-up approach, meaning patients are treated with milder therapies and given more aggressive drugs if their condition does not improve. When remission is achieved, patients are typically switched to a less harsh therapy with less side-effects. In the event of an inflammatory flare-up, the patient may be reverted back to a more impactful treatment. Recently, pediatric gastroenterologists have debated whether a top-down or step-up treatment plan is better for patients. Physicians used to initiate treatment mainly in response to acute inflammatory flare to improve clinical symptoms. Now, focus has shifted to preventing damage to the intestinal wall and to promote ‘mucosal healing’ (Rogler, 2013). However, the prevention of structural damage in the gut is usually associated with an early use of immunosuppressants and biologics, which are considered to be a more aggressive top-down treatment. Given the extreme heterogeneity of pediatric CD, a one-size-fits-all approach is likely unfeasible (Rogler, 2013). Pediatric CD patients should not be undertreated or remain on steroid treatment for prolonged periods if not needed. Clinicians should also avoid overtreating patients, putting them at risk of adverse effects without sufficient benefit. The goal of treatment for pediatric CD patients remains to induce and maintain clinical remission while optimizing quality of life and growth with adequate nutrition, with a focus on mucosal healing of the gut.

1.2 THE GUT MICROBIOME

The human microbiome is a dynamic microbial community that shares the human body space. Specifically, the gut microbiome refers to the set of organisms that occupy the gastrointestinal tract in a mutualistic relationship. It has been established that the gastrointestinal microbiota plays a major role in human health and disease in a tight connection with the human host (Althani et al., 2015). The gut microbiome was first

referred to as a “forgotten organ” in 2006 (O’Hara & Shanahan, 2006), as it contributes significant metabolic capability and immune regulation to the human host.

Microbial cells in the human body consist mainly of bacteria, but also protozoa, archaea, viruses, and fungi. The bacteria of the microbiome carry out many roles, including interactions with the immune system, vitamin production, and absorption of dietary nutrients (Moens & Veldhoen, 2012). There are at least as many microbial cells inside of and on the surface of the human body as there are somatic cells (Sender, Fuchs, & Milo, 2016), and studies estimate that 500-1000 unique species of bacteria exist in the human body at any given time (Turnbaugh et al., 2007). However, the number of unique genomes in the human microbiome are likely much more numerous than the number of estimated bacterial species (Locey & Lennon, 2016).

Individual human microbiomes have high interindividual variability. Twin studies have shown that some healthy twin pairs share less than 50% of bacterial taxa at the species level (Turnbaugh et al., 2010). Despite differences in the microbiome between people, the intraindividual variability of the gut microbiome is temporally stable in the absence of environmental perturbations to the gut. Furthermore, although taxonomic differences between individuals are vast, the variability in gene composition, and therefore functional capacity of the microbiome is much less variable (Human Microbiome Project Consortium, 2012; Jones et al., 2018).

The majority of colonization of the human microbiome occurs in the early years of life. The time of first colonization is disputed by many, whether it happens in the womb in the female placenta (Pelzer, Gomez-Arango, Barrett, & Nitert, 2017), or during pregnancy where an infant acquires microbes from the birth canal. Most microbiome researchers believe that the womb is a sterile environment, but some evidence suggests the presence of a placental microbiome (Aagaard et al., 2014). A case study looking at the gut microbes of an infant from birth until 2.5 years old demonstrated discrete steps of bacterial succession alongside life events, such as birth and introduction of table food (Koenig et al., 2011). The diversity of the gut microbiome gradually increases and shows chaotic shifts in early childhood, but becomes stable after weaning, leading to microbial stability in the adult gut (Gilbert et al., 2018).

The gut microbiota offers numerous benefits to the human body that complement its own metabolism and function. Bacteria in the large intestine ferment indigestible dietary carbohydrates into short chain fatty acids (SCFAs) and gases (Rowland et al., 2018). The three most abundant SCFAs found in human stool are acetate, propionate, and butyrate. These SCFAs perform unique but vital roles in the human body. Butyrate is a key energy source for colonocytes, has potential anti-cancer activity, and ability to regulate gene expression through inhibition of histone deacetylases (Steliou, Boosalis, Perrine, Sangerman, & Faller, 2012). Propionate plays a role in gluconeogenesis in the liver, and may play a role in satiety signalling in the gut through gut-brain neural circuits (De Vadder et al., 2014). Acetate is the most abundant SCFA in the gut, and is a necessary co-factor and metabolite for bacterial growth in the gut (Rowland et al., 2018). Additionally, certain strains of bacteria can *de novo* synthesize vitamins such as the B-vitamins and vitamin K (LeBlanc et al., 2013).

The microbiome is highly modifiable by environmental factors. A primary modifier of the gut microbiome is diet. Mice studies have shown that dietary changes can explain up to 57% of the variability in gut microbiota composition, with a much lesser influence from host genetics on the gut microbiota (Zhang et al., 2010). Observational studies in humans show that habitual dietary patterns (i.e. omnivorous vs vegetarian/vegan diets) harbor distinct bacterial profiles in the humans consuming them (Healey, Murphy, Brough, Butts, & Coad, 2017). A dietary intervention study by David et al. (2013) further demonstrated that consuming an animal- or plant-based diet significantly shifted the composition of taxa in the human gut. The animal-based diet consisting of meat, eggs, and cheese increased the abundance of bile-tolerant microbes (*Alistipes*, *Bilophila* and *Bacteroides*) and decreased taxa of the phylum Firmicutes that metabolize dietary plant polysaccharides (*Roseburia*, *Eubacterium rectale* and *Ruminococcus bromii*) (David et al., 2014). The plant-based diet was higher in fibre content, which have shown to be beneficial to the gut in both mice and human studies. In germ-free mice inoculated with a synthetic microbiota characteristic of humans, a low-fibre diet promoted expansion and activity of mucus-degrading bacteria. As a result, fibre-deficient mice had an eroded colonic mucus layer, promoting a “leaky” epithelium more easily accessible by pathogenic bacteria (Desai et al., 2016). In these mice, aggressive colitis ensued when

inoculated with the mucosal pathogen *Citrobacter rodentium*. Accordingly, in humans supplemented with dietary fiber, a downregulation of genes associated with mucin degradation were seen in those in the high fibre supplementation group (Tap et al., 2015). Furthermore, supplementation with the prebiotic fibre inulin showed increases of fecal Actinobacteria and *Bifidobacterium*, which are considered to be beneficial bacteria in the gut (Holscher et al., 2015). Taking this evidence in addition other diet studies with the microbiome into consideration, the evidence is overwhelming that fibre intake is beneficial for the gut, and that the microbiome can be rapidly shifted by food intake.

Although antibiotics have been a life-saving medicine for many years, their overprescription is rampant, and has led to the emergence of drug resistance in bacterial communities. Studies of the effects of antibiotics on the gut microbiota composition have generally found decreased levels of bacterial diversity, stereotypic declines and increases in abundances of select taxa, and several antibiotic-specific effects (Modi, Collins, & Relman, 2014). Ciprofloxacin, a broad-spectrum antibiotic commonly used for CD, has profound effects on gut microbiota. Within a few days of ciprofloxacin treatment, Dethlefsen et al. showed that the abundance of approximately a third of bacterial taxa in the gut were affected, and taxonomic richness was decreased (Dethlefsen, Huse, Sogin, & Relman, 2008). Nearly complete recovery of the microbiome was achieved by four weeks after treatment, but some compositional effects lasted up to six months. While antibiotic treatment can be beneficial for some cases of severe CD, the side effects of treatment include infection, increased antibiotic resistance, and intolerance to treatment (Nitzan, Elias, Peretz, & Saliba, 2016).

Non-antibiotic drugs also have an extensive impact on the human gut microbiome. A recent high-throughput screen of 1,197 compounds on 40 representative gut isolates revealed that 27% of non-antibiotic drugs tested exhibited anti-commensal activity (Maier et al., 2018). This demonstrates a potential risk of commonly prescribed non-antibiotic drugs like metformin, protein pump inhibitors, and nonsteroidal anti-inflammatory drugs in promoting antibiotic resistance. In a retrospective analysis of 2700 members of the TwinsUK study, gut microbiota associations were presented for 51 medications. The same study also showed significant associations between the gut microbiota and 38 common diseases (Jackson et al., 2018). Given the vast metabolic and

functional capability of the human gut microbiome, it is no surprise that compositional changes in the microbiome have been linked with a wide array of illnesses from IBD, to cancer, to depression (Gilbert et al., 2018). The human microbiome consists not only of bacteria, but also an expansive community of viruses that have only begun to be discovered.

1.3 THE GUT VIROME

While most studies have focused on the bacterial portion of the microbiome, there is a diverse and abundant community of viruses that also live in the gut environment. The advent of high-throughput next generation sequencing (NGS) technologies has permitted scientists to unravel the complexities of the human gut virome, which may be defined as the total population of viruses associated with the underlying gut microbial community (Ogilvie & Jones, 2015). The gut virome appears to predominantly consist of viruses that infect bacteria, called bacteriophage, or phage (Breitbart et al., 2003; Minot et al., 2011). The human gut phageome refers to the total population of bacteriophage in the gut virome (Manrique et al., 2016). Phages are the most abundant infectious entity with an estimated 10^{31} phages on the planet, and outnumber their bacterial hosts 10:1 in most habitats studied (Comeau et al., 2008).

Phages can lyse and kill host bacteria, and therefore play a crucial role in modulating bacterial community structure and function in the environment, and in animal microbiomes. Phages may facilitate gene transfer between bacterial taxa through transduction, or themselves encode functions that may benefit their hosts. These functions derived from phages to their bacterial hosts include toxin synthesis, virulence factor production, and metabolism related genes (Ogilvie & Jones, 2015). Phage-host relationships are dynamic interactions that affect the evolutionary rate of bacterial hosts (Paterson et al., 2010). They can integrate into a bacterial host's genome through the process of lysogeny, by which viral DNA can propagate through bacterial replication. Lytic phages replicate through the lytic cycle, where new viral progeny are released

following cell lysis, shortly after cellular infection (Mirzaei & Maurice, 2017). Phages have a finite host range with some that can infect several bacteria, but host range is an unfixed property of each species of bacteriophage that can evolve over time with unexpected shifts (Ross, Ward, & Hyman, 2016). While most phages are host-specific, many different phages can infect the same bacteria.

The phages of the gut consist of both DNA and RNA viruses, but the majority belong to the single-stranded or double-stranded DNA virus families. The most abundant phage families are the tailed double-stranded DNA (dsDNA) phages found in the order Caudovirales (Myoviridae, Podoviridae, and Siphoviridae families), and the single-stranded DNA (ssDNA) phages of the Microviridae family (Reyes et al., 2015). The RNA members of the gut phageome seem to be transient and likely originate from dietary sources such as plants (Zhang et al., 2006). Like the bacterial microbiome, it has been found gut phage populations are individually unique, stable over time (Manrique et al., 2016), and highly dynamic in the early periods of life (Breitbart et al., 2008; Koenig et al., 2011). Furthermore, gut phage populations from relatives and household members are more similar than those from unrelated individuals (Reyes et al., 2010).

Dutilh et al. (2014) demonstrated the presence of a novel phage named crAssphage (cross-Assembly phage), discovered through cross-assembly of virome samples from 12 individuals. The crAssphage genome represents up to 90% of gut viruses in some individuals and was discovered through bioinformatic approaches rather than culturing. crAssphage was recently cultured and was found to infect *Bacteroides intestinalis* through reductive *in vitro* culture techniques (Shkoporov, Khokhlova, et al., 2018). Despite the discovery of crAssphage, it is not overly abundant in some individuals, and the majority of the human gut virome remains uncharacterized.

Most phages in the gut have not been characterized beyond the family level, largely owing to limitations of isolating phage particles and annotating their genomes with known sequences. Due to a lack of robust markers for phage identification, low sequence annotation is a common problem in gut virome studies (Mirzaei & Maurice, 2017). Given cost- and time-prohibitive steps, RNA phages are often not considered in gut phage studies, and are also largely uncharacterized (Mirzaei & Maurice, 2017). New tools targeted at viral discovery and annotation have been recently developed to better define

these populations (Ren, Ahlgren, Lu, Fuhrman, & Sun, 2017; Roux, Enault, Hurwitz, & Sullivan, 2015; Shkoporov, Ryan, et al., 2018). Changes in the gut virome have been associated with colorectal cancer (Nakatsu et al., 2018), success of fecal transplantation treatment for *Clostridium difficile* infection (Zuo, Wong, et al., 2018), and IBD (Norman et al., 2015).

1.4 ANALYSING THE MICROBIOME

1.4.1 DNA SEQUENCING

Advances in DNA sequencing technologies have allowed researchers to first unravel the complexities of the human genome, and more recently the microbiome of complex communities from the deep ocean to the human gastrointestinal tract. There are many considerations when designing a microbiome study, including type of sequencing technology used, downstream bioinformatic analyses, and sample preparation. Before modern NGS technologies, early procedures like Sanger sequencing were time-intensive and expensive. Furthermore, primitive sequencing technologies were often limited to culturable bacterial species, and approximately 40% species in the human microbiome are uncultivable (Browne et al., 2016). New sequencing techniques allow us to characterize whole microbial communities, including bacteria, eukaryotes, and viruses.

The massively parallel, high-throughput sequencing offered by Illumina NGS technologies allows researchers to sequence the thousands of bacterial species that live in the gut. Illumina sequencing platforms use sequencing by synthesis (SBS) chemistry, which involves the simultaneous amplification of thousands of single-stranded DNA fragments on a flow cell. The sample is prepared by random fragmentation of the DNA, followed by adapter ligation on both ends of the DNA fragments. These adapters bind to complementary sites on the Illumina flow cell. Each fragment is amplified into distinct clusters for sequencing. The Illumina SBS technology uses a reversible-terminator method that detects single nucleotide bases as they are added to DNA template strands from the sample being analyzed. Each correct nucleotide added to the template emits a

distinct fluorescent signal according to the identity of the base being added (Illumina, 2017). Barcodes can be added to adaptor sequences, allowing for pooling of multiple samples on a single sequencing run, called ‘multiplexing’.

For analyzing the bacterial portion of the gut microbiome, the bacteriome, either a marker gene analysis or metagenomic approach can be used. Marker gene sequencing uses PCR primers that target the 16S ribosomal RNA (rRNA) gene in bacteria. The 16S rRNA gene contains conserved regions that can be bound by primers, which flank highly variable regions that can be used for identification of bacterial taxonomy (Knight et al., 2018). The 16S rRNA gene sequencing approach is rapid and cost-effective but often limited to genus level taxonomic resolution. Another approach is metagenomic sequencing (MGS), which is the sequencing of all microbial genomes in a sample. Metagenomics captures all DNA present in a sample, including viral and eukaryotic genetic material. MGS provides a more detailed account of genomic and taxonomic information than marker gene sequencing on its own, but is more expensive to prepare, sequence, and analyze (Knight et al., 2018). If a sample is sequenced to a sufficient depth (the number of sequencing reads per sample), taxonomic resolution to species or strain level, or assembly of whole microbial genomes from short DNA reads are possible. Furthermore, functional information about the microbial community can be identified since all genes in the sample are sequenced. However, there are bioinformatic approaches like PICRUSt that can infer the functional profile of a community based on its 16S rRNA genes (Langille et al., 2013).

1.4.2 BIOINFORMATIC APPROACHES

For marker gene analyses of bacteria, the computational method of oligotyping resolves Illumina sequencing data into exact sequence features called amplicon sequence variants (ASVs). Algorithms like Deblur (Amir et al., 2017) and DADA2 (Callahan et al., 2016) output a table of ASVs and their counts per sample. Next, bacterial taxonomy is assigned to ASVs using machine learning approaches such as the RDP classifier (Q. Wang, Garrity, Tiedje, & Cole, 2007). The popular microbiome analysis package QIIME

(Caporaso et al., 2010) provides support for ASV analysis, taxonomy assignment, and downstream analyses in a streamlined bioinformatic framework.

Viruses lack a universal marker like the 16S rRNA gene, therefore untargeted metagenomic sequencing approaches are needed to study the virome. Once a sample of either stool or intestinal biopsy is metagenomically sequenced, there are a variety of bioinformatic methods one can choose to analyze the virome. A reference-based approach involves mapping non-assembled sequencing reads to a reference database. Tools like BLAST (Altschul, Gish, Miller, Myers, & Lipman, 1990) and DIAMOND (Buchfink, Xie, & Huson, 2015) align sequencing reads to a reference database to provide information on which genes are present a sample. Output from DIAMOND can be input into MEGAN, which performs taxonomic and functional binning of the reads in the file, so that taxonomy of the sample can be easily interpreted (Huson et al., 2016). While this approach works well for assessment of bacterial taxonomy, analysis of viral reads in this manner is limited by reference databases that are incomplete and do not represent the full spectrum of viruses in the gut. Previous studies using a reference based approach have ranged from 14% to 87% of sequencing data being mapped to reference sequence databases (Minot et al., 2011; Norman et al., 2015). This has warranted the development of reference-independent programs for viral discovery.

The tools VirSorter and VirFinder have been developed to predict viral sequences from metagenomic reads. Taking assembled metagenomic reads as input, VirSorter predicts viral contigs using reference-dependent and -independent approaches (Roux et al., 2015). VirSorter detects viral regions through computation of six metrics, including: presence of viral hallmark genes, enrichment in uncharacterized genes, enrichment in short genes, and depletion of strand switching (the change of coding strand between two consecutive genes). VirFinder's approach is exclusively reference-independent, using k -mer frequency profiles (frequency of nucleotide sequences of length k) to discern viral sequences from host. VirSorter and VirFinder require the use of assembled sequencing reads produced by an assembler like metaSPAdes that assembles short reads into longer sequences, called contigs (Nurk, Meleshko, Korobeynikov, & Pevzner, 2017). These contigs can be binned into individual putative genomes using tools such as MetaBat (Kang, Froula, Egan, & Wang, 2015) in an effort to uncover viruses that may not be

present in databases. A recently released tool MARVEL (Metagenomic Analysis and Retrieval of Viral Elements) claims to have improved performance over both VirSorter and VirFinder, using a machine learning approach for viral discovery (Amgarten, Braga, da Silva, & Setubal, 2018). Difficulties remain in trying to identify viruses from MGS reads due to their vast genetic diversity and uncharacterized databases. However, bioinformatic methods are continually improving to optimize characterization of the virome.

1.5 THE MICROBIOME IN IBD

While the microbiome interacts with the central nervous system, liver, and other organ systems, its cooperative relationship with the intestinal epithelium is the most important in IBD pathology. The diversity of microbes in the gastrointestinal tract, or the abundance of distinct bacterial organisms, has been linked to several health outcomes. A reduced microbial diversity in addition to the presence or absence of certain bacterial species has been linked with IBD in both adult and pediatric analyses (Gevers et al., 2014; Kostic, Xavier, & Gevers, 2014). The role of microbial diversity in IBD pathology is substantiated by the use of antibiotics increasing the risk of ensuing IBD development in children, likely due to the depletion of bacterial diversity in the gut (Kronman, Zaoutis, Haynes, Feng, & Coffin, 2012). Patients with CD have circulating antibodies against several commensal bacterial antigens that are not present in healthy controls or UC patients (de Souza & Fiocchi, 2016). This finding is a factor in shifting the microbial contents of the gut in those with CD.

These structural imbalances in the gut, or dysbioses, have been well documented (Kostic et al., 2014). Several reports show a decrease in gut microbial diversity, known as alpha diversity or species richness in those with CD (Dicksved et al., 2008; Manichanh et al., 2006). While alpha diversity analyzes the number of species within a sample, beta diversity evaluates the bacterial composition between samples (Whittaker, 1972). Both of these measures, alpha and beta diversity, have shown to differ between CD patients and

controls (Harry Sokol, Lay, Seksik, & Tannock, 2008). Evidence of atypical gut microbiota has been shown in UC patients, however to a lesser degree than for CD (Andoh et al., 2011; de Souza & Fiocchi, 2016). The functional implications of this drop in diversity involve deleterious effects to metabolic functions such as reduced amino acid biosynthesis, increased oxidative stress, and increased toxin secretion (Kostic et al., 2014).

Although the role of the bacterial portion of the human gut microbiome in IBD has been well established, the same cannot be said for that of the gut virome. Changes to the gut virome has been associated with IBD in several adult and pediatric cohorts, but a causative role has not yet been established. A role for the gut virome in CD was first established when Lepage et al. found that CD patients (n = 19) had significantly greater counts of virus-like particles (VLPs) in their colonic biopsy samples compared to healthy controls (n = 10) (Lepage et al., 2008). Preliminary studies using metagenomic sequencing in small cohorts could not fully resolve differences between IBD patients and controls on the basis of viral community composition or abundance (Wagner et al., 2013a; Wang et al., 2015a). Perez-Brocal et al. observed a lower diversity but a greater variation in viral communities of CD samples compared with controls, but could not resolve differences between study groups based on VLP composition or abundance (Pérez-Brocal et al., 2013). The same research group later again showed no significant differences between the viral communities of CD patients and healthy controls, but did construct networks between viruses and bacteria for the first time in CD (Pérez-Brocal et al., 2015a).

The largest gut virome study to date was carried out by Norman et al. on three independent cohorts of IBD patients (n = 102) and healthy household controls (n = 54). They found an increased richness of Caudovirales phage in IBD patients compared to controls (Norman et al., 2015). Specifically, phages that infect *Lactococcus*, *Lactobacillus*, *Enterococcus*, *Clostridium*, and *Streptococcus* bacteria were shown to be associated with disease in the UK cohort studied. This increase in phage richness in IBD patients was associated with a corresponding decrease in bacterial diversity and richness. This inverse relationship between viral and bacterial richness was consistent across cohorts, but specific relationships between individual bacteriophages and individual

bacteria taxa could not be established. The researchers used a reference-dependent approach, where they mapped metagenomic sequencing reads to a database of known viral sequences. Only 15% of reads were able to be assigned viral taxonomy, warranting the use of improved bioinformatic techniques to uncover the taxonomy of unassigned reads. The aforementioned study did not analyze pediatric subjects, but a recent study by Fernandes et al. examined the virome of 12 children with IBD and 12 controls. Unlike the Norman et al. study, they did not find any measure of phage richness, diversity, or composition associated with disease (Fernandes et al., 2018). They found a greater relative abundance of Caudovirales compared to Microviridae phages in all subjects studied, and a greater richness of Microviridae in controls (n = 12) compared to CD patients (n = 7). These studies of the pediatric CD virome were not longitudinal in nature like the MAREEN study, missing the dynamic interaction of phages and bacteria that may take place over a long period of time. The aforementioned studies also did not take advantage of new reference-independent tools such as VirSorter or VirFinder, therefore some viral taxonomy may have been missed during analysis.

As previously discussed, exclusive enteral nutrition (EEN) is an effective therapy for induction and maintenance of remission in pediatric CD, but a causative mechanism for its action has yet to be elucidated (Hansen & Duerksen, 2018). Many studies have pointed to the gut microbiome as a mediator for the benefit of EEN in CD treatment, which was recently reviewed in detail by MacLellan et al. (2017). The emerging consensus from studies on EEN and the CD microbiome shows that EEN causes a general reduction in bacterial diversity in the gut and affects community-level metabolic functions, while improving disease activity. Paradoxically, this modulation involves a decrease in bacterial taxa thought to be beneficial for the gut (Dunn et al., 2016; Quince et al., 2015). A decrease in gut microbial diversity during EEN is thought to be a general consequence of treatment and is theorized to constitute a “reset” of the microbiome (Gerasimidis et al., 2014). Diversity has shown to be restored when patients resume their normal diet. An understanding of how EEN modifies the gut microbiome to improve disease status in children with CD would provide insight into the etiology of CD and development of microbiome-focused treatments.

1.6 OBJECTIVES

The main objective of this project was to characterize the role of viruses, specifically bacteriophages, in the gut microbiome of children with Crohn's Disease while receiving treatment. There has yet to be a study of this magnitude and duration analyzing the virome of children undergoing treatment for CD, specifically with EEN. First, we sought out to develop methods to identify viruses from metagenomic sequencing data of virus-like particle (VLP) preparations. Our second objective was to determine how the gut virome changes with treatment progression and clinical characteristics. Our final objective was to analyze how bacteria and viruses interact longitudinally. This research will provide a detailed and comprehensive analysis of the role of the human gut virome in CD and hopefully lead to effective treatment strategies in CD.

CHAPTER 2: MATERIALS AND METHODS

2.1 MAREEN STUDY RECRUITMENT

Twenty-two patients followed at the GI clinic at the IWK Health Center (Halifax, NS) consented to the MAREEN (A metagenomic approach to diagnosis, induction and maintenance of deep remission following exclusive enteral nutrition in pediatric Crohn's Disease) study. All patients were treated with EEN therapy via nasogastric/gastric tubing for at least 12 weeks. Patients provided a stool sample at weeks 4 and 8 post-induction of EEN treatment, and then every 12 weeks for up to 96 weeks. Stool samples were stored at -80° C following collection. Clinical data was collected at normal clinical follow-ups at the GI clinic every 12 weeks. All clinical metadata including FCP, disease activity, and serology results were collected from patient medical records. Clinical remission was defined in one of two ways: an FCP level of < 300 µg/g, or a weighted pediatric Crohn's Disease index (wPCDAI) < 7.5. The wPCDAI measures incorporates 8 symptomatic traits in addition to 3 laboratory values to derive a weighted score that represents disease activity at examination time (Turner et al., 2011). These clinical data were collected at exam time.

2.2 VIRUS LIKE PARTICLE ENRICHMENT PROTOCOL

Viral DNA was purified from 134 stool samples from twenty-two patients. A blank run negative control was also carried out in addition to a preparation of two EEN formulae used by the patients: Modulen and Isosource (Nestle Health Sciences). We used a protocol adapted from Norman et al. (2015). First, 150 mg of stool was homogenized in a sterile bead tube containing 650 µL of saline magnesium (SM) buffer (0.1 M NaCl, 0.008 M MgSO₄•7H₂O, 0.002% gelatin, 0.05 M Tris pH 7.5) by vortexing for 5 minutes. Stool suspensions were then cleared by centrifugation at 2500 x g for 5 minutes to remove debris, followed by centrifugation of the supernatant at 5000 x g for 15 minutes. Samples were passed through a 0.45 µM filter followed by a 0.22 µM filter to remove

cellular debris, which were each pre-treated with 500 μ L of SM buffer. The filtrate was treated with lysozyme (10 mg/mL at 37°C for 30 minutes) followed by chloroform (88 μ L at 21°C for 10 minutes) to degrade any bacterial or host cell membranes. Samples were centrifugated at 2500 x g for 5 minutes. The aqueous phase was treated with 50 μ L of 0.2 U/mL Turbo DNaseI (Invitrogen, Catalog #: AM2238) for 60 minutes at 37° C. To inactivate DNase, samples were incubated at 65° C for 15 minutes, then treated with 30 μ L of 10% SDS and 3 μ L of 20 mg/mL Proteinase K to lyse viral particles. Following treatment with CTAB plus 0.5 M NaCl solution (65° C for 10 min), viral nucleic acid was isolated using an equal volume of TRIzol (Invitrogen, Catalog #: 15596026). The aqueous fraction was washed with an equal volume of chloroform and spun at 8000 x g for 5 minutes, purified on a DNeasy column (QIAGEN, Catalog #: 69504), then concentrated to a volume of 40 μ L (Microcon DNA fast flow column, Milipore, Catalog #: MRCF0R100). Viral DNA was amplified for 18 hours using Phi29 polymerase (GenomiPhi V2 kit, GE Healthcare, Catalog #: 45-001-222). Two independent reactions using 2 μ L template DNA were performed for each sample then pooled and diluted for sequencing.

2.3 LIBRARY PREPARATION AND SEQUENCING

One nanogram of each purified PCR product sample was then subjected to Nextera XT (Illumina) library preparation, as per the manufacturer's instructions except clean-up and normalization were completed using the Charm Just-a-Plate Purification and Normalization Kit (Charm Biotech). Complete libraries were then pooled and sequenced in a portion of a 150+150 bp PE NextSeq run (Illumina Hi-Output v2 300 cycle kit).

2.4 VIRAL SEQUENCE PROCESSING AND ANALYSIS

Analysis of metagenomic sequencing reads was carried out on a laboratory server. The Microbiome Helper metagenomic workflow (Comeau, Douglas, & Langille, 2017) was optimized for analysis of reads from virally enriched particles and can be found on

GitHub (https://github.com/LangilleLab/microbiome_helper/wiki/Viromics-Pipeline).

The KneadData tool was used for sequence pre-processing, which uses Trimmomatic (Bolger, Lohse, & Usadel, 2014) to remove low quality sequences and Bowtie2

(Langmead & Salzberg, 2012) to screen out human and PhiX contaminant sequences.

Trimmomatic removes reads smaller than 50 base pairs, and those with low quality scores (PHRED $q < 20$).

Quality filtered reads were cross-assembled using metaSPAdes (Nurk et al., 2017) using k-mer lengths of 21, 33, 55, 77, 99, and 127. Viral contigs were predicted from assembly using VirFinder (Ren et al., 2017) and VirSorter (Roux et al., 2015). Contigs predicted by VirFinder with a p-value < 0.05 and false positive rate < 0.1 were combined with contigs predicted by VirSorter. The combined VirSorter and VirFinder file was filtered for redundancy with CD-HIT (Fu, Niu, Zhu, Wu, & Li, 2012) to remove duplicate contigs. Contigs predicted by VirSorter and VirFinder were together used as input for genome binning with MetaBat2 (Kang et al., 2015). Genome bins that contained conserved single copy bacterial genes identified by CheckM (Parks, Imelfort, Skennerton, Hugenholtz, & Tyson, 2015) were removed. Abundances of viral genome bins were identified by mapping sample reads to viral contigs in each genome bin using Bowtie2 (version 2.3.0, default parameters). Read counts were summed and normalized based on nucleotide length of bin and total number of reads in a sample.

We also conducted a reference-dependent analysis of the virome using DIAMOND (Buchfink et al., 2015). Quality filtered reads from each sample were searched against the RefSeq viral protein database using DIAMOND BLASTx. Briefly, this program compares the six-frame translation products of the nucleotide query against the viral protein database. Reads were assigned taxonomy using MEGAN Community Edition (v 6.12.6) (Huson et al., 2016). Absolute read counts were exported from MEGAN, normalized by sequencing depth, and imported into STAMP (Parks, Tyson, Hugenholtz, & Beiko, 2014) and QIIME2 (Bolyen et al., 2018) for statistical analyses and diversity calculations.

The aforementioned virome bioinformatic analyses were carried out on virally-enriched samples ($n = 134$) and normally prepared metagenomic samples ($n = 141$) to compare viral annotation of both sample preparation protocols. To assess proportion of

bacterial sequences in each dataset, reads were mapped to the SILVA 16S rRNA database (Quast et al., 2012) using sortmeRNA (Kopylova, Noé, & Touzet, 2012).

2.5 16S WETLAB PROTOCOL

DNA was isolated from 139 fecal samples for bacterial identification using the Stool DNA Isolation Kit (NORGEN Biotek). Briefly, 200 mg of stool was vortexed in a bead tube with lysis buffer for 5 minutes to homogenize the sample, then centrifuged for 2 minutes at 20,000 x g. Non-DNA organic and cellular debris are precipitated following centrifugation. The clean supernatant was incubated for 10 minutes on ice with a binding buffer, then centrifuged for 2 minutes at 20,000 x g. Clean supernatant was combined with an equal volume of 70% ethanol and applied to a spin-filter column with a silica membrane. After being bound to the column and washed, purified DNA was eluted in a low-salt solution and stored at -20° C.

Variable regions V4-V5 of the bacterial 16S ribosomal rRNA gene were amplified from extracted DNA using PCR conditions and custom primers as described in the Microbiome Helper protocol (Comeau et al., 2017). The forward and reverse primers used Nextera Illumina index tags and sequencing adapters fused to the 16S sequences. Each sample was amplified with a different combination of index tags to allow for sample identification after multiplex sequencing. Following amplification, paired-end 300+300 bp v3 sequencing was performed for all samples on the Illumina MiSeq.

2.6 16S RIBOSOMAL RNA SEQUENCING ANNOTATION

Analysis of 16S sequencing data was carried out on a laboratory server using the Microbiome Helper workflow (Comeau et al., 2017) specific to 16S analysis obtained from GitHub (https://github.com/LangilleLab/microbiome_helper/wiki/Amplicon-SOP-v2). First, primer sequences were removed from sequencing reads using cutadapt (v 1.14) (Martin, 2011) and primer-trimmed files were imported into QIIME2 for microbiome analysis (Bolyen et al., 2018). Forward and reverse paired-end reads were joined using

VSEARCH (v 2.9.0) (Rognes, Flouri, Nichols, Quince, & Mahé, 2016), and input into Deblur (Amir et al., 2017) to correct reads and obtain amplicon sequence variants (ASVs). ASVs that had a frequency of less than 0.1% of the mean sample depth were removed from downstream analysis. MAFFT (v 7.407) (Kato & Standley, 2013) was used to build a multiple-sequence alignment of ASVs, and input into FastTree (Price, Dehal, & Arkin, 2009) to construct a phylogenetic tree. Taxonomy was assigned to ASVs using the SILVA rRNA gene database (Quast et al., 2012) and the “feature-classifier” option in QIIME2. Diversity metrics and plots were generated using QIIME2. To determine whether levels of individual bacterial taxa were significantly different between clinical characteristics, the graphical software package STAMP was used (Parks et al., 2014).

2.7 STATISTICAL METHODS

All microbiome diversity metrics were generated using QIIME2 (Bolyen et al., 2018). Species richness (alpha diversity) was calculated for bacteria by counting number of observed ASVs in each sample and calculated for viruses using Shannon diversity. Shannon diversity accounts for both the abundance and evenness (distribution across samples) of the feature in question. Differences in alpha-diversity were calculated between groups using the Kruskal-Wallis non-parametric statistical test.

Beta diversity (community composition) was assessed in bacteria using weighted UniFrac principal coordinate analysis (PCoA) plots in QIIME2. Linear equations are fitted to the data in a way that each explains the most amount of community variation possible (principal components). The three principal components (PCs) that explain the most variation are assigned to each axis of a three-dimensional plot. Each individual sample is assigned a value based on their PCs and plotted in 3D space, with their relative proximity to other samples representing sample similarity. For viral populations, a non-phylogenetic measure of composition was used, called Bray-Curtis dissimilarity. Bray-Curtis values are bounded between 0 and 1, where 0 means the two sites have identical taxonomic composition, and 1 means they share no taxa in common. Bray-Curtis values can be used as input into a PCoA and represented as a distance between samples.

The Adonis test in QIIME2 tests for differences in beta-diversity between groups of samples. It is a non-parametric multivariate analysis of variance that compares the abundance of each taxon in a sample to its abundance in other samples. The R^2 value represents the effect size and indicates the percent variation explained by the tested variable (Anderson, 2001).

Tables of viral and bacterial taxonomy were exported from QIIME2 and input into STAMP (Parks et al., 2014). STAMP can compare groups of samples based on their metadata using robust statistical associations. To test for difference in individual taxa between groups of samples, a Welch's two-sided t-test with Benjamin-Hochberg false discovery rate (FDR) testing was used. The tool MaAsLin (Multivariate Association with Linear Models) was used to identify associations between microbial taxa and clinical metadata (Morgan et al., 2012). MaAsLin uses boosted, additive linear models between metadata and microbial abundances to select metadata that show potential for association with each microbial taxon inputted. These metadata are then used in a linear model to determine associations between microbial abundances and clinical outcomes. Bacterial taxa that showed associations with clinical remission status were then chosen for further investigation with their respective viral taxa. To discover associations between viral and bacterial taxa, a custom R script to conduct Spearman correlations between all viral and all bacterial features was conducted. To search for documented bacteriophage-host interactions, the Virus-Host DB (Mihara et al., 2016) was manually searched.

2.8 MACHINE LEARNING

To identify microbial taxa important in determining changes of FCP levels, we used a form of machine learning called supervised learning, where microbial features (either viral or bacterial taxa) are used to predict the clinical metadata of a sample. The machine learning algorithm Random Forests (RF) was used to determine the variable importance of microbial taxa in predicting FCP levels of each sample (Breiman, 2001). Microbial features that had non-zero counts in less than 10% of samples were removed to improve statistical power of the randomForest model. Pre-processed data were then input into R for analysis with the randomForest package (Liaw & Wiener, 2002). The

randomForest package outputs a variable importance metric of each significant microbial feature in classifying metadata. Microbial data were regressed against continuous FCP levels to assess which taxa are most important in determining FCP levels. The machine learning tutorial found on Microbiome Helper was used (Comeau et al., 2017).

CHAPTER 3: THE GUT VIROME

In this study, we investigate the gut microbiome of children undergoing treatment for Crohn’s Disease as part of the MAREEN study. In total, 22 patients were recruited for the study, with 21 providing stool samples for analysis. Patients ranged in age from 8 to 15 years old at baseline, and the cohort was 76% male. Patient demographics are shown in Table 1.

Table 1 MAREEN Study cohort information

Study ID	Sex	Age at baseline	wPCDAI	Baseline Calprotectin (µg/g stool)	Samples Sequenced
MRN-001	Male	13.2	0	2018	6
MRN-002	Male	15.2	35	6555	3
MRN-003	Male	12.1	75	6000	4
MRN-004	Male	14	50	3781	7
MRN-005	Male	10	10	4407	5
MRN-007	Male	14.7	80	6000	8
MRN-008	Male	11.3	40	2292	5
MRN-009	Male	14.7	50	745	6
MRN-010	Male	10.4	55	2745	8
MRN-011	Male	10.1	27.5	3906	8
MRN-012	Female	12.8	80	n/a	1
MRN-014	Male	11.5	37.5	3844	8
MRN-015	Male	13.7	15	2874	1
MRN-016	Female	10.2	45	n/a	8
MRN-017	Female	10.9	62.5	2691	9
MRN-018	Female	14.1	72.5	4663	2
MRN-019	Male	12.9	30	2012	9
MRN-020	Male	8.4	30	746	7
MRN-021	Female	10.7	42.5	n/a	5
MRN-022	Male	12	57.5	5801	3
MRN-023	Male	14.6	32.5	5496	8

3.1 VALIDATION OF VLP PREPARATION

We metagenomically sequenced virally-enriched DNA of 134 stool samples from 22 MAREEN study patients, however 2 samples were discarded from further analysis due to failed sequencing. Illumina NextSeq sequencing produced an average of 5,383,847 reads on average after quality filtering steps in the remaining 132 samples. An additional 141 samples that were not virally enriched were also sequenced on the NextSeq platform, producing an average of 5,304,747 reads per sample. As a proxy for the viral enrichment, the bacterial proportion in comparison to total reads was obtained by mapping to the SILVA 16S rRNA gene database. An average of 0.73% of virally-enriched reads and 0.46 % of normally prepared metagenomes mapped to the database. The virally-enriched samples had significantly more reads mapping to the 16S rRNA bacterial database ($p = 0.013$, T-test).

Sequencing reads were then searched against the RefSeq viral protein database. An average of 4.2% unique sequencing reads mapped to the database in the virally-enriched samples, ranging between 0.09 to 36.5% (**Figure 1**). The normally prepared metagenomes mapped 1.3% on average to the viral protein database, ranging between 0.30% to 17.76%. The VLP prepared samples had a significantly greater amount of its sequences virally annotated compared to the normal metagenomes ($p = 8.07 \times 10^{-7}$).

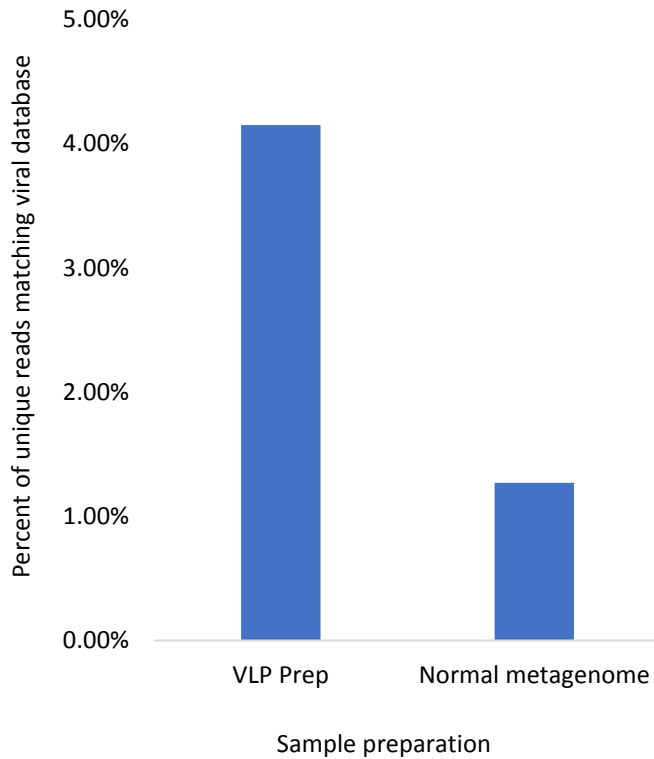


Figure 1 Virally-enriched samples show a greater percent read mapping to viral protein database. Samples enriched for viral DNA sequencing showed a significantly greater mapping to the RefSeq viral protein database compared to non-enriched samples ($p = 8.07 \times 10^{-7}$).

Three controls were processed to ensure quality of sample preparation and sequencing. A blank run (i.e. negative control) consisted of a full VLP preparation without the addition of any stool. Of the 1,884,016 reads sequenced from the blank run sample that consisted of only virome protocol reagents and no stool, 17.1% mapped to the RefSeq viral database. Most of the sample was dominated by *Burkholderia* virus BcepF1 (94.1%) (**Figure 2**).

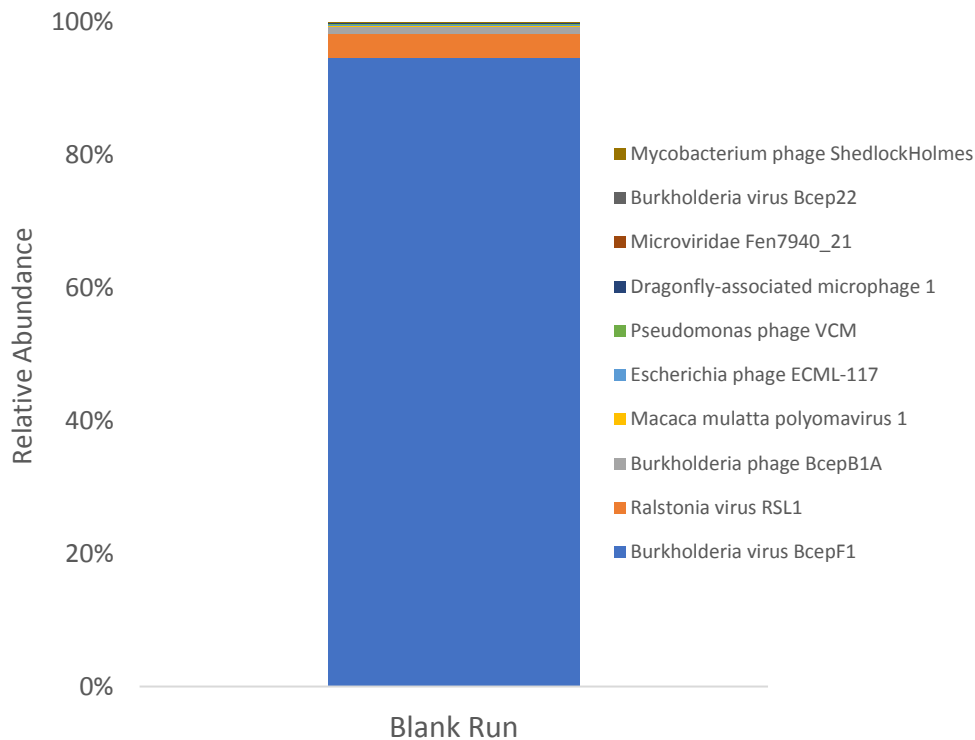


Figure 2 Stacked bar chart depicting relative abundance of top 10 most abundant phages in a blank run sample. Metagenomic sequencing was performed on DNA from a VLP preparation without any added stool. Reads were mapped to the RefSeq viral database and input into MEGAN to classify viral taxonomy. The sample was dominated by reads from *Burkholderia virus BcepF1*, suggesting this virus may be a contaminant.

VLP preparations were carried out on two samples of EEN formula: Modulen and Isosource. The DNA extracted from Isosource did not successfully sequence, but DNA extracted from Modulen produced 3,109,164 sequencing reads, 1.2% of which mapped to the RefSeq viral database. Similar to the blank run, the EEN sample was dominated by *Burkholderia virus BcepF1* (**Figure 3**).

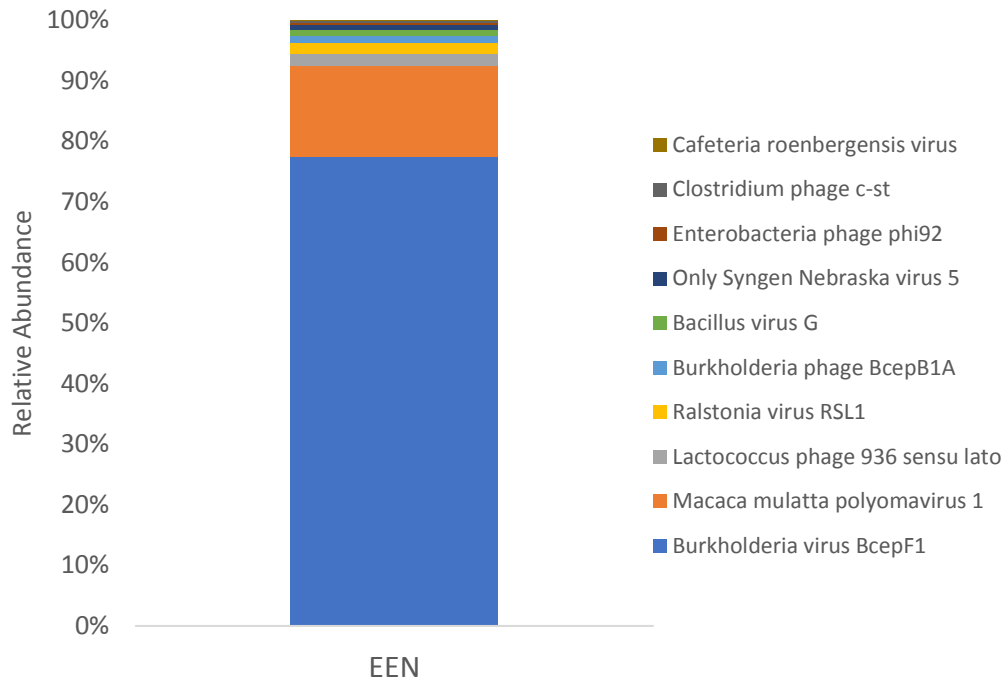


Figure 3 Stacked bar chart depicting relative abundance of 10 most abundant viruses in Modulen EEN formula. Metagenomic sequencing was performed on DNA from a VLP preparation carried out on Modulen EEN formula (Nestle Health Science). Reads were mapped to the RefSeq viral database and input into MEGAN to classify viral taxonomy. The sample was also dominated by reads from *Burkholderia* virus BcepF1 like the blank run (Figure 2).

Due to the high level of *Burkholderia* virus BcepF1 in the blank run and EEN, as well as its presence in patient samples (see below) we considered it to be a contaminant and was removed from further assessment of the gut virome in human samples.

3.2 COMPARISON OF BIOINFORMATIC METHODS FOR VIRAL ANNOTATION

Viral sequences were annotated using three methods: a reference database search with DIAMOND, and the reference-independent processes of VirSorter (VS) and VirFinder (VF). VirSorter and VirFinder results were combined and used as input for genome binning with MetaBat in an effort to define unique viral populations without the

use of a reference database. The reference-independent programs annotated a greater percentage of viral reads combined compared to the DIAMOND reference search (7.03% vs 4.30%). However, DIAMOND outperformed the reference-independent tools in terms of total unique base pairs annotated as viral (**Figure 4**).

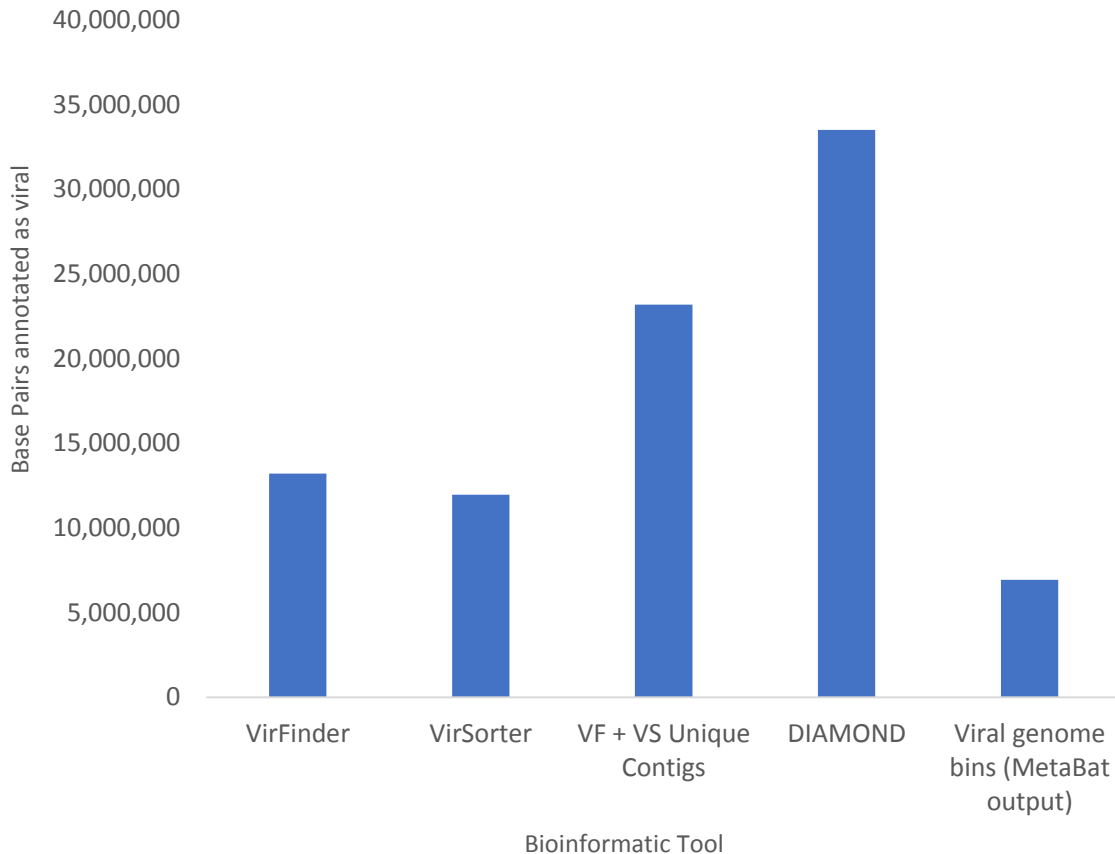


Figure 4 Performance of selected bioinformatic tools of annotating viral metagenomic sequencing reads in virally enriched samples. Viral sequences were annotated using reference-independent (VirFinder, VirSorter) and a reference-dependent method (DIAMOND). DIAMOND outperformed reference-independent methods in annotating viral reads, even when VirFinder (VF) and VirSorter (VS) results were combined (VF+VS Unique Contigs). When contigs from VF and VS were binned with MetaBat, 6,938,113 base pairs were annotated compared to 33,502,000 from DIAMOND.

3.3 REFERENCE-INDEPENDENT APPROACHES ARE UNINFORMATIVE FOR CLASSIFYING THE GUT VIROME

The reference-independent processes of VF and VS annotated 7.03% of the cross-assembled VLP prepared metagenome samples as viral. These sequences were reconstructed into genome bins with MetaBat, producing 315 unique genome bins. Bins were mapped back to each individual sample, to provide relative abundances of each bin. The relative abundance of each genome bin ranged from 0.01% to 20.9% when mapped back to each individual sample. The genome bin with the highest relative abundance was present in 66 of 122 samples (54.1%) and with an average relative abundance of 20.9%. It ranged in relative abundance from 0% to 95.6%. To assess taxonomy of this viral genome bin, it was searched against the RefSeq viral protein database. The top hits of the bin were against proteins of *Burkholderia* virus BcepF1, therefore was considered as contamination and removed from further analyses.

The alpha diversity of viral genome bins was significantly different between each individual patient and time point (**Figure 5**), with no significant trends related to alpha diversity observed amongst clinical characteristics. Similarly, there was no significant difference in alpha-diversity of viral genome bins between week 0 and week 12.

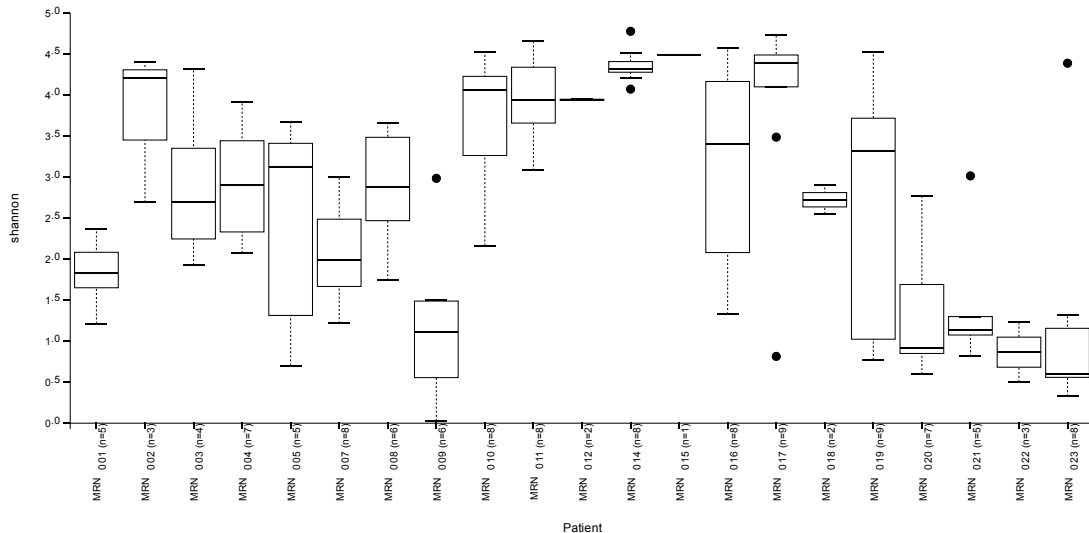


Figure 5. Shannon Diversity of each MAREEN patient across all timepoints. The Shannon diversity of viral genome bins found using VF and VS are shown for each MAREEN patient (n = 22) across all timepoints studied.

We then classified samples into sustained remission (SR) or non-remission (NR) based on fecal calprotectin (FCP) levels. Samples with an FCP level of 300 $\mu\text{g/g}$ stool or under were considered SR. Only 27 samples reached SR, and no difference was found in viral genome bin alpha-diversity between SR and NR groups (**Figure 6**). There was also no trend shown between alpha-diversity and remission classified by the weighted pediatric CD activity index (wPCDAI).

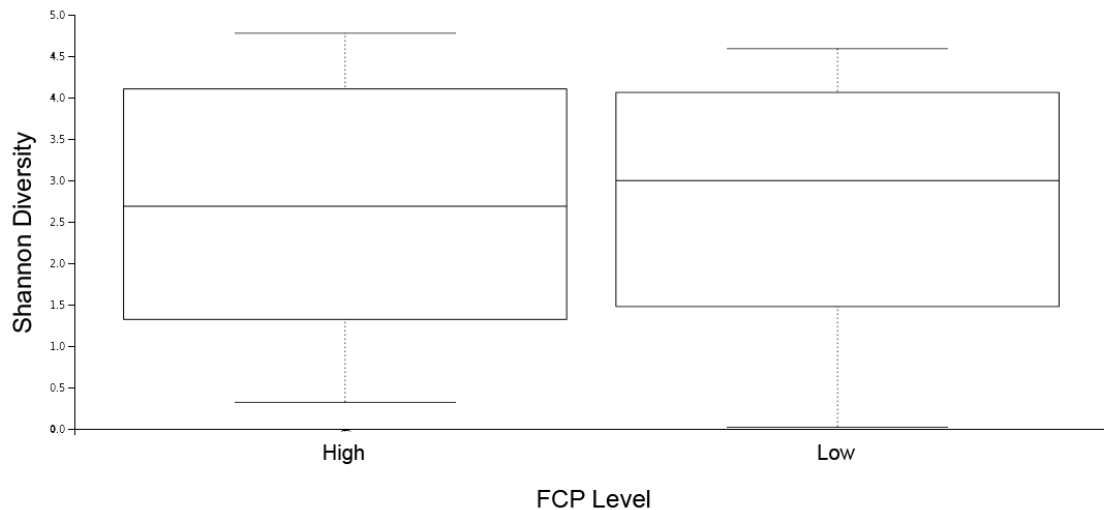


Figure 6 Viral genome bin diversity does not differ between samples with low FCP and high FCP levels. There is no significant difference in Shannon diversity of viral genome bins between samples with low (n=27) and high FCP levels (n=91) ($p = 0.73$, Kruskal-Wallis).

The beta-diversity of the viral genome bins was assessed using a non-phylogenetic measurement of diversity, called Bray-Curtis dissimilarity. Groups of samples based on remission, week sampled, and other clinical values did not have significantly different viral genome compositions based on an anosim test ($p > 0.05$). Viral genome bins were highly dissimilar between all patients assessed ($p = 0.001$, $R = 0.54$) (**Figure 7**).

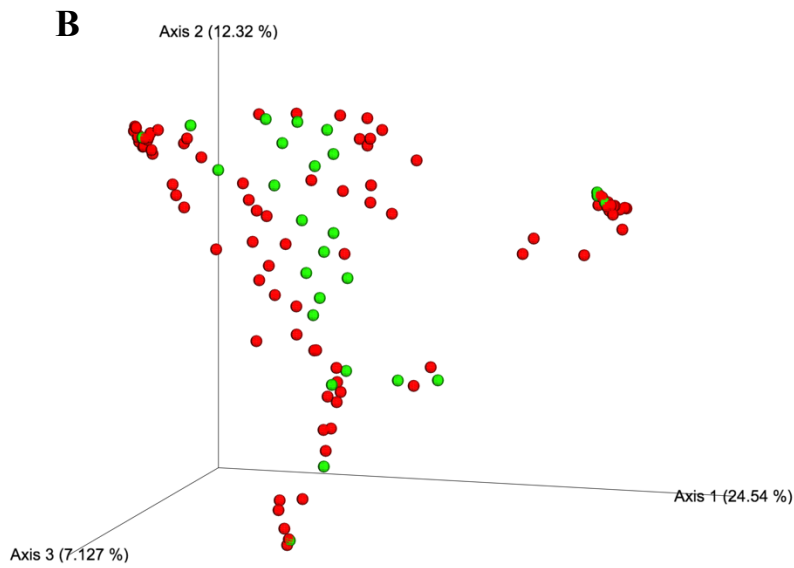
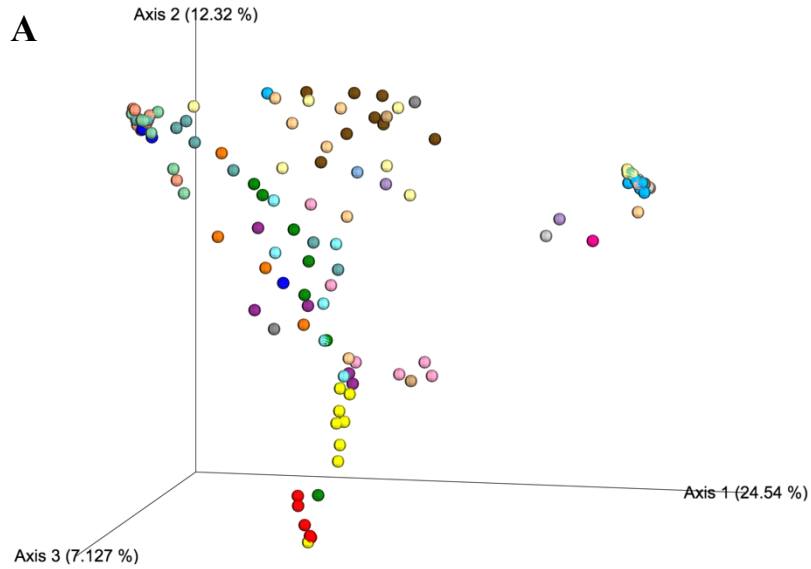


Figure 7 Viral genome bin dissimilarity in all MAREEN patients. Bray-Curtis dissimilarity principal coordinate axis plot comparing viral genome bin composition of all samples assessed. No significant differences in genome bin composition was found between any clinical characteristics (anosim test, $p > 0.05$). Samples are colored by patient ID (A) and by remission status as defined by FCP levels (B).

No significant differences in viral genome bin abundance were found between SR and NR groups, week of sample, or other clinical characteristics after multiple test correction in STAMP.

3.4 REFERENCE SEARCHES PRODUCE VALUABLE INSIGHT ON THE PEDIATRIC CD GUT VIROME

A reference-database search of the RefSeq viral database using DIAMOND annotated 4.03% of all VLP prepared metagenomic sequencing reads as viral. This approach identified 1747 viral species and 2266 viral taxa overall. The most abundant viral species found through reference searching was the *Burkholderia* virus BcepF1, being present in 82 of 122 samples (67.2%) with an average relative abundance of 28.2%. Due to being identified as a likely sample contaminant, all reads matching an instance of “*Burkholderia*” were removed from the reference search output files. After removal of *Burkholderia* phages from the reference database, an average of 2.75% reads were annotated as viral. The reference search identified 4 viral orders, 42 families, 301 genera, and 1810 species. Taxa present in only one sample were removed from further analyses, leaving 4 viral orders, 39 viral families, 278 genera, and 1416 species. The most abundant viral families observed were the Myoviridae, Siphoviridae, and Podoviridae, all part of the *Caudovirales* order (**Figure 8**).

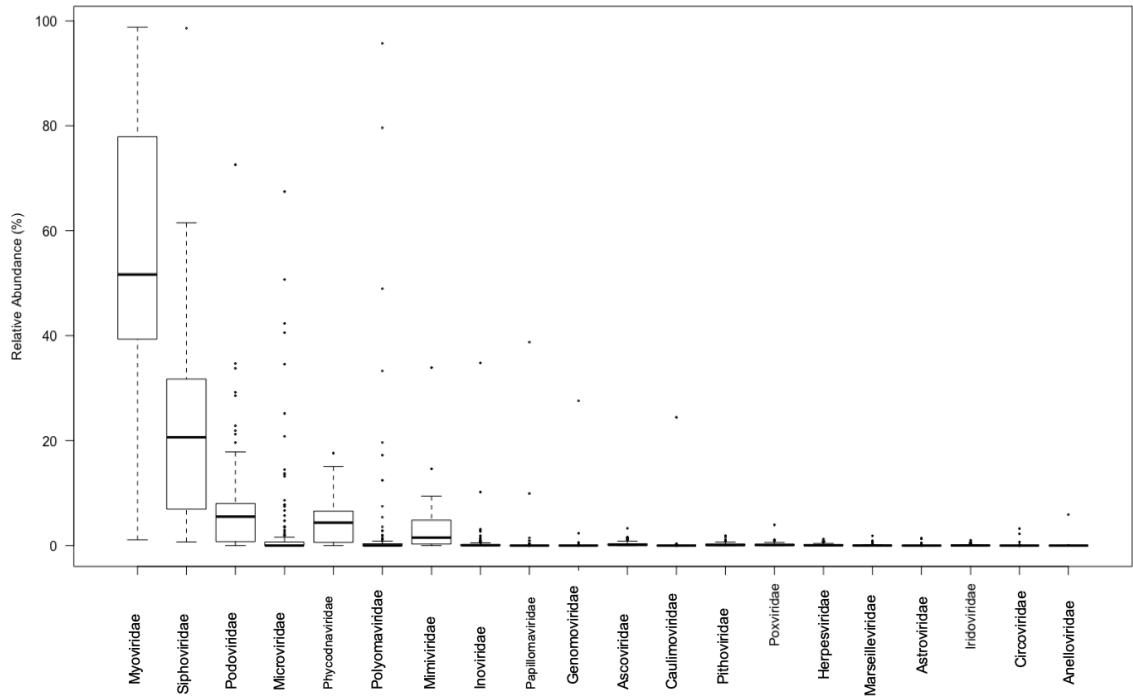


Figure 8 Boxplot of most abundant viral families observed in MAREEN patients.

MAREEN patients (n = 22) provided a total of 134 stool samples for virome analysis.

Abundances of viral families were identified using a viral reference database search. The four most abundant viral families (Myoviridae, Siphoviridae, Podoviridae, Microviridae) are bacteriophages. Minimal levels of eukaryotic viruses (Phycodnaviridae, Polyomaviridae, Mimiviridae) were observed.

Alpha diversity was assessed at each taxonomic level independently. At the species level, there was no significant difference in Shannon diversity between samples grouped by FCP and CRP levels. Samples classified as being in remission by disease activity levels had a significantly lower Shannon diversity of viral species compared to NR samples ($p = 0.0125$) (**Figure 9**). Shannon diversity was weakly correlated with FCP levels (Spearman $R = 0.1725$, $p = 0.049$) but more strongly correlated with hemoglobin ($R = -0.46$, $p = 0.001$) and hematocrit ($R = -0.46$, $p = 0.001$) levels. Species diversity did not correlate with wPCDAI levels ($R = 0.169$, $p = 0.1229$) (**Table 2**).

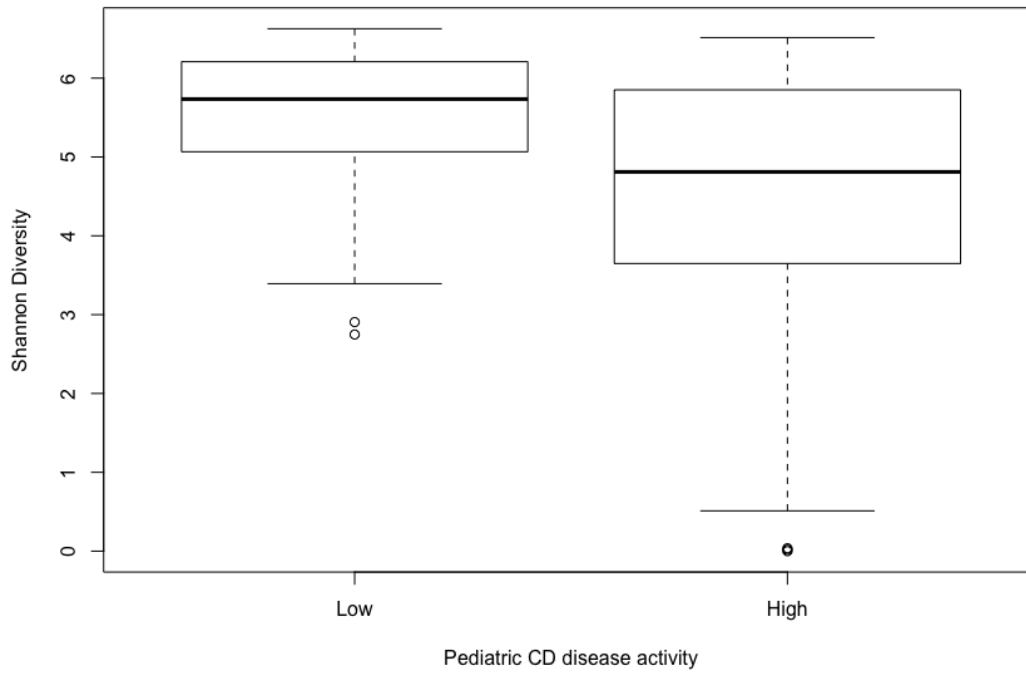


Figure 9 Viral species diversity is higher in samples with low disease activity compared to samples with high disease activity. Shannon diversity of viral species is higher in low (≤ 7.5 PCDAI) disease activity samples compared to those with high (> 7.5 PCDAI) disease activity ($p = 0.0125$, Kruskal-Wallis).

Table 2 Shannon diversity is correlated with selected clinical phenotypes. NS = not significant (p-value > 0.05).

Variable	Sample Size	P-value	Spearman Rho
Hemoglobin	64	0.0001	-0.4616
Hematocrit	48	0.0011	-0.4581
Albumin	88	0.0094	-0.2755
Fecal calprotectin	130	0.0497	0.1725
ESR	82	NS	0.1332
White blood cell count	60	NS	0.2110
wPCDAI	84	NS	0.1696

Although associations with alpha diversity and clinical outcomes were made, there were no significant differences after multiple test correction in abundance of individual taxa between any time point, remission status, or other clinical observations (**Figure 10**).

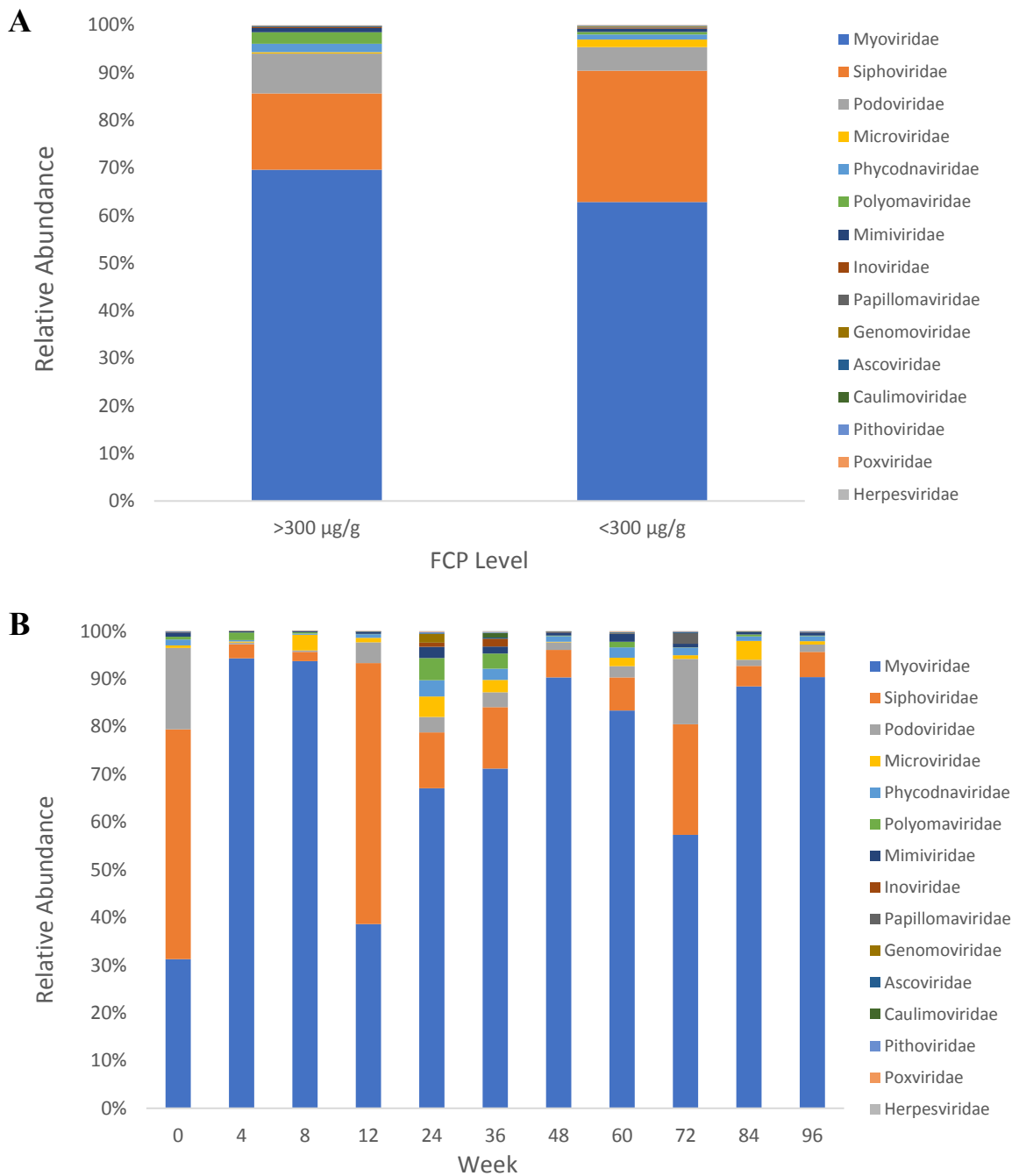


Figure 10 Stacked bar charts of viral families grouped by inflammation level and time sampled. Relative abundance of viral families from each sample was averaged and grouped into FCP levels (A) and week sampled (B). There were no individual taxa that were significantly different between FCP levels and timepoints after multiple test correction.

The viral composition of patients who underwent sustained remission after EEN administration were compared to those who did not (**Figure 11**). Patients who underwent sustained remission (SR) (n=4) had an insignificant increase in Podoviridae from week 0 to week 12 following EEN administration. NR patients (n=18) had a decrease of Myoviridae following EEN administration at week 12. These trends were not statistically significant.

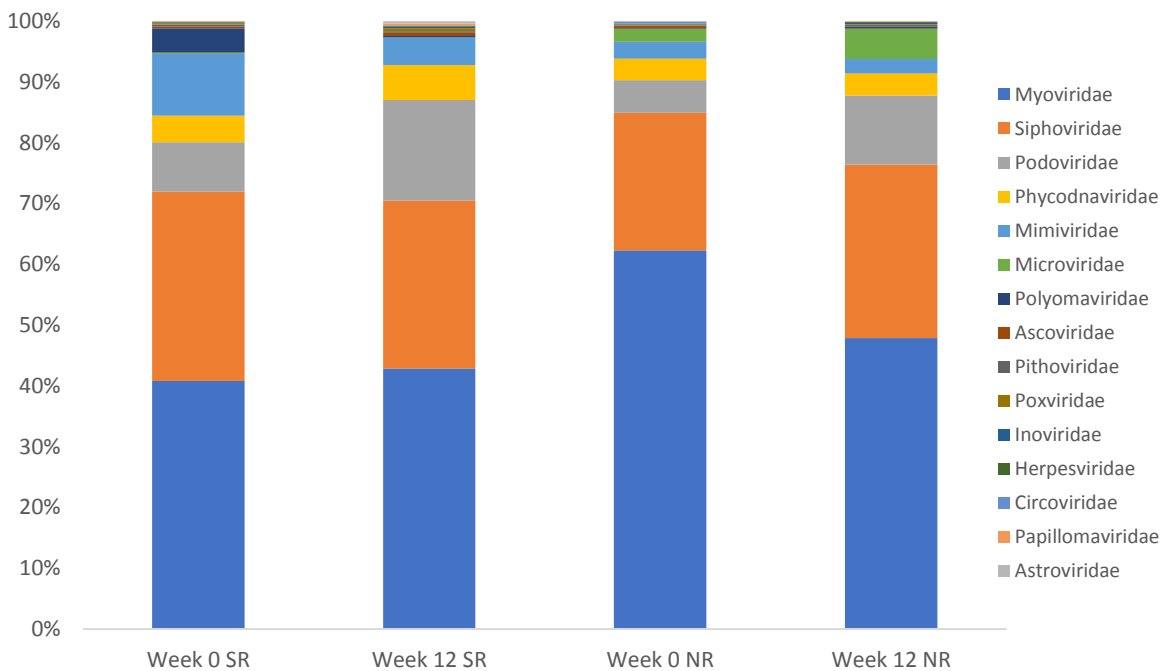


Figure 11 Patients that undergo sustained remission after EEN administration do not have distinct gut virome compositions. Stacked bar chart of gut virome composition at the family level is shown between patients that achieved sustained remission (SR) (n = 4) and those that did not (NR) (n=18) based on FCP levels. Relative abundance of Myoviridae appears to be enriched in week 0 NR samples but was not significantly different than in week 0 SR samples after multiple test corrections (p = 1.104).

Beta-diversity of the gut virome was assessed using Bray-Curtis dissimilarity at the viral species level. Samples were not significantly dissimilar based on FCP or

wPCDAI status, but the virome of individual patients were significantly different from one another as assessed by an anosim statistical test ($p = 0.001$, $R = 0.323$) (**Figure 12**).

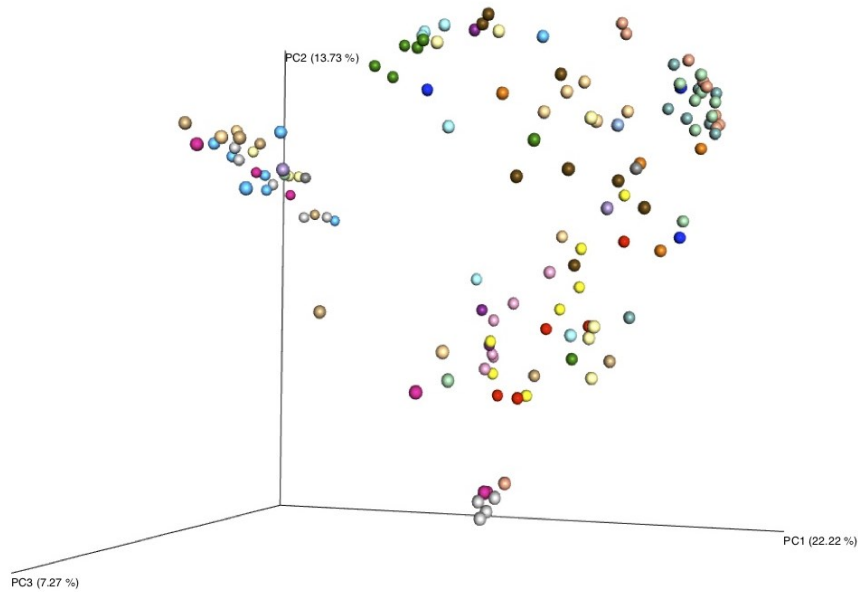


Figure 12 The gut virome significantly varies by patient. Bray-Curtis dissimilarity principal coordinate axis plot comparing viral species composition of all samples assessed. Each circle represents an individual sample and is colored by patient ID. No significant differences in viral community composition was found between any clinical characteristics (anosim test, $p > 0.05$), but patients were significantly different ($p = 0.001$, $R = 0.323$).

Although no viral taxa showed differential abundance between clinical metadata groups at any taxonomic level, several phages approached significance. *Halocynthia* phage JM-2012 and *Pseudomonas* virus KPP25 trended toward greater abundance in high FCP samples, and crAssphage trended towards higher abundance in low FCP samples (**Figure 13**).

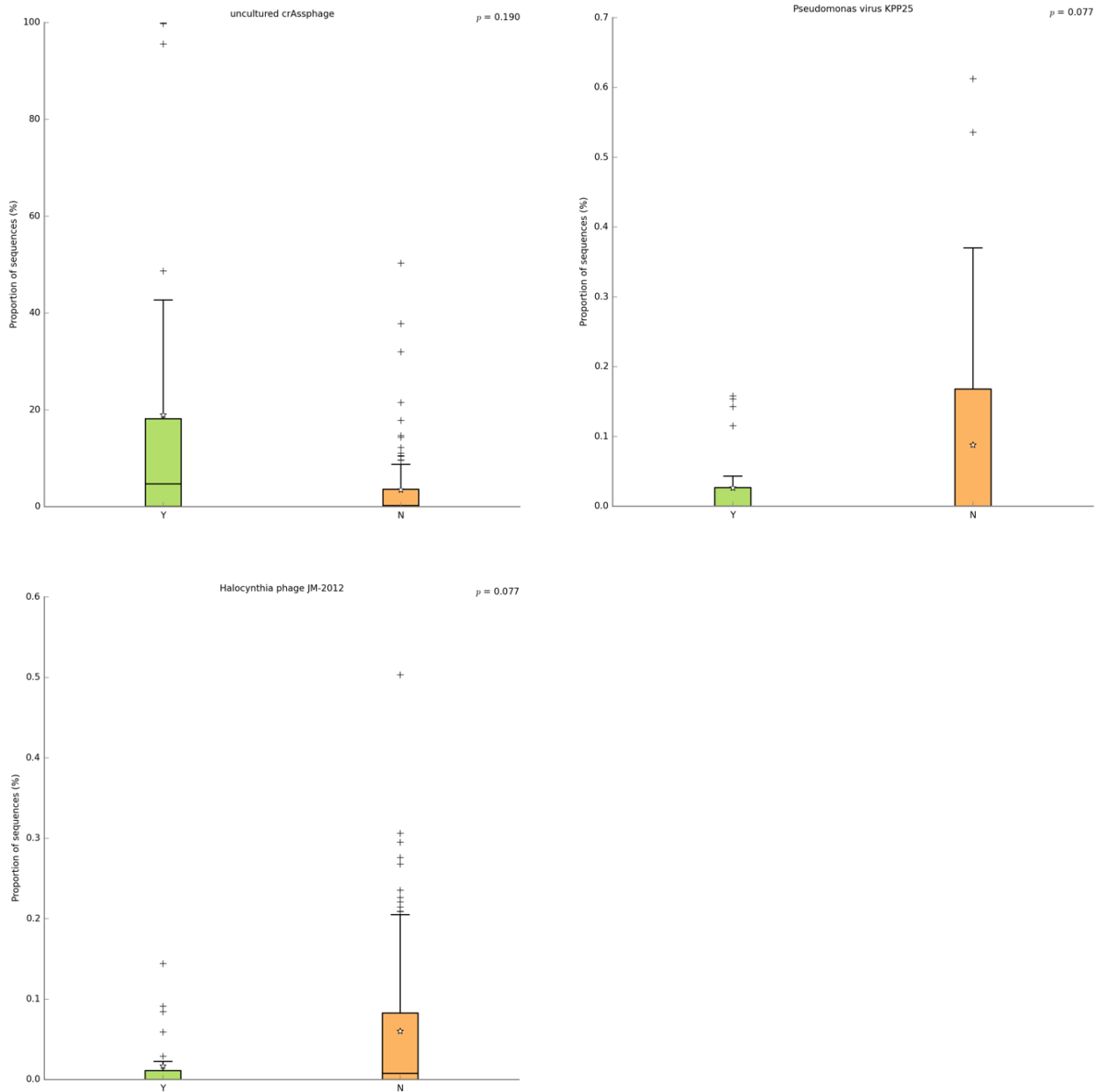


Figure 13 Selected phage trend toward significant difference between high and low FCP levels. Two group comparison using Benjamin-Hochberg FDR shows a trend in significant differences in crAssphage, *Pseudomonas* virus KPP25, and *Halocynthia* phage JM2012 between samples grouped by FCP level.

3.5 MACHINE LEARNING IDENTIFIES SUBTLE CHANGES IN THE GUT VIROME RELATED TO CLINICAL OUTCOMES

Viral species abundance data were input into a random forests (RF) model for regression with FCP levels to determine predictive power of viral taxa. The RF model explained 8.07% of the variance in the FCP data and was significant ($p = 0.002$, $R^2 = 0.091$). The top 20 viral taxa important for regression of FCP levels are shown in **Table 3**. Phages that infect *Clostridium*, *Escherichia*, and *Salmonella sp.* were represented twice each in the top 20 taxa.

Table 3 Top 20 viral species important for regression of FCP levels. Percent increase in mean square error (%IncMSE) is a measure of predictive accuracy of each feature in the random forests model.

Viral Species	%IncMSE
Clostridium phage phiMMP04	79201.9047
crAssphage	63828.6309
Escherichia virus Min27	55169.8066
Orpheovirus IHUMI LCC2	50952.8005
Bathycoccus sp RCC1105 virus BpV	48954.7598
Salmonella phage SEN34	45355.1537
Bacillus phage Shbh1	43091.5939
Synechococcus phage S CBS2	40273.222
Staphylococcus phage StB20	29084.1404
Megavirus chiliensis	28245.199
Phage Gifsy 1	26896.5311
Edwardsiella virus eiAU	25978.1707
Clostridium virus phiCD119	25516.5968
Escherichia virus N15	24748.5607
Hamiltonella virus APSE1	23073.8433
Salmonella phage vB_SosS_Oslo	22378.0416
Arthrobacter virus Mudcat	21654.8477
Geobacillus phage GBSV1	17863.606
Pseudomonas virus KPP25	17375.4326
Vibrio phage 11895 B1	15967.4168

A multivariate analysis of viral taxon abundance with clinical metadata was carried out using MaAsLin (Multivariate Association with Linear Models). At the viral species level, 52 viruses had significant associations with clinical data, but linear coefficients of association were weak (data not shown). Only three phage were associated with FCP levels.

3.6 SUMMARY

We found that a viral enrichment of stool samples resulted in improved annotation of viral reads, but virally enriched samples also had increased bacterial contamination as assessed by 16S rRNA. A viral contaminant (*Burkholderia* virus BcepF1) was uncovered by sequencing of a blank run and EEN samples and was removed from downstream analyses. Reference-independent approaches were not as informative as expected compared to the reference-database search of viral sequences. The reference-dependent approach revealed that bacteriophage from the Caudovirales order and Myoviridae family were the most abundant in the MAREEN samples. This approach found that viral species diversity was higher in samples with low pediatric CD disease activity compared to high activity, and that diversity of viral species correlates with several clinical phenotypes including FCP. No independent viral taxa were differentially abundant between any clinical outcome or time point after multiple test corrections. After validating our wet-lab protocol, we found subtle associations overall between the gut virome and clinical outcomes.

CHAPTER 4: THE BACTERIOME OF PEDIATRIC CD

4.1 ALPHA DIVERSITY OF THE BACTERIOME DIFFERS WITH FECAL CALPROTECTIN LEVELS

Taxonomic differences in the gut microbiome of 22 patients with CD were compared through profiling of the 16S rRNA gene of bacteria. Of the 139 samples sequenced for 16S rRNA gene sequencing, 11 samples were discarded from downstream analysis due to low sequencing depth. A total of 21,757 ASVs were identified across all 128 samples. After filtering out all ASVs that had a frequency of less than 0.1% of the mean sample depth, 1,053 ASVs remained for downstream analyses.

There was a significant difference in alpha diversity (species richness) between samples with high FCP ($>300 \mu\text{g/g}$ stool) and low FCP (**Figure 14**) overall ($p = 0.039$). There was no difference found in alpha diversity in samples with low disease activity compared to high disease activity.

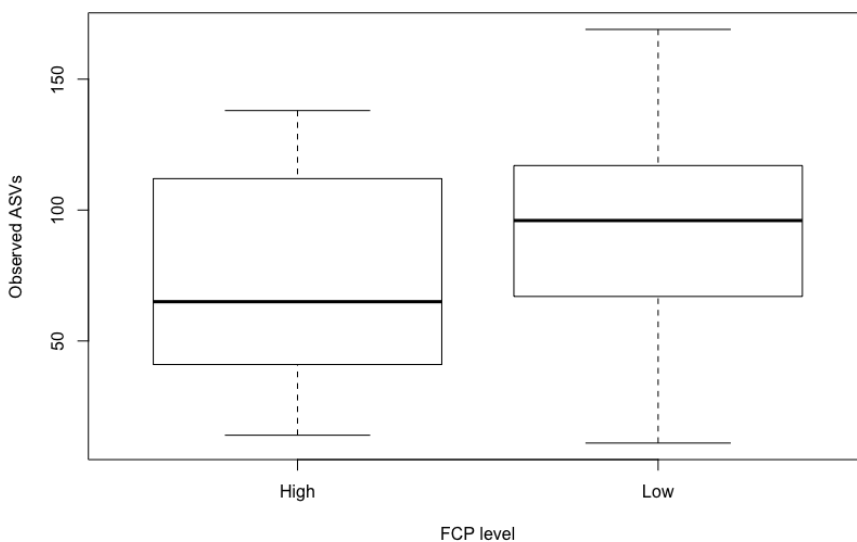


Figure 14 Alpha diversity of the bacteriome is significantly higher in samples with low FCP compared to high FCP. Samples were classified based on FCP level (above or below $300 \mu\text{g/g}$ stool) ($p = 0.039$, Kruskal-Wallis).

Stacked bar charts representing sample taxonomy were plotted for all samples at baseline (**Figure 15**). Bacterial composition was significantly different between patients (**Figure 16**) but did not differ based on FCP or other clinical groupings (Adonis test, $p = 0.109$).

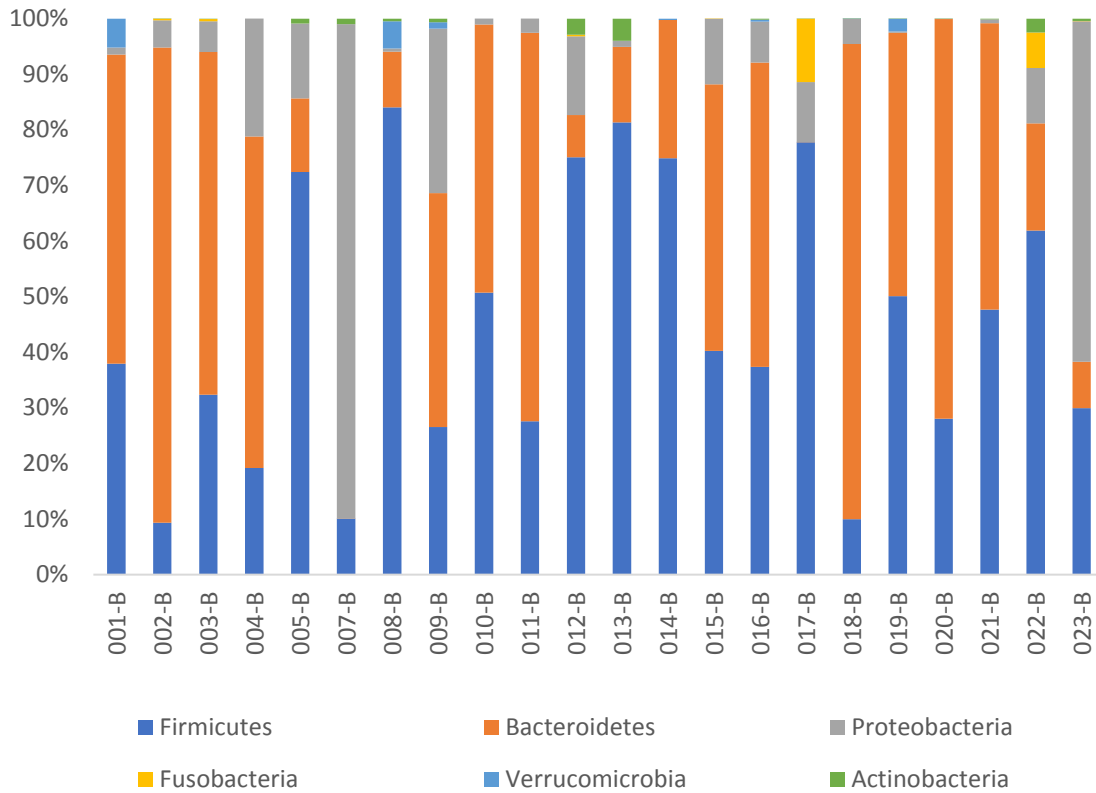


Figure 15 Stacked bar charts of baseline gut bacteriome in pediatric CD patients.

The relative abundance of bacterial phyla is shown for each patient at week 0 before EEN treatment, showing the variability in baseline bacterial microbiomes in MAREEN patients.

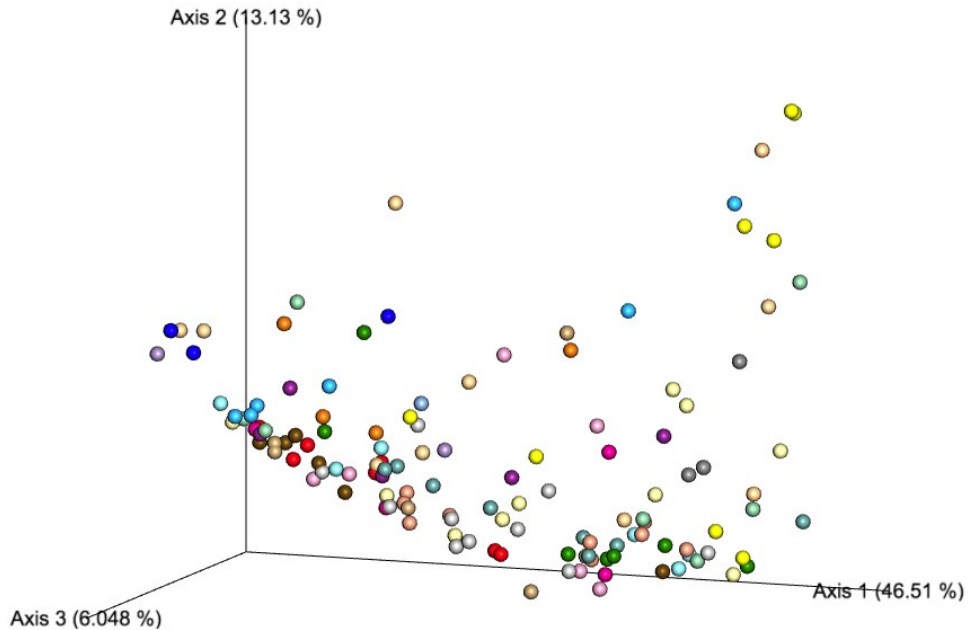


Figure 16 Weighted UniFrac principal coordinates axis plot of gut bacteriome.

Samples are colored by patient. Patients had significantly different bacterial taxa composition compared to one another ($p = 0.001$, $R^2 = 0.18$, Adonis test).

4.2 BACTERIAL TAXA DIFFER WITH FECAL CALPROTECTIN LEVELS AND REMISSION STATUS

Using a Welch's two-sided t-test in STAMP, four bacterial taxa were shown to be significantly differentially abundant after Benjamin-Hochberg FDR correction between samples of high and low FCP (**Table 4**). Enterobacteriaceae was 2.8% more abundant on average in high FCP samples but was not significantly different in abundance between groups after multiple test correction.

Table 4 Differentially abundant bacterial taxa between high and low fecal calprotectin samples. High FCP was defined as >300 µg/g stool. Benjamin-Hochberg FDR corrected p-values are shown. All four taxa had greater abundance in high FCP samples compared to low.

Phylum	Class	Order	Family	Genus	p-value	Percent difference in abundance
Firmicutes	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae	Holdemania	0.0078	0.071%
Actinobacteria	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Eggerthella	0.0220	0.037%
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Tyzzera	0.0223	0.140%
Firmicutes	Clostridia	Clostridiales	Eubacteriaceae	Anaerofustis	0.0374	0.005%

The bacterial composition of patients who underwent sustained remission after EEN administration were compared to those who did not (**Figure 17**). At baseline, those that did not undergo sustained remission (NR) had a greater proportion of Proteobacteria, which persisted until after treatment at week 12. In sustained remission patients (SR), the Proteobacteria present at baseline were not present at week 12.

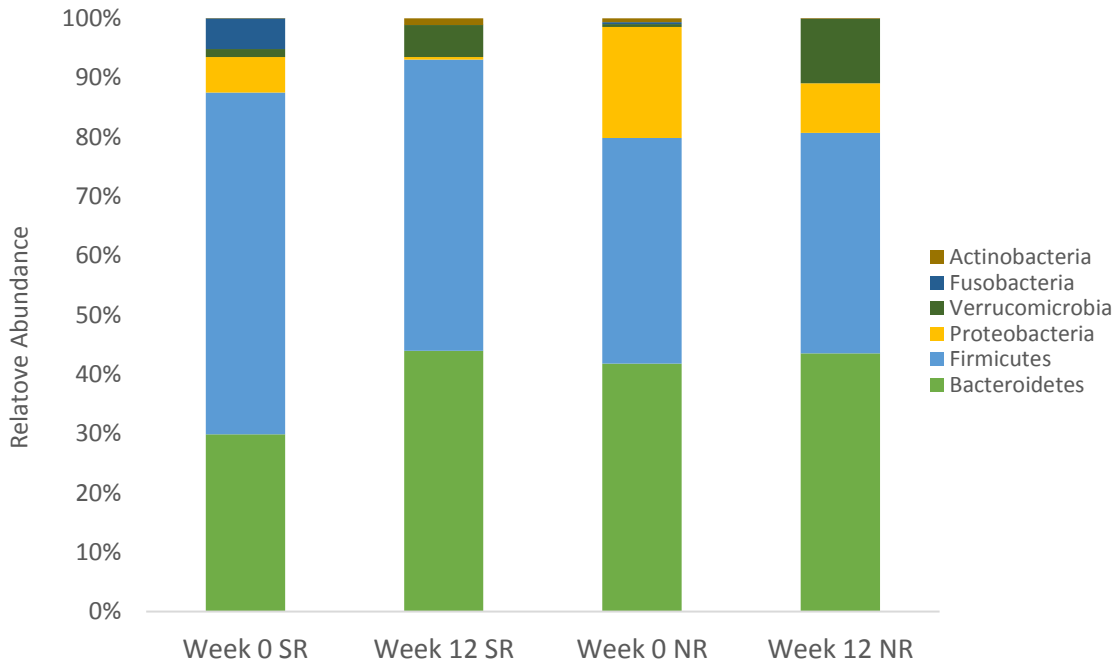


Figure 17 Patients that undergo sustained remission after EEN administration have distinct bacterial signatures. Sustained remission (SR) patients (n=4) have Proteobacteria at baseline that are not present after EEN administration. Non-remission (NR) patients (n=18) have Proteobacteria that persist at week 12.

4.3 BACTERIAL TAXA SIGNIFICANTLY MODEL FCP LEVELS

Abundance data of ASVs was input into a randomForests model and regressed against FCP levels to determine predictive power of individual taxa. The model was significant in regression against FCP levels ($p = 0.001$, $R^2 = 0.17$) and explained 17% of the variance in FCP levels. The top 20 taxa important in predicting FCP levels are shown in **Table 5**. The majority of informative taxa belonged to the phylum Firmicutes (12), followed by Proteobacteria (4), Bacteroides (2) and one taxon each from Fusobacteria and Actinobacteria. Eight taxa from the Lachnospiraceae family were in the top 20 taxa, including the two most informative features (*Lachnoclostridium sp.* and *Ruminococcus gnavus* group).

Table 5 Top 20 bacterial taxa important for regression of FCP levels. Percent increase in mean square error (%IncMSE) is a measure of predictive accuracy of each feature in the random forests model.

Phylum	Class	Order	Family	Genus	%IncMSE
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Lachnoclostridium	321714.9657
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	[Ruminococcus] gnavus group	188266.1645
Fusobacteria	Fusobacteriia	Fusobacteriales	Fusobacteriaceae	Fusobacterium	74972.5557
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	Escherichia-Shigella	74071.15665
Proteobacteria	Gammaproteobacteria	Pasteurellales	Pasteurellaceae	Haemophilus	62145.41177
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Agathobacter	56882.64904
Firmicutes	Negativicutes	Selenomonadales	Veillonellaceae	Veillonella	56658.72479
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Blautia	50443.94307
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	46129.4017
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Tyzzereella	38367.74869
Firmicutes	Clostridia	Clostridiales	Ruminococcaceae	Faecalibacterium	34542.4851
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	undefined	27238.83679
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburia	26208.9894
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	Roseburia	20467.06074
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	[Ruminococcus] torques	19772.1741
Firmicutes	Clostridia	Clostridiales	Lachnospiraceae	group	18635.57432
Actinobacteria	Coriobacteriia	Coriobacteriales	Eggerthellaceae	Eggerthella	16255.3413
Bacteroidetes	Bacteroidia	Bacteroidales	Bacteroidaceae	Bacteroides	14508.1527
Firmicutes	Erysipelotrichia	Erysipelotrichales	Erysipelotrichaceae	Erysipelotrichaceae UCG-003	14040.23128
Proteobacteria	Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	undefined	11990.00701

Bacterial abundance data were also input to MaAsLin to identify associations with clinical metadata. Only 5 associations were found between bacterial families and clinical measures, but 47 associations were found using bacterial genera (data not shown). Coefficients of association were low (all <0.06), similar to that of viral species results. The majority of associations were made with platelet counts (18) and hemoglobin levels (6).

4.4 SUMMARY

We found a significant increase in bacterial diversity in samples with low FCP level compared to high FCP level. Four bacterial taxa were significantly differentially abundant between high and low FCP samples and showed distinct differences between sustained remission (SR) and non-remission (NR) patients from baseline to week 12. Using RF modelling, bacterial taxa abundances explained 17% of the variance of FCP levels in a significant model. Most informative species belonged to the phylum Firmicutes, followed by Proteobacteria.

CHAPTER 5: VIROME-BACTERIOME INTERACTIONS

5.1 HIGH LEVEL BACTERIA – VIROME ASSOCIATIONS

The top 50 informative ASVs from the random forests model were input into a sample by sample Spearman correlation analysis with all viral species using a custom R script. The correlation analysis revealed 5986 significant associations between phage and ASVs, with 4047 remaining after Benjamin-Hochberg multiple test correction. The value of Spearman rho ranged from 0.17 to 0.48. Many of these correlations were between bacteria and phage with low abundance, with 3497 (86.4%) of the associations being between phage and bacteria with less than 0.1% average relative abundance in MAREEN samples. We narrowed our approach of phage-bacteria dynamics by focusing on taxa that were shown to be informative in machine learning models.

To identify meaningful bacteria-bacteriophage relationships, the top 20 informative ASVs from the random forests model (**Table 6**) were tested for Spearman correlations with their respective phage. The only significant association after multiple test correction was *Escherichia-Shigella sp.* with *Enterobacteria* phage mEpX1, a phage shown to infect *E. coli* ($p = 0.080$, Spearman Rho = 0.32). This correlation was especially prominent in patient MRN-004 ($p = 0.089$, Spearman Rho = 0.70) (**Figure 18**). No phage infecting these differentially abundant taxa between high and low FCP groups (**Table 4**) were found in our database therefore no correlations could be made.

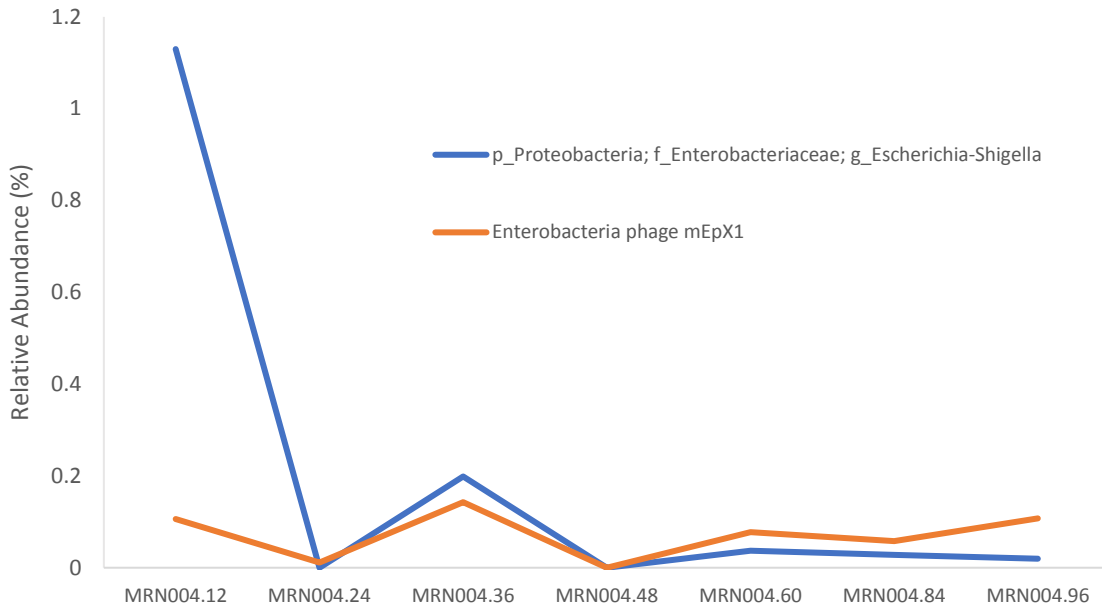


Figure 18 The abundance of *Enterobacteria phage mEpX1* is significantly correlated with its bacterial host, *Escherichia-Shigella sp.* Relative abundance of *Enterobacteria phage mEpX1* and its bacterial host is shown in patient MAREEN-004 during the study period ($p = 0.089$, Spearman Rho = 0.70).

A Spearman correlation analysis was also carried out with the top 50 most abundance viral species and the top 20 most informative ASVs in regression of FCP in the RF model (**Table 6**). A correlation plot visualizing the relationships between these viral and bacterial taxa is shown in **Figure 19**. Only ten bacterial-viral associations were significant after multiple test correction, and none represented a documented phage-host relationship after a search of the Virus-Host DB. An identical correlation analysis was attempted between differentially abundant ASVs (**Table 5**) and their respective bacteriophage, however no matching bacteriophage were found.

5.2 KNOWN PHAGE-BACTERIA INTERACTIONS DO NOT CORRELATE WELL IN PEDIATRIC CD PATIENTS

We attempted to validate previously identified bacteriophage-bacteria relationships from the literature in the MAREEN cohort. We first looked at the interaction between crAssphage and its host *Bacteroides intestinalis*. crAssphage had an average relative abundance of 6.9% across all samples and was present in 116 of 130 samples studied. It was the third most abundant phage in the MAREEN cohort. There were 37 ASVs that matched *Bacteroides*. There were no significant Spearman correlations between *Bacteroides*. ASV and crAssphage after multiple test correction (FDR p-value > 0.16). We also attempted to validate other bacteriophage-bacterial host interactions previously identified through NGS applications (Parmar, Gaikwad, Dhakephalkar, Kothari, & Singh, 2017). We found no significant correlations in the MAREEN cohort of published bacteria-bacteriophage interactions after multiple test corrections (FDR p-value > 0.09). Several bacteria like *Pseudomonas sp.* were present in a low percentage of samples, weakening correlation significance.

5.3 ENTEROCOCCUS PHAGE IS ASSOCIATED WITH ITS HOST AND DISEASE ACTIVITY IN ONE PATIENT

Despite a lack of significance across all samples, Enterococcus phage EF62 phi was significantly associated with its host *Enterococcus sp.* (p-value = 0.04, Spearman Rho = 0.53) in one patient (**Figure 20A**). Although correlations were not significant, abundance of both taxa trend with FCP levels (**Figure 20B**) and disease activity (**Figure 20C**).

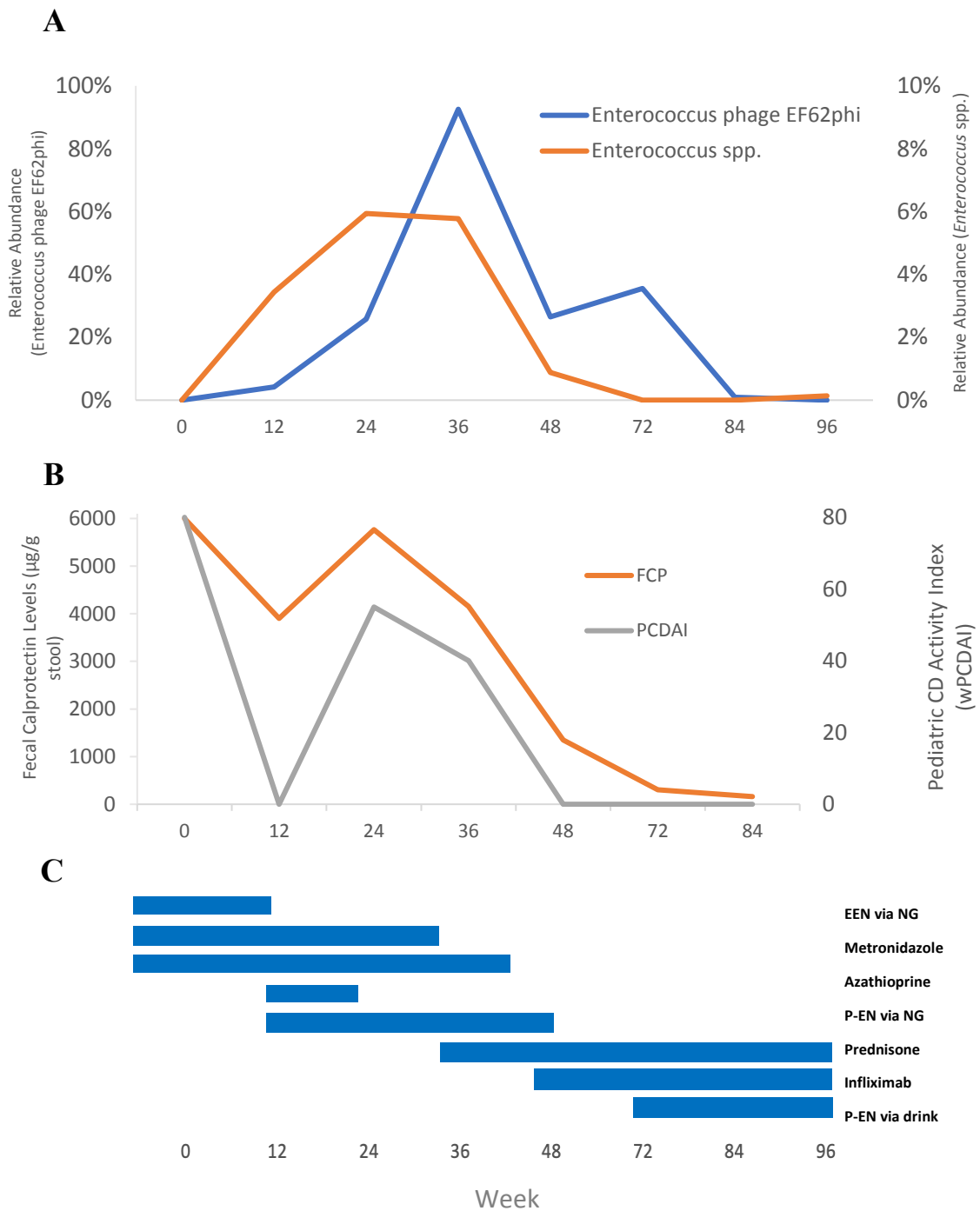


Figure 20 Abundance of *Enterococcus* phage EF62phi is significantly associated with abundance of *Enterococcus* sp. and trends with clinical diagnostic values in one patient. *Enterococcus* phage EF62phi and its host (A) were significantly correlated in one patient (p-value = 0.04, Spearman Rho = 0.53), and trended with clinical measures of inflammation and disease activity (B). The patient underwent a 12-month treatment of EEN before being exposed to other treatments to control disease activity (C).

5.4 SUMMARY

Bacterial features and viral species abundances were input into a Spearman correlation model for identification of prospective phage-host interactions in the pediatric CD gut. Over 4000 significant associations were found between viruses and bacteria, many of which were not between phage and their respective hosts. A strong correlation between *Escherichia-Shigella sp.* and Enterobacteria phage P4 was found, especially in one MAREEN patient. We were unable to validate previously found phage-bacteria relationships such as crAssphage and Pseudomonas phages with their respective hosts. We found a significant association between Enterococcus phage EF62 phi and its host *Enterococcus sp.* abundances that trended with both disease activity and gut inflammation in one patient.

CHAPTER 6: DISCUSSION

6.1 VIRUS LIKE PARTICLE PREPARATIONS

6.1.1 VIRAL STOOL ENRICHMENT IMPROVES BIOINFORMATIC ANNOTATION OF THE GUT VIROME

The advent of NGS sequencing technologies has allowed the examination of the gut microbiome and host genetics in great detail. Most studies on the microbiome to date have been focused on the bacterial portion of the microbiome, the gut “bacteriome”. Wet lab procedures for analysis of the gut bacteriome are more straightforward than that of the gut virome. Stool preparation kits such as the one used in this study are commercially available and well-validated. Preparation for 16S rRNA or metagenomic sequencing of the microbiome is not time or labour-intensive. Conversely, stool preparations for analysis of the gut virome are still in their infancy, are time-intensive, and lack standardized procedures.

In the present study we used a VLP stool preparation protocol adapted from Norman et al. (2015) and showed that viral enrichment results in 2.9% improved viral annotation compared to a typical stool preparation for metagenomic sequencing (**Figure 1**). In addition, the virally-enriched samples had 0.73% reads map to the 16S rRNA database to assess bacterial contamination, compared to 0.46% in the normally prepared metagenomes. The protocol for viral enrichment of stool involves an initial filtration step to isolate viral capsids, and subsequent steps that attempt to remove any bacteria that may remain. These include chloroform treatment to destabilize bacterial membranes, followed by incubation with DNase to remove any bacterial DNA that may persist.

A 2015 comparison of sample preparations for viral metagenomics by Conceição-Neto and colleagues proposed a reproducible sample preparation protocol for human stool called NetoVIR. They modified several steps of a typical virome isolation protocol to unravel the effects of filtration, homogenization, and treatment with various reagents on recovery of viral and bacterial sequences. They found that homogenisation with ceramic beads reduces number of viral particles isolated, which is the first step in our protocol. They also found that despite efficient removal of bacteria, 0.45 μm and 0.22 μm

filtration results in loss of 99% of mimivirus and 90% of herpesvirus (Conceição-Neto et al., 2015). Although herpesvirus is not a bacteriophage, it has been found in the colonic tissue of those with IBD (Shimada et al., 2017) and therefore may have been detectable in MAREEN samples. Finally, treatment with chloroform was shown to effectively remove bacteria, but altered abundance of certain viruses. The greatest effect of chloroform was shown on enveloped viruses such as coronavirus, rotavirus, herpesvirus and the non-enveloped rotavirus. Therefore chloroform may not be a preferred reagent to use for viral amplification (Conceição-Neto et al., 2015).

Given these discrepancies in sample preparation, further assessments of the gut virome should perhaps incorporate parts of the NetoVIR protocol to optimize viral enrichment. In addition, the NetoVIR protocol takes approximately 8 hours to complete from stool filtration to DNA library preparation, compared to the three-day protocol used in the current study. Despite lower viral annotation compared to other studies, our viral enrichment protocol showed improved viral annotation compared to normal MGS preparations in our cohort.

6.1.2 BURKHOLDERIA IS A COMMON DNA EXTRACTION KIT CONTAMINANT

We conducted a blank run of the viral enrichment protocol containing all reagents except for stool sample, in addition to separate viral enrichments on samples of EEN formulae. Both the blank run (**Figure 2**) and EEN samples (**Figure 3**) were expected to have negligible amounts of reads mapping to the viral protein database, but instead had 7.2% and 1.2% read mapping, respectively. The dominant virus in these samples was *Burkholderia* virus BcepF1, which was removed from downstream analysis of patient samples due to being considered a contaminant. Little is known about this specific phage aside from its host being *Burkholderia ambifaria*, a member of Proteobacteria that was discovered in 2001 (Laevens et al., 2001). No ASVs in the MAREEN samples mapped to any species in the *Burkholderia* genus, however 9 members of the Burkholderiaceae family that *B. ambifaria* is a part of.

There has yet to be an assessment in the literature focused on contamination of virome samples, but assessments of contamination in 16S rRNA studies have shown *Burkholderia* to be a significant contaminant in DNA extraction kits. In sequencing of negative “blank” controls similar to ours, *Burkholderia* was found as a contaminant in three different sequencing centers (Salter et al., 2014). The authors state that *Burkholderia* is commonly found in water and is likely a contaminant of the DNA extraction kit itself. Although the extraction kit used by Salter et al. was different from that used in the present study, contamination of similar DNA extraction kit reagents has been previously reported (Mohammadi, Reesink, Vandenbroucke-Grauls, & Savelkoul, 2005). While this does not provide a definitive answer for the contamination of *Burkholderia* virus BcepF1 in our samples, this evidence proposes that this phage may have been present alongside its host bacterium as a contaminant in the DNA extraction kit used.

BcepF1 was the most prominent virus in the EEN sample (75.8% relative abundance), but other notable viruses were also present. *Macaca mulatta* polyomavirus 1, or simian virus 40 (SV40) was shown to be 14.8% abundant in the EEN Modulen sample (**Figure 3**). This virus was also found in 55% of patients sampled, with a relative abundance of up to 84.4% in one patient (data not shown). SV40 is a polyomavirus found in both monkeys (*Macaca mulatta*, aka Rhesus macaque) and humans. Infection with SV40 has been linked to several primary brain cancers, tumors, and malignant mesotheliomas (Vilchez & Butel, 2004). Infection in rhesus macaques causes a silent but lifelong infection that persists in the kidney but can be reactivated in immunosuppressive conditions. Published data of SV40 infections are inconsistent, but prevalence is low in clean, developed environments (Butel, 2012). In those that do harbour polyomaviruses like SV40, their persistent infection may result in sporadic shedding of the virus in feces over time, explaining viral presence in MAREEN patient samples. Other studies from across the globe have detected SV40 in stool samples of healthy children and adults, with an overall low prevalence in randomly collected human samples (Butel, 2012). It may be suggested that quantitative PCR tests for SV40 should be done on EEN formulae to ensure its safety for consumption.

6.2 REFERENCE BASED BIOINFORMATICS ARE OPTIMAL FOR ANNOTATION OF THE GUT VIROME

Due to the infancy of viral metagenomics, bioinformatic methods for analyzing the viral component of the gut microbiome are not as advanced as that for bacteria. The most straightforward method for analyzing viromes is to search a viral protein database for sequences that match those in samples of interest and assign these sequences taxonomy based on their match. This reference-dependent method is straight forward and well-validated but is limited to only the sequences in the reference database, which is not all-encompassing of viral diversity. Reference independent methods such as VirSorter and VirFinder predict viral sequences based on unique features of viral genomes, such as *k*-mer profiling, enrichment in short genes, and depletion of strand switching (Ren et al., 2017; Roux et al., 2015). Although these reference-independent approaches seem promising for identifying more viral sequences and novel viruses not found in the viral protein database, we found that the reference database search annotated more overall base pairs as viral compared to the reference independent approach (**Figure 4**).

We employed a reference independent approach adapted from Anderson and colleagues that identified 2243 viral populations in rumen stool (C. L. Anderson, Sullivan, & Fernando, 2017). In our analysis we uncovered only 315 unique viral genome bins using identical parameters. The study by Anderson et al. used long-read sequencing on the Ion Torrent platform in addition to Illumina sequencing, which likely improved their ability to assembly complete viral genomes compared to Illumina sequencing on their own. Assembly-based analyses are not necessarily optimal for assessment of stool samples, since highly diverse samples like stool may yield poor assemblies and conceal taxa from downstream analyses (Knight et al., 2018). Although the reference database search performed better than these reference-independent approaches, future virome work should attempt to non-redundantly combine these two data outputs to gain more insight into the gut virome.

Our search of the RefSeq viral protein database assigned taxonomy to 4.2% of MGS reads compared to an average of 14% in the Norman et al. gut virome study that used similar extraction and sequencing protocols. Other gut virome studies of IBD

patients have ranged in read annotation from 1% to 17.72% using similar methods as ours (Pérez-Brocal et al., 2015b; Wagner et al., 2013b; W. Wang et al., 2015b). Our comparatively low annotation of viral sequences is likely due to some aspect of the sample preparation, as our bioinformatic analyses are largely identical to that of Norman et al (2015). The Norman et al. sample preparation protocol filtered stool through a 0.2 μm filter twice, while we carried out this step only once in our protocol, perhaps leading to a diminished viral enrichment.

6.3 VIRAL DIVERSITY IS ASSOCIATED WITH SEVERAL CLINICAL CHARACTERISTICS

We assessed differences in the gut virome using both reference-dependent and -independent approaches. There were no significant results between any clinical metadata using the bins as input. The reference-based approach produced several significant results with clinical metadata, including increased Shannon diversity in low disease activity samples (**Figure 9**) and correlations of diversity with several clinical outcomes (**Table 2**). The insignificant results from the viral genome bins may be in part due to their lower overall assignment as viral compared to the reference dependent approach, thus representing a smaller percentage of the gut virome in MAREEN patients. In both datasets, the gut virome was highly unique across individuals and timepoints, which is consistent with previous studies on the diversity of the human gut virome (Minot et al., 2011; Norman et al., 2015).

The majority of viruses assigned by the reference-based approach belonged to the Caudovirales order, consisting of the Myoviridae, Siphoviridae, and Podoviridae families (**Figure 8**). These phage families are the most abundant in the human gut, and infect members of the Firmicutes, Bacteroidetes, Proteobacteria, and Actinobacteria phyla (Mirzaei & Maurice, 2017). The Caudovirales have also been shown to be the most abundant bacteriophages in several studies of the IBD gut virome (Fernandes et al., 2018; Norman et al., 2015), and had increased richness in IBD patients compared to controls in two studies (Norman et al., 2015; Wagner et al., 2013a). The Caudovirales are tailed, dsDNA phages that target bacterial cells receptors through similar mechanisms, recently

reviewed in detail by Nobrega et al. (2018). Also present were phages from the ssDNA non-tailed Microviridae family, which were the fourth-most abundant phage family in MAREEN patients. A recent study of 24 children showed that Microviridae richness was greater in healthy controls compared to CD patients (Fernandes et al., 2018). Since there were no healthy controls in the MAREEN study, we were unable to make similar case-control comparisons as is frequently done in microbiome studies.

6.3.1 THE ROLE OF THE GUT VIROME IN IMMUNOMODULATION

We did find that diversity of viral species as a whole was higher in samples with lower disease activity compared to high disease activity (**Figure 9**) using the weighted pediatric CD disease activity index (wPCDAI). A lower diversity in the gut bacteriome is generally associated with adverse health outcomes and disease, like IBD and obesity (David et al., 2014; Qin et al., 2010). Our finding that a lower diversity of viral species is associated with increased disease activity is in line with this notion found in bacteria. Bacteriophage are thought to offer the gut immunity through interactions with the mucosal layer of epithelial cells. A 2013 study showed that bacteriophage T4 adhere to mucus in human gut samples through binding of mucin glycoproteins through their capsid proteins (Barr et al., 2013). This adherence of phage to the gut mucosa may create an antimicrobial layer that minimizes bacterial attachment to the mucus layer. Therefore, a more diverse community of bacteriophage in the gut may provide a protective barrier of the gut mucosa from pathogens that could harm the integrity of the mucosa and advance disease activity. Since we analyzed stool in the present study, it is difficult to assess the interaction between phage and bacteria at the mucosal surface, but nonetheless this analysis provides insight on phage dynamics in the gut.

The supposed protective role of phage at the mucosal surface is not supported by our finding that viral species diversity in MAREEN patients is correlated, albeit weakly, with gut inflammation (**Table 2**). There are a multitude of sensors in the gut that can set off a cascade of inflammation when a virus is sensed (Metzger, Krug, & Eisenächer, 2018). For example, toll-like receptors (TLRs) located on endosomes can recognize viral nucleic acids, such as ssDNA. Ligand binding of TLRs results in downstream induction

of pro-inflammatory cytokine production (e.g. IFN- β), thus influencing inflammation in the gut lumen. *In vitro*, phage have been shown to rapidly transcytose across cell layers from the gut and other tissues, meaning that phages from the gut may have the ability to act intracellularly, or pass into circulation (Nguyen et al., 2017). When in circulation, T4 phage also trigger an antibody response in the gut and the blood, substantiating the weak immunogenicity of phage in the human body (Majewska et al., 2015). Furthermore, the lytic killing of host bacteria by phage would be expected to release bacterial antigens that may trigger further inflammation. A recent review suggests it is likely that the major role of gut bacteriophages is mediated through the modulation of the bacteriome, especially at the intestinal mucosal surface (Metzger et al., 2018). Taking all evidence into account, the biological evidence of phages and immunomodulation is abundant (Górski et al., 2017), and may explain the weak link between FCP and phage diversity in our study.

The wPCDAI uses several symptomatic measures of well-being in addition to laboratory values to model disease activity in pediatric CD. The symptomatic measures used include abdominal pain, stool frequency, general well-being and an abdominal examination, while the lab values include hematocrit, ESR, and albumin (Turner et al., 2011). It is interesting to note that viral species diversity correlated with hematocrit and albumin levels that are used in the wPCDAI but did not significantly correlate with wPCDAI or ESR levels on their own. Hematocrit measures the volume of red blood cells compared to total blood volume, and hemoglobin is a measure of the protein responsible for delivery of oxygen to tissues (Billett, 1990). A low hematocrit and hemoglobin level may be indicative of anemia, and are predictive of CD complications or CD related surgery (Rieder et al., 2014). Low albumin levels are associated with inflammation and may be seen in CD due to the body being unable to properly absorb and digest protein. The negative correlation of viral species diversity with these serological measures of disease activity for the first time in pediatric CD patients suggests that the gut virome may be playing a pathogenic role in disease.

The phages that trended toward differential abundance between high and low FCP samples (**Figure 13**) and the top hits in random forest modeling (**Table 3**) have not yet been associated with disease in the literature. These phages may provide a starting point

for future mechanistic studies analyzing the role of phage in CD and in modulating the gut microbiome.

6.4 BACTERIOME OF CD

There has been considerable research done on the changes of the microbiome that are linked to IBD, namely that there is a decrease in alpha-diversity, or species richness in IBD patients, as well as a differential composition of bacteria in those with IBD compared to healthy controls (Kostic et al., 2014; Gevers et al., 2014; Hansen et al., 2012). Although our study does not involve the use of healthy controls, we see a decrease in alpha diversity in samples with high FCP levels (**Figure 14**). Higher FCP levels (>300 µg/g stool) have been associated with IBD patients, especially during disease recurrence compared to remission (Pascal et al., 2017). A recent study of 313 IBD patients and 582 controls found no significant association with bacterial abundance and FCP levels (Imhann et al., 2018), but deep MGS of 1135 Dutch study participants found that that microbial diversity and metabolic pathways were robustly associated with calprotectin levels (Zhernakova et al., 2016). However, we show this result for the first time in pediatric CD patients undergoing treatment.

We also found four taxa significantly enriched in high FCP samples compared to low FCP samples (**Table 4**). *Eggerthella sp.* are anaerobic, Gram-positive bacteria that are observed in severe bacteremia cases requiring antibiotic treatment (Gardiner et al., 2015), but have been observed as being present in the healthy human microbiome (F. Li, Hullar, Schwarz, & Lampe, 2009). It has yet to be associated with IBD, but research suggests that it may be a harmful pathogen. *Tyzerella sp.* has been previously associated with UC in a cohort of 22 patients (Qiu et al., 2017). *Anaerofustis* has only one species (*Anaerofustis stercorihominis*), which has yet to be discussed in the context of the human microbiome.

Although several studies analyzing the microbiome during EEN treatment have observed a decrease in alpha diversity and shift in bacterial composition after a 12-week period of treatment (MacLellan et al., 2017), we did not observe a significant shift of diversity or composition in MAREEN patients. We did however observe a distinct

signature of patients that underwent successful treatment (SR) compared to those that did not (NR) (**Figure 17**). This is consistent with the findings of Dunn and colleagues who showed a general enrichment of Proteobacteria in patients that did not achieve SR in response to EEN therapy (Dunn et al., 2016).

Even when patients in our analysis were split into SR and NR groupings, no significant change in alpha- or beta-diversity was found (data not shown). This particular result conflicts with the Dunn et al. study, which consisted of several patient samples from the MAREEN study and therefore patients from our analysis. To define SR and NR groupings, we grouped samples based on their FCP levels, while the Dunn et al study amongst others have used the PCDAI. Analysis of FCP gives direct insight into the inflammation going on in the gut at time of stool sampling, and therefore may provide a better picture of gut dynamics compared to the assessments used in scoring of the PCDAI. We also used new bioinformatic methods to analyze the bacteriome (i.e. amplicon sequence variants vs. operational taxonomic units) in the present analysis, which may have potentially affected our results. Additionally, the Dunn et al study employed the use of the tool BioMiCo, which uses Bayesian modeling to infer microbial communities (Shafiei et al., 2015). Use of tools like BioMiCo may uncover further information on the gut bacteriome of the full MAREEN cohort presented in this study.

Although there were limited taxonomic differences between high and low FCP samples, RF modeling with ASV abundances explained 17% of the variance in FCP levels (**Table 5**). The top taxonomic hit in the RF model, *Ruminococcus gnavus*, was represented in the most common 57 species of the gut microbiome in >90% of people (Qin et al., 2010), but has been associated with an increased abundance in several cohorts of CD patients (Joossens et al., 2011; Willing et al., 2010). The Willing et al. (2010) study also showed the disappearance of *Faecalibacterium* and *Roseburia*, and increased abundance of *Enterobacteriaceae* and *R. gnavus*, which are all in the top 20 most informative taxa of our RF model. *Faecalibacterium prausnitzii* is widely accepted to be a healthy member of the gut microbiome, and its reduction has been associated with IBD. Oral administration of *F. prausnitzii* has been shown to attenuate the severity of murine colitis, partly attributed to its metabolites blocking NF- κ B activation and downstream pro-inflammatory cytokine production (H. Sokol et al., 2008). Temperate phages of *F.*

prausnitzii were not characterized until recently, and therefore were likely missed in our analysis of the virome (Cornuault et al., 2018).

6.5 THE GUT VIROME AND BACTERIOME DISPLAY MODEST INTERACTIONS IN THE PEDIATRIC CD GUT

We sought out to analyze how bacteria and viruses interact longitudinally. First, the top 50 informative ASVs from the RF model were input into a Spearman correlation analysis with all viral species, resulting in 4047 significant associations between bacteria and phage. To isolate meaningful relationships from this large data set, we employed a similar approach as Norman et al., who found that a reduction in bacterial diversity of 18 differentially abundant bacterial taxa was inversely correlated with alterations in Caudovirales bacteriophages. No significant relationships were found between the 4 differential bacterial taxa between high and low FCP groupings in our analysis and all viral species, so we turned to the top informative taxa from the RF model to correlate with changes in their respective gut phage.

We found a significant correlation between *Escherichia-Shigella sp.* and *Enterobacteria* phage mEpX1, which has been shown to infect *E. coli* (**Figure 18**). This specific interaction has yet to be characterized in the human gut, but a recent analysis of the murine gut virome found a significant enrichment of Enterobacteriaceae-infecting phages in colitis-induced mice compared to healthy controls. These phages included those infecting *Escherichia*, *Salmonella*, *Shigella*, and *Enterococcus* taxa (Breck A. Duerkop et al., 2018).

We also found a significant correlation between abundances of Enterococcus phage EF62 phi and its host *Enterococcus sp.* in one patient, which trended with FCP and PCDAI levels (**Figure 20**). To our knowledge, this phage has yet to be associated with IBD in any microbiome study to date, while *Enterococcus sp.* has been associated with CD (Nishino et al., 2018; Zhou et al., 2016). This phage-bacteria interaction was recently modeled in mice using *Enterococcus faecalis*. When *E. faecalis* V583 colonizes the murine gut, it produces lytic phages derived from integrated prophage genes (B. A. Duerkop, Clements, Rollins, Rodrigues, & Hooper, 2012). Production of these lytic

phage particles conferred an advantage to *E. faecalis* V583 in competition with other enterococcus strains. This study indicated that phage can influence composition of gut bacterial communities, even from internally encoded prophage. Interestingly, an increase in a quorum sensing molecule from *E. faecalis* can induce prophages and enable the spread of pathogenicity genes to commensal gut enterococci (Rossmann et al., 2015).

The interaction between *Enterococcus sp.* and Enterococcus phage EF62 phi shows an example of kill-the-winner (KTW) dynamics. This model of phage-bacteria interaction is characterized by where a bacterial host increases in abundance, followed by an increase in the phage population (Mirzaei & Maurice, 2017). Infection by phage then causes a respective decrease in bacterial abundance, observable in our case study in week 48 to 96 (**Figure 20A**). It has been shown that these dynamics occur during early childhood (Koenig et al., 2011; Sharon et al., 2013), however we show here that they persist into adolescence as this MAREEN patient was aged 14 at baseline. It is suggested that these dynamics are not as commonly observable in healthy adults, in whom lysogeny of phage dominates (Mirzaei & Maurice, 2017). The dynamics observed between *Escherichia-Shigella sp.* and *Enterobacteria* phage mEpX1 (**Figure 18**) do not resemble that of KTW. Instead, they may be an example of the recently proposed piggyback-the-winner (PTW) model, which predicts that lysogeny may dominate when bacteria are abundant and are growing rapidly (Mirzaei & Maurice, 2017).

Phages affect the fitness and diversity of bacteria through direct killing of their prey, but also through horizontal gene transfer between bacteria. For example, phage isolated from cattle were found to be capable of transduction of genes related to antibiotic resistance to an isolate of *Salmonella typhimurium* (Y. Zhang & LeJeune, 2008). Most phages in the gut are stably integrated as latent prophage (Mirzaei & Maurice, 2017) in bacterial genomes. However, when stimulated by environmental triggers they may become lytic. A 2006 study showed that β -lactam antibiotic treatment stimulated prophage induction in *S. aureus* bacteria. This led to replication of lytic phage and a high-frequency transfer of staphylococcal pathogenicity genes, suggesting that antibiotics may promote spread of virulence factors through bacteriophage-mediated transduction (Maiques et al., 2006). In theory, phages could change the functional properties of the microbiota through gene transfer, impacting host metabolism and immunity.

Here, we describe interactions of phage that infect *Enterococcus faecalis* and *Escherichia coli*. Phage targeting both of these bacterial species have been investigated for phage therapy usage *in vivo*. Galtier and colleagues targeted adherent invasive *E. coli* (AIEC) strains that are frequently dominant in CD mucosa with virulent phages. They found that oral administration of three virulent *E. coli* phage significantly decreased the number of AIEC in stool and mucosal biopsies (Galtier et al., 2017). In addition, the phage cocktail also reduced colitis symptoms in conventional mice over a two-week period. Gelman et al. (2018) studied the efficacy of an *E. faecalis* phage cocktail in reducing severity of septic murine peritonitis. They demonstrated that a single injection of the two-phage cocktail resulted in a complete reversal a mortality trend caused by antibiotic resistant *E. faecalis*, without any harm to the commensal microbiome (Gelman et al., 2018). There are many pharmacological, safety, and feasibility barriers that must be overcome for phage therapy to emerge, but these initial investigations are promising for their development.

6.6 LIMITATIONS AND FUTURE DIRECTIONS

Our study is subject to several limitations. Although many clinical metadata are collected by the MAREEN study investigators, it is difficult to fully capture all potentially confounding factors such as diet into our analysis. There were also several missing metadata and stool samples due to study design and patient compliance. For example, there was serology data for only half of the stool samples since patients did not give a blood sample at each clinic visit. Our bioinformatic analyses are also limited to the analysis of stool rather than biopsy, which can provide insight on microbial interactions at the mucosal surface.

The limitations of our viral stool enrichment protocol were briefly discussed in 6.1.1. Another source of error from the VLP enrichment may arise from different technicians carrying out the protocol. We also observed a faulty Phi29 polymerase, which resulted in dozens of samples having to be re-processed before sequencing. A robust assessment of VLP protocols like NetoVIR (Conceição-Neto et al., 2015) should be carried out before proceeding with a protocol that works best for stool. This protocol was

also limited to the analysis of DNA viruses, which is not all encompassing of all viruses in the gut. Infection with RNA viruses like norovirus have been thought to alter the disease progression of IBD (Khan et al., 2009), and mice studies have shown that infection with murine norovirus in mice with mutations in CD susceptibility gene *Atg16L1* results in a CD-like pathology (Cadwell et al., 2010). Our viral enrichment also attempts to isolate only viral particles, which leaves out the opportunity to analyze prophage stably integrated into bacterial genomes.

There are several bioinformatic limitations surrounding viral MGS analysis, such as limited database sizes, prophage analysis, and sequence-independent approaches. The integrated microbial genome/virus (IMG/VR) system is an online platform dedicated to viral genomics with a growing database several magnitudes larger than the RefSeq viral database used in this study. As of November 2018, IMG/VR contains over 760,000 viral genomes clustered into viral operational taxonomic units (vOTUs) for simplified analysis (Paez-Espino et al., 2018). Other resources like the pVOG database are assisting with the improvement of viral sequence databases (Grazziotin, Koonin, & Kristensen, 2017). PHASTER is an online server that facilitates the rapid annotation of prophage sequences within bacterial genomes (Arndt et al., 2016). Future analyses of the gut virome should take advantage of these tools to maximize annotation of viral sequences, especially of prophage sequences that may have been missed by through traditional viromics analyses.

A limitation of sequencing-based microbiome analysis is that it quantifies microbial taxa as a fraction of what was found in a sample. These relative abundance data are informative but lack the power of determining the extent or directionality of changes in microbial abundance. A recent study that profiled the microbiome using quantitative methods measured absolute bacterial load in stool samples using quantitative PCR and fluorescence-activated cell sorting (FACS) (Vandeputte et al., 2017). They found an overall reduced cell count of bacteria in CD patients compared to controls, and that *Bacteroides* is higher in CD patients when using a relative abundance approach, but not a quantitative approach. Techniques assessing absolute bacterial load may become more mainstream in microbiome analyses to gain an improved understanding of microbial shifts in the gut.

Bioinformatic methods themselves are continually improving, in addition to growing sequence databases. The recently released tool MARVEL was shown to achieve better performance than VirSorter and VirFinder on recall of true positive viral sequences (Amgarten et al., 2018). This longitudinal study would also benefit from a bioinformatic method that can handle time series data. Ananke is a software package that clusters marker-gene data based on their time series profiles, facilitating the analysis of longitudinal microbiome sequencing data (Hall, Rohwer, Perrie, McMahon, & Beiko, 2017). Adapting Ananke for use of viral MGS reads may highlight important longitudinal associations that were missed in the present analysis.

6.7 CONCLUSIONS

In this study, we sought out to characterize the role of bacteriophages in the gut microbiome of children undergoing treatment for Crohn's Disease. We made novel associations with diversity of the gut virome and clinical metadata. Through machine learning models, we also identified both viral and bacterial species that may be important in achieving remission after treatment with EEN. Our integrated examination of the gut virome and bacteriome identified specific bacteria-phage dynamics for the first time in pediatric CD. This analysis of the gut virome provides insight into bacteriophage that may be important for disease progression of CD and may lay the foundation for future phage therapy work. With continually improving bioinformatic methods and sequencing technologies, precision medicine focused on the gut microbiome is getting closer to a possibility.

REFERENCES

- Aagaard, K., Ma, J., Antony, K. M., Ganu, R., Petrosino, J., & Versalovic, J. (2014). The Placenta Harbors a Unique Microbiome. *Science Translational Medicine*, 6(237), 237ra65-237ra65. <https://doi.org/10.1126/SCITRANSLMED.3008599>
- Akobeng, A. K., & Gardener, E. (2005). Oral 5-aminosalicylic acid for maintenance of medically-induced remission in Crohn's disease. In A. K. Akobeng (Ed.), *Cochrane Database of Systematic Reviews*. Chichester, UK: John Wiley & Sons, Ltd. <https://doi.org/10.1002/14651858.CD003715.pub2>
- Al Nabhani, Z., Dietrich, G., Hugot, J.-P., & Barreau, F. (2017). Nod2: The intestinal gate keeper. *PLOS Pathogens*, 13(3), e1006177. <https://doi.org/10.1371/journal.ppat.1006177>
- Althani, A., Marei, H. E., Hamdi, W. S., Nasrallah, G. K., El Zowalaty, M. E., Al Khdor, S., ... Abdel-Aziz, H. (2015). Human Microbiome and Its Association With Health and Diseases. *Journal of Cellular Physiology*, n/a-n/a. <https://doi.org/10.1002/jcp.25284>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Amgarten, D., Braga, L. P. P., da Silva, A. M., & Setubal, J. C. (2018). MARVEL, a Tool for Prediction of Bacteriophage Sequences in Metagenomic Bins. *Frontiers in Genetics*, 9, 304. <https://doi.org/10.3389/fgene.2018.00304>
- Amir, A., McDonald, D., Navas-Molina, J. A., Kopylova, E., Morton, J. T., Zech Xu, Z., ... Knight, R. (2017). Deblur Rapidly Resolves Single-Nucleotide Community Sequence Patterns. *MSystems*, 2(2), e00191-16. <https://doi.org/10.1128/mSystems.00191-16>
- Amre, D. K., D'Souza, S., Morgan, K., Seidman, G., Lambrette, P., Grimard, G., ... Seidman, E. G. (2007). Imbalances in Dietary Consumption of Fatty Acids, Vegetables, and Fruits Are Associated With Risk for Crohn's Disease in Children. *The American Journal of Gastroenterology*, 102(9), 2016–2025. <https://doi.org/10.1111/j.1572-0241.2007.01411.x>
- Ananthkrishnan, A. N. (2015). Epidemiology and risk factors for IBD. *Nature Reviews Gastroenterology & Hepatology*, 12(4), 205–217. <https://doi.org/10.1038/nrgastro.2015.34>
- Ananthkrishnan, A. N., Higuchi, L. M., Huang, E. S., Khalili, H., Richter, J. M., Fuchs, C. S., & Chan, A. T. (2012). Aspirin, nonsteroidal anti-inflammatory drug use, and risk for Crohn disease and ulcerative colitis: a cohort study. *Annals of Internal Medicine*, 156(5), 350–359. <https://doi.org/10.7326/0003-4819-156-5-201203060-00007>

- Ananthakrishnan, A. N., Khalili, H., Higuchi, L. M., Bao, Y., Korzenik, J. R., Giovannucci, E. L., ... Chan, A. T. (2012). Higher Predicted Vitamin D Status Is Associated With Reduced Risk of Crohn's Disease. *Gastroenterology*, *142*(3), 482–489. <https://doi.org/10.1053/j.gastro.2011.11.040>
- Ananthakrishnan, A. N., Khalili, H., Konijeti, G. G., Higuchi, L. M., de Silva, P., Korzenik, J. R., ... Chan, A. T. (2013). A prospective study of long-term intake of dietary fiber and risk of {Crohn}'s disease and ulcerative colitis. *Gastroenterology*, *145*(5), 970–977. <https://doi.org/10.1053/j.gastro.2013.07.050>
- Anderson, C. L., Sullivan, M. B., & Fernando, S. C. (2017). Dietary energy drives the dynamic response of bovine rumen viral communities. *Microbiome*, *5*(1), 155. <https://doi.org/10.1186/s40168-017-0374-3>
- Anderson, M. J. (2001). A new method for non-parametric multivariate analysis of variance. *Austral Ecology*, *26*(1), 32–46. <https://doi.org/10.1111/j.1442-9993.2001.01070.pp.x>
- Andoh, A., Imaeda, H., Aomatsu, T., Inatomi, O., Bamba, S., Sasaki, M., ... Fujiyama, Y. (2011). Comparison of the fecal microbiota profiles between ulcerative colitis and {Crohn}'s disease using terminal restriction fragment length polymorphism analysis. *Journal of Gastroenterology*, *46*(4), 479–486. <https://doi.org/10.1007/s00535-010-0368-4>
- Aparicio-Pagés, M. N., Verspaget, H. W., Hafkenschied, J. C., Crama-Bohbouth, G. E., Peña, A. S., Weterman, I. T., & Lamers, H. W. (1990). Inhibition of cell mediated cytotoxicity by sulphasalazine: effect of in vivo treatment with 5-aminosalicylic acid and sulphasalazine on in vitro natural killer cell activity. *Gut*, *31*(9), 1030–1032. <https://doi.org/10.1136/GUT.31.9.1030>
- Arndt, D., Grant, J. R., Marcu, A., Sajed, T., Pon, A., Liang, Y., & Wishart, D. S. (2016). PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic Acids Research*, *44*(W1), W16–21. <https://doi.org/10.1093/nar/gkw387>
- Atreya, I., & Neurath, M. F. (2008). Azathioprine in inflammatory bowel disease: Improved molecular insights and resulting clinical implications. *Expert Review of Gastroenterology and Hepatology*, *2*(1), 23–34. <https://doi.org/10.1586/17474124.2.1.23>
- Aujnarain, A., Mack, D. R., & Benchimol, E. I. (2013). The Role of the Environment in the Development of Pediatric Inflammatory Bowel Disease. *Current Gastroenterology Reports*, *15*(6), 326. <https://doi.org/10.1007/s11894-013-0326-4>
- Bantel, H., Berg, C., Vieth, M., Stolte, M., Kruis, W., & Schulze-Osthoff, K. (2000). Mesalazine inhibits activation of transcription factor NF-kappaB in inflamed mucosa of patients with ulcerative colitis. *The American Journal of Gastroenterology*, *95*(12), 3452–3457. <https://doi.org/10.1111/j.1572-0241.2000.03360.x>

- Barr, J. J., Auro, R., Furlan, M., Whiteson, K. L., Erb, M. L., Pogliano, J., ... Rohwer, F. (2013). Bacteriophage adhering to mucus provide a non-host-derived immunity. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(26). <https://doi.org/10.1073/pnas.1305923110>
- Benchimol, E. I., Bernstein, C. N., Bitton, A., Carroll, M. W., Singh, H., Otley, A. R., ... Kaplan, G. G. (2017). Trends in Epidemiology of Pediatric Inflammatory Bowel Disease in Canada: Distributed Network Analysis of Multiple Population-Based Provincial Health Administrative Databases. *The American Journal of Gastroenterology*. <https://doi.org/10.1038/ajg.2017.97>
- Benchimol, E. I., Fortinsky, K. J., Gozdyra, P., Van den Heuvel, M., Van Limbergen, J., & Griffiths, A. M. (2011). Epidemiology of pediatric inflammatory bowel disease: a systematic review of international trends. *Inflammatory Bowel Diseases*, *17*(1), 423–439. <https://doi.org/10.1002/ibd.21349>
- Berni Canani, R., Terrin, G., Rapacciuolo, L., Miele, E., Siani, M. C., Puzone, C., ... Troncone, R. (2008). Faecal calprotectin as reliable non-invasive marker to assess the severity of mucosal inflammation in children with inflammatory bowel disease. *Digestive and Liver Disease*, *40*(7), 547–553. <https://doi.org/10.1016/J.DLD.2008.01.017>
- Billett, H. H. (1990). *Hemoglobin and Hematocrit. Clinical Methods: The History, Physical, and Laboratory Examinations*. Butterworths. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21250102>
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C., Al-Ghalith, G. A., ... Caporaso, J. G. (2018). QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. <https://doi.org/10.7287/peerj.preprints.27295v1>
- Bousvaros, A., Antonioli, D. A., Colletti, R. B., Dubinsky, M. C., Glickman, J. N., Gold, B. D., ... Russo, P. A. (2007). Differentiating Ulcerative Colitis from Crohn Disease in Children and Young Adults: Report of a Working Group of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition and the Crohn's and Colitis Foundation of America. *Journal of Pediatric Gastroenterology and Nutrition*, *44*(5), 653–674. <https://doi.org/10.1097/MPG.0b013e31805563f3>
- Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breitbart, M., Haynes, M., Kelley, S., Angly, F., Edwards, R. A., Felts, B., ... Rohwer, F. (2008). Viral diversity and dynamics in an infant gut. *Research in Microbiology*, *159*(5), 367–373. <https://doi.org/10.1016/J.RESMIC.2008.04.006>

- Breitbart, M., Hewson, I., Felts, B., Mahaffy, J. M., Nulton, J., Salamon, P., & Rohwer, F. (2003). Metagenomic analyses of an uncultured viral community from human feces. *Journal of Bacteriology*, *185*(20), 6220–6223. <https://doi.org/10.1128/JB.185.20.6220-6223.2003>
- Browne, H. P., Forster, S. C., Anonye, B. O., Kumar, N., Neville, B. A., Stares, M. D., ... Lawley, T. D. (2016). Culturing of ‘unculturable’ human microbiota reveals novel taxa and extensive sporulation. *Nature*, *533*(7604), 543–546. <https://doi.org/10.1038/nature17645>
- Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. <https://doi.org/10.1038/nmeth.3176>
- Butel, J. S. (2012). Patterns of polyomavirus SV40 infections and associated cancers in humans: a model. *Current Opinion in Virology*, *2*(4), 508–514. <https://doi.org/10.1016/j.coviro.2012.06.004>
- Cadwell, K., Patel, K. K., Maloney, N. S., Liu, T.-C., Ng, A. C. Y., Storer, C. E., ... Virgin, H. W. (2010). Virus-Plus-Susceptibility Gene Interaction Determines Crohn’s Disease Gene Atg16L1 Phenotypes in Intestine. *Cell*, *141*(7), 1135–1145. <https://doi.org/10.1016/j.cell.2010.05.009>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. <https://doi.org/10.1038/nmeth.3869>
- Cantorna, M. T., Munsick, C., Bemiss, C., & Mahon, B. D. (2000). 1,25-Dihydroxycholecalciferol Prevents and Ameliorates Symptoms of Experimental Murine Inflammatory Bowel Disease. *The Journal of Nutrition*, *130*(11), 2648–2652. <https://doi.org/10.1093/jn/130.11.2648>
- Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., ... Knight, R. (2010). {QIIME} allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335–336. <https://doi.org/10.1038/nmeth.f.303>
- Cezard, J.-P., Munck, A., Mouterde, O., Morali, A., Lenaerts, C., Lachaux, A., ... Mary, J.-Y. (2009). Prevention of relapse by mesalazine (Pentasa) in pediatric Crohn’s disease: a multicenter, double-blind, randomized, placebo-controlled trial. *Gastroenterologie Clinique et Biologique*, *33*(1 Pt 1), 31–40. <https://doi.org/10.1016/j.gcb.2008.07.007>
- Chan, E. S. L., & Cronstein, B. N. (2013). Mechanisms of action of methotrexate. *Bulletin of the Hospital for Joint Diseases*, *71*(Suppl 1), S5-8. [https://doi.org/10.1016/S0162-3109\(00\)00189-2](https://doi.org/10.1016/S0162-3109(00)00189-2)
- Cho, J. H. (2008). The genetics and immunopathogenesis of inflammatory bowel disease. *Nature Reviews Immunology*, *8*(6), 458–466. <https://doi.org/10.1038/nri2340>

- Clerici, C., & Matthay, M. A. (2003). Transforming growth factor- α 1 regulates lung epithelial barrier function and fluid transport. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 285(6), L1190–L1191. <https://doi.org/10.1152/ajplung.00230.2003>
- Comeau, A. M., Douglas, G. M., & Langille, M. G. I. (2017). Microbiome Helper: a Custom and Streamlined Workflow for Microbiome Research. *MSystems*, 2(1), e00127-16. <https://doi.org/10.1128/mSystems.00127-16>
- Comeau, A. M., Hatfull, G. F., Krisch, H. M., Lindell, D., Mann, N. H., & Prangishvili, D. (2008). Exploring the prokaryotic virosphere. *Research in Microbiology*, 159(5), 306–313. <https://doi.org/10.1016/J.RESMIC.2008.05.001>
- Conceição-Neto, N., Zeller, M., Lefrère, H., De Bruyn, P., Beller, L., Deboutte, W., ... Matthijnssens, J. (2015). Modular approach to customise sample preparation procedures for viral metagenomics: a reproducible protocol for virome analysis. *Scientific Reports*, 5(1), 16532. <https://doi.org/10.1038/srep16532>
- Connors, J., Basseri, S., Grant, A., Giffin, N., Mahdi, G., Noble, A., ... Van Limbergen, J. (2017). Exclusive enteral nutrition therapy in pediatric Crohn's disease results in long-term avoidance of corticosteroids: results of a propensity-score matched cohort analysis. *Journal of Crohn's and Colitis*, 1–8. <https://doi.org/10.1093/ecco-jcc/jjx060>
- Consortium, H. M. P. (2012). Structure, function and diversity of the healthy human microbiome. *Nature*, 486(7402), 207–214. <https://doi.org/10.1038/nature11234>
- Cornuault, J. K., Petit, M.-A., Mariadassou, M., Benevides, L., Moncaut, E., Langella, P., ... De Paepe, M. (2018). Phages infecting *Faecalibacterium prausnitzii* belong to novel viral genera that help to decipher intestinal viromes. *Microbiome*, 6(1), 65. <https://doi.org/10.1186/s40168-018-0452-1>
- David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., ... Turnbaugh, P. J. (2014). Diet rapidly and reproducibly alters the human gut microbiome. *Nature*, 505(7484), 559–563. <https://doi.org/10.1038/nature12820>
- Day, A. S., Whitten, K. E., Lemberg, D. A., Clarkson, C., Vitug-Sales, M., Jackson, R., & Bohane, T. D. (2006). Exclusive enteral feeding as primary therapy for Crohn's disease in Australian children and adolescents: A feasible and effective approach. *Journal of Gastroenterology and Hepatology*, 21(10), 1609–1614. <https://doi.org/10.1111/j.1440-1746.2006.04294.x>
- De Vadder, F., Kovatcheva-Datchary, P., Goncalves, D., Vinera, J., Zitoun, C., Duchampt, A., ... Mithieux, G. (2014). Microbiota-Generated Metabolites Promote Metabolic Benefits via Gut-Brain Neural Circuits. *Cell*, 156(1–2), 84–96. <https://doi.org/10.1016/J.CELL.2013.12.016>

- de Jong, N. S. H., Leach, S. T., & Day, A. S. (2007). Polymeric Formula Has Direct Anti-Inflammatory Effects on Enterocytes in an in Vitro Model of Intestinal Inflammation. *Digestive Diseases and Sciences*, 52(9), 2029–2036. <https://doi.org/10.1007/s10620-006-9449-x>
- De Lange, K. M., Moutsianas, L., Lee, J. C., Lamb, C. A., Luo, Y., Kennedy, N. A., ... Barrett, J. C. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nature Genetics*, 49(2), 256–261. <https://doi.org/10.1038/ng.3760>
- de Souza, H. S. P., & Fiocchi, C. (2016). Immunopathogenesis of IBD: current state of the art. *Nature Reviews Gastroenterology & Hepatology*, 13(1), 13–27. <https://doi.org/10.1038/nrgastro.2015.186>
- Desai, M. S., Seekatz, A. M., Koropatkin, N. M., Kamada, N., Hickey, C. A., Wolter, M., ... Martens, E. C. (2016). A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility. *Cell*, 167(5), 1339–1353.e21. <https://doi.org/10.1016/J.CELL.2016.10.043>
- Dethlefsen, L., Huse, S., Sogin, M. L., & Relman, D. A. (2008). The Pervasive Effects of an Antibiotic on the Human Gut Microbiota, as Revealed by Deep 16S rRNA Sequencing. *PLoS Biology*, 6(11), e280. <https://doi.org/10.1371/journal.pbio.0060280>
- Dicksved, J., Halfvarson, J., Rosenquist, M., Järnerot, G., Tysk, C., Apajalahti, J., ... Jansson, J. K. (2008). Molecular analysis of the gut microbiota of identical twins with Crohn's disease. *The ISME Journal*, 2(7), 716–727. <https://doi.org/10.1038/ismej.2008.37>
- Duerkop, B. A., Clements, C. V., Rollins, D., Rodrigues, J. L. M., & Hooper, L. V. (2012). A composite bacteriophage alters colonization by an intestinal commensal bacterium. *Proceedings of the National Academy of Sciences*, 109(43), 17621–17626. <https://doi.org/10.1073/pnas.1206136109>
- Duerkop, B. A., Kleiner, M., Paez-Espino, D., Zhu, W., Bushnell, B., Hassell, B., ... Hooper, L. V. (2018). Murine colitis reveals a disease-associated bacteriophage community. *Nature Microbiology*, 3(September). <https://doi.org/10.1038/s41564-018-0210-y>
- Dunn, K. A., Moore-Connors, J., MacIntyre, B., Stadnyk, A. W., Thomas, N. A., Noble, A., ... Van Limbergen, J. (2016). Early Changes in Microbial Community Structure Are Associated with Sustained Remission After Nutritional Treatment of Pediatric Crohn's Disease. *Inflammatory Bowel Diseases*, 22(12), 2853–2862. <https://doi.org/10.1097/MIB.0000000000000956>

- Dutilh, B. E., Cassman, N., McNair, K., Sanchez, S. E., Silva, G. G. Z., Boling, L., ... Edwards, R. A. (2014). A highly abundant bacteriophage discovered in the unknown sequences of human faecal metagenomes. *Nature Communications*, 5. <https://doi.org/10.1038/ncomms5498>
- Ellinghaus, D., Jostins, L., Spain, S. L., Cortes, A., Bethune, J., Han, B., ... Franke, A. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nature Genetics*, 48(5), 510–518. <https://doi.org/10.1038/ng.3528>
- Escuro, A. A., & Hummell, A. C. (2016). Enteral Formulas in Nutrition Support Practice. *Nutrition in Clinical Practice*, 31(6), 709–722. <https://doi.org/10.1177/0884533616668492>
- Fagerberg, U. L., Lööf, L., Myrdal, U., Hansson, L.-O., & Finkel, Y. (2005). Colorectal inflammation is well predicted by fecal calprotectin in children with gastrointestinal symptoms. *Journal of Pediatric Gastroenterology and Nutrition*, 40(4), 450–455. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/15795593>
- Fell, J. M., Paintin, M., Arnaud-Battandier, F., Beattie, R. M., Hollis, A., Kitching, P., ... Walker-Smith, J. A. (2000). Mucosal healing and a fall in mucosal pro-inflammatory cytokine mRNA induced by a specific oral polymeric diet in paediatric Crohn's disease. *Alimentary Pharmacology & Therapeutics*, 14(3), 281–289. <https://doi.org/10.1046/J.1365-2036.2000.00707.X>
- Fernandes, M. A., Verstraete, S. G., Phan, T. G., Deng, X., Stekol, E., LaMere, B., ... Delwart, E. (2018). Enteric Virome and Bacterial Microbiota in Children with Ulcerative Colitis and Crohn's Disease. *Journal of Pediatric Gastroenterology and Nutrition*. <https://doi.org/10.1097/MPG.0000000000002140>
- Fjellbirkeland, L., Cambier, S., Broaddus, V. C., Hill, A., Brunetta, P., Dolganov, G., ... Nishimura, S. L. (2003). Integrin $\alpha\beta 8$ -mediated activation of transforming growth factor- β inhibits human airway epithelial proliferation in intact bronchial tissue. *American Journal of Pathology*, 163(2), 533–542. [https://doi.org/10.1016/S0002-9440\(10\)63681-4](https://doi.org/10.1016/S0002-9440(10)63681-4)
- Friswell, M., Campbell, B., & Rhodes, J. (2010). The Role of Bacteria in the Pathogenesis of Inflammatory Bowel Disease. *Gut and Liver*, 4(3), 295–306. <https://doi.org/10.5009/gnl.2010.4.3.295>
- Froicu, M., & Cantorna, M. T. (2007). Vitamin D and the vitamin D receptor are critical for control of the innate immune response to colonic injury. *BMC Immunology*, 8(1), 5. <https://doi.org/10.1186/1471-2172-8-5>
- Froicu, M., Zhu, Y., & Cantorna, M. T. (2006). Vitamin D receptor is required to control gastrointestinal immunity in IL-10 knockout mice. *Immunology*, 117(3), 310–318. <https://doi.org/10.1111/j.1365-2567.2005.02290.x>

- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. (2012). CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150–3152. <https://doi.org/10.1093/bioinformatics/bts565>
- Galtier, M., De Sordi, L., Sivignon, A., de Vallée, A., Maura, D., Neut, C., ... Debarbieux, L. (2017). Bacteriophages targeting adherent invasive *Escherichia coli* strains as a promising new treatment for Crohn's disease. *Journal of Crohn's and Colitis*, 840–847. <https://doi.org/10.1093/ecco-jcc/jjw224>
- Galvez, J., Rodríguez-Cabezas, M. E., & Zarzuelo, A. (2005). Effects of dietary fiber on inflammatory bowel disease. *Molecular Nutrition & Food Research*, *49*(6), 601–608. <https://doi.org/10.1002/mnfr.200500013>
- Gardiner, B. J., Tai, A. Y., Kotsanas, D., Francis, M. J., Roberts, S. A., Ballard, S. A., ... Korman, T. M. (2015). Clinical and microbiological characteristics of *Entergerthella lenta* bacteremia. *Journal of Clinical Microbiology*, *53*(2), 626–635. <https://doi.org/10.1128/JCM.02926-14>
- Gelman, D., Beyth, S., Lerer, V., Adler, K., Poradosu-Cohen, R., Copenhagen-Glazer, S., & Hazan, R. (2018). Combined bacteriophages and antibiotics as an efficient therapy against VRE *Enterococcus faecalis* in a mouse model. *Research in Microbiology*, 1–9. <https://doi.org/10.1016/j.resmic.2018.04.008>
- Gerasimidis, K., Bertz, M., Hanske, L., Junick, J., Biskou, O., Aguilera, M., ... Edwards, C. A. (2014). Decline in presumptively protective gut bacterial species and metabolites are paradoxically associated with disease improvement in pediatric Crohn's disease during enteral nutrition. *Inflammatory Bowel Diseases*, *20*(5), 861–871. <https://doi.org/10.1097/MIB.0000000000000023>
- Gevers, D., Kugathasan, S., Denson, L. A., Vázquez-Baeza, Y., Van Treuren, W., Ren, B., ... Xavier, R. J. (2014). The Treatment-Naive Microbiome in New-Onset Crohn's Disease. *Cell Host & Microbe*, *15*(3), 382–392. <https://doi.org/10.1016/j.chom.2014.02.005>
- Gilbert, J. A., Blaser, M. J., Caporaso, J. G., Jansson, J. K., Lynch, S. V., & Knight, R. (2018). Current understanding of the human microbiome. *Nature Medicine*, *24*(4), 392–400. <https://doi.org/10.1038/nm.4517>
- Gorelik, L., Constant, S., & Flavell, R. A. (2002). Mechanism of transforming growth factor beta-induced inhibition of T helper type 1 differentiation. *The Journal of Experimental Medicine*, *195*(11), 1499–1505. <https://doi.org/10.1084/JEM.20012076>
- Gorelik, L., Fields, P. E., & Flavell, R. A. (2000). Cutting edge: TGF-beta inhibits Th type 2 development through inhibition of GATA-3 expression. *Journal of Immunology (Baltimore, Md. : 1950)*, *165*(9), 4773–4777. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/11045997>

- Górski, A., Dąbrowska, K., Międzybrodzki, R., Weber-Dąbrowska, B., Łusiak-Szelachowska, M., Jończyk-Matysiak, E., & Borysowski, J. (2017). Phages and immunomodulation. *Future Microbiol*, (10), 905–914. <https://doi.org/10.2217/fmb-2017-0049>
- Grazziotin, A. L., Koonin, E. V., & Kristensen, D. M. (2017). Prokaryotic Virus Orthologous Groups (pVOGs): a resource for comparative genomics and protein family annotation. *Nucleic Acids Research*, 45(D1), D491–D498. <https://doi.org/10.1093/nar/gkw975>
- Greenfield, S. M., Hamblin, A. S., Shakoor, Z. S., Teare, J. P., Punchard, N. A., & Thompson, R. P. (1993). Inhibition of leucocyte adhesion molecule upregulation by tumor necrosis factor alpha: a novel mechanism of action of sulphasalazine. *Gut*, 34(2), 252–256. <https://doi.org/10.1136/GUT.34.2.252>
- Griffiths, A., Koletzko, S., Sylvester, F., Marcon, M., & Sherman, P. (1993). Slow-release 5-aminosalicylic acid therapy in children with small intestinal Crohn's disease. *Journal of Pediatric Gastroenterology and Nutrition*, 17(2), 186–192. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/8229546>
- Hall, M. W., Rohwer, R. R., Perrie, J., McMahon, K. D., & Beiko, R. G. (2017). Ananke: Temporal clustering reveals ecological dynamics of microbial communities. *PeerJ Preprints*, 5, 2879v1. <https://doi.org/10.7287/PEERJ.PREPRINTS.2879V1>
- Hansen, T., & Duerksen, D. R. (2018). Enteral nutrition in the management of pediatric and adult crohn's disease. *Nutrients*, 10(5). <https://doi.org/10.3390/nu10050537>
- Healey, G. R., Murphy, R., Brough, L., Butts, C. A., & Coad, J. (2017). Interindividual variability in gut microbiota and host response to dietary interventions. *Nutrition Reviews*, 75(12), 1059–1080. <https://doi.org/10.1093/nutrit/nux062>
- Henderson, P., Hansen, R., Cameron, F. L., Gerasimidis, K., Rogers, P., Bisset, M. W., ... Wilson, D. C. (2012). Rising incidence of pediatric inflammatory bowel disease in Scotland*. *Inflammatory Bowel Diseases*, 18(6), 999–1005. <https://doi.org/10.1002/ibd.21797>
- Holscher, H. D., Bauer, L. L., Gourineni, V., Pelkman, C. L., Fahey, G. C., & Swanson, K. S. (2015). Agave Inulin Supplementation Affects the Fecal Microbiota of Healthy Adults Participating in a Randomized, Double-Blind, Placebo-Controlled, Crossover Trial. *The Journal of Nutrition*, 145(9), 2025–2032. <https://doi.org/10.3945/jn.115.217331>
- Huson, D. H., Beier, S., Flade, I., G??rska, A., El-Hadidi, M., Mitra, S., ... Tappu, R. (2016). MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLoS Computational Biology*, 12(6), 1–12. <https://doi.org/10.1371/journal.pcbi.1004957>

- Illumina. (2017). *An introduction to Next-Generation Sequencing Technology*. Retrieved from https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf
- Imhann, F., Vich Vila, A., Bonder, M. J., Fu, J., Gevers, D., Visschedijk, M. C., ... Weersma, R. K. (2018). Interplay of host genetics and gut microbiota underlying the onset and clinical presentation of inflammatory bowel disease. *Gut*, *67*(1), 108–119. <https://doi.org/10.1136/gutjnl-2016-312135>
- Jackson, M. A., Verdi, S., Maxan, M. E., Shin, C. M., Zierer, J., Bowyer, R. C. E., ... Steves, C. J. (2018). Gut microbiota associations with common diseases and prescription medications in a population-based cohort. *Nature Communications*, *9*(1), 1–8. <https://doi.org/10.1038/s41467-018-05184-7>
- Jones, R. B., Zhu, X., Moan, E., Murff, H. J., Ness, R. M., Seidner, D. L., ... Shrubsole, M. J. (2018). Inter-niche and inter-individual variation in gut microbial community assessment using stool, rectal swab, and mucosal samples. *Scientific Reports*, *8*(1), 4139. <https://doi.org/10.1038/s41598-018-22408-4>
- Joossens, M., Huys, G., Cnockaert, M., De Preter, V., Verbeke, K., Rutgeerts, P., ... Vermeire, S. (2011). Dysbiosis of the faecal microbiota in patients with Crohn's disease and their unaffected relatives. *Gut*, *60*(5), 631–637. <https://doi.org/10.1136/gut.2010.223263>
- Jostins, L., Ripke, S., Weersma, R. K., Duerr, R. H., McGovern, D. P., Hui, K. Y., ... Cho, J. H. (2012). Host–microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*, *491*(7422), 119–124. <https://doi.org/10.1038/nature11582>
- Kang, D. D., Froula, J., Egan, R., & Wang, Z. (2015). MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, *3*, e1165. <https://doi.org/10.7717/peerj.1165>
- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Khan, R. R., Lawson, A. D., Minnich, L. L., Martin, K., Nasir, A., Emmett, M. K., ... Udall, J. N. (2009). Gastrointestinal Norovirus Infection Associated With Exacerbation of Inflammatory Bowel Disease. *Journal of Pediatric Gastroenterology and Nutrition*, *48*(3), 328–333. <https://doi.org/10.1097/MPG.0b013e31818255cc>
- Kirschner, B. S. (2016). Indeterminate Colitis/Inflammatory Bowel Disease Unclassified (IBD-U). In *Textbook of Pediatric Gastroenterology, Hepatology and Nutrition* (pp. 335–340). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-17169-2_29

- Knight, R., Vrbanac, A., Taylor, B. C., Aksenov, A., Callewaert, C., Debelius, J., ... Dorrestein, P. C. (2018, July 23). Best practices for analysing microbiomes. *Nature Reviews Microbiology*. Nature Publishing Group. <https://doi.org/10.1038/s41579-018-0029-9>
- Koenig, J. E., Spor, A., Scalfone, N., Fricker, A. D., Stombaugh, J., Knight, R., ... Ley, R. E. (2011). Succession of microbial consortia in the developing infant gut microbiome. *Proceedings of the National Academy of Sciences*, *108*(Supplement 1), 4578–4585. <https://doi.org/10.1073/PNAS.1000081107>
- Kopylova, E., Noé, L., & Touzet, H. (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics*, *28*(24), 3211–3217. <https://doi.org/10.1093/bioinformatics/bts611>
- Kostic, A. D., Xavier, R. J., & Gevers, D. (2014). The Microbiome in Inflammatory Bowel Disease: Current Status and the Future Ahead. *Gastroenterology*, *146*(6), 1489–1499. <https://doi.org/10.1053/j.gastro.2014.02.009>
- Kronman, M. P., Zaoutis, T. E., Haynes, K., Feng, R., & Coffin, S. E. (2012). Antibiotic exposure and {IBD} development among children: a population-based cohort study. *Pediatrics*, *130*(4), e794--803. <https://doi.org/10.1542/peds.2011-3886>
- Kunzmann, S., Mantel, P.-Y., Wohlfahrt, J. G., Akdis, M., Blaser, K., & Schmidt-Weber, C. B. (2003). Histamine enhances TGF-beta1-mediated suppression of Th2 responses. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *17*(9), 1089–1095. <https://doi.org/10.1096/fj.02-1008com>
- Laevens, S., Gillis, M., LiPuma, J. J., Henry, D., Vandamme, P., Mahenthiralingam, E., ... Coenye, T. (2001). *Burkholderia ambifaria* sp. nov., a novel member of the *Burkholderia cepacia* complex including biocontrol and cystic fibrosis-related isolates. *International Journal of Systematic and Evolutionary Microbiology*, *51*(4), 1481–1490. <https://doi.org/10.1099/00207713-51-4-1481>
- Lahad, A., & Weiss, B. (2015). Current therapy of pediatric Crohn's disease. *World Journal of Gastrointestinal Pathophysiology*, *6*(2), 33–42. <https://doi.org/10.4291/wjgp.v6.i2.33>
- Langille, M. G. I., Zaneveld, J., Caporaso, J. G., McDonald, D., Knights, D., Reyes, J. A., ... Huttenhower, C. (2013). Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature Biotechnology*, *31*(9), 814–821. <https://doi.org/10.1038/nbt.2676>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. <https://doi.org/10.1038/nmeth.1923>

- LeBlanc, J. G., Milani, C., de Giori, G. S., Sesma, F., van Sinderen, D., & Ventura, M. (2013). Bacteria as vitamin suppliers to their host: a gut microbiota perspective. *Current Opinion in Biotechnology*, *24*(2), 160–168. <https://doi.org/10.1016/J.COPBIO.2012.08.005>
- Lepage, P., Colombet, J., Marteau, P., Sime-Ngando, T., Doré, J., & Leclerc, M. (2008). Dysbiosis in inflammatory bowel disease: a role for bacteriophages? *Gut*, *57*(3), 424–425. <https://doi.org/10.1136/gut.2007.134668>
- Li, F., Hullar, M. A. J., Schwarz, Y., & Lampe, J. W. (2009). Human Gut Bacterial Communities Are Altered by Addition of Cruciferous Vegetables to a Controlled Fruit- and Vegetable-Free Diet. *The Journal of Nutrition*, *139*(9), 1685–1691. <https://doi.org/10.3945/jn.109.108191>
- Li, Y. R., Li, J., Zhao, S. D., Bradfield, J. P., Mentch, F. D., Maggadottir, S. M., ... Hakonarson, H. (2015). Meta-analysis of shared genetic architecture across ten pediatric autoimmune diseases. *Nature Medicine*, *21*(9), 1018–1027. <https://doi.org/10.1038/nm.3933>
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*(3), 18–22. Retrieved from https://www.r-project.org/doc/Rnews/Rnews_2002-3.pdf
- Liu, J. Z., van Sommeren, S., Huang, H., Ng, S. C., Alberts, R., Takahashi, A., ... Weersma, R. K. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics*, *47*(9), 979–986. <https://doi.org/10.1038/ng.3359>
- Locey, K. J., & Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(21), 5970–5975. <https://doi.org/10.1073/pnas.1521291113>
- Lodes, M. J., Cong, Y., Elson, C. O., Mohamath, R., Landers, C. J., Targan, S. R., ... Hershberg, R. M. (2004). Bacterial flagellin is a dominant antigen in Crohn disease. *The Journal of Clinical Investigation*, *113*(9), 1296–1306. <https://doi.org/10.1172/JCI20295>
- MacDermott, R. P., Kane, M. G., Steele, L. L., & Stenson, W. F. (1986). Inhibition of cytotoxicity by sulfasalazine. I. Sulfasalazine inhibits spontaneous cell-mediated cytotoxicity by peripheral blood and intestinal mononuclear cells from control and inflammatory bowel disease patients. *Immunopharmacology*, *11*(2), 101–109. [https://doi.org/10.1016/0162-3109\(86\)90030-5](https://doi.org/10.1016/0162-3109(86)90030-5)
- MacLellan, A., Connors, J., Grant, S., Cahill, L., Langille, M., Van Limbergen, J., ... Van Limbergen, J. (2017). The Impact of Exclusive Enteral Nutrition (EEN) on the Gut Microbiome in Crohn’s Disease: A Review. *Nutrients*, *9*(5), 0447. <https://doi.org/10.3390/nu9050447>

- Mahid, S. S., Minor, K. S., Soto, R. E., Hornung, C. A., & Galandiuk, S. (2006). Smoking and inflammatory bowel disease: a meta-analysis. *Mayo Clinic Proceedings*, *81*(11), 1462–1471. <https://doi.org/10.4065/81.11.1462>
- Maier, L., Pruteanu, M., Kuhn, M., Zeller, G., Telzerow, A., Anderson, E. E., ... Typas, A. (2018). Extensive impact of non-antibiotic drugs on human gut bacteria. *Nature*, *555*(7698), 623–628. <https://doi.org/10.1038/nature25979>
- Maiques, E., Ubeda, C., Campoy, S., Salvador, N., Lasa, I., Novick, R. P., ... Penadés, J. R. (2006). beta-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *Journal of Bacteriology*, *188*(7), 2726–2729. <https://doi.org/10.1128/JB.188.7.2726-2729.2006>
- Majewska, J., Beta, W., Lecion, D., Hodyra-Stefaniak, K., Kłopot, A., Kaźmierczak, Z., ... Dąbrowska, K. (2015). Oral Application of T4 Phage Induces Weak Antibody Production in the Gut and in the Blood. *Viruses*, *7*(8), 4783–4799. <https://doi.org/10.3390/v7082845>
- Malmborg, P., Grahnquist, L., Lindholm, J., Montgomery, S., & Hildebrand, H. (2013). Increasing incidence of paediatric inflammatory bowel disease in northern Stockholm County, 2002-2007. *Journal of Pediatric Gastroenterology and Nutrition*, *57*(1), 29–34. <https://doi.org/10.1097/MPG.0b013e31828f21b4>
- Maltzman, J. S., & Koretzky, G. A. (2003). Azathioprine: old drug, new actions. *Journal of Clinical Investigation*, *111*(8), 1122–1124. <https://doi.org/10.1172/JCI18384>
- Manichanh, C., Rigottier-Gois, L., Bonnaud, E., Gloux, K., Pelletier, E., Frangeul, L., ... Dore, J. (2006). Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut*, *55*(2), 205–211. <https://doi.org/10.1136/gut.2005.073817>
- Manrique, P., Bolduc, B., Walk, S. T., van der Oost, J., de Vos, W. M., & Young, M. J. (2016). Healthy human gut phageome. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(37), 10400–10405. <https://doi.org/10.1073/pnas.1601060113>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. <https://doi.org/10.14806/ej.17.1.200>
- McGovern, D. P. B., Kugathasan, S., & Cho, J. H. (2015). Genetics of {Inflammatory} {Bowel} {Diseases}. *Gastroenterology*, *149*(5), 1163--1176.e2. <https://doi.org/10.1053/j.gastro.2015.08.001>
- Metzger, R., Krug, A., & Eisenächer, K. (2018). Enteric Virome Sensing—Its Role in Intestinal Homeostasis and Immunity. *Viruses*, *10*(4), 146. <https://doi.org/10.3390/v10040146>

- Mihara, T., Nishimura, Y., Shimizu, Y., Nishiyama, H., Yoshikawa, G., Uehara, H., ... Ogata, H. (2016). Linking Virus Genomes with Host Taxonomy. *Viruses*, 8(3), 66. <https://doi.org/10.3390/v8030066>
- Minot, S., Sinha, R., Chen, J., Li, H., Keilbaugh, S. A., Wu, G. D., ... Bushman, F. D. (2011). The human gut virome: inter-individual variation and dynamic response to diet. *Genome Research*, 21(10), 1616–1625. <https://doi.org/10.1101/gr.122705.111>
- Mirzaei, M. K., & Maurice, C. F. (2017). Menage a trois in the human gut: interactions between host, bacteria and phages. *Nature Reviews Microbiology*, 15(7), 397–408. <https://doi.org/10.1038/nrmicro.2017.30>
- Modi, S. R., Collins, J. J., & Relman, D. A. (2014). Antibiotics and the gut microbiota. *The Journal of Clinical Investigation*, 124(10), 4212–4218. <https://doi.org/10.1172/JCI72333>
- Moens, E., & Veldhoen, M. (2012). Epithelial barrier biology: good fences make good neighbours. *Immunology*, 135(1), 1–8. <https://doi.org/10.1111/j.1365-2567.2011.03506.x>
- Mohammadi, T., Reesink, H. W., Vandenbroucke-Grauls, C. M. J. E., & Savelkoul, P. H. M. (2005). Removal of contaminating DNA from commercial nucleic acid extraction kit reagents. *Journal of Microbiological Methods*, 61(2), 285–288. <https://doi.org/10.1016/J.MIMET.2004.11.018>
- Morgan, X. C., Tickle, T. L., Sokol, H., Gevers, D., Devaney, K. L., Ward, D. V., ... Huttenhower, C. (2012). Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biology*, 13(9), R79. <https://doi.org/10.1186/gb-2012-13-9-r79>
- Mow, W. S., Vasiliauskas, E. A., Lin, Y.-C., Fleshner, P. R., Papadakis, K. A., Taylor, K. D., ... Targan, S. R. (2004). Association of antibody responses to microbial antigens and complications of small bowel Crohn's disease. *Gastroenterology*, 126(2), 414–424. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14762777>
- Nakatsu, G., Zhou, H., Wu, W. K. K., Wong, S. H., Coker, O. O., Dai, Z., ... Yu, J. (2018). Alterations in Enteric Virome Are Associated With Colorectal Cancer and Survival Outcomes. *Gastroenterology*, 155(2), 529–541.e5. <https://doi.org/10.1053/j.gastro.2018.04.018>
- Neal, T. M., Winterbourn, C. C., & Vissers, M. C. M. (1987). Inhibition of neutrophil degranulation and superoxide production by sulfasalazine: Comparison with 5-aminosalicylic acid, sulfapyridine and olsalazine. *Biochemical Pharmacology*, 36(17), 2765–2768. [https://doi.org/10.1016/0006-2952\(87\)90262-0](https://doi.org/10.1016/0006-2952(87)90262-0)
- Nestlé Health. (2018). Modulen IBD®. Retrieved September 28, 2018, from <https://www.nestlehealthscience.ca/en/brands/modulen/modulen-ibd>

- Nguyen, S., Baker, K., Padman, B. S., Patwa, R., Dunstan, R. A., Weston, T. A., ... Barr, J. J. (2017). Bacteriophage Transcytosis Provides a Mechanism To Cross Epithelial Cell Layers. *MBio*, 8(6), e01874-17. <https://doi.org/10.1128/mBio.01874-17>
- Nielsen, O. H., Vainer, B., & Rask-Madsen, J. (2001). The treatment of inflammatory bowel disease with 6-mercaptopurine or azathioprine. *Alimentary Pharmacology and Therapeutics*, 15(11), 1699–1708. <https://doi.org/10.1046/j.1365-2036.2001.01102.x>
- Nishino, K., Nishida, A., Inoue, R., Kawada, Y., Ohno, M., Sakai, S., ... Andoh, A. (2018). Analysis of endoscopic brush samples identified mucosa-associated dysbiosis in inflammatory bowel disease. *Journal of Gastroenterology*, 53(1), 95–106. <https://doi.org/10.1007/s00535-017-1384-4>
- Nitzan, O., Elias, M., Peretz, A., & Saliba, W. (2016). Role of antibiotics for treatment of inflammatory bowel disease. *World Journal of Gastroenterology*, 22(3), 1078–1087. <https://doi.org/10.3748/wjg.v22.i3.1078>
- Nobrega, F. L., Vlot, M., de Jonge, P. A., Dreesens, L. L., Beaumont, H. J. E., Lavigne, R., ... Brouns, S. J. J. (2018). Targeting mechanisms of tailed bacteriophages. *Nature Reviews Microbiology*, 16(12), 760–773. <https://doi.org/10.1038/s41579-018-0070-8>
- Norman, J. M., Handley, S. A., Baldridge, M. T., Droit, L., Liu, C. Y., Keller, B. C., ... Virgin, H. W. (2015). Disease-Specific Alterations in the Enteric Virome in Inflammatory Bowel Disease. *Cell*, 160(3), 447–460. <https://doi.org/10.1016/j.cell.2015.01.002>
- Nurk, S., Meleshko, D., Korobeynikov, A., & Pevzner, P. A. (2017). metaSPAdes: a new versatile metagenomic assembler. *Genome Research*, 27(5), 824–834. <https://doi.org/10.1101/gr.213959.116>
- O'Hara, A. M., & Shanahan, F. (2006). The gut flora as a forgotten organ. *EMBO Reports*, 7(7), 688–693. <https://doi.org/10.1038/sj.embor.7400731>
- Ogilvie, L. A., & Jones, B. V. (2015). The human gut virome: A multifaceted majority. *Frontiers in Microbiology*, 6(SEP), 1–12. <https://doi.org/10.3389/fmicb.2015.00918>
- Owen, J., Punt, J., & Stranford, S. (2009). *Kuby Immunology* (Seventh). New York: W.H Freeman and Company.
- Paez-Espino, D., Roux, S., Chen, I.-M. A., Palaniappan, K., Ratner, A., Chu, K., ... Kyrpides, N. C. (2018). IMG/VR v.2.0: an integrated data management and analysis system for cultivated and environmental viral genomes. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gky1127>

- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., & Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Research*, *25*(7), 1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Parks, D. H., Tyson, G. W., Hugenholtz, P., & Beiko, R. G. (2014). STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics*, *30*(21), 3123–3124. <https://doi.org/10.1093/bioinformatics/btu494>
- Parmar, K. M., Gaikwad, S. L., Dhakephalkar, P. K., Kothari, R., & Singh, R. P. (2017). Intriguing Interaction of Bacteriophage-Host Association: An Understanding in the Era of Omics. *Frontiers in Microbiology*, *8*, 559. <https://doi.org/10.3389/fmicb.2017.00559>
- Pascal, V., Pozuelo, M., Borruel, N., Casellas, F., Campos, D., Santiago, A., ... Manichanh, C. (2017). A microbial signature for Crohn's disease. *Gut*, *66*(5), 813–822. <https://doi.org/10.1136/gutjnl-2016-313235>
- Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., ... Brockhurst, M. A. (2010). Antagonistic coevolution accelerates molecular evolution. *Nature*, *464*(7286), 275–278. <https://doi.org/10.1038/nature08798>
- Pelzer, E., Gomez-Arango, L. F., Barrett, H. L., & Nitert, M. D. (2017). Review: Maternal health and the placental microbiome. *Placenta*, *54*, 30–37. <https://doi.org/10.1016/J.PLACENTA.2016.12.003>
- Pérez-Brocal, V., García-López, R., Nos, P., Beltrán, B., Moret, I., & Moya, A. (2015a). Metagenomic Analysis of Crohn's Disease Patients Identifies Changes in the Virome and Microbiome Related to Disease Status and Therapy, and Detects Potential Interactions and Biomarkers. *Inflammatory Bowel Diseases*, *21*(11), 2515–2532. <https://doi.org/10.1097/MIB.0000000000000549>
- Pérez-Brocal, V., García-López, R., Nos, P., Beltrán, B., Moret, I., & Moya, A. (2015b). Metagenomic Analysis of Crohn's Disease Patients Identifies Changes in the Virome and Microbiome Related to Disease Status and Therapy, and Detects Potential Interactions and Biomarkers. *Inflammatory Bowel Diseases*, *21*(11), 2515–2532. <https://doi.org/10.1097/MIB.0000000000000549>
- Pérez-Brocal, V., García-López, R., Vázquez-Castellanos, J. F., Nos, P., Beltrán, B., Latorre, A., & Moya, A. (2013). Study of the viral and microbial communities associated with Crohn's disease: a metagenomic approach. *Clinical and Translational Gastroenterology*, *4*, e36. <https://doi.org/10.1038/ctg.2013.9>

- Perminow, G., Brackmann, S., Lyckander, L. G., Franke, A., Borthne, A., Rydning, A., ... The IBSEN-II Group, T. I.-I. (2009). A characterization in childhood inflammatory bowel disease, a new population-based inception cohort from South-Eastern Norway, 2005–07, showing increased incidence in Crohn's disease. *Scandinavian Journal of Gastroenterology*, 44(4), 446–456. <https://doi.org/10.1080/00365520802647434>
- Peyrin-Biroulet, L., Oussalah, A., & Bigard, M.-A. (2009). Crohn's disease: The hot hypothesis. *Medical Hypotheses*, 73(1), 94–96. <https://doi.org/10.1016/J.MEHY.2009.01.022>
- Prefontaine, E., Sutherland, L. R., Macdonald, J. K., & Cepoiu, M. (2009). Azathioprine or 6-mercaptopurine for maintenance of remission in Crohn's disease. *The Cochrane Database of Systematic Reviews*, (1), CD000067. <https://doi.org/10.1002/14651858.CD000067.pub2>
- Price, M. N., Dehal, P. S., & Arkin, A. P. (2009). FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. *Molecular Biology and Evolution*, 26(7), 1641–1650. <https://doi.org/10.1093/molbev/msp077>
- Qin, J., Li, R., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., ... Wang, J. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, 464(7285), 59–65. <https://doi.org/10.1038/nature08821>
- Qiu, Z., Yang, H., Rong, L., Ding, W., Chen, J., & Zhong, L. (2017). Targeted Metagenome Based Analyses Show Gut Microbial Diversity of Inflammatory Bowel Disease patients. *Indian Journal of Microbiology*, 57(3), 307–315. <https://doi.org/10.1007/s12088-017-0652-6>
- Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., ... Glöckner, F. O. (2012). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Research*, 41(D1), D590–D596. <https://doi.org/10.1093/nar/gks1219>
- Quince, C., Ijaz, U. Z., Loman, N., Eren, A. M., Saulnier, D., Russell, J., ... Gerasimidis, K. (2015). Extensive Modulation of the Fecal Metagenome in Children With Crohn's Disease During Exclusive Enteral Nutrition. *The American Journal of Gastroenterology*, 110(12), 1718–1729. <https://doi.org/10.1038/ajg.2015.357>
- Ren, J., Ahlgren, N. A., Lu, Y. Y., Fuhrman, J. A., & Sun, F. (2017). VirFinder: a novel k-mer based tool for identifying viral sequences from assembled metagenomic data. *Microbiome*, 5(1), 69. <https://doi.org/10.1186/s40168-017-0283-5>
- Reyes, A., Blanton, L. V., Cao, S., Zhao, G., Manary, M., Trehan, I., ... Gordon, J. I. (2015). Gut DNA viromes of Malawian twins discordant for severe acute malnutrition. *Proceedings of the National Academy of Sciences of the United States of America*, 112(38), 11941–11946. <https://doi.org/10.1073/pnas.1514285112>

- Reyes, A., Haynes, M., Hanson, N., Angly, F. E., Heath, A. C., Rohwer, F., & Gordon, J. I. (2010). Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature*, *466*(7304), 334–338. <https://doi.org/10.1038/nature09199>
- Rieder, F., Paul, G., Schnoy, E., Schleider, S., Wolf, A., Kamm, F., ... Klebl, F. (2014). Hemoglobin and hematocrit levels in the prediction of complicated Crohn's disease behavior--a cohort study. *PloS One*, *9*(8), e104706. <https://doi.org/10.1371/journal.pone.0104706>
- Roberts, C. L., Keita, A. V., Duncan, S. H., O'Kennedy, N., Söderholm, J. D., Rhodes, J. M., & Campbell, B. J. (2010). Translocation of Crohn's disease *Escherichia coli* across M-cells: contrasting effects of soluble plant fibres and emulsifiers. *Gut*, *59*(10), 1331–1339. <https://doi.org/10.1136/gut.2009.195370>
- Rogler, G. (2013). Top-Down or Step-Up Treatment in Crohn's Disease? *Digestive Diseases*, *31*(1), 83–90. <https://doi.org/10.1159/000347190>
- Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. <https://doi.org/10.7717/peerj.2584>
- Rook, G. A. W. (2012). Hygiene Hypothesis and Autoimmune Diseases. *Clinical Reviews in Allergy & Immunology*, *42*(1), 5–15. <https://doi.org/10.1007/s12016-011-8285-8>
- Ross, A., Ward, S., & Hyman, P. (2016). More Is Better: Selecting for Broad Host Range Bacteriophages. *Frontiers in Microbiology*, *7*, 1352. <https://doi.org/10.3389/fmicb.2016.01352>
- Rossmann, F. S., Racek, T., Wobser, D., Puchalka, J., Rabener, E. M., Reiger, M., ... Huebner, J. (2015). Phage-mediated Dispersal of Biofilm and Distribution of Bacterial Virulence Genes Is Induced by Quorum Sensing. *PLOS Pathogens*, *11*(2), e1004653. <https://doi.org/10.1371/journal.ppat.1004653>
- Roux, S., Enault, F., Hurwitz, B. L., & Sullivan, M. B. (2015). VirSorter: mining viral signal from microbial genomic data. *PeerJ*, *3*, e985. <https://doi.org/10.7717/peerj.985>
- Rowland, I., Gibson, G., Heinken, A., Scott, K., Swann, J., Thiele, I., & Tuohy, K. (2018). Gut microbiota functions: metabolism of nutrients and other food components. *European Journal of Nutrition*, *57*(1), 1–24. <https://doi.org/10.1007/s00394-017-1445-8>
- Salter, S. J., Cox, M. J., Turek, E. M., Calus, S. T., Cookson, W. O., Moffatt, M. F., ... Walker, A. W. (2014). Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. *BMC Biology*, *12*(1), 87. <https://doi.org/10.1186/s12915-014-0087-z>

- Sartor, R. B., & Wu, G. D. (2017). Roles for Intestinal Bacteria, Viruses, and Fungi in Pathogenesis of Inflammatory Bowel Diseases and Therapeutic Approaches. *Gastroenterology*, *152*(2), 327–339.e4. <https://doi.org/10.1053/J.GASTRO.2016.10.012>
- Sender, R., Fuchs, S., & Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *PLoS Biology*, *14*(8), 1–14. <https://doi.org/10.1371/journal.pbio.1002533>
- Shafiei, M., Dunn, K. A., Boon, E., MacDonald, S. M., Walsh, D. A., Gu, H., & Bielawski, J. P. (2015). BioMiCo: a supervised Bayesian model for inference of microbial community structure. *Microbiome*, *3*(1), 8. <https://doi.org/10.1186/s40168-015-0073-x>
- Sharon, I., Morowitz, M. J., Thomas, B. C., Costello, E. K., Relman, D. A., & Banfield, J. F. (2013). Time series community genomics analysis reveals rapid shifts in bacterial species, strains, and phage during infant gut colonization. *Genome Research*, *23*(1), 111–120. <https://doi.org/10.1101/gr.142315.112>
- Shimada, T., Nagata, N., Okahara, K., Joya, A., Hayashida, T., Oka, S., ... Gatanaga, H. (2017). PCR detection of human herpesviruses in colonic mucosa of individuals with inflammatory bowel disease: Comparison with individuals with immunocompetency and HIV infection. *PLOS ONE*, *12*(9), e0184699. <https://doi.org/10.1371/journal.pone.0184699>
- Shkoporov, A. N., Khokhlova, E. V, Fitzgerald, C. B., Stockdale, S. R., Draper, L. A., Ross, R. P., & Hill, C. (2018). ΦCrAss001, a member of the most abundant bacteriophage family in the human gut, infects Bacteroides. *BioRxiv*, 1–10. <https://doi.org/10.1101/354837>
- Shkoporov, A. N., Ryan, F. J., Draper, L. A., Forde, A., Stockdale, S. R., Daly, K. M., ... Hill, C. (2018). Reproducible protocols for metagenomic analysis of human faecal phageomes. *Microbiome*, *6*(1), 1–17. <https://doi.org/10.1186/s40168-018-0446-z>
- Sokol, H., Lay, C., Seksik, P., & Tannock, G. W. (2008). Analysis of bacterial bowel communities of {IBD} patients: what has it revealed? *Inflammatory Bowel Diseases*, *14*(6), 858–867. <https://doi.org/10.1002/ibd.20392>
- Sokol, H., Pigneur, B., Watterlot, L., Lakhdari, O., Bermudez-Humaran, L. G., Gratadoux, J.-J., ... Langella, P. (2008). Faecalibacterium prausnitzii is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patients. *Proceedings of the National Academy of Sciences*, *105*(43), 16731–16736. <https://doi.org/10.1073/pnas.0804812105>
- Soon, I. S., Molodecky, N. A., Rabi, D. M., Ghali, W. A., Barkema, H. W., & Kaplan, G. G. (2012). The relationship between urban environment and the inflammatory bowel diseases: a systematic review and meta-analysis. *BMC Gastroenterology*, *12*(1), 51. <https://doi.org/10.1186/1471-230X-12-51>

- Stahn, C., Löwenberg, M., Hommes, D. W., & Buttgereit, F. (2007). Molecular mechanisms of glucocorticoid action and selective glucocorticoid receptor agonists. *Molecular and Cellular Endocrinology*, 275(1–2), 71–78. <https://doi.org/10.1016/J.MCE.2007.05.019>
- Steliou, K., Boosalis, M. S., Perrine, S. P., Sangerman, J., & Faller, D. V. (2012). Butyrate Histone Deacetylase Inhibitors. *BioResearch Open Access*, 1(4), 192–198. <https://doi.org/10.1089/biores.2012.0223>
- Tap, J., Furet, J.-P., Bensaada, M., Philippe, C., Roth, H., Rabot, S., ... Leclerc, M. (2015). Gut microbiota richness promotes its stability upon increased dietary fibre intake in healthy adults. *Environmental Microbiology*, 17(12), 4954–4964. <https://doi.org/10.1111/1462-2920.13006>
- Tiede, I., Fritz, G., Strand, S., Poppe, D., Dvorsky, R., Strand, D., ... Neurath, M. F. (2003). CD28-dependent Rac1 activation is the molecular target of azathioprine in primary human CD4+ T lymphocytes. *The Journal of Clinical Investigation*, 111(8), 1133–1145. <https://doi.org/10.1172/JCI16432>
- Turnbaugh, P. J., Ley, R. E., Hamady, M., Fraser-Liggett, C. M., Knight, R., & Gordon, J. I. (2007). The Human Microbiome Project. *Nature*, 449(7164), 804–810. <https://doi.org/10.1038/nature06244>
- Turnbaugh, P. J., Quince, C., Faith, J. J., McHardy, A. C., Yatsunencko, T., Niazi, F., ... Gordon, J. I. (2010). Organismal, genetic, and transcriptional variation in the deeply sequenced gut microbiomes of identical twins. *Proceedings of the National Academy of Sciences of the United States of America*, 107(16), 7503–7508. <https://doi.org/10.1073/pnas.1002355107>
- Turner, D., Griffiths, A. M., Walters, T. D., Seah, T., Markowitz, J., Pfefferkorn, M., ... Levine, A. (2011). Mathematical Weighting of the Pediatric Crohn's Disease Activity Index (PCDAI) and Comparison with Its Other Short Versions. <https://doi.org/10.1002/ibd.21649>
- Turner, D., Grossman, A. B., Rosh, J., Kugathasan, S., Gilman, A. R., Baldassano, R., & Griffiths, A. M. (2007). Methotrexate following unsuccessful thiopurine therapy in pediatric Crohn's disease. *The American Journal of Gastroenterology*, 102(12), 2804–12; quiz 2803, 2813. <https://doi.org/10.1111/j.1572-0241.2007.01474.x>
- Uhlig, H. H., & Muise, A. M. (2017). Clinical Genomics in Inflammatory Bowel Disease. *Trends in Genetics*, 33(9), 629–641. <https://doi.org/10.1016/J.TIG.2017.06.008>
- Van Limbergen, J., Russell, R. K., Drummond, H. E., Aldhous, M. C., Round, N. K., Nimmo, E. R., ... Wilson, D. C. (2008). Definition of Phenotypic Characteristics of Childhood-Onset Inflammatory Bowel Disease. *Gastroenterology*, 135(4), 1114–1122. <https://doi.org/10.1053/j.gastro.2008.06.081>

- Vandeputte, D., Kathagen, G., D'hoel, K., Vieira-Silva, S., Valles-Colomer, M., Sabino, J., ... Raes, J. (2017). Quantitative microbiome profiling links gut community variation to microbial load. *Nature*, *551*(7681), 507–511. <https://doi.org/10.1038/nature24460>
- Vilchez, R. A., & Butel, J. S. (2004). Emergent human pathogen simian virus 40 and its role in cancer. *Clinical Microbiology Reviews*, *17*(3), 495–508, table of contents. <https://doi.org/10.1128/CMR.17.3.495-508.2004>
- Wagner, J., Maksimovic, J., Farries, G., Sim, W. H., Bishop, R. F., Cameron, D. J., ... Kirkwood, C. D. (2013a). Bacteriophages in Gut Samples From Pediatric Crohn's Disease Patients. *Inflammatory Bowel Diseases*, *19*(8), 1598–1608. <https://doi.org/10.1097/MIB.0b013e318292477c>
- Wagner, J., Maksimovic, J., Farries, G., Sim, W. H., Bishop, R. F., Cameron, D. J., ... Kirkwood, C. D. (2013b). Bacteriophages in Gut Samples From Pediatric Crohn's Disease Patients. *Inflammatory Bowel Diseases*, *0*(0), 1. <https://doi.org/10.1097/MIB.0b013e318292477c>
- Wahl, C., Liptay, S., Adler, G., & Schmid, R. M. (1998). Sulfasalazine: a potent and specific inhibitor of nuclear factor kappa B. *The Journal of Clinical Investigation*, *101*(5), 1163–1174. <https://doi.org/10.1172/JCI992>
- Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. <https://doi.org/10.1128/AEM.00062-07>
- Wang, W., Jovel, J., Halloran, B., Wine, E., Patterson, J., Ford, G., ... Mason, A. L. (2015a). Metagenomic analysis of microbiome in colon tissue from subjects with inflammatory bowel diseases reveals interplay of viruses and bacteria. *Inflammatory Bowel Diseases*, *21*(6), 1419–1427. <https://doi.org/10.1097/MIB.0000000000000344>
- Wang, W., Jovel, J., Halloran, B., Wine, E., Patterson, J., Ford, G., ... Mason, A. L. (2015b). Metagenomic analysis of microbiome in colon tissue from subjects with inflammatory bowel diseases reveals interplay of viruses and bacteria. *Inflammatory Bowel Diseases*, *21*(6), 1419–1427. <https://doi.org/10.1097/MIB.0000000000000344>
- Whittaker, R. H. (1972). Evolution and {Measurement} of {Species} {Diversity}. *Taxon*, *21*(2/3), 213–251. <https://doi.org/10.2307/1218190>
- Wild, C. P. (2005). Complementing the genome with an “exposome”: The outstanding challenge of environmental exposure measurement in molecular epidemiology. *Cancer Epidemiology Biomarkers and Prevention*, *14*(8), 1847–1850. <https://doi.org/10.1158/1055-9965.EPI-05-0456>

- Willing, B. P., Dicksved, J., Halfvarson, J., Andersson, A. F., Lucio, M., Zheng, Z., ... Engstrand, L. (2010). A Pyrosequencing Study in Twins Shows That Gastrointestinal Microbial Profiles Vary With Inflammatory Bowel Disease Phenotypes. *Gastroenterology*, *139*(6), 1844–1854.e1. <https://doi.org/10.1053/j.gastro.2010.08.049>
- Wilson, D. C., & Russell, R. K. (2017). Overview of paediatric IBD. *Seminars in Pediatric Surgery*, *26*(6), 344–348. <https://doi.org/10.1053/j.sempedsurg.2017.10.002>
- Xavier, R. J., & Podolsky, D. K. (2007). Unravelling the pathogenesis of inflammatory bowel disease. *Nature*, *448*(7152), 427–434. <https://doi.org/10.1038/nature06005>
- Yamamoto, T., Nakahigashi, M., Umegae, S., Kitagawa, T., & Matsumoto, K. (2005). Impact of Elemental Diet on Mucosal Inflammation in Patients with Active Crohn's Disease: Cytokine Production and Endoscopic and Histological Findings. *Inflammatory Bowel Diseases*, *11*(6), 580–588. <https://doi.org/10.1097/01.MIB.0000161307.58327.96>
- Yu, Y. R., & Rodriguez, J. R. (2017). Clinical presentation of Crohn's, ulcerative colitis, and indeterminate colitis: Symptoms, extraintestinal manifestations, and disease phenotypes. *Seminars in Pediatric Surgery*, *26*(6), 349–355. <https://doi.org/10.1053/j.sempedsurg.2017.10.003>
- Zhang, C., Zhang, M., Wang, S., Han, R., Cao, Y., Hua, W., ... Zhao, L. (2010). Interactions between gut microbiota, host genetics and diet relevant to development of metabolic syndromes in mice. *The ISME Journal*, *4*(2), 232–241. <https://doi.org/10.1038/ismej.2009.112>
- Zhang, T., Breitbart, M., Lee, W. H., Run, J. Q., Wei, C. L., Soh, S. W. L., ... Ruan, Y. (2006). RNA viral community in human feces: Prevalence of plant pathogenic viruses. *PLoS Biology*, *4*(1), 0108–0118. <https://doi.org/10.1371/journal.pbio.0040003>
- Zhang, Y., & LeJeune, J. T. (2008). Transduction of blaCMY-2, tet(A), and tet(B) from *Salmonella enterica* subspecies *enterica* serovar Heidelberg to *S. Typhimurium*. *Veterinary Microbiology*, *129*(3–4), 418–425. <https://doi.org/10.1016/j.vetmic.2007.11.032>
- Zhernakova, A., Kurilshikov, A., Bonder, M. J., Tigchelaar, E. F., Schirmer, M., Vatanen, T., ... Fu, J. (2016). Population-based metagenomics analysis reveals markers for gut microbiome composition and diversity. *Science (New York, N.Y.)*, *352*(6285), 565–569. <https://doi.org/10.1126/science.aad3369>
- Zhou, Y., Chen, H., He, H., Du, Y., Hu, J., Li, Y., ... Nie, Y. (2016). Increased *Enterococcus faecalis* infection is associated with clinically active Crohn disease. *Medicine*, *95*(39), e5019. <https://doi.org/10.1097/MD.0000000000005019>

Zuo, T., Kamm, M. A., Colombel, J.-F., & Ng, S. C. (2018). Urbanization and the gut microbiota in health and inflammatory bowel disease. *Nature Reviews Gastroenterology & Hepatology*, *15*(7), 440–452. <https://doi.org/10.1038/s41575-018-0003-z>

Zuo, T., Wong, S. H., Lam, K., Lui, R., Cheung, K., Tang, W., ... Ng, S. C. (2018). Bacteriophage transfer during faecal microbiota transplantation in *Clostridium difficile* infection is associated with treatment outcome. *Gut*, *67*(4), 634–643. <https://doi.org/10.1136/gutjnl-2017-313952>